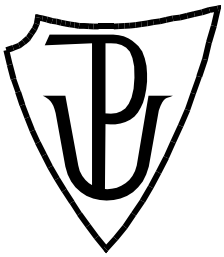


PALACKÝ UNIVERSITY

FACULTY OF MEDICINE AND DENTISTRY

Program DSP: Pediatrics

PROTEOMICS IN BIOMARKER DISCOVERY

Jana Václavková, M.Sc.

Supervising department:

Institute of Molecular and Translational Medicine,

Faculty of Medicine and Dentistry,

Palacký University and University Hospital in Olomouc

Supervisor:

Džubák Petr M.D., Ph.D.

Olomouc 2020

I declare that this thesis has been written solely by myself and that all the relevant resources are cited and included in the references part. The research was carried out at the Laboratory of Experimental Medicine, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University in Olomouc and in cooperation with University Hospital in Olomouc.

In Olomouc

October 2020

Jana Václavková, M.Sc.

Acknowledgments

At first, I would like to thank my supervisor, Džubák Petr M.D., Ph.D. for his advice, patience, help and all the time spent in scientific discussions during my Ph.D. studies. Next, I would like to say thank you to the director of IMTM, associate professor Marián Hajdúch M.D., Ph.D. for his advice during progress reports meetings at IMTM and for the opportunity to realize my Ph.D. studies at IMTM laboratories and take advantage of its equipment. I want to thank all my colleagues at IMTM which helped me during my studies, especially to our technician Eva Kašniarová M.Sc. for her support and help with sample preparation and my colleague from proteomics laboratory Dušan Holub M.Sc., Ph.D. for cooperation and initial training for mass spectrometry measurements at the beginning of my Ph.D. studies.

Thanks also belong to Department of Pediatrics and Department of Respiratory Medicine, University Hospital in Olomouc and Faculty of Medicine and Dentistry, Palacký University in Olomouc. Here, I would like to mention professor František Kopřiva M.D., Ph.D., Tatiana Gvozdiaková M.D., Jana Závodská, professor Vítězslav Kolek M.D., Ph.D., Juraj Kultan M.D. who helped me a lot during my work on our corporate projects.

I would also like to thank Proteome Discover developers from Thermo Fisher Scientific company, especially Michaela Scigelova, Ph.D., for help with software issues and search method development.

And finally, I would like to thank my family and friends who have supported and encouraged me during my studies. Special thanks belong to my mother, Zdenka Václavková M.A., for grammar corrections of all my works written in the Czech language.

The work on projects related to this thesis was supported by grants from the Ministry of Health of the Czech Republic (16-32302A and 16-32318A), from the Czech Ministry of Education, Youth and Sports (LM2018133 – EATRIS-CZ and LM2018130 – CZ-OPENSCREEN), European Regional Development Fund - Project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868) and by the internal grants of Palacký University Olomouc.

In Olomouc

October 2020 Jana Václavková, M.Sc.

Bibliographical identification:

Author's name and surname	Jana Václavková
Title	Proteomics in biomarker discovery
Type of thesis	Dissertation
Department	Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University and University Hospital in Olomouc
Supervisor	Petr Džubák, MD, PhD
The year of presentation	2020
Keywords	Proteomics, biomarker identification, mass spectrometry, exhaled breath condensate, treatment, cellular response.
Language	English

Abstract

This thesis aimed to characterize the proteome of various biological matrices using proteomic methods. We have focused mainly on proteomic approaches to design and validate biomarkers of multiple diseases. Mass spectrometric methods and methods based on electrophoresis play a crucial role in this thesis. We have characterized the process of discovering biomarkers, and the difficulties in this process were mentioned. Biomarkers of diseases could be searched in many biological matrices, and there is a need to look for non-invasively taken ones. Once introduced into clinics, non-invasive diagnostics would facilitate the investigation and diagnosis of many diseases and for patients, make sampling less burdening than the current invasive methods. Exhaled breath condensate (EBC) was selected to be described and further studied as a promising matrix that is collected non-invasively and is suitable for detecting biomarkers of various respiratory and systemic diseases. Biomarkers of response to treatment could be so-called biomarkers of drug toxicity and efficacy, and it is necessary to identify them to understand the mechanism of drug action. Ways to search for and identify these biomarkers have also been described.

In the experimental part, a gel-free mass spectrometry approach was used for in-depth characterization of the EBC proteome with high reproducibility. This approach includes high resolution mass spectrometry and gel-free sample processing. Because EBC is the body fluid proximal to lung and airways, it could be successfully used to detect biomarkers of lung and other respiratory diseases. In our work, the age-related changes in protein profiles of EBC have been reported in healthy children and adults. It has been suggested that proteins that are differentially expressed in different age groups are predictive markers of age-related diseases and resistance or susceptibility to certain diseases. In the second part of the experimental work, proteomic methods for the biomarker validation were implemented. The cellular response to oxaliplatin treatment was assessed by mass spectrometry-based proteome profiling. The identified biomarkers of this cellular response were successfully validated by Western blot. Furthermore, the Drug Affinity Responsive Target Stability (DARTS) method is useful in the biomarker validation. It has been used to confirm that the NPL4 protein is a molecular target of the old alcohol aversion drug disulfiram (Antabuse) when used to treat cancer.

Bibliografická identifikace:

Jméno a příjmení autora	Jana Václavková
Název práce	Proteomika ve vývoji biomarkerů
Typ práce	Disertační
Pracoviště	Ústav molekulární a translační medicíny, Lékařská fakulta Univerzity Palackého a Fakultní nemocnice v Olomouci
Vedoucí práce	MUDr. Petr Džubák, Ph.D.
Rok obhajoby práce	2020
Klíčová slova	Proteomika, identifikace biomarkerů, hmotnostní spektrometrie, léčba, buněčná odpověď.
Jazyk	Anglický

Abstrakt

Cílem předkládané dizertační práce bylo charakterizovat proteom různých biologických matric s využitím proteomických metod. Zaměřili jsme se zejména na proteomické přístupy pro navržení a validaci biomarkerů různých onemocnění. Stěžejní část této práce tvoří hmotnostně spektrometrické metody, ale i metody založené na elektroforéze. Byl popsán proces hledání biomarkerů a byly zmíněny okolnosti komplikující tento proces. Biomarkery mohou být identifikovány v mnoha různých biologických matricích a snahou je využívat takový biologický materiál, jenž je možné odebrat neinvazivně. V případě zavedení nového biomarkeru do klinické praxe může neinvazivní způsob odběru usnadnit vyšetření a diagnostiku řady onemocnění a odběr vzorku bude pro pacienta méně zatěžující než invazivní odběr. Za slibnou biologickou matricí pro neinvazivní diagnostiku respiračních a systémových onemocnění je považován i kondenzát vydechovaného vzduchu (EBC). Ten byl v práci detailně popsán a dále studován. Biomarkery buněčné odpovědi na léčbu jsou také chápány jako biomarkery toxicity a účinků léčiva. Tyto biomarkery je důležité identifikovat pro pochopení mechanismu účinku daného léčiva a jsou popsány také metody, jak je hledat a identifikovat.

V experimentální části práce byl zaveden nový přístup pro detailní a vysoce reprodukovatelnou charakterizaci proteomu EBC. Tento přístup využívá hmotnostní spektrometrii o vysokém rozlišení a zpracování vzorku bez předchozí elektroforetické separace. Vzhledem k tomu, že EBC je proximální tekutinou plic a dýchacích cest, může být využíván pro úspěšnou detekci biomarkerů plicních onemocnění a respiračního traktu. V naší práci byly popsány změny v proteinovém profilu EBC v závislosti na věku u zdravých dětí a dospělých jedinců. Předpokládá se, že proteiny, které byly rozdílně exprimovány v EBC jedinců různých věkových kategorií, jsou prediktivními biomarkery onemocnění spojených s věkem a s rezistencí, případně s vnímavostí k některým onemocněním. Ve druhé kapitole experimentální části práce jsou popsány proteomické metody pro validaci biomarkerů. Buněčná odpověď na ošetření oxaliplatinou byla analyzována pomocí proteomického profilování založeného na hmotnostní spektrometrii. Identifikované biomarkery buněčné odpovědi na léčbu byly úspěšně validovány metodou Western blot. Dále byla zavedena metoda DARTS (Drug Affinity Responsive Target Stability), která se ukázala být velice užitečnou pro validaci biomarkerů. Tato metoda byla použita pro potvrzení skutečnosti, že protein NPL4 je molekulárním cílem disulfiramu (Antabus), pokud se toto léčivo, dříve předepisované k léčbě závislosti na alkoholu, využije pro terapii rakoviny.

TABLE OF CONTENTS

1 INTRODUCTION – PROTEOMICS IN THE BIOMARKER DISCOVERY

1.1. Introduction to discovery of the protein biomarkers

Biomarkers are indicators of a biological process. It could be proteins as well as genes, small molecules, or metabolites (Hale *et al.*, 2003). The biomarker discovery could be focused on identifying the diagnostic and prognostic biomarkers of the diseases (Drabovich, Martínez-Morillo and Diamandis, 2019) or to report a drug efficacy and toxicity as a key step in a drug development process (Hale *et al.*, 2003). The significant progress in a protein biomarker discovery was connected with the development of mass spectrometry, which allowed the accurate and sensitive identification of proteins in the biological matrix and is a crucial method in many of the publications cited in this work. Mass spectrometry, together with other methods connected with this thesis, is summarized in the following chapter.

1.2. Proteomic methods for biomarker discovery identification and validation

1.2.1. Electrophoretic methods

Electrophoretic methods are used to separate proteins based on their charge and size. High-resolution separation of proteins according to their charge was described by disc electrophoresis, displacement electrophoresis (isotachophoresis) and isoelectric focusing (IEF). Separation of by molecular weight is performed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The two-dimensional gel electrophoresis (2-DE) combines the gel IEF, followed by SDS-PAGE of the proteins previously separated by IEF, and thus affords the high resolution with up to several thousand spots per gel (Vesterberg, 1993).

Nowadays, SDS-PAGE is the most widely used method for protein separation. This could be used to estimate protein size, verify protein purity and quantity, monitor protein integrity, compare protein profiles of different samples or analyze the number and size of polypeptide subunits. It could be followed by mass spectrometry-based shotgun proteomics, Western blot or by the second dimension of 2-DE (Righetti and Candiano, 2011). Proteins separated by SDS-PAGE or analyzed by 2-DE could be visualized and analyzed in the gel by several approaches. However, it is usually used as a separation technique for other techniques. For spot detection, Coomassie Blue is typically used because of its simplicity as a regressive staining approach. The limitations of Coomassie Blue staining are the poor detection sensitivity and small linear dynamic range. Silver staining techniques are based upon saturating gels with silver ions. Silver staining methods are able to detect a single nanogram of protein. However, it is a multistep process that must be stopped at some arbitrary time point to avoid overdevelopment. Radiolabeling is quite sensitive but hazardous and expensive. The use of radioisotopes is now limited to *in vivo* experiments. Fluorescent detection is sensitive and suitable for incorporation into integrated proteomics platforms. It is sensitive that 4 - 8 ng of protein can be detected and its linear dynamic range is superior to colorimetric staining methods. Fluorescent SYPRO^{TM} dyes were designed for a simple, rapid, one-step staining procedure which does not involve a destaining step. Reverse stain methods are used for in-gel visualization followed by subsequent MS detection and include Coomassie Blue staining, potassium chloride, copper chloride, and zinc chloride – imidazole methods. Copper chloride staining is slightly more sensitive, while potassium chloride staining is significantly less sensitive than Coomassie Blue staining. Zinc

chloride–imidazole staining is the most sensitive (Patton, 2002). Destaining method of silver-stained gels was developed for subsequent MS identification (Shevchenko *et al.*, 1996).

Although 2-DE was very popular in the past, it is displaced by the high-performance mass spectrometry-based methods (Lee, Saraygord-Afshari and Low, 2020). However, there are many works where the 2-DE is combined with spot excision and mass spectrometry detection in biomarker discovery research (Gianazza *et al.*, 2004; Fumagalli *et al.*, 2008; Paul *et al.*, 2016; Lee, Saraygord-Afshari and Low, 2020). A combination of 2-DE and mass spectrometry spot identification was successfully used for proteomic profiling studies. For example, this approach has led to description mechanisms of drug resistance of T-lymphoblastic leukemia cells to anticancer agent bohemin (Skalnikova *et al.*, 2011), cancer cells response to anthracyclines (Tyleckova *et al.*, 2012), proteins responsible for drug-induced apoptosis after paclitaxel treatment (MacKeigan *et al.*, 2003). Up to date, the 2-DE is used in clinical proteomics for comparative studies, and for this purpose, an algorithm for automatic spot alignment was developed (Pérès *et al.*, 2008). Another approach for comparative proteomics is difference gel electrophoresis (DIGE) which is used for the detection of differences between two protein samples in a single 2-DE gel. It was described as a reproducible, sensitive method which can detect the difference between two *Drosophila* embryo extracts at nanogram levels. Two protein samples were prelabeled with two cyanine dyes Cy3 and Cy5, and thus his approach enables to run two different samples on the same gel in both dimensions. Protein spots can be detected by fluorescence imaging immediately after electrophoresis with sensitivity equal to silver staining (Ünlü, Morgan and Minden, 1997). Two-dimensional electrophoresis has been shown to distinguish different charges or mass forms of a single protein. Thus, it could be applied to characterize post-translational modifications, including glycosylation, phosphorylation, oxidation, and nitrosylation with high resolution and reproducibility. One of the applications of 2-DE in post-translational modifications analysis was the characterization of post-translational modifications of mammalian 20S proteasome complexes (Zong *et al.*, 2008).

1.2.2. Western blot

Western blot is a method for the subsequent identification of proteins separated in the first step by gel electrophoresis. The separated proteins are then transferred to a membrane, which forms a band for each protein. The membrane is then incubated with labeled antibodies specific for the protein of interest (Mahmood and Yang, 2012). It is one of the most commonly used laboratory techniques for protein identification and semi-quantitative protein level measurement. Western blotting can also be used for absolute protein quantification, but this requires a standard linear curve of the purified target protein. Thus, the Western blot is rarely used for absolute quantification and the main application is a semi-quantitative estimation of protein levels (Ghosh, Gilda and Gomes, 2014). Western blot is widely used to validate protein biomarkers even in recent publications (Xu *et al.*, 2015; Paul *et al.*, 2016; Madda *et al.*, 2020).

For the first time, the method was described by Harry Towbin's group as the electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets. Immobilized proteins were detected by immunological methods. Additional nitrocellulose binding capacity was blocked by excess protein, then a specific antibody was bound and a secondary antibody against the primary antibody was applied. The secondary antibody was either radiolabeled or conjugated to fluorescein or peroxidase. The specific protein was then detected either by autoradiography, under UV light, or by the peroxidase reaction product. The Towbin's method described the quantitative transfer from urea-containing gels and not quantitative transfer without loss of resolution from sodium dodecyl sulfate gels (Towbin, Staehelin and Gordon, 1979). Western blotting comes from W. Neal Burnette, who adapted the protocol to ensure the complete and quantitative transfer of most proteins from SDS gels to nitrocellulose without loss of electrophoretic resolution (Burnette, 1981). Efficient protein transfer with high protein yields has also been described using polyvinylidene difluoride (PVDF) membranes instead of nitrocellulose (Mozdzanowski, Hembach and Speicher, 1992). Unlike a nitrocellulose membrane, the PVDF membrane is hydrophobic and an initial soaking of the PVDF membrane in methanol is required before blotting (Mansfield, 1995). Soaking in transfer buffer is another important step, as higher concentrations of SDS could inhibit protein transfer. Higher concentration of methanol in transfer buffer also reduce protein recovery (Mozdzanowski, Hembach and Speicher, 1992). After transfer, membranes could be stained to monitor the transfer efficiency before subsequent immunodetection or to visualize proteins. There are several options for staining the membrane. The staining with amido black or Coomassie blue R-250 showed similar sensitivity (more than 50 ng of protein per band), but amido black is staining faster and Coomassie staining could not be used for nitrocellulose

membranes due to high concentrations of methanol in the staining solvent that resulted in the dissolution of nitrocellulose. Once visualized, the amido black staining and Coomassie blue R-250 staining appears as dark blue bands on a light blue background. Ponceau S staining appears as red bands on a pink background and is widely used because it is simple, rapid, and easily reversible. However, its sensitivity is low (200 ng protein per band). Another reversible staining is MemCode reversible protein stain, which is more sensitive (25 ng protein per band) but less fast and appears as blue bands on a white background. Colloidal gold is highly sensitive (2 ng protein per band), the commercial staining kits are available, and it appears as red bands on a pink background. Colloidal silver staining is less sensitive (5 ng protein per band) than colloidal gold, but it is more economical and appears as black bands on a brown background. India ink staining showed a sensitivity and appearance (black bands on a gray background) similar to colloidal silver. Fluorescent dyes such as SYPRO Ruby or fluorescamine are highly sensitive (about 2 ng of protein per band) and the membranes can be used for subsequent immunodetection (Goldman, Harper and Speicher, 2016). At the time of its discovery, immunodetection on membranes was described as sensitive that 100 pg of protein was clearly detectable (Towbin, Staehelin and Gordon, 1979). However, using alkaline phosphatase-labeled secondary antibodies and enhanced chemifluorescence reagents, a sensitivity of less than one femtomole (i.e. hundreds of attomoles) has been achieved (Coorssen *et al.*, 2002). Specific detection of proteins is performed using primary antibodies. The membrane is blocked after the electroblotting to reduce the likelihood of nonspecific binding of the primary antibody. Blocking can be done with skim milk powder or bovine serum albumin. The protein – primary antibody complex is visualized by a secondary antibody targeted against the primary antibody and is associated with a label (e.g. chemiluminescent or fluorescent). Then, the signal of the label is developed (Bass *et al.*, 2017).

Both 1-D and 2-D gels can be used for Western blot transfer. In the case of 2-DE, isoforms and post-translationally modified target proteins with similar molecular masses can be identified (Ghosh, Gilda and Gomes, 2014). It is suitable for denatured and native proteins. A nondenaturing polyacrylamide gel electrophoresis (PAGE) system coupled to Western blotting was developed to detect native proteins. A system with a very low (0.1%) SDS concentration was used to keep the proteins in the native state. Proteins were successfully blotted onto a nitrocellulose membrane and detected by monoclonal antibodies (Cohen *et al.*, 1986). It is also possible to use Western blot transfer blue native PAGE gel, where the gel is incubated in a transfer buffer supplemented with low SDS concentrations (such as 0.037%) before transfer by standard Western blot (Dewson, 2015).

The Western blot's main limitation is the need for a primary antibody against the protein of interest. In the case of specific post-translational modifications, particular antibodies should be used. Other limitations include the non-specificity of some antibodies, the need to optimize the protocol independently for each antibody, relatively high inter-operator variability or transfer efficiency. At the same time, the membrane may not retain small proteins (< 10kDa), and large proteins (> 140kDa) may not be transferred to the membrane (Ghosh, Gilda and Gomes, 2014).

1.2.3. Mass spectrometry

Mass spectrometry-based techniques are widely used in analytical chemistry since they were developed more than a hundred years ago. Applications in proteomics came with development of soft ionization techniques in the 1980s (Cañas Montalvo *et al.*, 2006). Currently, MALDI (matrix-assisted laser desorption/ionization; Karas and Hillenkamp, 1988) and ESI (electrospray ionization; Yamashita and Fenn, 1984) are used for protein analysis. Since 1980s, these two ionization techniques have remained the most important methods for ionization of biomolecules in mass spectrometry applications. In MALDI, the analyte is spotted to a MALDI plate with a typically acidic matrix that strongly absorbs UV light. The molecules are excited by a short laser pulse; parts of the matrix are heated rapidly and are vaporized and ionized together with the analyte. In ESI, an electric field is applied to the analyte solution flowing through the capillary. At the fine tip of the capillary, the liquid is emitted towards the counter electrode. The solvent evaporates, the droplet size decreases, and this results in a Coulomb explosion and smaller droplets with charged ions are generated (Nadler *et al.*, 2017). The complementarity of these two ionization techniques has been reported (Medzihradszky *et al.*, 2001; Nadler *et al.*, 2017). Each ionization technique allows the detection of significantly different peptides. Differences between the resulting peptide sets were observed with respect to amino acid composition, charge-related parameters, hydrophobicity, and modifications of the detected peptides (Nadler *et al.*, 2017). Several mass analyzers, such as ion trap (IT) a quadrupole (Paul and Steinwedel, 1953), time-of-flight mass spectrometer (TOF; Wiley and McLaren, 1955), mass reflectron time-of-flight mass spectrometer (Mamyrin *et al.*, 1973), Fourier transform ion cyclotron resonance spectroscopy (FT-ICR; Comisarow and Marshall, 1974), Orbitrap (Makarov, 2000) could be used and combined to gain the best results and optimal resolution and sensitivity. Liquid chromatography (LC) coupled to tandem mass spectrometry (called LC-MS/MS) is widely used to achieve efficient separation of biological materials and sensitive identification of individual components by mass spectrometry. This is a powerful strategy for the analysis of peptides and proteins. Electrospray ionization technique is used to combine liquid chromatography with mass spectrometry. In a typical LC-MS/MS experiment, the analyte is eluted from the reverse phase column to separate the peptides by hydrophobicity and is ionized and transferred with high efficiency to a mass spectrometer for analysis. Prior to LC-MS/MS, methods such as 1-D or 2-D electrophoresis, immunoprecipitation, or other protein purification techniques could be used to obtain better results (Mann, Hendrickson and Pandey, 2001).

1.2.3.1. Shotgun proteomics

Shotgun proteomics is most frequently used approach for biomarker identification using mass spectrometry methods. Shotgun proteomics is an approach in which a protein mixture is proteolytically cleaved to obtain a mixture of peptides, which is then separated by single- or multi- dimensional high-pressure liquid chromatography (LC or LC/LC) and identified by tandem mass spectrometry (MS/MS) analysis. This approach is suitable for complex samples. Despite gel-based techniques, shotgun proteomics shows advantages in speed, sensitivity, scope of analysis, and dynamic range. However, in the case of complex protein samples, shotgun proteomics is usually combined with pre-processing by gel-based methods. Despite the fact that these approaches are strong and constantly improving, there is no approach that can detect all proteomic challenges (McDonald and Yates, 2002).

1.2.3.2. Data acquisition modes

Data-dependent acquisition (DDA) has been widely used in proteomics since the beginning of this millennium. In DDA, the subset of the most abundant ions reaching the mass spectrometer detector during the MS1 scan is individually isolated and fragmented in the MS2 sequential scans, and each MS2 scan can be analyzed using a database search algorithm. Recent improvements in mass spectrometer design and bioinformatics algorithms have resulted in the rediscovery and development of data-independent acquisition (DIA), which is able to sample each peptide in a protein digest in a comprehensive and repeatable manner and create a complex set of mass spectra that is usually interpreted by external spectral libraries (Wolf-Yadlin, Hu and Noble, 2016).

1.2.3.3. Orbitrap technology in shotgun proteomics

Orbitrap mass spectrometry is the latest technology in mass analysis, which has led to increased selectivity and reliability of analyzes and is still evolving. The measurements provide accurate mass and high resolution, and these properties are similar to FT ICR instrumentation. The Fourier transformation is also used in Orbitrap to obtain oscillation frequencies for ions with different masses, which leads to an accurate reading of their m/z (Scigelova and Makarov, 2006; Zubarev and Makarov, 2013). The Orbitrap mass spectrometer allowed high coverage proteome analyzes. Draft map of the human proteome was identified using Orbitrap technology. The proteomic profile was analyzed from 30 histologically normal human samples, including 17 adult tissues, 7 fetal tissues, and 6 purified primary hematopoietic cells. In this study, a complex mixture of peptides obtained by tryptic digestion of gel bands from SDS–PAGE or

basic RPLC fractions was separated on a micro-capillary RPLC column, ionized by nanoflow electrospray ionization and analyzed on LTQ-Orbitrap Elite or LTQ-Orbitrap Velos mass spectrometers using HCD fragmentation. The analysis resulted in the identification of 17,294 genes, which is approximately 84% of the total annotated genes encoding proteins in humans (Kim *et al.*, 2014). In-depth proteomic analyzes that have identified more than 5000 proteins in samples usually require long nano-spray HPLC columns (>50 cm) and long MS machine run times (up to several days or more per sample run). Thus, the effort has led to a faster and more convenient method for deep proteome analysis of large-scale samples. One step towards better identification technology was the introduction of the Orbitrap Fusion, the first model of the Orbitrap tribrid mass spectrometer. Using Orbitrap Fusion, the approach for deep proteome analysis was optimized. With the optimized method and off-line high pH separation, more than 5000 proteins were identified using an in-house packed 10 cm C18 RP column in a single one hour LC-MS run and 8493 proteins with 6 orders of dynamic range in only 10 hours of MS run time in HeLa cells. More than 4500 phosphopeptides were identified in one hour LC-MS run after optimization. These results supported the fact that Orbitrap Fusion technology with an optimized method is suitable for system-wide proteomic profiling and protein post-translational modification studies and would facilitate the discovery of proteomic biomarkers (Nie *et al.*, 2016). Orbitrap technology is still evolving; the latest technology is called Orbitrap Exploris, which is quadrupole-orbitrap mass spectrometer. This instrument is equipped with a front-end High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) interface which acts as an ion selection device and an electrospray filter that prevents neutrals from entering the orifice of the mass spectrometer while reducing chemical background noise. The FAIMS interface increases the sensitivity and number of identifications. It has also been designed for in-depth proteome profiling and single cell proteomics with high sensitivity and high proteome coverage. Both data-dependent acquisition (DDA) and data-independent acquisition (DIA) can be performed. In the case of the DDA approach in combination with FAIMS, more proteins but fewer peptides were identified (i.e. resulting in lower sequence coverage), while in the case of DIA more proteins were identified and the number of peptide identifications was maintained (Bekker-Jensen *et al.*, 2020).

1.2.3.4. Targeted proteomics

Targeted proteomics methods are used to complement shotgun proteomics as a next step to validate potential biomarkers (Lange *et al.*, 2008). Targeted proteomics detects proteins of interest with high sensitivity, quantitative accuracy and reproducibility. The triple quadrupole

mass spectrometer which was originally developed in the late 1970s still remains a widely used tool in targeted proteomics. The triple quadrupole is used for an approach called Selected Reaction Monitoring (SRM), also known as Multiple Reaction Monitoring (MRM), which focuses on a mass spectrometer to detect a preselected group of analytes. It is possible to detect even several thousand peptides in a single sample. A thousands of samples could be analyzed using this approach (Marx, 2013).

Multiple reaction monitoring (MRM) LC-MS/MS has been the standard for targeted proteomics when accurate quantification and validation were the main objectives (Wolf-Yadlin, Hu and Noble, 2016). Sensitivity and reproducibility are the main advantages of MRM, and once MRM assays are generated, it is much faster than a typical discovery-based experiment (Marx, 2013). Selectivity and dynamic range is another advantage of this approach and it is possible to detect low-abundance proteins in highly complex mixtures (Lange *et al.*, 2008). There was an evidence of MRM analysis in 1978 when the method was applied for direct analysis of metabolites in complex samples. Here, the mass-analyzed ion kinetic energy (MIKE) spectrometry was used and fragmentation spectra for MS/MS were obtained by collision induced dissociation (Kondrat, McClusky and Cooks, 1978). In the MRM experiment, a predefined precursor ion and one of its fragments are selected by the two mass filters of a triple quadrupole instrument and monitored over time for accurate quantification. A series of transitions (which means precursor and fragment ion pairs) in combination with the retention time of the targeted peptide are set in the assay. The first and the third quadrupoles are used as filters to specifically select predefined m/z values corresponding to the peptide ion and the specific fragment ion of the peptide, and the second quadrupole serves as a collision cell (Lange *et al.*, 2008).

Using the DIA approach, parallel reaction monitoring (PRM) has been introduced as a newer version of MRM (Wolf-Yadlin, Hu and Noble, 2016). This approach was developed for a new generation of quadrupole-equipped high resolution and accurate mass instruments such as the quadrupole equipped Orbitrap MS. In the PRM experiment, all products of a target peptide are simultaneously monitored under conditions that offer high resolution and high mass accuracy. The precursor selection and fragmentation in the collision cell is performed in the same way as in the case of MRM. However, the PRM monitors all product ions of a mass-selected peptide target in parallel with a single ion injection and full mass range by Orbitrap mass analysis. Spectra are highly specific because all potential product ions of a peptide are monitored instead of just 3–5 transitions, which allowed a higher probability of peptide identification. Then, higher resolution and scanning of full fragmentation spectrum allowed

better chance to distinguish the co-isolated background peptides. It is not necessary to preselect the target transitions before analysis. Much of the effort required to develop and optimize the traditional MRM assay is eliminated in PRM. Data analysis of the PRM results is facilitated and automated compared to MRM. Additionally, PRM provided quantitative data in a wider dynamic range than the MRM method (Peterson *et al.*, 2012). A sometime before the publication of PRM method, a DIA approach was developed on a LTQ-Orbitrap instrument without a quadrupole. In this so-called Fourier Transform-All Reaction Monitoring (FT-ARM), the need for precursor mass selection was replaced with high-resolution mass measurement on all fragment ions. The ions are not physically filtered during the analysis as in the MRM assay. All ions are fragmented at each scan during LC separation to give a total ion chromatogram and complex fragmentation spectrum of all ions is performed and compared to hypothetical peptide fragmentation spectra (Weisbrod *et al.*, 2012).

Another approach using the DIA acquisition is SWATH analysis (Marx, 2013). In this approach, data were acquired on a fast, high resolution quadrupole-quadrupole time-of-flight (TOF) instrument by repeated cycling through 32 consecutive 25-Da precursor isolation windows (swaths). The analysis consists of the consecutive acquisition of high resolution, accurate mass fragment ion spectra during the entire chromatographic elution (retention time) range by repeated stepping through 32 discrete precursor isolation windows of 25-Da width across the 400 –1200 m/z range. In a single sample injection, time-resolved fragment ion spectra are recorded for all analytes detectable within this precursor range and in a user-defined retention window. A series of isolation windows acquired for a given precursor mass range and across the LC is referred to as a swath. The cycle time is defined as the time required returning to the acquisition of the same precursor isolation window. This method combines the advantages of quantifying of as many compounds as those typically identified by conventional shotgun proteomics, with the accuracy and reproducibility of MRM across many samples (Gillet *et al.*, 2012).

1.2.3.5. Quantitative mass spectrometry

It is possible to introduce methods of absolute and relative quantification. Absolute quantification employed preparation of synthetic peptides and is basically used for targeted proteomics. Relative quantification methods include metabolic labeling of proteins in cell culture, and chemical reagent labeling of proteins in a sample or cell lysate is commonly used for shotgun proteomic analysis (Gygi *et al.*, 1999; Ong *et al.*, 2002; Gerber *et al.*, 2003; Thompson *et al.*, 2003; Ross *et al.*, 2004; Choe *et al.*, 2007).

The absolute quantification of proteins and their modification states is called AQUA. Peptides are synthesized with stable isotopes (^{13}C or ^{15}N) incorporated as ideal internal standards to mimic native peptides generated by proteolysis. These synthetic peptides can also be prepared by post-translational modification. The synthetic internal standard peptides are then used to accurately and quantitatively measure the absolute levels of proteins and post-translationally modified proteins after proteolysis using a selected reaction monitoring analysis in a tandem mass spectrometer (Gerber *et al.*, 2003).

A method of quantitative proteomics based on stable isotope quantification called SILAC was designed for the metabolic *in vivo* incorporation of specific amino acids into all mammalian proteins. Mammalian cell lines are grown in standard growth media and in parallel, in media lacking a standard essential amino acid, but supplemented with a non-radioactive, isotopically labeled form of this amino acid. It was shown that there is no difference in cell morphology, doubling time, and the ability to distinguish between cells grown with a standard or modified amino acid. Complete incorporation of the modified amino acid occurred after five doublings in the cell lines and proteins. Cells from experimental and control samples are mixed in 1:1 ratio immediately after harvesting. This method is widely used for mass spectrometric identification and relative quantitation of changes in complex protein mixtures. Originally, the method was published using deuterated leucine (Leu-d3) as an isotopically labeled form of the amino acid (Ong *et al.*, 2002). Later, $^{13}C_6$-labeled arginine was used rather than Leu-d3 in the SILAC method (Ong, Kratchmarova and Mann, 2003). Several SILAC amino acids could be used and are commercially available with different numbers of ^{13}C, ^{15}N or ^{2}H. Amino acids labeled by ^{13}C and ^{15}N are more expensive, however, they can simplify data analysis because the SILAC pairs do not separate as much in LC–MS. Ideally, the SILAC amino acid should be essential for survival of the cultured cells, because in this case the heavy amino acid in media is the only source of that particular amino acid. Leucine, lysine and methionine are the essential amino acids that have been used in SILAC. Arginine is not an essential amino acid, but it is mostly obtained through diet in whole organisms and has been shown to be essential in many cultured cell lines. Thus, arginine and has been successfully used for SILAC labeling (Ong and Mann, 2007). For comparative analysis of prostate cancer cells with varying metastatic potential, $^{12}C_6$- or $^{13}C_6$- l-lysine was introduced (Everley *et al.*, 2004). A combination of $^{13}C_6$-arginine and $^{13}C_6$-lysine is usually used as well. This approach has also been extended to the quantification of phosphoproteins (Ibarrola *et al.*, 2003; Zhang and Neubert, 2009). For identification and quantification of *in vivo* methylation sites, labeling with [$^{13}CD_3$] methionine was implemented. Heavy [$^{13}CD_3$] methionine is converted intracellularly to [$^{13}CD_3$] S-adenosyl

methionine which transfers the heavy methyl groups to methyl acceptors such as proteins. Heavy methyl groups are fully incorporated into methylation sites *in vivo*, and thus directly label the post-translational modified methyl sites (Ong, Mittler and Mann, 2004).

Isotopic labels in proteomics were for the first time used for bacterial proteins analysis, where uniformly ^{15}N-labeled (>96%) rich cell growth media was used for labeling. This approach has led to the incorporation of ^{15}N into all amino acids in bacterial proteins (Oda *et al.*, 1999; Lahm and Langen, 2000). However, this method is limited to use in case of microorganisms because ^{15}N-substituted media are difficult and expensive to make for mammalian system. Another limitation of ^{15}N-label is that the degree of incorporation is not necessarily 100% and interpretation of the resulting mass spectra would be difficult (Ong *et al.*, 2002).

The first of the approaches based on chemical reagents was termed isotope-coded affinity tags (ICAT) and protein identification and quantification by tandem mass spectrometry. Isotope-coded affinity tags (ICAT) is composed of three functional elements: a specific chemical reactivity, an isotopically coded linker, and an affinity tag. The cysteines in proteins are reduced and the tag is bound via a specific chemical reactivity group with specificity toward sulfhydryl groups. One sample is derivatized with the isotopically light form of the ICAT reagent, the second with the isotopically heavy reagent. The two samples are combined and enzymatically cleaved to generate peptides. The tagged (cysteine-containing) peptides are isolated by affinity chromatography and the isolated peptides are separated and analyzed by LC-MS/MS. However, the main disadvantage of this approach is that only highly abundant proteins are detected when complex sample such as total lysate is measured and is limited to cysteine-containing peptides (Gygi *et al.*, 1999).

Amine-reactive isobaric tagging reagents were developed for multiplex analysis where more than two samples should be compared. The derivatized peptides are indistinguishable in MS, but exhibit intense low-mass MS/MS signature ions that support quantitation. Up to 4-fold multiplexing was enabled simultaneously for relative protein measurements. An isobaric reagent consists of amine reactive group and isobaric tag. During MS2, the isobaric tag is fragmented to the reporter group, which ranges in mass from m/z 114.1 to 117.1, and a balance group which makes the mass constant (145.1 Da) when bound to peptide, and is lost during the MS2 as a neutral loss. The tags are added to tryptic digest to label the peptides. The option to determinate absolute levels of a target protein was reported using the comparative analysis of three peptide mixture labeled with 114, 115 and 116 tag and synthetic peptide standards labeled with 117 isobaric tag (Ross *et al.*, 2004). This new class of isobaric reagents was

commercialized by the same group at Applied Biosystems in 2004 as iTRAQ™ Reagents (Zieske, 2006). An 8-plex version of an isobaric reagent for the quantitation of proteins using shotgun methods was published by the same group which developed iTRAQ labels. The 8-plex version of the reagent was based on amine-labeling chemistry of peptides similar to 4-plex reagents. Reporter ions for MS/MS were used at 113, 114, 115, 116, 117, 118, 119, and 121 m/z are used to quantify protein expression in cerebrospinal fluid of subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease (Choe *et al.*, 2007).

A quantification strategy for comparative analysis of complex protein mixtures by MS/MS isotopomer labels, referred to as tandem mass tags (TMT). These tags have the same overall mass and are designed to ensure that identical peptides labeled with different TMTs exactly comigrate in all separations. Each tag comprises a sensitization group and a mass differentiated group that together comprises the TMT fragment that is actually detected. The TMT fragment is linked to a mass-normalization group that ensures that each tag in a pair of tags shares the same overall mass and atomic composition. The first and second generation tags are distinguished by the presence of an additional fragmentation-enhancing group, proline, in the second generation tag. The tags will additionally comprise a reactive functionality to enable the tag to be coupled to any peptide. These tags were originally developed for differential analysis of two samples (Thompson *et al.*, 2003). The peptides are also labeled by TMT after proteolytic digestion (Dayon *et al.*, 2008). The TMT fragment is detached from a mass-normalization group during MS fragmentation by CID (Thompson *et al.*, 2003). During MS/MS fragmentation, quantification information is obtained through the losses of the reporter ions. Firstly, 6-plex isobaric mass tagging technology was performed with respective reporter ions at m/z 126.1, 127.1, 128.1, 129.1, 130.1 and 131.1 (Dayon *et al.*, 2008). Up to date, the standard TMT labels with dimethylpiperidine reporter are used up to 11-plex assays. Recently, the TMTpro reagents were synthesized for multiplexing up to 16 samples. The new reporter structure provided a set of 16 tags for use with resolution of 6.3 mDa mass differences in high-resolution mass spectrometers and a set of 9 reagents with 1 Da spacing between reporter ions for single dalton analysis using 9 heavy nuclei per tag (Thompson *et al.*, 2019). Using 16 isobaric reagents based on an isobutyl-proline immonium ion reporter structure (TMTpro), more than 8,800 proteins per replicate were quantified with an analysis time of only 1.1 h per proteome. These reagents have similar characteristics to existing tandem mass tag reagents but with increased fragmentation efficiency and signal. The use of TMTpro reagents led to quantification with no missing values across the 16 samples and no loss in quantitative integrity (Li *et al.*, 2020). The method using 27-plex TMT was described combining commercially available TMT 11-plex

and TMTpro 16-plex. Labeling with TMT reagents was coupled with two-dimensional liquid chromatography (LC/LC) for extensive peptide fractionation and high-resolution tandem mass spectrometry (MS/MS) for peptide quantification and the approach was applied to evaluate the complex human brain proteome of Alzheimer's disease (Wang *et al.*, 2020).

Despite the recent advances in the field of isotope labeling methods, these methods have some limitations such as increased time and complexity of sample preparation, the requirement for higher sample concentration, high cost of the reagents or incomplete labeling (Zhu, Smith and Huang, 2010). Therefore, the methods of label-free proteomics are still widely used.

One of the label-free approaches is the measurement of ion intensity changes based on peptide peak areas or peak heights in chromatography. The second methodology is based on the spectral counting of identified proteins after MS/MS analysis. Peptide peak intensity or spectral count is measured for individual LC-MS/MS or LC/LC-MS/MS runs and changes in protein abundance are calculated via a direct comparison between different analyses (Zhu, Smith and Huang, 2010).

1.2.4. Drug Affinity Responsive Target Stability (DARTS)

A method, drug affinity responsive target stability (DARTS) was described for identifying molecular targets, global mapping of protein–metabolite interaction networks and in the label-free screening of unlimited varieties of compounds for development as molecular imaging agents (Lomenick *et al.*, 2009). In the context of the thesis, DARTS was used for validation of the biomarker of the treatment response.

The method is based on a reduction in the protease susceptibility of the target protein upon drug binding. The both, protein in complex with small molecule and protein alone are lysed by protease. When a small molecule compound is bound to a protein, the interaction stabilizes the target protein's structure and thus, the protein becomes protease resistant and no proteolysis occurs. In the case of protein without small molecule bound, nothing is protected and the protein is digested. The two variants are always performed, one with small molecule and the second serves as control (without small molecule). The differences are compared with the methods such as Western blotting, SDS-PAGE, 2D-PAGE, and gel-free or gel-based mass spectrometry proteomics. The use of DARTS method is universal because is independent of the mechanism of drug action and it is not limited by synthetic chemistry, which means that it does not require modification or immobilization of the small molecule. Additionally, the DARTS is applicable to any sample, from pure proteins to complex protein mixture while there is no need to reduce sample complexity (Lomenick *et al.*, 2009, 2011).

1.3. Identification of protein biomarkers of the diseases

1.3.1. Introduction to identification of protein biomarkers of the diseases

The vision of biomedical and translational science is to discover diagnostic and prognostic biomarkers of almost all diseases including cancer, cardiovascular and neurodegenerative diseases. This is facilitated using high-throughput -omics technologies such as genomics, epigenomics, transcriptomics, and proteomics. Using these approaches, the analyses resulted in obtaining a global profile of genes, mRNA, and proteins and distinguishing healthy condition from the disease (Drabovich, Martínez-Morillo and Diamandis, 2019). Additionally, single-cell omics methods can play an important role in cancer biomarker research. It is evident that individual tissues, especially tumor tissues, are heterogeneous structures. Continuing advancements in single-cell technology can lead to the identification of intra-tumor changes and to the development of more personalized cancer treatment (Ortega *et al.*, 2017).

Biomarker discovery is a very perspective field of scientific work and the number of publications uncovering potential biomarkers is continuously increasing. However, the number of biomarkers approved by the USA Food and Drug Administration (FDA) is still low. For instance, no major cancer biomarker for screening or early diagnosis was approved by FDA in the last 25 years. Only a few biomarkers could satisfy the requirements for the ideal biomarker. The major problem is low specificity (Drabovich, Martínez-Morillo and Diamandis, 2019). Prostate specific antigen could serve as a good example. Prostate specific antigen (PSA), which is nowadays used in clinical practice, was approved by FDA in 1986 to monitor prostate cancer and in 1994 PSA was approved as a tool for prostate cancer detection in men older than 50 years. The screening of PSA has significantly increased the detection of prostate cancer (Duskova and Vesely, 2015). Prostate specific antigen is used as a marker for prostate cancer screening, monitoring of treatment efficacy and prediction of cancer recurrence (Freedland *et al.*, 2003). Expression of PSA is prostate gland specific but it is increased in many both malignant and benign diseases (Duskova and Vesely, 2015). Elevated expression of PSA was detected in inflammatory prostate diseases (Azab, Osama and Rafaat, 2012). Another example is human epididymis protein 4 (HE4) which was found to be highly expressed in the early stages of ovarian cancer. Human epididymis protein 4 was described as a promising marker of the early diagnosis and potential important early indicator of the recurrence of the disease. The patients with ovarian cancer demonstrated high levels of HE4 with 96.9% sensitivity and there was 3.7% positivity in the other pathologies (Anastasi *et al.*, 2010). However, HE4 was

approved by the FDA only for monitoring of the cancer recurrence (Drabovich, Martínez-Morillo and Diamandis, 2019).

The biomarker identification is a multistep process beginning with gene transcript or protein identification and candidate qualification, continuing with verification, development of pre-clinical assay, clinical validation and is terminated with assay approval by health agencies. The integrated processes protein identification and candidate qualification provide dozens of candidate proteins for the verification phase. The sensitive, complicated and low-throughput technologies are displaced by easy and high-throughput techniques and the number of patients samples increasing towards the end of the whole process. The whole process timeline could exceed many years and its cost could include millions of dollars (Drabovich, Martínez-Morillo and Diamandis, 2019).

1.3.2. Biological matrices which could serve as a source of biomarkers

The tumor tissue itself is good, but still limited, tool for cancer diagnosis. Nevertheless, the tumor biopsy remains the gold standard for cancer detection, analysis of suspicious mass, diagnosis and possible treatment indications. Tissue biopsy could be performed only in the case that the tumor mass is visible, so the detection is not as early as needed and thus the probability of treatment success is potentially decreased. Other complications include that the tissue biopsies are invasive, cause pain and distress for the patients and that certain tumors are not accessible to biopsies because of it locations. The subsequent analysis of tissue biopsy is complicated because of tumor tissue heterogeneity (Mannelli, 2019).

The biomarker research of tumor biopsies includes mainly the analysis of formalin-fixed paraffin-embedded (FFPE) tissue which is stable over time and is usually associated with rich clinical and phenotypic data. Analysis of FFPE tissue samples was performed on the genomic, transcriptomic and proteomic levels. Immunohistochemistry (IHC) is the widely used method for protein analysis on FFPE tissue samples. Immunohistochemical analysis is applied for FFPE samples since 1991 when the heat-induced antigen retrieval (HIAR) technique for IHC assay was developed (Zhu *et al.*, 2019). Since then, recombinant antibody microarrays for sensitive and multiplex protein expression profiling were developed (Pauly *et al.*, 2013) as well as several mass spectrometric approaches (Nazarian *et al.*, 2008; Byrum *et al.*, 2011; Holub *et al.*, 2019; Zhu *et al.*, 2019). The high-throughput proteomic analyses are used, these include recently developed a PCT – SWATH (Pressure Cycling Technology – Sequential Windowed Acquisition of All Theoretical Fragment Ion) mass spectrometry workflow (Zhu *et al.*, 2019). The data dependent analysis on the Orbitrap Elite Hybrid Ion Trap-Orbitrap mass spectrometer was successfully used for amyloid typing in FFPE tissue samples (Holub *et al.*, 2019).

The barriers of tissue biopsies could be overcome by the liquid biopsies which mean the analysis of body fluids. Liquid biopsy is less invasive and could be implemented in the clinical setting for screenings, diagnosis, and targeted therapies and to monitor disease evolution and progression. The most common kind of liquid biopsy is the analysis of circulating tumor DNA (ctDNA) in the blood (Mannelli, 2019).

Blood is generally the most common body fluid used for biomarker identification. It is because of the simplicity of obtaining blood samples. However, identifying protein biomarkers from blood show limits which are the consequence of blood sample complexity. Serum and plasma are very complex mixtures of proteins and some of them are very abundant. The most abundant proteins form 95% of the total protein amount but only 0.1% of total protein species. Thus, the detection of low abundant biomarkers will be difficult and the strategies to reduce the

serum or plasma complexity are usually introduced (Tessitore *et al.*, 2013). Additionally, the blood is not organ specific and each organ secretes the specific body fluid where the biomarker concentration should be higher than in human blood. Among these specific body fluids, urine, sputum, sperm, cervical mucus, cerebrospinal fluid, pancreatic juice, bile, amniotic fluid, tears, saliva, nasal mucus, bronchoalveolar lavage or exhaled breath condensate could be mentioned. Some of them such as cerebrospinal fluid, amniotic fluid or bronchoalveolar lavage are used in clinics nowadays but the collection is invasive, not pleasant for patients and can be accompanied by serious complications. Therefore, intensive research work is conducted to develop non-invasive methods for biomarker detection. The experimental part of this thesis and related articles are focused on non-invasive collection and analysis of exhaled breath condensate as a promising source of biomarkers of respiratory diseases. Thus, the exhaled breath condensate analysis will be further described in detail.

1.3.3. Exhaled breath condensate analysis

The collection of exhaled breath condensate is a non-invasive and easy method to obtain lung samples from human adults as well as preschool children (Horváth *et al.*, 2005; Rosias *et al.*, 2008). When collected, exhaled breath is directed through a cooling device, resulting in a liquid or solid phase accumulation of exhaled breath constituents depending on the condenser temperature. It is a safe method, while no adverse events have been reported in over 10,000 measurements performed in different laboratories with different devices (Horváth *et al.*, 2005). In this thesis, the EcoScreen® collection device and TurboDECCS device was used (Figure 1). TurboDECCS consists of a portable Turbo Unit (Turbo is an acronym for Transportable Unit for Research of Biomarkers Obtained) and a Disposable Exhaled Condensate Collection System (DECCS). DECCS is equipped with a mouthpiece, a non-return valve, a tube, and a collecting cell inserted in an electrical cooling system of the Peltier type (Konstantinidi *et al.*, 2015).

Figure 1. Collection devices and sample collection. (A) EcoScreen® collection device, (B) TurboDECCS collection device, (C) sample collection using TurboDECCS device. The photos were taken in IMTM and Department of Pediatrics, respectively.

Exhaled breath condensate (EBC) collection does not have any influence on lung function or mediator levels, and can be repeated several times with short intervals (minutes) between measurements (Horváth *et al.*, 2005). The human adult is able to provide 1 – 3 ml of condensed breath during 10 minutes of quiet breathing. Most of the condensate consists of water vapor which is 99.9% of the condensate volume and the analytes are very diluted (Konstantinidi *et al.*, 2015). Exhaled breath condensate also contains various volatile and nonvolatile molecules from the lower respiratory tract (Mutlu *et al.*, 2001). This thesis is focused on protein biomarkers in EBC.

The characterization of the EBC proteome is not easy even with very sensitive methods such as mass spectrometry. The low protein concentration is the major problem in the study of

the EBC proteome (Bloemen *et al.*, 2009). One of the first EBC analyses was performed using high-resolution 2-D electrophoresis. The proteins in exhaled breath condensates were detected in subnanogram amounts. The proteins that were spotted are supposed to be a surfactant protein A (SP-A), complement C3 and albumin but no confirmation of the protein identity was done (Griese, Noss and Von Bredow, 2002). Cytokines and growth factors were detected in asthmatic children using enzymatic immunoassays (Shahid *et al.*, 2002; Leung *et al.*, 2005) and multiplex immunoassays (Rosias *et al.*, 2008). The first study using mass spectrometry to analyze the proteome of EBC was done and published by Gianazza *et al.*, 2004. They analyzed individual samples from 25 smokers compared to 21 non-smokers using 1D electrophoresis. Protein spots were detected by 1D and 2-D gel electrophoresis. For mass spectrometry analysis, the samples from each group (smoking and non-smoking) were pooled and analyzed by 2-D electrophoresis (2-DE), excised, in-gel digested and analyzed by HPLC-MS / Q-TOF mass spectrometer, resulting in the identification of human keratins (cytokeratin 9 and cytokeratin 1), while some 2-DE spots remained unidentified. The study was not able to distinguish smokers from non-smokers based on proteins identified by MS. Gel electrophoresis showed that smokers' breath condensate samples are more abundant in proteins compared to nonsmokers (Gianazza *et al.*, 2004). The presence of cytokeratins (CK) was confirmed by another study involving the identification of three cytokeratins: CK-2, 9 and 10 in 30 mechanically ventilated patients and ten healthy volunteers. Gel electrophoretic analysis was performed and specific bands that distinguished patients from controls were evaluated by MALDI-TOF MS (Gessner *et al.*, 2008). Cytokeratin identification was the result of SELDI-TOF MS analysis in 20 patients with pulmonary emphysema (associated with α1 antitrypsin deficiency) and 25 non-smoking volunteers with peaks corresponding to CK-10, 9 and 1. Protein profiles also evidenced some less intensive peaks of interleukin 15, 2, and γ- interferon. No signals corresponding to proteins were identified in the EBC of the individual controls. Some peaks were described (5 ± 1 peak), but no protein was reported, in the pooled controls (Fumagalli *et al.*, 2008). It has been hypothesized that the presence of cytokeratins is caused by mechanical stress and may be a marker of lung injury (Fumagalli *et al.*, 2008; Gessner *et al.*, 2008). In another study in patients with pulmonary emphysema, samples were analyzed by 1D electrophoresis followed by in-gel digestion and HPLC/MS analysis (HPLC/ESI-ION TRAP LCQ-Advantage). This approach, like the previous study, showed poor MS signal quality and the search against human protein database was not successful. Furthermore, protein separation was performed by 2-DE for each individual and for pooled sample consisting of three controls; differential spots were identified by HPLC/MS as cytokeratins. Where 2-DE failed to delineate the complete protein profile,

protein coverage in each sample was improved by using μ-HPLC fractionation. In-solution trypsin digestion of separately collected peaks allowed obtaining additional information on the number of cytokeratins present in each patient's EBC. All analyses performed in this study revealed that cytokeratins were observed in all twenty patient samples, but were not detected in controls unless the controls were pooled from at least three individuals (Fumagalli *et al.*, 2008).

The methods for characterization of EBC proteome are constantly improving and protein biomarkers of lung diseases in the EBC are currently uncovered by number of scientific groups. Efforts to characterize the EBC proteome and suggest biomarkers of various lung diseases are described in detail below. A summary of recently proposed biomarkers in the EBC is provided (Table 1).

Table 1: A summary of proteins identified in the EBC of healthy subjects and patients with various lung diseases published to date.

disease	Proteins	comment	reference
healthy subjects	cytokeratins; calgranulin A and B; hemoglobin; lysozyme C		Bloemen *et al.*, 2009
	cytokeratins: CK 1, 2, 5, 6, 9, 10, 14, 16, 17, 25, keratin type II; dermcidin; prostaglandin H2 D-isomerase; cystatin A; α1-microglobulin/bikunin precursor; ubiquitin		Kurova *et al.*, 2009
	cytokeratins: CK 1, 2, 9, 6B, 5, 16, 14, 17, 10, 13, 18, 23, Hb6, 1b, 3, 4, 72, 78, 79, 80; actin; cystatin; desmoplakin; filaggrin; hornerin; calgranulin; zinc-alpha-2-glycoprotein		Muccilli *et al.*, 2015
	cytokeratins; serpin B3; fatty acid-binding protein; cystatin A; mucin 5B; DMBT1 protein; α-1-antitrypsin; lysozyme; lactoferrin		Lacombe *et al.*, 2018
	albumin; surfactant proteins (including SP-A); various immunoglobulins; serotransferrin; Clara cell protein (CC16); lysozyme C; proteases and inhibitors (including α1-antitrypsin); annexins; complement factors		Bredberg *et al.*, 2012
asthma	cytokeratins: CK 1, 2, 9, 10, 5, 6, 8, 14, 16; albumin; actin; hemoglobin; lysozyme; dermcidin; calgranulin B	not asthma specific, present both in asthma and controls	Bloemen *et al.*, 2011
COPD	type I cytokeratins: CK 9, 14, 26; type II cytokeratins: CK 1, 2, 5, 6B; inflammatory cytokines: IL-1α, IL-1β, IL-2, IL-12 α and β subunits, IL-15, interferon α and γ; tumor necrosis factor α; complement C3	predominant in EBC of patients	Fumagalli *et al.*, 2012
	fibroblast growth factor receptor 4; sushi domain-containing protein; cytoplasmic actin 1; LIM domain containing preferred translocation partner in lipoma isoform CRA_e; tropomodulin-2; spondin-1; tubulin β-chain; glutamine-tRNA ligase; histone H1.0; 60S ribosomal protein L24; heterochromatin protein 1-binding protein 3; filamin-A; cysteine and glycine-rich protein 2; vinexin; transcriptional coactivator YAP1	upregulated in patients	Sun *et al.*, 2019
	protein deglycase DJ-1; non-erythrocytic spectrin α-chain 1; T-complex protein 1 subunit β; tumor suppressor p53-binding protein; moesin; RWD domain-containing protein 4; inositol-3-phosphate synthase 1; heat shock cognate 71 kDa protein; polyubiquitin-C	downregulated in patients	
intubated newborns	keratins type I cytoskeletal: 9, 10, 12, 16, 18, 24; type II cytoskeletal 1, 1b, 2, 5	not specific for any group	Kononikhin *et al.*, 2017
	complement C4-B cluster; cluster of Ig-α-1 chain C region; vitronectin; basement membrane specific heparan sulfate proteoglycan core protein; prostaglandin-H2 D-isomerase; osteopontin; Ig κ-chain C region; fibronectin; clusterin; extracellular matrix protein 1; immunoglobulin J chain; prothrombin; α-2-HS-glycoprotein; cluster of Ig mu chain C region	elevated in pneumonia	
	dermcidin; cystatin A; calmodulin-like protein 5; cluster of ubiquitin-40S ribosomal protein S27a; caspase 14; protein Shroom3; mucin-like protein 1; suprabasin; protein S100-A7; prolactin inducible protein; thioredoxin	elevated in LCDH	
astronauts after the spaceflights	keratins: 1, 2, 5, 6A, 6C, 9, 10, 14; dermcidin	identified in all samples	Kononikhin *et al.*, 2019
	caspase-14; desmoglein-1; desmoplakin; junction plakoglobin; keratin 6B; plakophilin-1; 14-3-3 protein sigma; protein-glutamine gamma-glutamyltransferase K; protein-glutamine gamma-glutamyltransferase E; actin; cytoplasmic 2; tubulin alpha-1C chain; tubulin alpha-4A chain; tubulin beta chain; tubulin beta-8 chain; 14-3-3 protein zeta/delta	typical for astronauts after spaceflights	
lung cancer	mitochondrial pantothenate kinase 2; dihydropyrimidinase-related protein 2; dihydropyrimidinase-related protein 5; probable ATP-dependent RNA helicase DDX20; septin 7; mitochondrial ATPase inhibitor; TBC1 domain family member 1; TBC1 domain family member 4; nuclear ubiquitous casein and cyclin-dependent kinase substrate 1; POTE ankyrin domain family member E; serine/arginine-rich splicing factors: 1, 3, 4, 6; WD-repeat-containing protein 13; synapsin-1; protein Spindly; BSD domain-containing protein 1; heat shock protein HSP 90-α	increased in cancer patients	Fedorchenko *et al.*, 2016
	cytokeratins 6A, 6B, 6C; hemoglobin subunit β; histones: H2A, H2B, H4	increased in cancer patients	López-Sánchez *et al.*, 2017
	epidermal growth factor	increased in NSCLC patients	Chen *et al.*, 2019

1.3.3.1. Improvement of sample preparation and MS methods in groups of healthy subjects

One of the improvements brought an approach with a gel-free method followed by MALDI-TOF MS. This study compared EBC from six healthy smokers and non-smokers. Proteomic MS analysis identified the more abundant proteins in smokers as cytokeratins, which were discussed as a sign of airway epithelial damage. In addition to cytokeratins, calgranulin A and B, hemoglobin and lysozyme C were also identified. Calgranulin B was specific marker of smokers (Bloemen *et al.*, 2009). Similarly to the work of Gianazza *et al.*, 2004, all identified proteins were more abundant in smokers compared to non-smokers. Compared to the most abundant protein, cytokeratin 10, the calgranulin B concentration was approximately 20 times lower, suggesting an improvement in the sensitivity of the used gel-free method (Bloemen *et al.*, 2009). Cytokeratins and dermcidin were identified in another study using the gel-free method followed by HPLC-MS/MS with a 7-Tesla Finnigan LTQ-FT mass spectrometer. This study evaluated a breath condensate from 17 young healthy non-smoking donors. The MS revealed that the major identified proteins are cellular keratins. In addition to keratins, at least 30% of condensate donors experienced dermcidin, prostaglandin H2 D-isomerase, α_1-microglobulin/bikunin precursor, ubiquitin and cystatin A. Proteome EBC samples obtained with and without a dust filter from three individuals were compared and the results were similar. Room air was collected and condensed for 5 h and 20 h and this analysis identified the contaminants which should be excluded (Kurova *et al.*, 2009). Another research group (Muccilli *et al.*, 2015) analyzed pooled EBC samples from nine healthy subjects. For relative quantification of the protein composition of the EBC, they used a high-resolution HPLC tandem mass spectrometry (Orbitrap Elite) method coupled with electrophoretic separation. This approach led to the identification of 167 unique gene products. Cytokeratins were again the most abundant proteins in EBC samples (Muccilli *et al.*, 2015). In general, the cytokeratin profile was very similar to that described above (Kurova *et al.*, 2009). In addition to keratins, some of the proteins including actin, cystatin, desmoplakin, filaggrin, hornerin, calgranulin, and zinc-alpha-2-glycoprotein are located at the epithelial lining level (Muccilli *et al.*, 2015). In another study (Lacombe *et al.*, 2018), condensates were collected from 20 healthy subjects, and two pooled samples from 10 individuals were analyzed. The pooled EBC samples were lyophilized, in-gel digested and analyzed by nanoLC-MS/MS LTQ-Orbitrap Velos Pro. A total of 229 unique proteins were identified in the two pooled EBC samples, 153 proteins were common to both pools. They also identified keratin major components of EBC and 10 cytokeratins were reliably identified. Moreover, they identified 49 proteins expressed only in the respiratory tract, some are mainly expressed in the upper respiratory tract such as serpin B3,

others are more abundant in the lower pulmonary airways such as the fatty acid-binding protein, which is strongly expressed in the lung macrophages. Proteins such as cystatin A are expressed throughout the respiratory tract. Proteins present in airway mucus have been identified including mucin 5B, deleted in malignant brain tumors 1 (DMBT1) protein, α-1-antitrypsin, lysozyme and lactoferrin, which are the two most abundant antibacterial proteins, together with plenty of immune system proteins (Lacombe *et al.*, 2018). In comparison with other studies employing pooled samples of healthy subjects, Lacombe's group reported 27 proteins common in all 3 studies (Bredberg *et al.*, 2012; Muccilli *et al.*, 2015; Lacombe *et al.*, 2018), 50 proteins common in Muccilli *et al.* study, 9 proteins common in Bredberg *et al.* study, and 59 unique proteins (Lacombe *et al.*, 2018). Bredberg's group collected not directly EBC but endogenous particles present in exhaled air. The individuals exhaled into a sampling device that uses a three-stage impactor to collect endogenous particles. Samples collected from 6 healthy subjects in the first session and 10 healthy subjects in the second session were analyzed. Samples were analyzed by 1D SDS-PAGE, digested and measured by high-resolution HPLC/MS-LTQ-FT-ICR (linear quadrupole ion trap - Fourier transform ion cyclotron resonance). 124 proteins were identified in 2 pooled samples (32 proteins in the first sample, 116 in the second sample, 24 proteins common in both samples). There were no keratins and the most abundant proteins were: albumin, surfactant proteins (including SP-A), various immunoglobulins, serotransferrin, Clara cell protein (CC16), lysozyme C, proteases and inhibitors (including α_1-antitrypsin), annexins and complement factors (Bredberg *et al.*, 2012). Albumin and SP-A protein were confirmed in another study by ELISA (Larsson *et al.*, 2012).

1.3.3.2. Exhaled breath condensate analysis in patients with asthma, COPD and other specific groups such as mechanically ventilated newborns and cosmonauts after long-term spaceflights.

Another milestone in the protein breath condensate was achieved by Bloemen *et al.* who analyzed 40 children with asthma and compared them to 30 healthy controls. EBC proteins were concentrated on reverse-phase POROS^{TM} R2 beads, followed by tryptic cleavage into peptides and analyzed by MALDI TOF/TOF MS. The analysis was performed for individual samples in the study and, to identify as many proteins as possible additional pooled samples from healthy volunteers were analyzed using this approach. In order to focus on less abundant proteins, present in the samples, exclusion lists containing peptide masses of more abundant proteins have been used for which two positive matches have already been found. Additionally, inclusion lists containing the peptide masses selected in the pattern were used and the settings

were changed for different samples to obtain more identifications. Spectra from all samples were compared to the positive matches, to identify proteins based on the presence of masses and retention times. They tried to construct a model that could differentiate asthma patients and healthy controls based on proteolytic peptide pattern. By applying proteomics combined with support vector machine (SVM) statistical analysis, they were able to select a signature based on 10 peptide masses, making it possible to distinguish between partially or uncontrolled asthmatic patients and healthy controls. Moreover, a model based on 11 peptide masses was able to predict whether the overall pattern of abundance of these peptides more closely resembled the typical pattern observed in the control group or in the controlled patient group. Based on 14 peptides a 98% correct classification between all asthma patients and healthy controls was obtained. An attempt to distinguish between the five groups (suspicious asthma, controlled asthma, moderate asthma, uncontrolled asthma, and healthy individuals) using one model yielded no useful results, probably due to the inclusion of five different groups, which were all rather small. Regarding protein identification, the cytokeratins were found in most individual samples. Additionally, cytokeratins and some less abundant proteins, such as albumin, actin, hemoglobin, lysozyme, dermcidin, and calgranulin B, were identified in the pooled samples. In individual samples, corresponding peptides were present. Identification of these proteins was based, respectively, on a maximum 8, 3, 3, 3, 2, and 2 different peptides. However, none of these peptides could discriminate between the asthma group and controls (Bloemen *et al.*, 2011).

Concerning COPD patients, pooled samples of controls, chronic obstructive pulmonary disease (COPD) patients without emphysema and patients with pulmonary emphysema associated with α_1-antitrypsin deficiency (AATD) were analyzed by gel-free approach. Samples were vacuum dried and prepared for HPLC-MS/MS measurement (LTQ mass spectrometer and LTQ Orbitrap mass spectrometer). All of the 44 proteins that were identified were found in the pool of healthy controls, 17 of them in COPD patients and 15 of them in AATD subjects (Fumagalli *et al.*, 2012). The analysis with individual COPD patients (not samples pools) included 19 COPD patients and 19 healthy controls by matching with age, sex and smoking history was performed using tandem mass tags (TMT 6-plex) quantitative mass spectrometry (nano-UPLC-MS/MS- Orbitrap Q Exactive). This approach was used for the first time in the analysis of individual EBC samples. A total of 257 proteins were identified and 24 proteins were differentially expressed in COPD patients (Sun *et al.*, 2019). These two studies differ in reported proteins. In pooled samples, the most abundant proteins found in LC-MS were some type I and type II cytokeratins. Cytokeratins have been found mainly in EBC of healthy

individuals, contradictory to the hypothesis that cytokeratins are markers of bronchial epithelial damage. Several inflammatory cytokines and complement C3 were detected as predominant in EBC of patients. Lysozyme C was discussed as detected only in AATD with emphysema, but according to the results, it was identified in controls as well. The study included proteins that have been previously reported such as cytokeratins, cystatin A and surfactant proteins (Fumagalli *et al.*, 2012). The second study (Sun *et al.*, 2019) reported different proteins than those identified in previous studies, but these proteins also played a predominant role in inflammatory response. Of the 257 proteins identified in individual samples, 15 were upregulated and 9 downregulated. The pathway analysis showed that most of the annotated pathways had been acknowledged in COPD including MAPKs signaling pathway, phagosome, Ras signaling pathway, PI3K-Akt signaling pathway, leukocyte migration, nuclear damage and apoptosis that is in concordance with COPD mechanism. Six proteins (upregulated vinexin, filamin-A, fibroblast growth factor receptor 4; and downregulated, non-erythrocytic spectrin α-chain 1, heat shock cognate 71 kDa protein, polyubiquitin-C) were annotated as participants of MAPK signaling pathway so they could be novel biomarkers of COPD. Similarly, the identified cytoskeleton-associated proteins were likely to be biomarkers of COPD (Sun *et al.*, 2019).

The EBC of intubated newborns (Kononikhin *et al.*, 2017) was directly used for in-solution sample preparation for HPLC analysis coupled with 7-T LTQ-FT Ultra high-resolution mass spectrometer. EBC samples were collected from 24 mechanically ventilated intubated newborns, 12 with congenital pneumonia and 12 infants with a left-sided congenital diaphragmatic hernia (LCDH) used as a control group. Both proteomic and metabolomics analyses were performed. Using label-free semi-quantitative approach, 119 proteins in 64 clusters in EBC of intubated preterm newborns were identified. Invariant part of EBC proteins belonged to different types of keratins, but the cytokeratin putative identification was not specific for infant's group studied. Some non-keratin proteins were found as markers significantly elevated in pneumonia and these proteins included mainly immune system related proteins. For LCDH group the significantly elevated markers were dermcidin, cystatin A, calmodulin-like protein 5, cluster of ubiquitin-40S ribosomal protein S27a, caspase 14, protein Shroom3, mucin-like protein 1, suprabasin, protein S100-A7, prolactin inducible protein and thioredoxin. They concluded that a single biomarker could not reflect the pathology of premature inborn. They suggested to use the combination of the suggested biomarkers, which should be validated on a larger cohort (Kononikhin *et al.*, 2017). Another recent and exciting study compared the EBC of five Russian astronauts taken one month before the spaceflight, immediately upon landing after 169-199 days' lasting spaceflight (R0) and seventh day after

landing (R+7). The HPLC-MS- TiMS TOF analysis revealed 164 different proteins, 150 of them were unique for R0 time point, 10 were unique for the background samples (one month before the spaceflight), 4 were unique for the R+7 time point and 9 proteins (8 cytokeratins and dermcidin) were common for all collections. All of these proteins are typical for EBC, despite the condition of the pulmonary system. However, CK2 increased and CK9 decreased after spaceflight. Pathways enrichment analysis showed that most proteins detected after spaceflight play a role in innate immune processes and neutrophil degranulation; some of the proteins are part of keratinization. Enrichment analysis using KEGG database revealed significant participation of detected proteins in pathogenic *E. coli* infection. Another sign of infection was detection of dermcidin, CK1 and CK6A which are involved in the immune response. Microbial contamination was confirmed by microbiological cultivation assays that revealed the highest microbial contamination by *E. coli* and *S. aureus* during spaceflight. During the rehabilitation, the quantitative and species microflora was normalized as well as protein composition (Kononikhin *et al.*, 2019).

1.3.3.3. Lung cancer patients' analysis

Three scientific groups (Fedorchenko *et al.*, 2016; López-Sánchez *et al.*, 2017; Núñez-Naveira, Mariñas-Pardo and Montero-Martínez, 2019) published the proteomic analysis of lung cancer biomarkers in the EBC. The first study on early diagnosis of lung cancer by EBC was a comparative study of 4 donor groups: 26 subjects diagnosed with lung cancer (9 patients in first/second stage, 17 patients in third/fourth stages), 17 patients with chronic obstructive pulmonary disease (COPD), 13 patients with community-acquired pneumonia and 23 healthy non-smokers. The samples were processed by the gel-free method and analyzed by high-resolution HPLC MS/MS – ICR mass spectrometry. Of the more than 300 identified proteins, 21 proteins turned out to be common to all donor groups, and 19 of these have been found in EBC of patients with early-stage of lung cancer and are potentially important as biomarkers (Fedorchenko *et al.*, 2016). In the second study concerning early diagnosis of lung cancer, a higher number of samples was used, thereby increasing the coverage of the EBC proteome. EBC was collected from 192 individuals: 49 controls, 49 smokers, 46 COPD and 48 lung cancer patients. Using nanoLC in conjunction with QqTOF tandem mass spectrometry, 348 different proteins showed a different pattern among the groups. Importantly, more proteins (approximately 2-fold higher protein counts) were identified from lung cancer patients in EBC compared to other groups. Most (146) of the total 348 proteins were unique in the lung cancer group (López-Sánchez *et al.*, 2017). In the third, most recent work, a 100 peptides and 18

proteins were detected while EBC was collected from 30 healthy volunteers (15 smokers and 15 non-smokers), processed by gel-free approach and HPLC/MS-MALDI TOF/TOF (Núñez-Naveira, Mariñas-Pardo and Montero-Martínez, 2019).

The main components of the EBC proteome were keratins. Several isoforms of cytokeratin 6 (6A, 6B and 6C) have the potential to be biomarkers of lung cancer. Quantification of most keratins (particularly CK1, 6C, 9 and 10) showed significant positive correlation with tumor size. When a prediction model was constructed to diagnose lung cancer, it turned out that the β hemoglobin subunit was clearly the most important discriminative variable followed by keratin type I cytoskeletal 16 (López-Sánchez *et al.*, 2017). The highest incidence was found for dermcidin, a peptide with proteolytic and antimicrobial activities (Fedorchenko *et al.*, 2016; López-Sánchez *et al.*, 2017) and hornerin (López-Sánchez *et al.*, 2017). No correlation with tumor size was observed with these proteins, despite the fact in the case of dermcidin a higher number of peptides of this protein was identified in patients with lung cancer. Hornerin (protein S100A2) was more frequently detected in EBC from controls than in patients; its incidence is associated with the worst prognosis, shorter survival and high risk of metastasis and is downregulated in most cancers (López-Sánchez *et al.*, 2017). In a study by Núñez-Naveira *et al.*, the candidate proteins for lung cancer biomarkers were selected from identified proteins according to literature. Dermcidin, protein S100A9, and cathepsin G were selected and quantified by ELISA in EBC samples of 9 lung cancer patients and nine healthy non-smokers. Dermcidin and S100A9 expressions were significantly higher in lung cancer than in healthy volunteers. Cathepsin G was barely detectable with statistically significant differences between lung cancer patients and control group (Núñez-Naveira, Mariñas-Pardo and Montero-Martínez, 2019).

Proteins commonly identified in lung cancer patients were hemoglobin subunit β and several isoforms of histones: H2A, H2B, H4 (López-Sánchez *et al.*, 2017). In the EBC of patients with lung cancer in the first/second stage, 42 proteins of a non-keratin origin were identified, which were absent in the EBC of healthy non-smoking control group, as well as in the EBC of the patients with COPD and community-acquired pneumonia. Of these candidates, 19 proteins were selected as a diagnostic panel for lung cancer in EBC. Expression of all represented proteins is increased in cancer patients (Fedorchenko *et al.*, 2016). Among these proteins, the following ones were represented in both Fedeorchenko and López-Sánchez lung cancer studies. The heat shock protein HSP 90-α was identified only in lung cancer patients and POTE ankyrin domain family members (E, F, I, J) were identified in controls and lung cancer but disappeared in COPD. The number of peptides identified in the POTE ankyrin domain

family member E increased in lung cancer (López-Sánchez *et al.*, 2017). In addition to the 19 potential lung cancer biomarkers, the HMG-I/Y protein family and lactoferrin in the EBC of lung cancer patients can be noted; this may be associated to capillary network overgrowth induced by tumors and to immune response to the activity of cancer cells (Fedorchenko *et al.*, 2016). Epidermal growth factor (EGF), a known biomarker of lung cancer was detected with 80 % sensitivity and 89.6 % specificity by ELISA in the EBC of non-small lung cancer (NSCLC) patients. EGF levels were higher in patients than in controls, EBC quantity was higher in the group of smokers, and in later stages III and IV. Patients with better survival exhibited lower EGF levels in EBC (Chen *et al.*, 2019).

1.3.3.4. Keratin and saliva proteins contamination

Despite the fact that most previous studies have claimed that there is a significant difference between patients and the control group, there is still some controversy about keratins (López-Sánchez *et al.*, 2017).

The first mass spectrometry-based proteomic studies claimed that the studies were designed to exclude external keratin contamination, and therefore identified keratins are associated with mechanical stress during ventilation and are increased in lung cancer EBC (Gianazza *et al.*, 2004; Fumagalli *et al.*, 2008; Gessner *et al.*, 2008). One of the later studies included two technical control samples (containing distilled water only) as blank samples to check possible protein contamination during pre-analytical procedure. Identified keratins were absent in technical controls and claimed as a valid component of EBC (Lacombe *et al.*, 2018). Keratins were described in bronchial and alveolar epithelial cells but not in alveolar septa, capillaries and other non-epithelial elements (Blobel *et al.*, 1984; Gessner *et al.*, 2008). Several isoforms of cytokeratin 6 have the potential to be biomarkers of lung cancer and the most keratins in EBC samples showed a significant positive correlation with tumor size (López-Sánchez *et al.*, 2017).

By contrast, Fedorchenko *et al.* did not report any keratins as biomarkers and claimed that some cytoskeletal type I and type II keratins, were invariant for all the samples. According to human protein catalog, cytokeratins CK 1, 2, 9 and 10 are of epidermal origin and have been considered to be exogenous keratins from the air (Fedorchenko *et al.*, 2016). However, the same group in the following work concerning effect of spaceflights on EBC of cosmonauts reported significant changes in keratin quantity and admitted that the keratins could be a real part of the EBC and are involved in cellular signaling (Kononikhin *et al.*, 2019). Núñez-Naveira´s group excluded keratins or salivary and submucosal cells related proteins from the study and

considered them to be contamination (Núñez-Naveira, Mariñas-Pardo and Montero-Martínez, 2019).

In some studies, α-amylase was measured to check saliva contamination, and most of the studies reported zero amylase activity and no saliva contamination (Gianazza *et al.*, 2004; Fumagalli *et al.*, 2008, 2012; Gessner *et al.*, 2008; Bloemen *et al.*, 2009; Kononikhin *et al.*, 2017). In the first EBC study with just the electrophoretic detection of the proteins, saliva contamination of the samples was excluded, however, about two-thirds of the spots detected in EBC were also present in the matched saliva samples (Griese, Noss and Von Bredow, 2002). In one study (Bloemen *et al.*, 2011) the samples with proven saliva contamination were excluded. Lacombe's group compared the EBC protein profile with the salivary proteome described in previous studies (Lacombe *et al.*, 2018).

1.3.3.5. Sample concentration methods

As the protein level in the exhaled breath condensate is low, the preconcentration methods and their effect on the quality of mass spectrometry measurements have been the subject of some studies. Protein precipitation by trichloroacetic acid, pyrogallol red, and precipitation on POROS-20-R2-RP beads, ultrafiltration, ultracentrifugation, lyophilization and vacuum concentration methods were studied. The precipitation by pyrogallol red was proven to be a reliable method with high recovery, but this dye could not be removed upfront of MS protein analysis. Therefore, a protein capture on beads was alternatively chosen. After ultrafiltration, most proteins were irreversibly lost on the membrane and it was noted that vacuum concentration might result in protein instability, while freeze-drying was found to be highly irreproducible (Bloemen *et al.*, 2009). In contrast to these findings, the freeze-drying was chosen as the gentlest and rapid procedure among the above-mentioned protein concentration methods by another study (Kurova *et al.*, 2009). According to further EBC study measured by the MALDI TOF/TOF, neither ultracentrifugation nor freeze-drying was suitable, and no proteins were detected. The reversed-phase chromatography was chosen as an effective method (Núñez-Naveira, Mariñas-Pardo and Montero-Martínez, 2019). In this case, however, suppression of the signal by the present ions can play an important role.

1.4. Identification of protein biomarkers of the cellular response to treatment

1.4.1. Introduction to identification of protein biomarkers of the cellular response to treatment

Biomarkers of drug efficacy and toxicity are one of the key steps in drug development. Protein biomarkers provide predictive information of the response to treatment and could be used in monitoring the effects of drug treatments (Hale *et al.*, 2003).

Many efforts lead to evaluate the biomarkers of drug toxicity and biomarkers which help to uncover the molecular mechanism of action. Toxicity caused by toxic substances including drugs induces significant pathological responses. Identification of toxicity biomarkers, potential drug targets and characterization of toxicity mechanisms is an important step towards drug safety. Drug safety is a major point of concern during the development phase and clinical application. The methods of quantitative proteomics were successfully used for the identification of the toxicity biomarkers (Gao, Holland and Yu, 2009). Each anticancer treatment induces stress to the cells. The cellular response to this stress is a key determinant of drug activity. It has been shown that the checkpoint proteins which drive the cell progression into the next step of the cell cycle plays the important role in this response. The checkpoint proteins are tightly connected to the activation of any repair mechanisms or with triggering cell death. The understanding of these mechanisms could lead to the increased activity and specificity of anticancer agents (Damia and Broggini, 2004).

In this thesis, the identification of protein biomarkers of the response to treatment includes cooperation in searching for specific biomarkers in proteomic profiling and drug repurposing studies. For such studies, the biomarkers are searched for in the biological matrices treated by a drug that is already routinely used in clinical practice. In proteomic profiling study, cellular response to an antitumor treatment was evaluated. In the second case, the biomarkers of the treatment response were searched for in the cancerous biological matrices treated by a drug which was previously used for the treatment of completely different disease.

1.4.2. Proteomic profiling of the cellular response to antitumor treatment

Proteomic profiling is increasingly being used for molecular stratification of cancer patients and cell-line panels (Ali *et al.*, 2018). Mass spectrometry is a widely used tool for protein profiling studies in both cell lines and patients' samples. Despite the fact that the uncovering of global cellular response to already approved antineoplastic agents by modern high-resolution proteomic approaches should be highly beneficial, only a few publications focused on this. Here, the summary includes the experimental efforts which utilize mass spectrometry for proteomic profiling and were published over the last decade.

The proteomic profile of two breast cancer cell lines treated with the chemotherapeutic agent doxorubicin followed by the death receptor ligand TRAIL was measured using iTRAQ labeling coupled with multidimensional LC−MS/MS-QqTOF of the enriched mitochondria and endoplasmic reticulum fraction. Peptidyl-prolyl cis−trans isomerase B, neuroblast differentiation-associated protein, solute carrier family 1 member 5 were detected as biomarkers for assessing response to doxorubicin and TRAIL cancer treatment in both cancer cell lines. Upregulation of peptidyl-prolyl cis−trans isomerase B and downregulation of neuroblast differentiation-associated protein and solute carrier family 1 member 5 was confirmed by immunofluorescence followed by confocal microscopy visualization (Leong *et al.*, 2012). The response of human multiple myeloma cell line to bortezomib was evaluated by the combination of two methods for studying cancer cells, deep sequencing of RNA and iTRAQ LC-MS/MS quantitative proteomics. Each iTRAQ sample was then analyzed using two separate mass spectrometers, QqTOF and Orbitrap. Another step was the SRM assay for iTRAQ validation. Concerning deep sequencing methods, ribosome profiling and mRNAseq were performed; deep sequencing was performed on an Illumina HiSeq 2000. Bortezomib is a proteasome inhibitor and the mentioned approaches were used to globally examine the dynamic interplay between transcription, translation, and proteolytic degradation during bortezomib-induced apoptosis. Ribosome profiling revealed preferential translation and translational regulation of genes expected to reduce unfolded protein stress after bortezomib treatment. Despite these changes in translational control, only a few critical anti-apoptotic proteins were detectably increased in proteomic studies against a background of general translation inhibition. It was shown that degradation by caspases is largely independent of upstream bortezomib effects. Moreover, previously uncharacterized non-caspase proteolytic events also participate in cellular deconstruction. There is evidence that the combination of bortezomib with anti-apoptotic regulator HSF1 (transcription factor regulating the heat shock response) inhibition could be beneficial to promote cell death in myeloma therapy (Wiita *et al.*, 2013). A SILAC based

quantitative gel-based proteomic approach followed by LC-MS/MS-Orbitrap identification was used to evaluate global changes of cellular protein abundance in HeLa cells after paclitaxel treatment. Validation methods such as gene silencing by shRNA, immunoprecipitation, PCR methods and Western blotting were used. The MS method led to the identification of 347 proteins whose levels changed in response to paclitaxel treatment. Among them, it was confirmed that paclitaxel downregulates tumor suppressor programmed cell death 4 (PDCD4) and upregulates ubiquitin K11 linkage conjugating enzyme critical for mitotic exit (UBE2S). It was shown that PDCD4 is a cell-cycle regulated protein and that changes in its abundance are sufficient to alter sensitivity to paclitaxel in multiple human cancer cell lines. High levels of PDCD4 in lung cancer tissues are positively correlated with the longer overall survival time of the examined lung cancer patients. Tumor suppressor PDCD4 was also offered as a predictive biomarker for paclitaxel-based personalized chemotherapy in the treatment of lung cancer (Xu *et al.*, 2015). Proteomic profiling was also used for the discovery of etoposide chemoresistance markers in non-small cell lung carcinoma. Electrophoretic methods 2-DE and DIGE were coupled with MALDI-TOF/TOF and iTRAQ based quantitative proteome analysis and used to explore the differential proteome in etoposide sensitive and resistant lung cancer cell lines. The potential markers of chemoresistance were validated by RT PCR, Western blotting and MRM. The MS analysis results in the identification of 83 differentially expressed proteins, which are involved in various cellular processes like cytoskeleton organization, protein folding, metabolic regulation, and apoptosis. Of these candidates, 12 proteins were confirmed as markers of chemoresistance (Paul *et al.*, 2016). Two proteomic technologies, targeted reverse-phase protein array (RPPA) and global mass spectrometry (MS), were tested for their accuracy in predicting the sensitivity of cancer cells from NCI-60 cell lines panel to both cytotoxic chemotherapeutics and molecularly targeted anticancer compounds. These approaches were valuable for understanding and predicting drug sensitivities in cancer patients. The protein profiles obtained by MS and RPPA methods were complementary for the response prediction. Global MS data revealed plethora of proteins for drug response prediction but there is a common problem with missing values. When the number of proteins dropped dramatically from 8113 (original MS data) to 42 (COSMIC cancer census genes with no missing values), it resulted in maximal predictive accuracy for both cytotoxic and targeted compounds (Ali *et al.*, 2018). Using nanoUPLC-ESI-Q-TOF MS/MS, the significant markers from high-grade osteosarcoma in patients after cryotherapy and irradiation were revealed. This comparative proteomic analysis led to the identification of 131 differentially expressed proteins in high-grade osteogenic sarcomas treated versus untreated. Among these, 91 proteins were upregulated

and 40 proteins were downregulated in untreated/treated samples. The functional enrichment analysis revealed that the identified differentially expressed proteins have belonged to more than 10 different protein categories include cytoskeletal, extracellular matrix, immune, enzyme modulators, and cell signaling molecules. Extracellular matrix proteins significantly contributed to cancer progression and were significantly downregulated after freezing and irradiation treatments of high-grade osteosarcoma. Among them, fibronectin and protein S100 A4 were selected. These were two candidates that were upregulated with more than two folds in the untreated group and the MS results were confirmed using Western blot analysis (Madda *et al.*, 2020).

The study present in this thesis was focused on the proteomic profile of chemosensitive CCRF-CEM cell line to oxaliplatin treatment by HPLC-MS/MS-Orbitrap method followed by multiple validation steps (Ozdian *et al.*, 2017).

1.4.3. Drug repurposing

A drug repurposing is a strategy for identifying new uses for approved or investigational drugs (Pushpakom *et al.*, 2018). It is a useful strategy in drug research and development because these drugs have already been approved and tested in humans, and so detailed information is available on their pharmacology, formulation, dosing and potential toxicity (Collins, 2011). Drug repurposing is attractive mainly because the compounds are already de-risked and because of potentially lower overall development costs and shorter development time (Pushpakom *et al.*, 2018). Although many clinical properties have already been established, approved drugs are needed to be evaluated for novel therapeutic indications through the specification of the right dose for new application and through the discovery of biologically and clinically relevant affinities for new targets. Potential candidates for drug repurposing could be generic drugs that were already approved and are no longer protected by patent, then those drugs that have been through some stage of clinical development but are currently not on the market or eventually patented drugs, either approved or in the late stage of clinical development. Several approaches were used for drug repurposing including *in silico* modeling, searching for hits of the approved drugs against new diseases in cells using high-throughput screening platforms, as well as –omic methods such as transcriptomics, genomics, epigenetics and proteomics, both alone or in combination to multi –omics level (Cha *et al.*, 2018). These days, the role of drug repurposing is increasing because many drugs including antineoplastic agents are tested for repurposing against SARS-CoV-2 infection (Ciliberto, Mancini and Paggi, 2020).

Concerning drug repurposing to gain new anticancer agents, one of the possibilities is to test psychiatric drugs for the anticancer effect. The long history of use makes the psychiatric drugs ideal candidates for repurposing. For example, valproic acid is primarily used for bipolar disorder, epilepsy, and migraine headaches and has also been identified as a promising anti-neoplastic drug based on histone deacetylase inhibition. It is active against many malignancies including the advanced ones such as breast, rectal and cervical cancer, recurrent and metastatic nasopharyngeal carcinoma, gliomas or acute myeloid leukemia. However, the use of these drugs is accompanied by many side effects such as weight gain and associated metabolic disorders (Huang *et al.*, 2018). Recent research has shown a role of proteomics in drug repurposing studies. Using shotgun proteomics, the mechanisms of nitroxoline anticancer action were discovered. Nitroxoline is an antibiotic that is widely used in several countries from the 1960s and is able to chelate divalent metal ions such as Mg^{2+} and Mn^{2+}, an activity which is considered as a possible mechanism for its antibacterial action. It was found out that nitroxoline is highly effective against pancreatic cancer in several cellular models *in vitro* and

animal cancer models. The shotgun proteomics experiments identified proteins consistently deregulated by nitroxoline. Additionally, it was confirmed that nitroxoline induces downregulation of Na/K-ATPase pump and β-catenin, which associated with drastic impairment in cell growth, migration, invasion, increased ROS production and induction of DNA damage response. As well, nitroxoline induced previously unknown deregulation of molecules with a critical role in cell bioenergetics, which resulted in mitochondrial depolarization. It was suggested that the deregulation of cytosolic iron homeostasis and of co-translational targeting to membrane contribute to nitroxoline anticancer action (Veschi *et al.*, 2020). The antitumor activity of nitroxoline was reported before using methods such as flow cytometry, immunohistochemistry, Western blot, cell invasion assay, mouse models and magnetic resonance imaging. It was reported that nitroxoline induced apoptosis and inhibited the proliferation and invasion of glioblastoma cells both *in vivo* and *in vitro* (Lazovic *et al.*, 2015). A good example of drug repurposing is disulfiram (Antabuse) which is an old drug and is used as aversion therapy for alcoholics up to date. In the human body, disulfiram is rapidly metabolized to dithiocarb (diethyldithiocarbamate) which is a strong copper chelator and reacts with copper forming Cu(II)diethyldithiocarbamate complex (CuET). There was some evidence that disulfiram could be a new nonprofit drug for cancer (Cvek, 2012). Anti-cancer effects of disulfiram were reported in various malignant tumors. The first evidence was that sodium dithiocarb improves disease-free survival of patients with non-metastatic high-risk breast cancer (Dufour *et al.*, 1993). Breast cancer is the first and the most studied use of disulfiram for anticancer treatment. In breast cancer cell cultures and xenografts, it was shown that disulfiram induces apoptotic cell death and that the mechanism of action is the inhibition of proteasome activity (Chen *et al.*, 2006). Antitumor activity was demonstrated in lymphoid malignant cells *in vitro* and *in vivo* via ROS-related activation of the JNK pathway and inhibition of NF-κB and Nrf2 (Zha *et al.*, 2014). Disulfiram inhibits growth and induces apoptosis of prostate cancer cells where the growth inhibition is absolutely dependent on copper and ROS act as a mediator of the copper-dependent cytotoxic effects of disulfiram (Safi *et al.*, 2014). The copper-dependent anticancer activity was confirmed also in gliomas by inducing apoptosis and anti-angiogenic activity through the EGFR/Src/VEGF pathway (Li *et al.*, 2015). Anti-cancer effects in head and neck squamous cell carcinoma via autophagic cell death were confirmed *in vitro* as well as in xenograft animal model (Park *et al.*, 2018). Disulfiram was successfully tested in patients with newly diagnosed metastatic non-small cell lung cancer. The addition of disulfiram to combination treatment with cisplatin and vinorelbine was well tolerated and appeared to prolong survival (Nechushtan *et al.*, 2015). Additionally, enzymatic, cell-based, mass

spectrometric and mutagenesis experiments have shown that inhibition of phosphoglycerate dehydrogenase is one of the mechanisms of action of disulfiram (Spillier *et al.*, 2019). In the present study, it was shown that disulfiram acts as an anticancer drug towards breast cancer and its mechanism of action is executed via p97 segregase adaptor NPL4 (Skrott *et al.*, 2017).

2 EXPERIMENTAL PART

2.1. Aims

This thesis aimed to employ proteomic methods such as mass spectrometry, electrophoresis, Western blot and drug affinity responsive target stability (DARTS) to identify or validate biomarkers recently identified at IMTM. The exhaled breath condensate is a promising source of biomarkers that is collected noninvasively. Thus, it was chosen as a biological matrix where the biomarkers of individual physiological conditions and various respiratory diseases can be found. The biomarker validation is a subsequent step that should follow the proteomic profiling studies. Validation studies are performed independently on previous research and are done to confirm or displace potential biomarkers.

1. Exhale breath condensate (EBC) analysis
    a) Describe an approach to identify as many proteins as possible in the EBC.
    b) Evaluate the reproducibility of mass spectrometric measurements when introducing biological and technical replicates, comparing the protein profile between individual samples of healthy subjects, and describing the reproducibility between three technical replicates of a single sample.
    c) Evaluate healthy individuals' exhaled breath condensate to predict age according to proteomic signatures and describe differences in proteomic profile according to sex and smoking status.
2. Validation of the biomarkers of the response to treatment.
    a) Use Western blot to validate the potential biomarkers of the cellular response to oxaliplatin in T-lymphoblastic leukemia cells (CCRF-CEM).
    b) Use the DARTS method to validate that the Nuclear protein localization protein 4 homolog (NPL4) protein is affected by disulfiram in response to breast cancer treatment.

2.2. Exhaled breath condensate analysis

2.2.1. Introduction to exhaled breath condensate analysis

The exhaled breath is a promising source of biomarkers collected noninvasively and could be used to diagnose various respiratory diseases. As is described in the introduction in part **1.3.3. "Exhaled breath condensate analysis"**, there were many efforts to characterize EBC and to reveal the biomarkers of some respiratory diseases and condition of human lungs. We described a gel-free mass spectrometry-based approach which showed in-depth characterization of the EBC proteome.

2.2.2. Materials and methods

2.2.2.1. EBC collection and study cohorts

Exhaled breath condensates were collected using EcoScreen® collection device (Erich Jaeger GmbH, Hoechberg, Germany) or using Turbo 14 Turbo DECCS System (Medivac, Italy) respectively. The EBC was being collected for 10 min without wearing the nose clip. After collection, samples were immediately frozen and stored at -80°C until analysis.

In the first study, samples were taken from 3 individuals from one family. These individuals included two sisters, ages 8 and 11, and their father, aged 45. These samples were collected using Turbo 14 Turbo DECCS System (Medivac, Italy).

The second study included 248 healthy controls, of them 61 were pediatric healthy individuals (samples were collected using EcoScreen® collection device) and 187 adult healthy individuals (samples were collected using Turbo 14 Turbo DECCS System). The participants were classified according to age, sex and smoking status.

2.2.2.2. Sample preparation

The exhaled breath condensate samples were freeze-dried using a SpeedVac (Thermo Fisher Scientific). Proteins were solubilized in the TEZ buffer (10 mM Tris, pH 8.0; 1 mM EDTA; 0.002% Zwittergent 3-16), cleared at 14,000 g, at 20 °C for 10 min and the supernatants were transferred to 1.5 ml Eppendorf tubes. Heat-mediated antigen retrieval was performed at 98 °C for 90 min. Proteins were denatured by sonication in a water bath for 30 minutes at room temperature and digested with trypsin at 37 °C overnight. Trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega) was added at a final concentration of 0.5 µg per sample. Proteins were reduced with dithiothreitol (10 µM final concentration) and peptides were purified using Stage tips (Rappsilber, Ishihama and Mann, 2003). The purified samples were concentrated in a vacuum concentrator (Eppendorf) and dissolved in 50 µL solution of 1% acetonitrile, 0.05% trifluoroacetic acid (TFA).

2.2.2.3. Mass spectrometric (MS) analysis

In the first study, MS analysis was performed in 3 technical replicates using the Orbitrap Fusion™ mass spectrometer with the Proxeon EASY-Spray ionization source (Thermo Fisher Scientific) coupled to the Dionex UltiMate 3000 RSLC Nanoliquid chromatograph. The peptides were loaded on the Acclaim PepMap™ 100 nano trap column (nanoViper C18 100 µm

i.d. × 2 cm, 5 μm particle size, 100 Å pore size; Thermo Fisher Scientific), loading solvent was 1% acetonitrile with 0.05% TFA in water and sample loading volume was 10 μl using a partial-loop injection mode. The trap column was directly connected to analytical EASY-Spray column PepMap™ RSLC C18 (75 μm × 15 cm, 3 μm particle size, 100 Å pore size; Thermo Fisher Scientific) heated to 35 °C. The peptides were separated for 95 minutes; the starting organic phase gradient was set to 2% (0 to 5 min), then it increased gradually to 35% (5 to 65 min), then to 90% (65 to 73 min), and finally dropped quickly back to 2%. As the aqueous and organic mobile phases, water and acetonitrile were used, respectively, both with 0.1% formic acid. Concerning MS analysis, the static positive ion spray voltage was set to 2000 V, ion transfer tube temperature was 200 °C, master scan was performed in Orbitrap, positive mode, FTMS resolution was set to 120,000 and the m/z mass range for scanning of precursor ions was 400–1500. Cycle time data dependent mode was chosen, time between master scans was set to 3 s. Ion fragmentation was performed by the collision-induced dissociation (CID). The MS2 scan was performed in an ion trap, the rapid scan rate was set, the collision energy was 30% and the activation time was 10 ms. The profile data type was recorded in both master scan and MS2.

In the second study, MS analysis was performed in 2 technical replicates using Orbitrap Elite™ mass spectrometer with Proxeon Easy-Spray ionization source (Thermo Fisher Scientific) coupled to a Dionex UltiMate 3000 RSLC Nanoliquid chromatograph. The peptides were loaded on nano trap column Acclaim PepMap™ 100 (nanoViper C18 100 μm i.d. × 2 cm, 5 μm particle size, 100 Å pore size; Thermo Fisher Scientific), loading solvent was 1% acetonitrile with 0.05% TFA in water and sample loading volume was 10 μl using partial-loop injection mode. The trap column was directly connected to analytical EASY-Spray column PepMap™ RSLC C18 (75 μm × 15 cm, 3 μm particle size, 100 Å pore size; Thermo Fisher Scientific) heated to 35°C. The peptides were separated for 105 min total run time and the gradient of organic phase was increasing from 2% up to 10th min to 35% up to 65th min, to 90% up to 77th min and then it dropped quickly back to 2% of organic phase again. The aqueous mobile phase was water with 0.1% formic acid and organic mobile phase was acetonitrile with 0.1% formic acid. Concerning MS analysis, the MS1 was performed in Orbitrap, it was set to full scan type, positive mode, FTMS resolution was set to 120,000 and m/z mass range for scanning of precursor ions was 300 – 1700. The ten most intense ions were selected for fragmentation by collision-induced dissociation (CID). The MS2 scan was performed in an ion trap, normal scan rate was set, normalized collision energy was 35 V and activation time was 10 ms. Profile data type was recorded in both MS1 and MS2.

2.2.2.4. Data processing

In the first study, protein search was performed in beta version of Proteome Discoverer™ 2.5 Software (Thermo Fisher Scientific). The processing workflow includes a search against the complete human UniProtKB database, which included both Swiss-Prot reviewed and TrEMBL computationally analyzed but unreviewed proteins (https://www.uniprot.org/uniprot/, downloaded in January 2020). Sequest HT search engine was used where oxidation, N-terminal acetylation, methionine loss and, finally, methionine loss in combination with N-terminal acetylation were set as dynamic modifications. The Fixed Value PSM Validator was used.

In the second study, which contains older data but is still preparing for publication, protein search was performed in Proteome Discoverer™ software version 2.4 (Thermo Fisher Scientific). Processing workflow includes a search against the complete human UniProtKB database, which included both Swiss-Prot reviewed and TrEMBL computationally analyzed but unreviewed proteins (https://www.uniprot.org/uniprot/, downloaded in March 2019); chromatogram RT limit was set to 10 min from beginning and ending to decrease a size of the data set to make the software able to complete the next step (consensus workflow); Sequest HT and MS Amanda 2.0 search engines were used where oxidation and N-terminus acetylation were set as dynamic modifications and Fixed Value PSM Validator was used.

Consensus workflow included the same parameters for both studies. It was set as followed. Retention time alignment was enabled, only unique peptides were used for quantification, total peptide amount normalization was performed, no scaling was done, protein abundances were calculated using top 3 average method, protein ratio calculation was performed using protein abundance based method, low abundance resampling mode was chosen for missing values imputation and ANOVA (individual proteins) hypothesis test was used. Results were filtered only for high confidence peptides.

2.2.2.5. Statistical analysis

In the first study, a peptide/protein was considered to be present in a sample, if it had been detected at least in one of the technical replicates of this sample. For each sample and each peptide/protein, abundance was calculated as the log2 value of the median of the three values (from Proteome Discoverer) from technical replicates of this sample. Statistical analysis was performed using the R program, ver. 3.5.2 (R Core Team 3.5.1., 2018). Heatmaps were created with the Heatmap function (from the ComplexHeatmap library).

In the second study, technical replicates of each sample were aggregated by mean and transformed with the log2 function. For analyzes performed in R software (R Core Team 3.5.1.,

2018), a subset of 936 proteins detected at least in 2/3 of samples of at least one condition (age category, gender, smoking) subgroup was selected. The multinomial logistic regression model LASSO (glmnet R package, ver. 2.0-16*;* Friedman, Hastie and Tibshirani, 2010) was fitted on the balanced training dataset (n = 12 x 7 = 84 samples) and gender subsets. The proteins selected by LASSO models were used for building classification trees and graphically presented by heatmaps annotated with clinical characteristics and predictions calculated by each particular model. The relationship between the abundance values of each of the selected proteins and the age category was also analyzed by univariate analysis methods (Kruskal-Wallis test + multiple testing).

2.2.3. Results

STUDY 1.

Aim: Evaluate the reproducibility of mass spectrometric measurements when introducing biological and technical replicates, comparing the protein profile between individual samples of healthy subjects, and describing the reproducibility between three technical replicates of a single sample.

In the first study, we have focused on reproducibility of mass spectrometric measurements, benefits of introducing biological and technical replicates and comparison of EBC samples collected from one family members (**Václavková J**, P. Kouřilová, J. Vrbková, D. Holub, M. Hajdúch, P. Džubák. Proteomic Analysis of Exhaled Breath Condensate Samples: High Reproducibility of Mass Spectrometric Measurements. Chem. Listy 2020, 114, 470−479, ISSN 1213-7103, IF: 0.39; **see Appendix 8.1**).

The study cohort included 3 persons, namely, two children (daughters: D1 and D2) and an adult (father, F). Samples were taken at 3 time points in the interval of two hours from each person and each sample was measured in 3 technical replicates. Thus, 27 samples were analyzed. All these samples were evaluated by Proteome Discoverer software in one search. All the statistical analyses were performed at both protein and peptide level. The protein search led to the identification of 2797 proteins and 5006 peptides across all 27 samples and of them 2263 proteins and 4189 peptides were quantified.

The protein count per one technical replicate was equal to 1056 ± 354 proteins. A protein content of each sample (also called biological replicate) was achieved by combining of 3 technical replicates. Protein was considered to be detected in a sample, if it was detected in at least one of the technical replicates of the sample. The overlap between technical and biological individual replicates was high. However, introducing of technical replicates have increased a number of identifications. In our study, the number of detected proteins increased to 1198 ± 454 proteins per sample. A higher number of proteins (1603 ± 607) was identified in an adult (father's) sample compared to 996 ± 140 in daughter's samples. The number of proteins detected in individual biological and technical replicates is summarized in Table 2. Despite this observation, this is not the conclusion that the adult EBC contains more proteins than that from children generally because the study cohort was too small to make such a decision and is not the aim of our study. Instead, we showed high reproducibility of mass spectrometric measurements both in biological and technical replicates, advantages of introducing such replicates and we performed characterization and comparison of EBC samples taken from members of one family.

Table 2. Numbers of proteins detected in samples and technical replicates of samples. Samples were taken from three persons: father (F) and two daughters (D1 and D2). Three samples were taken from each person (the time delay between samples was about one hour), i.e. there were nine samples in the analysis. Each of the nine samples was measured in three technical replicates. A protein was considered to be detected in a sample if it was detected in at least one of the technical replicates of the sample (Václavková *et al.*, 2020).

	number of detected proteins in the sample (biological replicate)	number of detected proteins in the technical replicate (1st| 2nd| 3rd)
D1_01	1196	1052|1068|1069
D1_02	884	767| 794| 787
D1_03	888	753| 778| 803
D2_01	903	725| 772| 804
D2_02	956	793| 859| 764
D2_03	1148	986| 972|1006
F_01	2307	1869|1871|1900
F_02	1434	1276|1204|1213
F_03	1067	993| 949| 930

The analyses of protein and peptide count showed the same evidence, except the fact that the peptide number was higher than the protein number. The overlap between individual samples (biological replicates) of individual persons was relatively high. In the case of proteins, the overlap contained 644, 753 and 818 proteins for particular samples (Figure 2A). Of those overlapping proteins, 358 proteins were common to all three individuals. In the case of peptides, 955, 1147, and 1606 peptides were overlapping between individual biological replications and 532 of them were common to all three persons (Figure 2B).

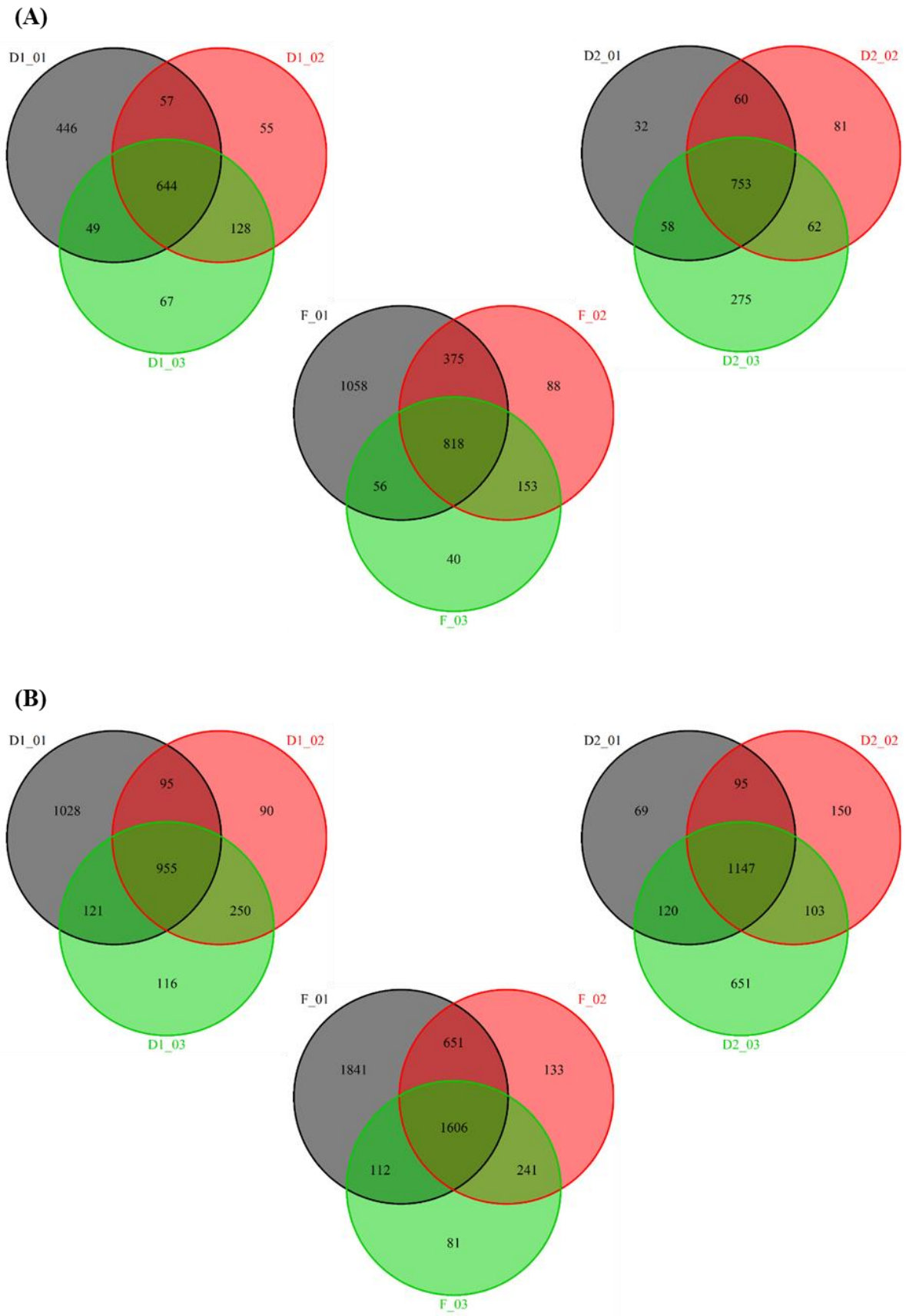


Figure 2. The number of (A) proteins and (B) peptides detected in each sample and the overlap between biological replicates of each person expressed by Venn diagrams (Václavková *et al.*, 2020).

A functional protein enrichment analysis was performed for 358 proteins common to all three persons in the study using STRING tool. It revealed that these proteins are involved in processes such as cornification, keratinocyte differentiation, skin development, keratinization, epidermis development, inter-mediate filament cytoskeleton organization, epithelial cell differentiation, intermediate filament organization, tissue development and epithelium development. However, all these processes are mediated by the same key proteins, mainly keratins. On the other hand, neutrophil degranulation, myeloid leukocyte activation, leukocyte-mediated immunity, cell activation, programmed cell death, regulated exocytosis, immune effector process, leukocyte activation, immune response, secretion by cell and immune system process, involved various proteins from the interaction network. All the processes reported were scored with the false discovery rate (FDR) < 0.001.

Heatmaps were constructed to express protein and peptide abundancies and to show the samples clustering (Figure 3). The number of proteins used for heatmap construction was lower than number reported above, because only the quantified proteins were utilized in the heatmaps. Proteins that were detected in some samples but not quantified in at least one of the samples were excluded because an abundance value is needed to draw a heatmap. Including values which are all imputed to the heatmap is useless and could be misleading even though the imputed values are very low. The cluster analysis of all nine samples (biological replicates) was performed to see whether the individual samples of the particular persons are clustered together or not. The heatmap has shown that the samples of individual family members are quite similar. Daughters' samples were clustered together with one exception. One of the daughters' samples (D1_01) clustered together with the father's samples (Figure 4A). This was the sample with the highest number of proteins and the reason why it belonged to the cluster with father's samples could be the fact that the father's samples generally include a higher number of proteins. The impact of the protein count could overbalance the impact of protein composition in such a situation when the samples are collected in one family and are quite similar. The same observations are reported on the peptide level (Figure 4B).

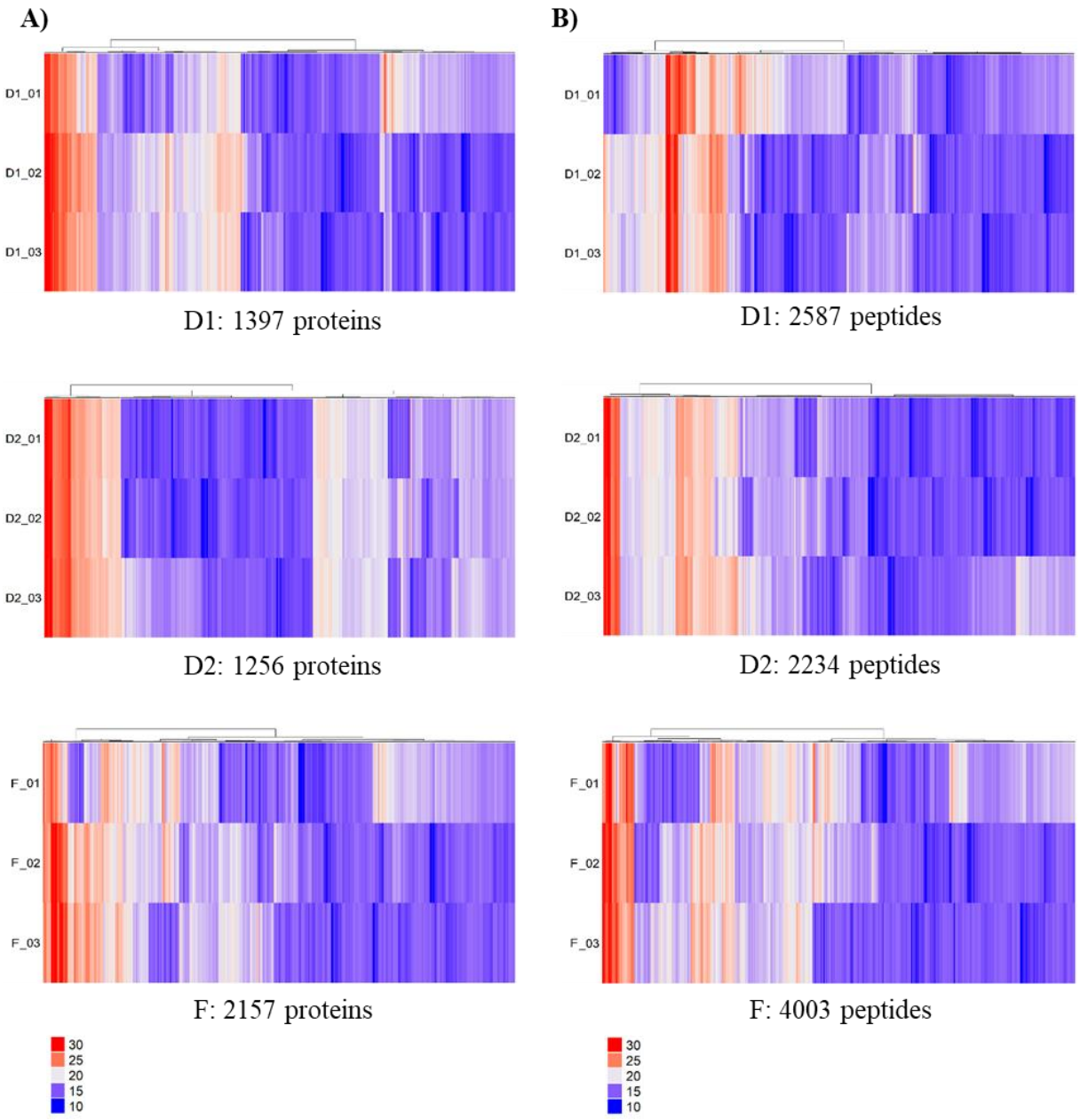


Figure 3. Heatmaps of abundance. (A) proteins and (B) peptides in three samples of each person. In each person-heatmap there are proteins or peptides that have been detected and quantified in at least one sample of that person (Václavková *et al.*, 2020).

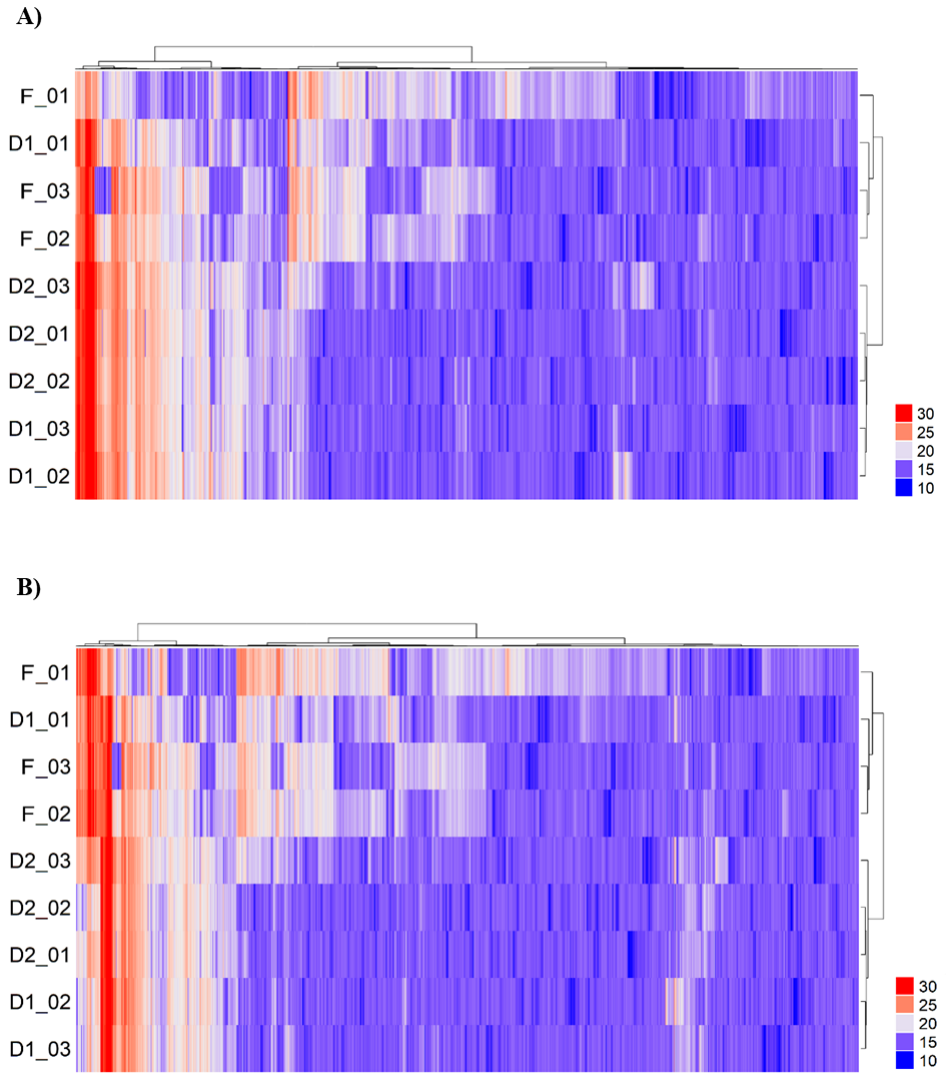
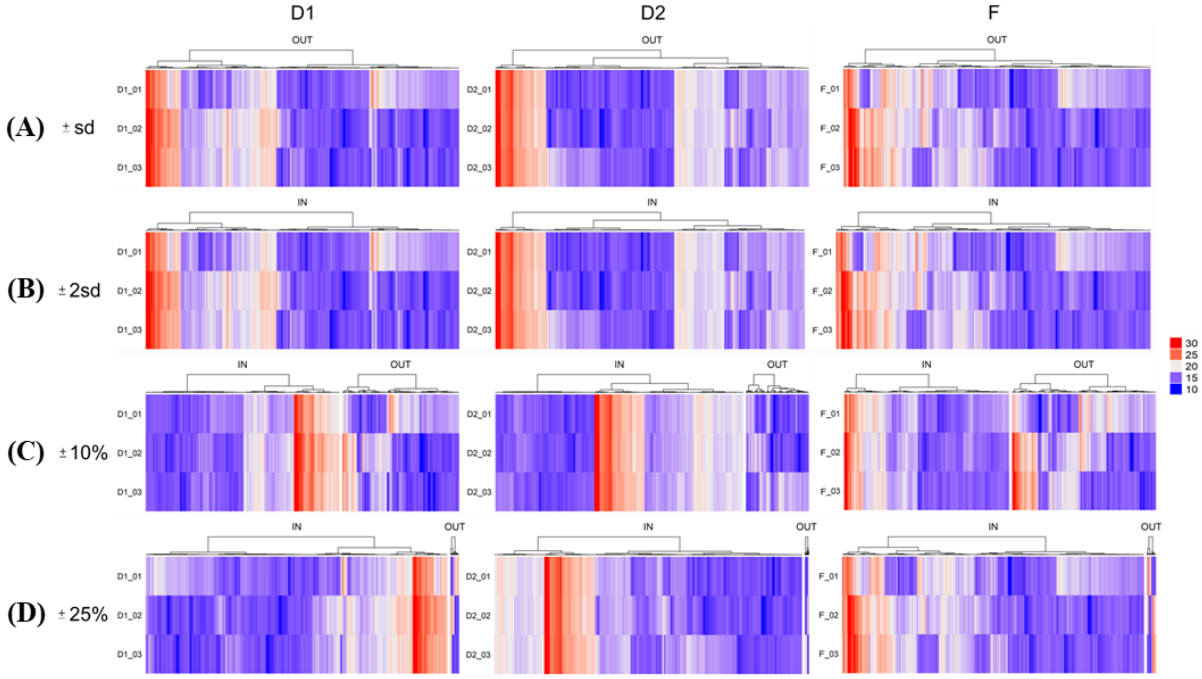


Figure 4. Heatmap expressing the abundance of all proteins and peptides across nine samples from individual family members. (A) Heatmap with all 2263 quantified proteins and (B) heatmap with all 4189 quantified peptides. The presented proteins and peptides are those that have been detected in at least one of the 9 samples. The columns represent individual proteins (respectively peptides), the rows represent the clustering of EBC samples of family members. The daughter's samples are clustered together except for the first replicate of D1, which contained the highest number of proteins and formed a cluster with the father samples (Václavková *et al.*, 2020).

To evaluate a degree of variability of mass spectrometric measurements, the distance of the abundancies form mean was calculated. This distance was calculated for each person, each protein. For each sample (biological replicate) and for each protein in the sample separately, the mean value of protein abundancies was counted and then, the difference of the three values of abundancies from the mean was calculated. The four intervals were established and it was evaluated whether the individual values belong to this interval or not. The 4 intervals for the values were following: ± sd (mean – sd; mean + sd); ± 2sd (mean – 2sd; mean + 2sd); ± 10%

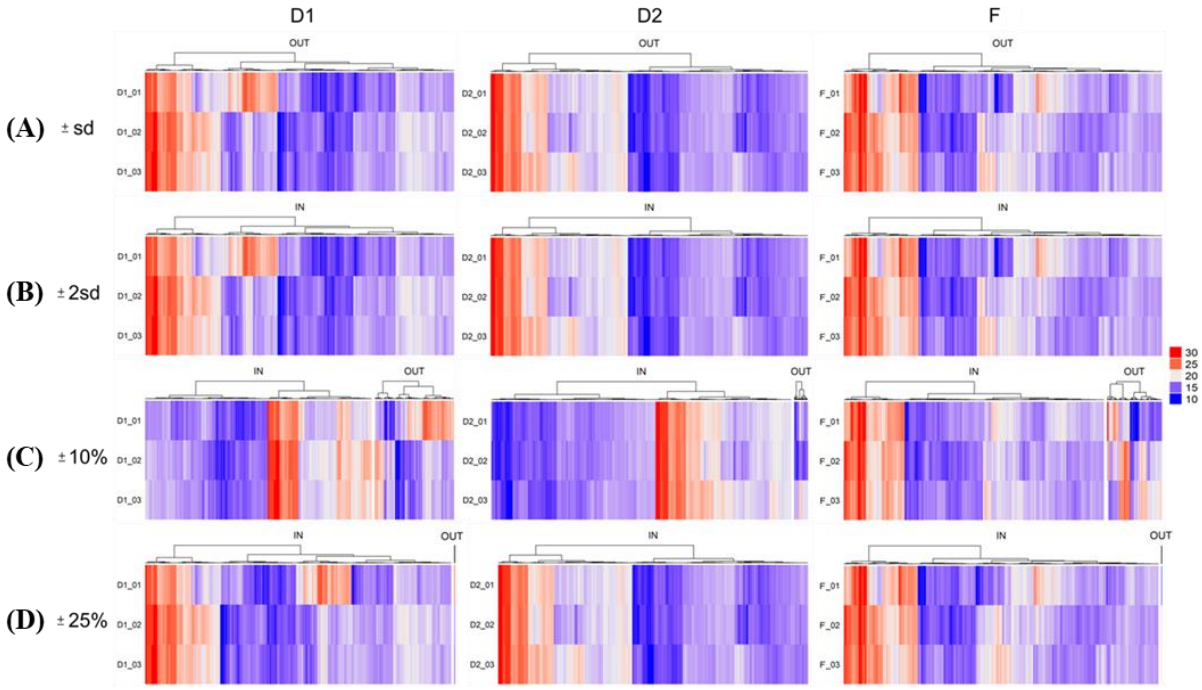
(0.9 × mean; 1.1 × mean); ± 25% (0.75 × mean; 1.25 × mean). In case that all 3 values of abundancies laid in the interval, it was considered as a stable value which is inside the interval and denoted as IN. In case that at least one of the 3 values of abundancies laid out of the interval, it was considered as a value out of the interval and was denoted as OUT (Figure 5). The same calculations, intervals and criteria were used for peptide analysis. There was only one difference that only the peptides, which were quantified in all technical and all biological replicates from the particular person, were included in the analysis (Figure 6).



(E)

	D1		D2		F	
	IN	OUT	IN	OUT	IN	OUT
±sd	0	1397	0	1256	0	2157
±2sd	1397	0	1256	0	2157	0
±10%	871	526	1002	254	1152	1005
±25%	1357	40	1240	16	2094	63
total	1397		1256		2157	

Figure 5. The distance of the abundancies of the three samples was analyzed for each person and each protein to evaluate the variability of the measured values. Only proteins that have been quantified in at least one sample of individual person are included in the heatmaps. Thus, the total number of proteins is lower than reported before. Four types of intervals were determined and the proteins were divided into two groups: IN = the abundance of all three samples lies in the respective interval; OUT = the abundance of at least one of the three samples is outside this interval. Heatmaps (A – D) expressing proteins that lie in or out of intervals. Selected intervals were: **(A)** ± sd (mean – sd; mean + sd), **(B)** ± 2sd (mean – 2sd; mean + 2sd), **(C)** ± 10% (0.9 × mean; 1.1 × mean), **(D)** ± 25% (0.75 × mean; 1.25 × mean). Table **(E)** lists the number of proteins which lie in and out of the respective intervals. Mean = mean of the abundancies of the protein of the three samples of each person; sd = standard deviation of the mean (Václavková *et al.*, 2020).



(E)

	D1		D2		F	
	IN	OUT	IN	OUT	IN	OUT
±sd	0	687	0	800	0	1128
±2sd	687	0	800	0	1128	0
±10%	509	178	765	35	931	197
±25%	685	2	800	0	1126	2
total	687		800		1128	

Figure 6. The distance of the abundancies of the three samples was analyzed for each person and each peptide. For each individual, peptides quantified in all three technical replicates and in all samples of the person were included, i.e. only peptides that were quantified in 9 replicates (technical and biological) of the individuals, no values imputed by Proteome Discoverer software were included. Four types of intervals were set and the peptides were divided into two groups: IN = the abundance of all three samples lies in the respective interval; OUT = the abundance of at least one of the three samples is outside this interval. Heatmaps (A – D) expressing peptides that lie in or out of intervals. Selected intervals were: **(A)** ± sd (mean – sd; mean + sd), **(B)** ± 2sd (mean – 2sd; mean + 2sd), **(C)** ± 10% (0.9 × mean; 1.1 × mean), **(D)** ± 25% (0.75 × mean; 1.25 × mean). Table **(E)** lists the number of peptides that lie within and out of the respective intervals. Mean = mean of the abundancies of the peptide in the three samples of each person; sd = standard deviation of the mean (Václavková *et al.*, 2020).

Analyses on both protein and peptide level have shown the same outcomes. Firstly, the criteria, that the distance of the abundancies from the mean should be lower than the difference of one standard deviation or lower than 10%, were too strict. In the case of one standard deviation, no protein or peptide laid in the perspective interval. In the case of 10% difference from the mean, more than half of the proteins and peptides were denoted as in the interval which is better but still not suitable. Secondly, when the criteria were relaxed, the doubling of the standard deviation (± 2 × sd) caused that all the proteins and peptides belonged to this interval.

When we analyzing proteins and peptides with the distance from the mean which is less than 25%, almost all values laid in this interval. Only a few tens of proteins and 2 peptides were denoted as outliers. The reason why the number of peptides is lower than number of proteins is that the peptides included in the analysis were reduced only to peptides which were in all technical and all biological replicates from the particular person. The number of peptides in the analysis was lower and thus, all the numbers were lower than in the case of all proteins. Protein data did not include the information whether the value was imputed or really quantified, so all the proteins have to be included.

Next, we were interested if the deviations are a consequence of variability of mass spectrometric method or biological variability of the samples. The commercial standard of bovine serum albumin (BSA) was chosen to evaluate the variability of MS method. The chosen commercial BSA standard (Pierce™ BSA Protein Digest Standard, LC/MS Grade, Thermo Fisher Scientific) is used routinely for quality control of the MS measurements in our labs. The same amount as for quality control measurement which is 100 fmol BSA was injected for all the measurements in this analysis. The mean of log2 abundance of 3 technical replicates for each peptide separately. The difference from the mean was calculated for all three values. The absolute value of the largest difference was chosen and expressed as a percentage of the mean value. Again, only the peptides which were quantified in all technical and all biological replicates from the particular person were included. For the BSA standard, it meant that all the identified peptides were always quantified and thus, included. The maximum distance for all the peptides was equal to 8.11%. This should be considered as a threshold for the variability of mass spectrometric measurements. This value is significantly lower than the variability of individual biological replicates. However, the commercial standard was used and it should be expected that the variability of technical replicates will be lower in the case of standardized mixture of peptides from one protein than in the case of complex protein mixture in biological matrix. To confirm this, we have evaluated variability of technical replicates of each person's samples. We applied the same criteria for peptide inclusion and calculated the maximum distances from the mean the same way as for BSA standard. The values of the distances from the mean abundance were calculated for all included peptides of all samples of all study participants. These values were compared and summarized in Table 3. The results have shown that all the maximum distances from the mean are above 20%. The maximum value achieved in the whole study was 19.69%.

Table 3. The maximum distances from the mean abundance value of the three technical replicates (expressed in %) for individual samples (Václavková *et al*., 2020).

	maximum distance (%) from the mean
D1_01	12.66
D1_02	19.69
D1_03	17.61
D2_01	13.13
D2_02	10.33
D2_03	8.31
F_01	9.04
F_02	13.39
F_03	13.44

Our data have shown that the differences in abundances below 10% are believed to be caused by the deviation of the MS measurement and the influence of biological variations is minimized. Then, the change in protein expression which is lower than 20% is a consequence of method variability and should not be considered as a change in protein expression when the label-free MS approach is used. And additionally, variability between individual biological samples of one person is greater than between technical replicates of one EBC sample. However, this difference is quite small.

STUDY 2.

Aim: Evaluate the exhaled breath condensate of healthy individuals to predict age according to proteomic signatures and describe differences in proteomic profile according to sex and smoking status.

In the second study, we focused on age prediction by proteomic signatures in exhaled breath condensate (Václavková *et al.*, unpublished data). We have detected 3202 proteins across the whole individuals' cohort; of them 2769 proteins were quantified. By protein filtering 936 proteins were selected and evaluated. Protein abundance and clustering were expressed by heatmap (Figure 7).

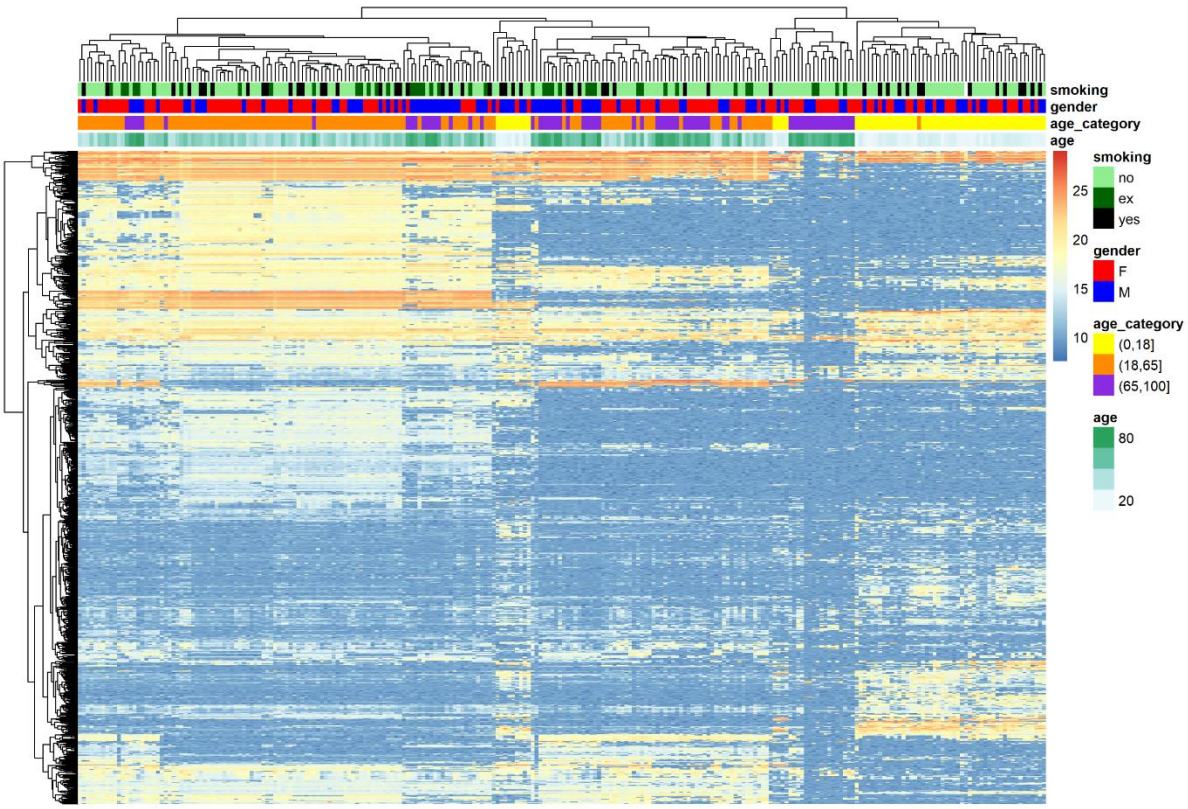


Figure 7. Protein profile clustering of the 936 selected proteins across all samples expressed by heatmap.

Number of detected proteins in individual samples in each age category (children and young adults, adults, elderly group) was summarized in a Table 4. Both in males and females, the statistically significant increase in number of proteins detected was observed between adult group (18 – 65 years old) compared to children and young adults (less than 18 years old) and elderly group (more than 65 years old). However, when compared children and young adults and elderly group there was not statistically significant difference.

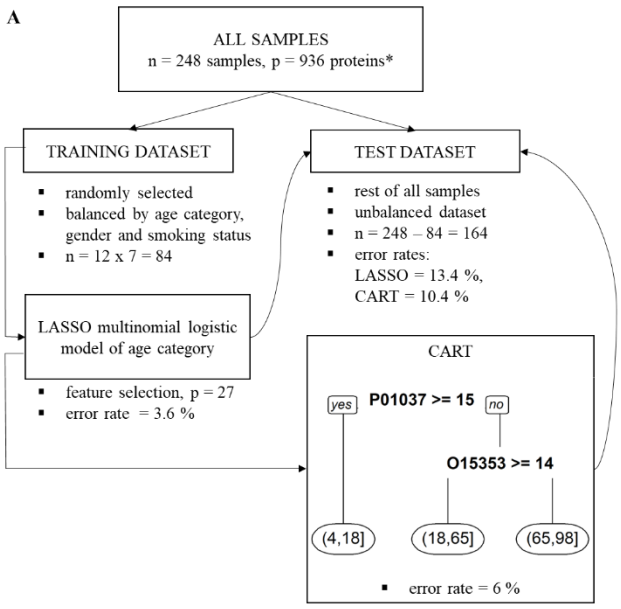
Table 4: Number of samples by age category, sex and smoking status (smoking in family in the age category "≤ 18") and number of detected proteins per sample. The adult group showed significantly higher number of proteins than children and young adults or elderly group. There was not significant difference in protein count between children and young adults and elderly individual group.

number of subjects (no-smoking|smoking) number of detected proteins per sample median (min|max)		children and young adults (4,18]	adults (18,50]	adults (50,65]	adults subtotal	elderly (65,98]	total
gender	female	31 (24|7)	53 (36|17)	29 (19|10)	82 (55|27)	30 (21|9)	143 (100|43)
		540 (365|1085)	1147 (417|1402)	1144 (507|1447)	1145 (417|1447) #	447 (119|1008)	711 (119|1447)
	male	30 (19|10)*	19 (5|14)	20 (15|5)	39 (20|19)	36 (15|21)	105 (54|50)*
		568 (336|1081)	1085 (455|1580)	1024 (404|1529)	1052 (404|1580) #	690 (48|1218)	734 (48|1580)
total		61 (43|17)*	72 (41|31)	49 (34|15)	121 (75|46)	66 (36|30)	248 (154|93)*
		566 (336|1085)	1119 (417|1580)	1078 (404|1529)	1115 (404|1580) #	549 (48|1218)	717 (48|1580)

** one sample with missing smoking status*

 # significant difference between adults group and elderly and between adults group and children and young adults (post-hoc Nemenyi test, $p < 0.001$)

To avoid the influence of gender and smoking status in each age category, the LASSO multinomial logistic regression model (glmnet R package, ver. 2.0-16; Friedman, 2010) were fitted on the balanced training dataset randomly selected (dplyr R package, ver. 0.8.3) from each condition subgroup (n = 12 x 7 = 84 samples). The rest of the 164 samples were used for model testing with a simple error rate measure. The proteins selected by LASSO model were consequently used as input for building a classification tree (rpart R package, ver. 4.1-13, rpart.plot, ver. 3.0.8) on the same training dataset of 84 samples (Figure 8A). Prediction of age category was performed for males and females; smokers, ex-smokers and nonsmokers respectively. Error rates calculated for training and test dataset showing that the prediction is quite nice for children and young adults (≤ 18) and still good, but generally worse, in adults and elderly groups independently on gender and smoking status (Figure 8B).



B

			FEMALE/MALE	age category (4,18]	age category (18,65]	age category (65,98]	total
LASSO MODEL	TRAINING DATASET	prediction	(4,18]	14|14	0|0	0|0	14|14
			(18,65]	0|0	13|14	2|0	15|14
			(65,98]	0|0	1|0	12|14	13|14
		error rate (%)		*0|0*	*7.1|0*	*14.3|0*	*7.1|0*
			NO-SMOKING|EX|SMOKING	age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	14|-|14	0|0|0	0|0|0	14|0|14
			(18,65]	0|-|0	14|8|5	0|2|0	14|10|5
			(65,98]	0|-|0	0|0|1	14|10|2	14|10|3
		error rate (%)		*0|-|0*	*0|0|16.7*	*0|16.7|0*	*0|10|4.5*
	TEST DATASET		FEMALE/MALE	age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	16|16	1|0	0|0	17|16
			(18,65]	1|0	59|19	2|4	62|23
			(65,98]	0|0	8|6	14|18	22|24
		error rate (%)		*5.9|0*	*13.2|24*	*12.5|18.2*	*11.9|15.9*
			NO-SMOKING|EX|SMOKING	age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	28|-|3	0|0|1	0|0|0	28|0|4
			(18,65]	1|-|0	50|8|20	2|2|2	53|10|22
			(65,98]	0|-|0	11|1|2	20|11|1	31|12|3
		error rate (%)		*3.4|-|0*	*18|11.1|13*	*9.1|15.4|66.7*	*12.5|13.6|17.2*
CART MODEL	TRAINING DATASET		FEMALE/MALE	age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	14|14	0|0	0|0	14|14
			(18,65]	0|0	13|14	3|1	16|15
			(65,98]	0|0	1|0	11|13	12|13
		error rate (%)		*0|0*	*7.1|0*	*21.4|7.1*	*9.5|2.4*
			NO-SMOKING|EX|SMOKING	age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	14|-|14	0|0|0	0|0|0	14|0|14
			(18,65]	0|-|0	14|8|5	0|3|1	14|11|6
			(65,98]	0|-|0	0|0|1	14|9|1	14|9|2
		error rate (%)		*0|-|0*	*0|0|16.7*	*0|25|50*	*0|15|9.1*
	TEST DATASET		FEMALE/MALE	age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	16|16	1|0	0|0	17|16
			(18,65]	1|0	67|20	4|6	72|26
			(65,98]	0|0	0|5	12|16	12|21
		error rate (%)		*5.9|0*	*1.5|20*	*25|27.3*	*5.9|17.5*
			NO-SMOKING|EX|SMOKING	age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	28|-|3	0|0|1	0|0|0	28|0|4
			(18,65]	1|-|0	59|7|21	6|3|1	66|10|22
			(65,98]	0|-|0	2|2|1	16|10|2	18|12|3
		error rate (%)		*3.5|-|0*	*3.3|22.2|8.3*	*27.3|23.1|33.3*	*8|22.7|10.4*

Figure 8. Model construction and age category prediction. (A) Model construction. Samples were split into training and test dataset. The LASSO multinomial logistic regression model was fitted on the balanced training dataset randomly selected from each condition subgroup (n = 12 x 7 = 84 samples) to avoid the influence of gender and smoking status in each age category. The rest of 164 samples were used for model testing with a simple error rate measurement. The proteins selected by LASSO model were consequently used as an input for building a classification tree on the same training dataset of 84 samples. The CART (Classification And Regression Tree) chose 2 proteins, which are Cystatin-SN (P01037) and Forkhead box protein N1 (O15353) for age category prediction. (B) Age category prediction, the number of wrongly classified individuals and error rates of predicted age. The numbers of wrongly classified individuals and error rates were calculated for LASSO multinomial logistic regression model and for CART as well. In both cases, calculations were performed for training and test dataset separately. The number of wrongly classified individuals and error rates of predicted age was calculated; firstly, the values were counted for males and females separately and secondly for nonsmokers, ex-smokers and smokers separately. The numbers of wrongly classified individuals are written in red.

The LASSO multinomial logistic model chose 27 proteins which can be used for age category prediction (Figure 9A) and of them, Cystatin-SN and Forkhead box protein N1 were chosen by classification tree (Figure 9B). Cystatin SN is highly abundant in children and young adults compared to other groups and Forkhead box protein N1 disappears in the elderly group (Figure 9C).

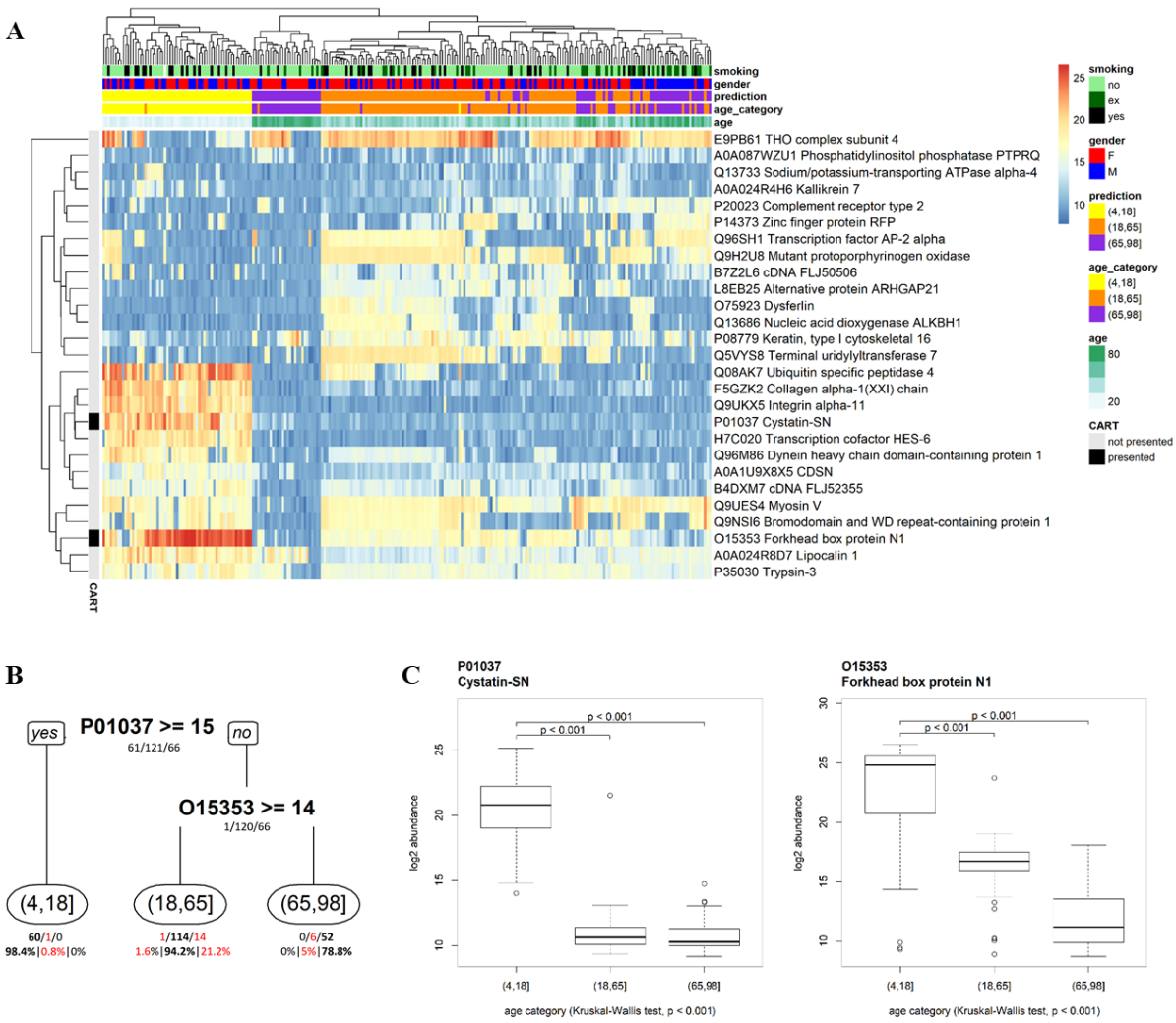


Figure 9. Proteins selected by LASSO multivariate analysis and CART. (A) The proteins (n = 27) selected by LASSO models for age category prediction and their relative abundance and clustering expressed by heatmap. Real age (green scale) and real age category (yellow, orange, violet) are marked. Age category predicted is marked (with the same yellow, orange, violet colors) above the real age category. Proteins selected by CART are labeled by black squares on the left side. (B) The CART for age category prediction – complete dataset (total number of 248 samples). The number of individuals who were classified into the respective group is written below each node. The number of wrongly classified individuals is labeled in red. (C) The relative abundance of these selected proteins in individual subgroups is graphically presented by boxplots. The univariate analysis and Kruskall-Wallis test were performed for significance determination. Significant results (p-value < 0.001) are labeled in the box plots.

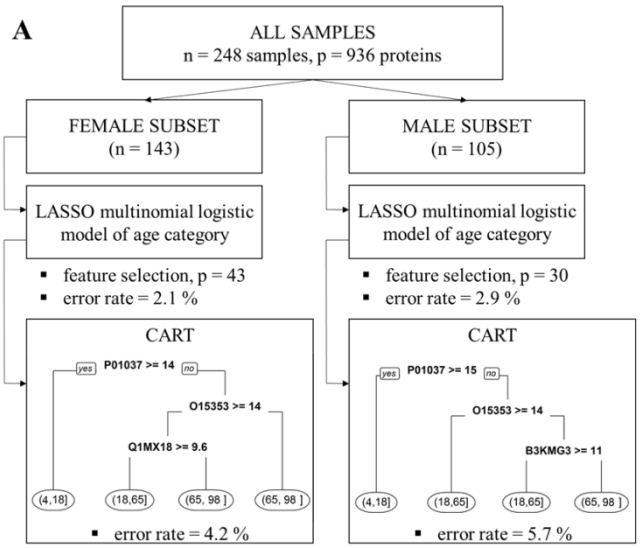
Then, the age prediction by proteomic signatures in exhaled breath condensate was performed for males and females separately. Because of the small size of the balanced training dataset of all male and female samples analyzed separately, the analyses were performed with whole gender subsets (without separation to the test-training dataset). Proteins selected by the LASSO multinomial logistic regression models were used for building classification trees and

similar univariate analyses were done as in the first training-test datasets analysis (Figure 10A). Prediction of age category was performed for male and female subset separately and was counted for the individual smoking conditions. The prediction was correct in all cases in the children group and again, was worse, but still good, in adults independently on smoking status (Figure 10B).

LASSO multinomial logistic model chose 43 protein predictors for age category estimation in females (Figure 11A) and those 30 for males (Figure 11B). Classification trees revealed Cystatin-SN and Forkhead box protein N1 again for both males and females with similar behavior as described above. For females, Protein inscuteable homolog is typically highly abundant in the adult group and was chosen to distinguish it from the elderly group. For males, cDNA FLJ10929 fis, clone OVARC1000479, highly similar to Cullin-associated NEDD8-dissociated protein 1 is differently abundant in the adult group and elderly group while the elderly group is characterized by lower levels of this protein.

These results have shown that we were able to detect changes in the protein profile of EBC induced by age. The biggest difference in the EBC protein profile was reported in the case of children and young adults whose EBC proteome was quite different from groups of adults. This could be useful for the identification of predictive markers of age-related diseases and resistance or susceptibility to some diseases.

A



B

FEMALE SUBSET	LASSO	TRAINING		age category (4,18]	age category (18,65]	age category (65,98]	total	NO-SMOKING|EX|SMOKING		age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	31	0	0	31	prediction	(4,18]	24|-|7	0|0|0	0|0|0	24|0|7
			(18,65]	0	82	3	85		(18,65]	0|-|0	55|9|18	1|2|0	56|11|18
			(65,98]	0	0	27	27		(65,98]	0|-|0	0|0|0	20|5|2	20|5|2
		error rate (%)		*0*	*0*	*10*	*2.1*	*error rate (%)*		*0|-|0*	*0|0|0*	*4.8|28.6|0*	*1|12.5|0*
	CART	FREQ.		age category (4,18]	age category (18,65]	age category (65,98]	total	NO-SMOKING|EX|SMOKING		age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	31	1	0	32	prediction	(4,18]	24|-|7	0|0|1	0|0|0	24|0|8
			(18,65]	0	77	1	78		(18,65]	0|-|0	54|9|14	1|0|0	55|9|14
			(65,98]	0	4	29	33		(65,98]	0|-|0	1|0|3	20|7|2	21|7|5
		error rate (%)		*0*	*6.1*	*3.3*	*4.2*	*error rate (%)*		*0|-|0*	*1.8|0|22.2*	*4.8|0|0*	*2|0|14.8*

MALE SUBSET	LASSO	FREQ.		age category (4,18]	age category (18,65]	age category (65,98]	total	NO-SMOKING|EX|SMOKING		age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	30	0	0	30	prediction	(4,18]	19|-|10	0|0|0	0|0|0	19|0|10
			(18,65]	0	37	1	38		(18,65]	0|-|0	20|8|9	0|1|0	20|9|9
			(65,98]	0	2	35	37		(65,98]	0|-|0	0|0|2	15|17|3	15|17|5
		error rate (%)		*0*	*5.1*	*2.8*	*2.9*	*error rate (%)*		*0|-|0*	*0|0|18.2*	*0|5.6|0*	*54|26|24*
	CART	FREQ.		age category (4,18]	age category (18,65]	age category (65,98]	total	NO-SMOKING|EX|SMOKING		age category (4,18]	age category (18,65]	age category (65,98]	total
		prediction	(4,18]	30	0	0	30	prediction	(4,18]	19|-|10	0|0|0	0|0|0	19|0|10
			(18,65]	0	39	6	45		(18,65]	0|-|0	20|8|11	1|3|2	21|11|13
			(65,98]	0	0	30	30		(65,98]	0|-|0	0|0|0	14|15|1	14|15|1
		error rate (%)		*0*	*0*	*1.7*	*5.7*	*error rate (%)*		*0|-|0*	*0|0|0*	*6.7|16.7|66.7*	*1.9|11.5|8.3*

Figure 10: Model analysis of all male and female samples separately – model construction and proteins were chosen for age category prediction. (A) Model construction. LASSO multinomial logistic regression model and CART was performed for the male and female subgroup. The error rate was calculated for LASSO model and for CART as well. For female subset classification tree assigned Cystatin-SN (P01037); Forkhead box protein N1 (O15353) and Protein inscuteable homolog (Q1MX18) as age category predictors. For the male subset, the age category predictors were Cystatin-SN (P01037); Forkhead box protein N1 (O15353) and cDNA FLJ10929 fis, clone OVARC1000479, highly similar to Cullin-associated NEDD8-dissociated protein 1 (B3KMG3). (B) Age category prediction, the number of wrongly classified individuals and error rates of predicted age for male and female subset evaluated separately. The numbers of wrongly classified individuals and error rates were counted for each subset, both females and males. In both cases, the values were counted for LASSO multinomial model and CART. Firstly, the whole female and a male subset was evaluated (left half of the tables) and secondly, the values were counted for nonsmokers, ex-smokers and smokers separately (right half of the tables).

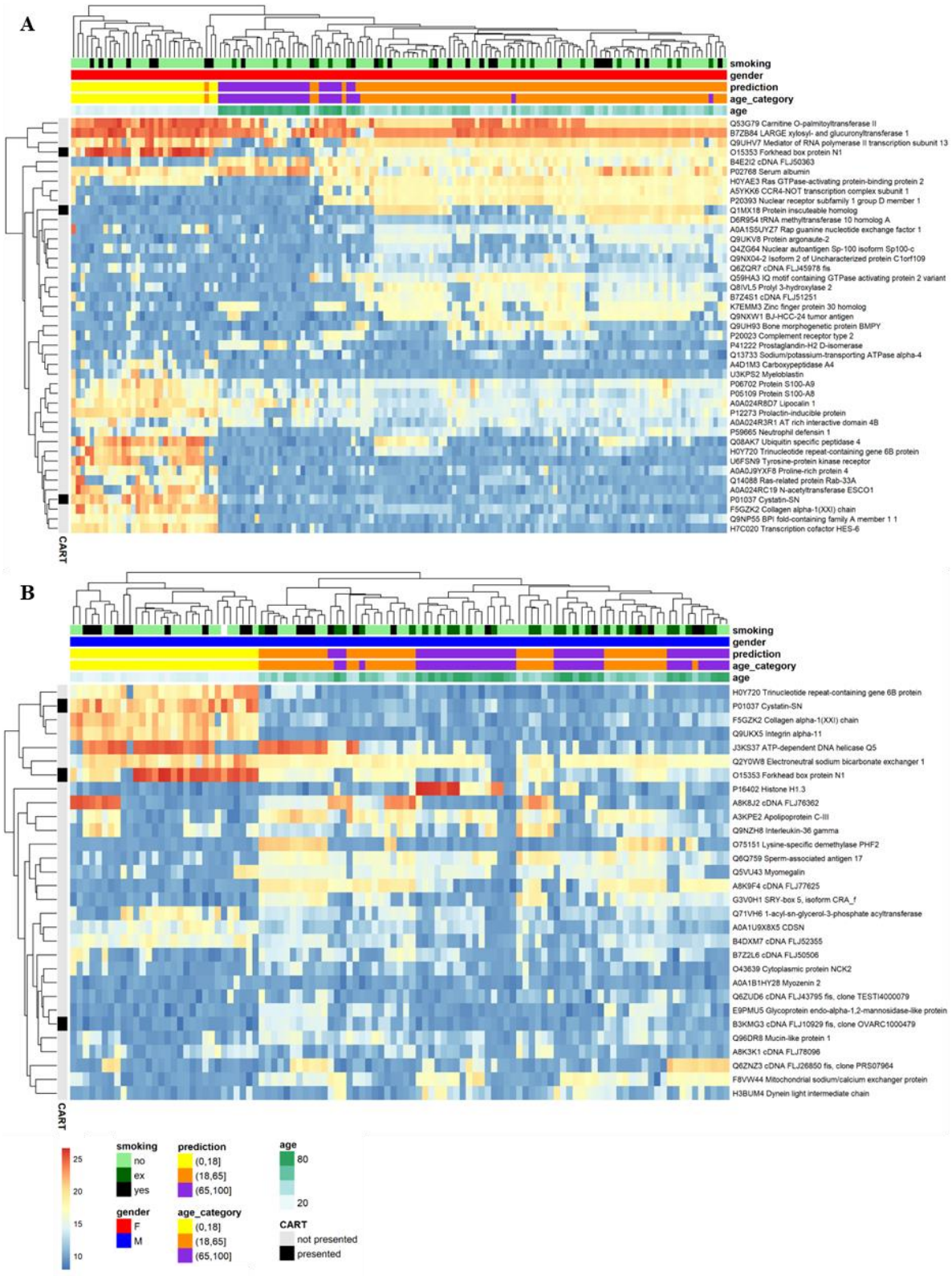


Figure 11: The performed as regressors in LASSO model in female and male subset for age category prediction and their relative abundance and clustering expressed by heatmap. (A) Proteins (n = 43) selected by LASSO for age category prediction in female subset and (B) proteins (n = 30) selected by LASSO for age category prediction in male subset. In heatmaps, real age (green scale) and real age category (yellow, orange, violet) is marked. The age category predicted is marked (with the same yellow, orange, violet colors) above the true age category.

2.2.4. Discussion

We have shown that our mass spectrometric method is sensitive and highly reproducible and could be implemented for in depth characterization of human EBC and biomarker discovery research of lung diseases.

In both studies reported in this thesis, the numbers of proteins detected (or quantified respectively) were higher than the protein numbers reported in any of the previously published studies, even if the studies used pooled samples. The studies using pooled samples are predominantly focused on in depth characterization of EBC from healthy subjects. Firstly, when endogenous particles in exhaled air were collected, 124 proteins were detected in two pooled samples from 12 volunteers (Bredberg *et al.*, 2012). In a pool of nine healthy subjects, 167 proteins were identified (Muccilli *et al.*, 2015). Analysis of two pooled samples where each consisted of 10 individuals led to the identification of 229 proteins (Lacombe *et al.*, 2018). Analyses that focused on biomarker discovery usually did not use the pooled samples but measured individual samples. These included the EBC analysis of 24 intubated newborns, where each sample was measured in triplicate and 119 proteins were identified (Kononikhin *et al.*, 2017). So far, the highest number of proteins was 348 different proteins identified across the cohort of 192 individuals in the study which compared lung cancer and COPD patients and control subjects and suggested the biomarkers of lung cancer and COPD (López-Sánchez *et al.*, 2017).

The second study has shown that the number of identified proteins was significantly higher in adults compared to children and elderly groups. This confirmed hypothesis that arose from the first study where, however, the study cohort was too small to make such a conclusion. Our hypothesis was that this is due to the larger lung surface in adults compared to children. The number of proteins dropped again in the elderly group, most probably due to physiological changes during lung aging. There was no significant difference between protein count in children and young adults and the elderly. However, the protein signature of children EBC was quite different than those in adults. This is supported by the fact that the cluster of children group was separated from the adults nicely in the heatmaps and the numbers of wrongly classified individuals were quite low in children during age prediction. Up to date, there is no evidence in the literature concerning age-dependent changes in protein count in the EBC.

Both studies had shown that keratins are the most abundant and most frequent proteins in human EBC. Looking to the past, keratins were the first proteins that were identified in the breath condensate using a mass spectrometric approach (Gianazza *et al.*, 2004). The keratins were also usually reported in most EBC studies. We were able to detect 37 keratins, 20 of these

were among 936 proteins which were further statistically evaluated in the second study. We have identified keratins type I cuticular: Ha1 and Ha6; keratins type II cuticular Hb4 and Hb5; keratins, type I cytoskeletal 9, 10, 14, 16, 17; keratin, type II cytoskeletal 1b, 2 (epidermal and oral), 3, 5, 6B, 75, 78, 80. In healthy controls keratins, type I cytoskeletal 9, 10, 14, 16, 17 and keratins, type II cytoskeletal 1b, 2, 3, 5, 6B, 78, 80 were previously identified (Gianazza *et al.*, 2004; Fumagalli *et al.*, 2008; Gessner *et al.*, 2008; Bloemen *et al.*, 2009; Kurova *et al.*, 2009; Muccilli *et al.*, 2015; Lacombe *et al.*, 2018). A study performed by Fumagalli *et al.*, 2012 analyzed COPD patients without emphysema and patients with pulmonary emphysema associated with α1-antitrypsin deficiency (AATD) but they found cytokeratins mainly in pooled EBC samples of healthy individuals. We have detected the similar keratin profile as Muccilli *et al.*, 2015, the same keratins included type I cytoskeletal 9, 10, 14, 16, 17; keratin, type II cytoskeletal 1b, 2, 3, 5, 6B, 78, 80. Muccilli's study was performed on the same mass spectrometer as our study, Orbitrap Elite. In contrast to our approach, Muccilli *et al.* used pooled control samples and gel-based method. Some scientific groups claimed that keratins are exogenous contamination and should be excluded from further analysis (Fedorchenko *et al.*, 2016; Núñez-Naveira, Mariñas-Pardo and Montero-Martínez, 2019). However, we confirmed that keratins play an important role when detected in EBC. Regarding patients' studies, the cytokeratins 2, 5, 6, 9, 10, 14 and 16 and, were found in various samples of healthy controls and asthma patients and additionally to our study, cytokeratins 1 and 8 were identified (Bloemen *et al.*, 2011). Keratins such as keratin type I cytoskeletal 9, 10, 12, 16, 18, 24; and type II cytoskeletal 1, 1b, 2, 5 were identified in the EBC of intubated newborns but was not specific for any of the infants group, neither congenital pneumonia or a left-sided congenital diaphragmatic hernia (Kononikhin *et al.*, 2017). Several isoforms of cytokeratin 6 (6A, 6B and 6C) have the potential to be biomarkers of lung cancer. Quantification of most keratins (particularly CK1, 6C, 9 and 10) showed a significant positive correlation with tumor size (López-Sánchez *et al.*, 2017). In another lung cancer study, cytoskeletal keratins type II 1, 2, 3, 4, 5, 6 and cytoskeletal keratins type I 9, 10, 14, 15 and 16 were invariant for all the samples which included lung cancer patients and healthy controls (Fedorchenko *et al.*, 2016).

As well, the identified proteins were included in the immune system processes. These were reported as the main processes in the EBC of intubated newborns (Kononikhin *et al.*, 2017) and were overrepresented in EBC compared to lung proteome in healthy controls which was considered as evidence that EBC was a relevant matrix to study the major physiological functions of the respiratory tract, particularly mucosal secretion, innate and adaptive antimicrobial defense mechanisms and clearance of inhaled particles (Lacombe *et al.*, 2018).

Concerning proteins that were identified as key age-related protein biomarkers in EBC, cystatin SN and forkhead box protein N1 were chosen for age category prediction. Cystatin SN is involved in inflammation, cell cycle, cellular senescence, tumorigenesis, and metastasis. Cystatin SN is also known to participate in signaling pathways like Wnt signaling pathway, GSK3 signaling pathway, Akt signaling pathway, and IL-6 signaling pathway. In healthy individuals, cystatin SN acts as the inhibitor of cysteine proteases in the oral environment and is involved in the control of the proteolytic events such as inhibiting the activity of papain *in vivo* (Liu and Yao, 2019). In contrast to our findings, the consistent and strong upregulation of Cystatin SN (CST1) expression was reported during cellular senescence on cellular models (Keppler *et al.*, 2011). However, later publication reported that knockdown of CST1 induces cellular senescence. They concluded that extracellular CST1 is a positive regulator of extracellular Cathepsin B and contributes to tumorigenesis by preventing cellular senescence through the inhibition of abnormal glycogen accumulation (Oh *et al.*, 2017). Cystatin SN was found in many cancer types, in cancer there was a higher expression of CST1. It was found to be an independent prognostic factor through univariate and multivariate analyses and was greatly related to the survival of cancer patients (Liu and Yao, 2019). In the study EBC study with Russian astronauts after spaceflights where the biological samples were collected 1 month before the spaceflight, immediately after landing and on the seventh day after landing, the Cystatin SN was detected in the individuals on the seventh day after landing (Kononikhin *et al.*, 2019) which could suggest that the CST1 is a marker of lungs reparation. Forkhead box protein N1 (FOXN1) is a ubiquitously expressed member that has been implicated in the embryo development, metabolism, aging and cancer. It was described as a tumor suppressor; the higher level of FOXN1 was associated with a better prognosis of non-small lung cancer (NSCLC) patients (Ji *et al.*, 2018). Experiments on mouse model showed that FOXN1 disappears during aging (Rode *et al.*, 2015) which is in concordance with our findings in EBC.

Cystatin SN and forkhead box protein N1 were selected for age prediction by decision tree for whole individuals group as well as for the female subgroup and male subgroup respectively. Cystatin SN was reported as decreasing with age in both females and males in another proteomic study using 2-D electrophoresis and MS identification of selected spots (Fleissig *et al.*, 2010). The decision tree selected Protein inscuteable homolog (Insc) as the third protein which was female-specific for prediction of age category 18 – 65 years old and disappearing in children and older adults too. Protein inscuteable is usually described in *Drosophila* where it plays role in asymmetric cell division and mitotic spindle orientation (Bowman *et al.*, 2006). Human Insc shows an overall amino acid similarity to 25%. Human Insc was detected in various

tissues with abundant expression in the lung, kidney, pancreas, and small intestine (Izaki *et al.*, 2006). Its role was described in the asymmetrical cell division of human neural stem cells and other stem cells (Katoh and Katoh, 2003). The role in asymmetric cell division and apico-basal spindle orientation was later confirmed by crystallographic assays and in vitro binding assays (Culurgioni *et al.*, 2011). When the same was performed for the men subgroup, the decision tree selected cDNA FLJ10929 fis, clone OVARC1000479, highly similar to Cullin-associated NEDD8-dissociated protein 1 as the third and male-specific protein. This protein was detected at the transcript level only. It is not present in children and younger adults but appears in older adults (> 65 years old).

2.3. Validation of biomarkers that indicate a response to treatment

2.3.1. Objectives

The validation of biomarkers that indicate a response to treatment was performed to independently verify the findings of in-depth studies conducted by other research groups at the Institute of Molecular and Translational Medicine.

STUDY 1 – Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells.

In this study, the biomarker validation was a part of the proteomic profiling study. The protein profile was analyzed to evaluate the cellular response to treatment by oxaliplatin, an antineoplastic agent widely used for cancer treatment alone or in combination with other drugs since its approval in 2005. This study included high resolution mass spectrometric approach with SILAC quantification method to reveal differential proteomic profile after oxaliplatin treatment. The identified protein biomarkers of the response to oxaliplatin treatment such as cyclin B1 (CCNB1), probable U3 small nucleolar RNA-associated protein 11 (UTP11L), WD repeat-containing protein 46 (WDR46), probable ATP-dependent RNA helicase DDX56 (DDX56), 40S ribosomal protein S19 (RPS19) and cellular tumor antigen p53 (TP53) were further validated by several methods including Western blot (Ozdian T, D. Holub, Z. Maceckova, L. Varanasi, G. Rylova, J. Rehulka, **J. Vaclavkova**, H. Slavik, P. Moudry, P. Znojek, J. Stankova, J. B. de Sanctis, M. Hajduch, P. Dzubak. Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells. Journal of Proteomics 2017 Jun; 162:73-85. DOI: 10.1016/j.jprot.2017.05.005. ISSN: 1874-3919; IF: 3.867; PMID: 28478306; **see Appendix 8.2**).

STUDY 2 - Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4. Validation of molecular targets by Drug Affinity Responsive Target Stability (DARTS).

In the second study, a repurposed drug disulfiram was tested as an anticancer agent. Disulfiram is rapidly metabolized to a strong cooper chelator dithiocarb (diethyldithiocarbamate) and forms a CuET (Cu(II)diethyldithiocarbamate) complex (Cvek, 2012). Several epidemiological and experimental approaches including experiments on cellular, *in vivo* animals, patient blood samples and purified protein level, led to discovering the mechanism of action of disulfiram that targets cancer via p97 segregase adaptor NPL4. Drug affinity responsive target stability (DARTS) was performed as another independent approach to validate the drug binding to NPL4 (Skrott Z, M. Mistrik, K. Andersen, S. Friis, D. Majera, J. Gursky, T. Ozdian, J. Bartkova, Z. Turi, P. Moudry, M. Kraus, M. Michalova, **J. Vaclavkova**, P. Dzubak, I. Vrobel, P. Pouckova, J. Sedlacek, A. Miklovicova, A. Kutt, J. Li, J. Mattova, C. Driessen, Q. Dou, J. Olsen, M. Hajduch, B. Cvek, R. Deshaies, J. Bartek. Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4. Nature 2017, 552 (7684), 194-199; ISSN: 0028-0836; IF: 40.137; PMID: 29211715; **see Appendix 8.3**).

2.3.2. Material and methods

2.3.2.1. Cell cultivation and lysis

The T-lymphoblastic leukemia-derived cell line (CCRF-CEM) obtained from ATCC were used for experiments. Cells were cultured in RPMI-1640 medium supplemented with 20% fetal calf serum and antibiotics (100 IU/ml penicillin and 10 μg/ml streptomycin). Cells (CCRF-CEM seeded to final cell concentration of 1×10^6/ml) were treated with a 29.3 μM of oxaliplatin (i. e. concentration equal to 5 × IC 50), which approximately corresponds to the IC 50 value when compared to cell count used in cytotoxic MTT assay. Cells were harvested and centrifuged at 90g for 5 min at 4 °C and then washed twice in phosphate buffer saline (PBS) with protease and phosphatase inhibitors (5 mM sodiumpyrophosphate, 1 mM sodiumorthovanadate, 5 mM sodium fluoride and 1 mM PMSF). After washing, the cells were lysed with SDS lysis buffer (62.5 mM Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 1% mercaptoethanol, 0.5% bromophenol blue), heated for 10 min at 95 °C, and incubated with 2 units of benzonase for one hour at room temperature to prepare a whole cell lysate (Ozdian *et al.*, 2017).

2.3.2.2. SDS-PAGE and Western blot analysis

Total protein concentration was measured using the Pierce 660 nm Protein Assay containing the Ionic Detergent Compatibility Reagent. Samples were loaded to SDS-PAGE with uniform loading 10 μg of total protein and were separated on 4% stacking and 12% separation acrylamide gels. After SDS-PAGE, the proteins were blotted onto a nitrocellulose membrane (0.2 μm pore size, Bio-Rad) using the TransBlot Turbo semi-dry system (Bio-Rad). Then, the membrane was blocked for 1 h with 5% non-fat dry milk in Tris-buffered saline with 0.1% Tween-20 (TBS/T) and incubated with appropriate primary antibodies in 5% v/v BSA and 0.1% sodium azide in TBS/T overnight at 4 °C with agitation. After incubation, the membranes were washed with TBS/T, incubated with a peroxidase-labeled secondary antibody, and visualized using the Luminata Forte peroxidase substrate. Chemiluminescence was recorded using a Li-Cor Odyssey FC instrument equipped with HCD camera. The image analysis and comparison with previous mass spectrometric data were performed (Ozdian *et al.*, 2017).

2.3.2.3. DARTS with purified protein

The target protein NPL4 was obtained in a purified form for the experiment. The purified protein with glutathione (GST) tag for purification was used for the validation experiment. The experiment was performed according to a modified published protocol for DARTS (Lomenick *et al.*, 2011).

Purified wild-type GST–NPL4(WT) and mutant GST–NPL4(MUT) proteins were diluted by 100 mM phosphate buffer, pH 7.4 to the final concentration of 0.03 μg/μl (the concentration was optimized on BSA-carborane system, data not shown). The proteins were treated with CuET (metabolite of disulfiram, Cu(II)diethyldithiocarbamate complex; final concentration of 5 μM; dissolved in DMSO) for 1 h and equal amounts of DMSO were added to the solutions, which served as control samples. Pronase (Sigma-Aldrich) was dissolved in TNC buffer (50 mM Tris-Cl, 50 mM NaCl, 10 mM $CaCl_2$, pH 7.5). The 0.025 μg of pronase was added to 50 μl of protein solution and incubated for 1 h at 37 °C. Samples without pronase served as the non-digested controls. The pronase reaction was stopped by the addition of 5× SDS loading buffer; the samples were boiled at 95 °C for 15 min and loaded on SDS–PAGE gels. After SDS–PAGE, gels were silver-stained and scanned on a GS-800 Calibrated Densitometer (Bio-Rad) or used for Western blot analysis (Skrott *et al.*, 2017).

2.3.3. Results

STUDY 1 - Biomarker validation by Western blot.

As discussed in the theoretical introduction to the methods in part **1.2.2. "Western blot"**, the Western blot validation is a widely used method for the biomarker validation. Here, the response to oxaliplatin treatment was evaluated by gel-based proteomics approach with mass spectrometry identification which revealed the significant regulation of protein biomarkers such as cyclin B1 (CCNB1), probable U3 small nucleolar RNA-associated protein 11 (UTP11L), WD repeat-containing protein 46 (WDR46), probable ATP-dependent RNA helicase DDX56 (DDX56), 40S ribosomal protein S19 (RPS19) and cellular tumor antigen p53 (TP53). The expression level of all of these proteins was evaluated by Western blot (Figure 12).

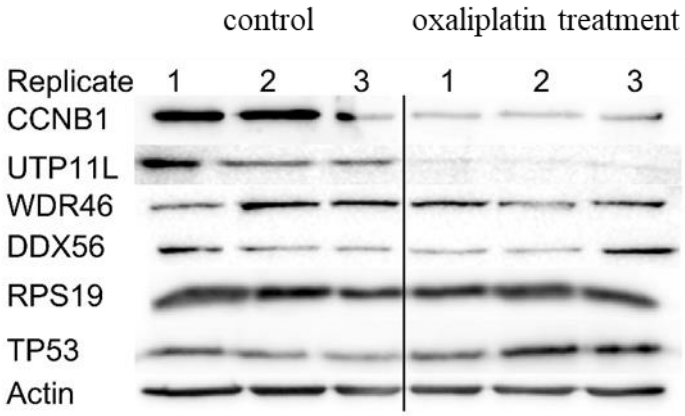


Figure 12: The expression level of studied proteins as a result of the Western blot method. The Western blot verification was performed to validate results obtained by MS. Expression levels of all the proteins showed the same trend (downregulation or upregulation) in both Western blot and MS experiments. Downregulated proteins were cyclin B1 (CCNB1), probable U3 small nucleolar RNA-associated protein 11 (UTP11L), WD repeat-containing protein 46 (WDR46), probable ATP-dependent RNA helicase DDX56 (DDX56) and 40S ribosomal protein S19 (RPS19). Cellular tumor antigen p53 (TP53) protein was upregulated after treatment. For each loading, the cell lysate was prepared independently and the replicates depicted are the three biological replicates. The results here are representative pictures from three Western blot experiments (Ozdian *et al.*, 2017).

In all cases, the results of the Western blot confirmed the data obtained by MS. All reported proteins except p53 were downregulated after oxaliplatin treatment. Cellular tumor antigen p53 was upregulated after oxaliplatin treatment (Figure 12). Further results in this article have shown that p53 is activated by a mechanism independent of the MDM2 pathway. Therefore the p53 protein is not involved in oxaliplatin-induced nucleolar and ribosomal stress in T-lymphoblastic leukemia cells (Ozdian *et al.*, 2017).

STUDY 2 – Biomarker validation by DARTS.

The present study demonstrated the activity of the old anti-alcohol aversion drug disulfiram against cancer cells, and *in vivo* models demonstrated preferential accumulation of the CuET complex in tumors. It was shown that the CuET complex acts as a proteasome inhibitor, which is caused by p97-dependent protein degradation. Inhibition of the p97 pathway is independent of ATPase activity but depends on the p97 segregase, NPL4 and UFD1 protein complex components. CuET has been shown to bind and immobilize NPL4.

The drug affinity responsive target stability (DARTS) is based on the fact that a ligand specifically binds to the target protein and sterically prevents complete digestion of protein by the protease. Our research on the identification and validation of biomarkers using DARTS included determining optimal conditions for the identification and validation of drug targets. It is a useful method to identify the interaction of the native protein with a small molecule. Here, DARTS was chosen as another independent approach to verify the binding of CuET to NPL4.

It was shown that NPL4 was differentially cleaved by pronase after the addition of CuET compared to the mock-treated sample and that full-length NPL4 is cleaved more in the mock-treated sample than in the CuET treated sample due to stabilization and steric protection of protein structure after drug binding. In the case of NPL4 mutant protein, there was no change in pronase cleavage (Figure 13). Concerning the obtained mutated protein, site-directed mutagenesis was used within the amino acid sequence of the putative zinc finger domain of NPL4. It was proved that the CuET does not bind the mutated protein, and thus, the tested zinc finger domain is responsible for the interaction between CuET and NPL4 protein. As before, DARTS was one of the methods that independently verified this result (Skrott *et al.*, 2017).

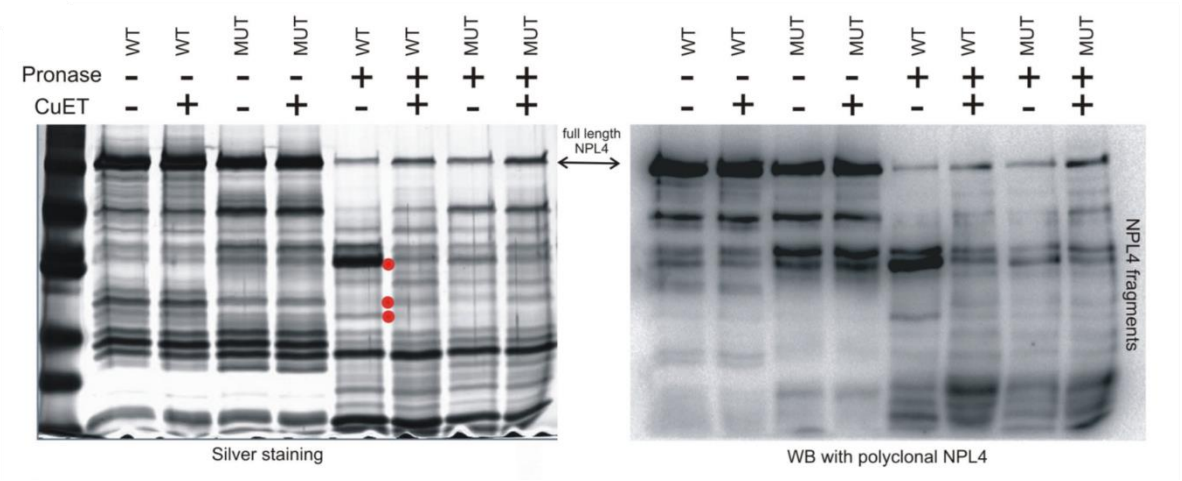


Figure 13. Drug affinity responsive target stability (DARTS) experiments with recombinant NPL4 proteins showed that differential pronase-mediated proteolysis after CuET addition is apparent for wild-type NPL4 protein (WT) but not for mutated protein (MUT). The differential cleavage by pronase was detected by either silver-stained SDS–PAGE (red dots mark the most prominent differential bands) or by blotting with an anti-NPL4 polyclonal antibody. The results depicted are representative of two independent experiments (Skrott *et al.*, 2017).

The application of DARTS successfully confirmed the NPL4 protein as a molecular target of disulfiram. These results showed us that DARTS is an advantageous and relatively fast method in the case of the availability of purified recombinant protein.

2.3.4. Discussion

Western blot is a method that is widely used to validate protein biomarkers previously identified by MS. This validation was also used in studies where, as in our study, biomarkers of cellular response to treatment were evaluated. Western blot was one of the methods used to validate identified proteins in a SILAC-based MS experiment that revealed mechanisms of cellular resistance to paclitaxel (Xu *et al.*, 2015) and also in an iTRAQ-based MS experiment to reveal etoposide resistance mechanisms (Paul *et al.*, 2016). Confirmation of MS results by Western blot was performed by validation of markers from high-grade osteosarcoma in patients after cryotherapy and irradiation (Madda *et al.*, 2020). As in these studies, we have decided to choose Western blot as a relatively fast and accurate validation approach. Immunoprecipitation, ELISA, flow cytometry, microscopic methods and PCR method were also used to reveal other mechanisms of action and protein interactions, Proteins for which the level of expression has been validated are those whose regulation triggers a response to DNA damage, and thus, nucleolar and ribosomal stress (Ozdian *et al.*, 2017).

Unlike the Western blot method, the DARTS has never been used to verify biomarkers or molecular targets. The original work published by the Lomenick's group (Lomenick *et al.*, 2009, 2011) was focused mainly on compounds with unknown molecular targets. Thus, molecular targets were identified by comparing proteolytic patterns of treated and untreated complex cellular lysates. Our work was focused on the interaction of purified proteins with a small molecule ligand. The application of DARTS to purified proteins was also mentioned in Lomenick's dissertation but was never published. However, we found this application very useful. A limitation of the method is that the efficiency of pronase digestion depends on the 2D and 3D structure of the protein, thus the digestion conditions are not universal and applicable to all proteins. Some proteins are more resistant to protease and no digestion is observed using the same conditions as for NPL4. In such cases, digestion conditions (temperature, the concentration of pronase, time of digestion) must be re-optimized.

3 SUMMARY

The presented thesis focuses on the application of proteomic methods such as mass spectrometry (MS), Western blot and Drug Affinity Responsive Target Stability (DARTS) for biomarker discovery and validation. In experimental work, we studied exhaled breath condensate as an excellent tool for monitoring the condition of human lungs and describing a method for discovering biomarkers of respiratory and systemic diseases. The second part of our work involved the validation of protein biomarkers that were recently identified at IMTM.

Exhaled breath condensate (EBC) is a biological matrix collected non-invasively that is a potential biomarkers source. In the proposed thesis, we developed a gel-free mass spectrometric approach for in-depth characterization of the EBC proteome (**section 2.2**). Our method involved freeze-drying sample concentration, digestion of proteins in-solution, purifying the peptides using Stage Tips and HPLC-MS/MS-Orbitrap identification. The search for proteins and peptides was performed in the Proteome Discoverer^{TM} tool (Thermo Fisher Scientific). The optimal parameters were set in cooperation with Thermo Fisher Scientific developers to obtain the most reliable identifications. Using our approach, we were able to identify many more proteins than reported in any of the studies published previously.

In the first EBC study, we reported 2797 proteins across 27 analyzed samples. It was almost ten times higher than in previous publications. To date, the highest number of identifications was 348 different proteins in a cohort of 192 individuals in a study which compared lung cancer and COPD patients and controls (López-Sánchez *et al.*, 2017). We have shown that our MS method is quite sensitive and highly reproducible between individual measurements and is a suitable tool for subsequent biomarkers studies. In tests of the variability of the reported results, we have shown that the difference in the abundances of proteins or peptides that are distant less than 10% from the mean is much more the consequence of the MS measurement deviation than the biological variability of the samples. In addition, the distance from the mean up to 20% could be caused by method variability. It should not be considered a change in protein expression when the data is evaluated using a label-free MS approach (**Václavková *et al.*, 2020**).

In the second EBC study, we analyzed the age-related changes in protein expression. We identified 3202 proteins across 248 individuals' samples. We have reported the changes in the protein profile of EBC according to age. The largest difference was observed in the children and young adult groups (under 18 years of age) and adult's groups (adults under 65 years of

age and the elderly over 65 years of age). Slight differences were found between adult and elderly groups, and age-related markers between the male and female subgroups were also reported. Incorrect age prediction of a group of adults and the elderly can be caused by different biological ages of individuals. However, age-related changes in protein profiles may predict resistance or susceptibility to age-related diseases (**Václavková *et al.*, unpublished data**).

The presented work continued with the verification of biomarkers (**section 2.3**). In the first validation study, we validated biomarkers of the cellular response to oxaliplatin treatment. These biomarkers were identified in a proteomic profiling study based on mass spectrometry. The study was performed on an oxaliplatin-treated T-lymphoblastic leukemia cell line. Following oxaliplatin treatment, Western blot confirmed the downregulation of cyclin B1, probable U3 small nucleolar RNA-associated protein 11, WD repeat-containing protein 46, probable ATP-dependent RNA helicase DDX56 and 40S ribosomal protein S19 and upregulation of the cellular tumor antigen p53 (Ozdian *et al.*, 2017).

The last study in the thesis and the second biomarker validation study provided independent evidence that the NPL4 protein is the target of disulfiram. Here, the old drug disulfiram was tested for anticancer activity. We used the DARTS method, which is based on the fact that the ligand binds specifically to the protein of interest and prevents it from being entirely digested by the protease. We applied the DARTS method to both wild-type and mutant forms of NPL4 protein. We reported that the wild-type NPL4 after disulfiram treatment showed a differential protease cleavage pattern compared to the mock-treated protein. In the case of the mutant protein, proteolytic cleavage showed the same pattern in disulfiram and the mock-treated protein. Thus, the DARTS approach confirmed that disulfiram binds the NPL4 protein and that mutations within the amino acid sequence of the putative zinc finger domain of NPL4 prevent disulfiram binding (Skrott *et al.*, 2017).

In conclusion, mass spectrometry methods are very sensitive and widely used in the discovery of protein biomarkers. Our MS-based approach for in-depth characterization of the EBC proteome has proven to be highly reproducible and applicable to biomarker discovery studies. We have achieved excellent proteome coverage thanks to samples' preconcentration, gel-free sample preparation, Orbitrap-based MS analysis, and a highly efficient protein search tool. We have also proved that exhaled breath condensate is a suitable matrix for the biomarker discovery. Non-invasive diagnostics of respiratory diseases from the human breath could replace or complement bronchoscopy or other recently used invasive techniques. Analysis of exhaled breath would be relatively inexpensive and much more pleasant for patients.

To conclude the biomarker validation studies, current methods are a relatively quick and inexpensive way to confirm the conclusions of long-running experiments that require expensive equipment. Western blot is a widely used method for the validation of biomarkers identified by mass spectrometry. On the other hand, DARTS has not been used to validate biomarkers in any current publication. It was originally used for molecular target discovery, where proteolytic patterns were compared between treated and untreated cell lysates. Our application of DARTS is an excellent tool for evaluation of drug – protein interactions and for independent validation whether the identified protein is the true target of the drug of interest.

4 SOUHRN

Předkládaná dizertační práce se zabývá využitím proteomických metod, jako jsou hmotnostní spektrometrie (MS), Western blot a metoda stability molekulárního cíle závislé na afinitě léčiva (Drug Affinity Responsive Target Stability; DARTS), k identifikaci a validaci biomarkerů. Experimentální část je zaměřena na kondenzát vydechovaného vzduchu, který je vynikajícím biologickým materiálem pro monitorování stavu lidských plic a který byl využit k vývoji metody pro objev biomarkerů respiračních a systémových onemocnění. Druhá část experimentální práce se týká validace biomarkerů, které byly nedávno identifikovány jinými výzkumnými skupinami na ÚMTM.

Kondenzát vydechovaného vzduchu (EBC) je biologická matrice, která je odebírána neinvazivně a je bohatým zdrojem biomarkerů. V této dizertační práci jsme vyvinuli hmotnostně spektrometrický přístup s metodou zpracování vzorků bez předchozí elektroforetické separace. Tento přístup vedl k detailní charakterizaci proteomu EBC (viz **kapitola 2.2**). Součástí naší metody je zakoncentrování vzorku vysušením mrazem, štěpení proteinů na peptidy v roztoku, přečištění peptidů s využitím Stage Tips technologie a měření na přístroji HPLC-MS/MS-Orbitrap. Vyhledání proteinů a peptidů z hmotnostně spektrometrických dat bylo provedeno v softwaru Proteome Discoverer^{TM} (Thermo Fisher Scientific), přičemž nastavení optimálních parametrů pro vyhledávání bylo provedeno ve spolupráci s developery softwaru z firmy Thermo Fisher Scientific tak, abychom získali co nejspolehlivější data. S využitím našeho přístupu se podařilo identifikovat více proteinů než ve studiích doposud publikovaných.

V první studii týkající se EBC jsme identifikovali 2797 proteinů napříč všemi 27 analyzovanými vzorky. Tento počet je téměř desetkrát vyšší než počet proteinů identifikovaných v předchozích publikacích. Doposud nejvyšší identifikovaný počet proteinů byl 348 a tento počet byl publikován ve studii, jež zahrnovala 192 jedinců a srovnávala proteinový profil pacientů s nádory plic, s chronickou obstrukční plicní nemocí (CHOPN) a zdravých kontrol (López-Sánchez *et al.*, 2017). Dokázali jsme, že naše MS metoda je opravdu citlivá a jednotlivá měření jsou vysoce reprodukovatelná. Tato metoda je také vhodným nástrojem pro další studie, které se zaměří na konkrétní biomarkery. Když jsme testovali variabilitu našich výsledků, dokázali jsme, že pokud je vzdálenost abundancí proteinů a peptidů od průměru menší než 10%, je to mnohem pravděpodobněji dáno odchylkou MS měření než biologickou variabilitou vzorků. Dále jsme dospěli k závěru, že pokud je vzdálenost od průměru

menší než 20%, je to pravděpodobně způsobeno variabilitou metody a tato změna by, pokud používáme label-free techniky hmotnostní spektrometrie, neměla být považována za změnu v expresi proteinu (**Václavková *et al.*, 2020**).

Ve druhé studii týkající se EBC jsme analyzovali a popsali změny v proteinové expresi v závislosti na věku. Identifikovali jsme 3202 proteinů napříč souborem 248 jedinců. Největší rozdíl byl pozorován mezi skupinou dětí a mladších dospělých (do 18 let věku) a dvěma skupinami dospělých (dospělými do 65 let věku a dospělými pokročilejšího věku nad 65 let). Pouze malé rozdíly v identifikovaných proteinových profilech byly patrné mezi skupinou dospělých a dospělých pokročilejšího věku. Dále byly patrné rozdíly v biomarkerech spojených s věkem mezi podskupinami žen a mužů. Nesprávná predikce věku mezi skupinami dospělých a dospělých pokročilejšího věku může být způsobena odlišným biologickým věkem studovaných jedinců. Ovšem změny v proteinovém profilu v závislosti na věku mohou predikovat rezistenci, nebo náchylnost k nemocem spojeným s věkem (**Václavková *et al.*, nepublikovaná data**).

Dizertační práce dále pokračuje projekty zaměřenými na validací identifikovaných biomarkerů (**kapitola 2.3**). V první validační studii jsme validovali biomarkery buněčné odpovědi na ošetření oxaliplatinou. Tyto biomarkery byly identifikovány jako součást studie, která byla zaměřena na proteomické profilování s využitím hmotnostní spektrometrie. V této studii byla buněčná linie odvozená od T-lymfoblastické leukémie ošetřena oxaliplatinou a byla zkoumána buněčná odpověď. Metoda Western blot potvrdila, že exprese proteinů CCNB1, UTP11L, WDR46, DDX56, RPS19 byla snížena a exprese nádorového antigenu p53 byla zvýšena (Ozdian *et al.*, 2017).

Poslední studie v dizertační práci a současně druhá studie týkající se validace biomarkerů přinesla nezávislý důkaz, že protein NPL4 je molekulárním cílem disulfiramu. Jde o součást studie, v níž byly zkoumány protinádorové účinky disulfiramu - Antabusu, který se využívá k léčbě alkoholismu. V tomto případě jsme použili metodu DARTS, která je založena na tom, že se malá molekula specificky váže na protein a tento cílový protein stericky chrání proti kompletnímu rozštěpení proteázou. Metodu DARTS jsme aplikovali na standardní i mutantní protein NPL4. Bylo zjištěno, že pokud se na standardní protein NPL4 naváže disulfiram, jsou po proteolytickém štěpení pozorovány jiné štěpné produkty než v případě proteinu NPL4 bez léčiva. V případě mutantního NPL4 proteinu jsou po proteolýzy pozorovány stejné štěpné produkty bez ohledu na to, zda je tento protein disulfiramem ošetřen, či nikoliv. S využitím metody DARTS jsme tedy dokázali, že se disulfiram váže na protein NPL4 a že mutace

v aminokyselinové sekvenci domény zinkového prstu proteinu NPL4 brání navázání disulfiramu (Skrott *et al.*, 2017)

Závěrem lze říci, že hmotnostně spektrometrické metody jsou velmi citlivé a jsou hojně využívány při objevování proteinových biomarkerů. Náš přístup založený na hmotnostní spektrometrii vedl k detailní charakterizaci proteomu EBC. Ukázalo se, že tento přístup je vysoce reprodukovatelný a lze jej aplikovat ve studiích zabývajících se identifikací nových biomarkerů. Díky zakoncentrování vzorků, přípravě vzorků bez předchozí elektroforetické separace, MS analýzy s využitím Orbitrapu a velmi efektivního nástroje pro vyhledávání dat (ProteomDiscover) jsme dosáhli velmi dobrého pokrytí proteomu. Také jsme potvrdili, že kondenzát vydechovaného vzduchu je vhodnou biologickou matricí pro hledání biomarkerů. Neinvazivní diagnostika respiračních onemocnění z lidského dechu může v budoucnu doplnit, nebo nahradit bronchoskopii a další používané invazivní techniky odběru vzorků od pacientů. Analýza vydechovaného vzduchu by mohla být poměrně levnou a pro pacienty příjemnější alternativou.

Co se týče shrnutí studií zaměřených na validaci biomarkerů, lze říci, že zmíněné metody jsou relativně rychlou a levnou cestou, jak potvrdit výsledky dlouhodobých experimentů, které vyžadují nákladné přístrojové vybavení. Western blot je velmi rozšířenou metodou používanou pro validaci biomarkerů, které byly identifikovány pomocí hmotnostní spektrometrie. Na druhou stranu, metoda DARTS ještě v žádné publikaci nebyla použita pro validaci identifikovaných biomarkerů. Tato metoda byla původně využita k identifikaci molekulárních cílů tak, že byly porovnávány výsledné produkty proteolytického štěpení mezi dvěma vzorky, z nichž jedním byl buněčný lyzát ošetřený léčivem a druhým buněčný lyzát neošetřený léčivem. Naše využití metody DARTS se ukázalo být skvělým nástrojem pro ověření interakce mezi léčivem a proteinem a pro nezávislou validaci, zda je identifikovaný protein opravdu cílem studovaného léčiva.

5 ABBREVIATIONS

$^{12}C_6$-	"light" form of amino acid, containing naturally occurring carbon isotope
$^{13}C_6$-	"heavy" form of amino acid, containing one proton heavier carbon isotope than naturally occurring one
$^{13}CD_3$	one carbon-13 and three deuterium atoms
^{15}N-	heavy" form of amino acid, containing one proton heavier nitrogen isotope than naturally occurring one
1-D	one-dimensional
2-D	two-dimensional
2-DE	two-dimensional gel electrophoresis
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
7-T LTQ-FT	7-Tesla linear ion trap quadrupole combined with a Fourier transform ion cyclotron resonance mass analyzer
AATD	α_1-antitrypsin deficiency
Akt	protein kinase B (Akt kinase); serine/threonine-specific protein kinase
ANOVA	analysis of variance
AQUA	absolute quantification of proteins and their modification states
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
BSA	bovine serum albumin
C18	18-carbon chain
CART	classification and regression tree
CC16	Clara cell protein
CCNB1	cyclin B1
CCRF-CEM	acute lymphoblastic leukemia cell line derived from human lymphoblasts from peripheral blood of a child with acute leukemia
cDNA	complementary DNA
CID	collision-induced dissociation
CK	cytokeratin

COPD	chronic obstructive pulmonary disease
COSMIC	catalogue of somatic mutations in cancer
CST1	cystatin SN
ctDNA	circulating tumor DNA
CuET	Cu(II)diethyldithiocarbamate complex
Da	Dalton unit
DARTS	drug affinity responsive target stability
DDA	data-dependent acquisition
DDX56	probable ATP-dependent RNA helicase DDX56
DIA	data-independent acquisition
DIGE	difference gel electrophoresis
DMBT1	deleted in malignant brain tumors 1 protein
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
E. coli	Escherichia coli
EBC	exhaled breath condensate
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
ELISA	enzyme-linked immunosorbent assay
ESI	electrospray ionization
ESI-ION TRAP LCQ-Advantage	electrospray ion-trap mass spectrometer with MS^n-capability
FAIMS	front-end high field asymmetric waveform ion mobility spectrometry
FDA	USA Food and Drug Administration
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded (tissue)
FOXN1	forkhead box protein N1
FT-ARM	Fourier transform – all reaction monitoring
FT-ICR	Fourier transform ion cyclotron resonance spectroscopy
FTMS	Fourier transform mass spectrometry
GSK3	glycogen synthase kinase 3
GST	glutathione
HCD	higher-energy collisional dissociation

HE4	human epididymis protein 4
HeLa cells	Henrietta Lacks cells, immortal cell line derived from cervical cancer cells
HIAR	heat-induced antigen retrieval
HMG-I/Y protein family	high mobility group I/Y protein family
HPLC	high performance liquid chromatography
HPLC-MS	high performance liquid chromatography coupled to mass spectrometry
HPLC-MS/MS	high performance liquid chromatography coupled to tandem mass spectrometry
HSF1	heat shock factor 1
i.d.	inner diameter
i.e.	id est; that is to say
IC 50	half maximal inhibitory concentration
ICAT	isotope-coded affinity tags
ICR	ion cyclotron resonance spectroscopy
IEF	isoelectric focusing
IHC	immunohistochemistry
IL-6	interleukin 6
IMTM	Institute of Molecular and Translational Medicine
Insc	protein inscuteable homolog
IT	ion trap
iTRAQ	isobaric tags for relative and absolute quantification
IU	international unit
JNK	c-Jun N-terminal kinase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LASSO	least absolute shrinkage and selection operator
LC	liquid chromatography
LC/LC	multidimensional liquid chromatography
LC/LC-MS/MS	multidimensional liquid chromatography coupled to tandem mass spectrometry
LCDH	left-sided congenital diaphragmatic hernia
LC-MS	liquid chromatography coupled to mass spectrometry
LC-MS/MS	liquid chromatography coupled to tandem mass spectrometry

Leu-d3	deuterated leucine
LTQ	linear trap quadrupole
LTQ-FT	linear ion trap quadrupole combined with a Fourier transform ion cyclotron resonance mass analyzer
m/z	mass to charge ratio
MALDI	matrix-assisted laser desorption/ionization
MAPK	mitogen-activated protein kinase
MDM2	mouse double minute 2 homolog; E3 ubiquitin-protein ligase Mdm2
MIKE	mass-analyzed ion kinetic energy
MRM	multiple reaction monitoring
mRNA	messenger RNA
mRNAseq	messenger RNA sequencing
MS	mass spectrometry, mass spectrometer
MS/MS	tandem mass spectrometry
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide; tetrazolium salt used for cytotoxicity assay
MUT	mutant
Na/K-ATPase	sodium–potassium adenosine triphosphatase
nanoLC/nanoUPLC	(ultra performance) liquid chromatography with nanoliter flow
NCI-60	National Cancer Institute – 60 human cancer cell lines for the screening of compounds to detect potential anticancer activity
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NPL4	nuclear protein localization protein 4 homolog
Nrf2	nuclear factor erythroid 2-related factor 2
NSCLC	non-small lung cancer
p97	ATPase, also called valosin-containing protein (VCP)
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PCT – SWATH	pressure cycling technology – sequential windowed acquisition of all theoretical fragment ion
PDCD4	programmed cell death protein 4
pH	decimal logarithm of the reciprocal of the hydrogen ion activity
PI3K	phosphatidylinositol 3-kinase

PMSF	phenylmethylsulfonyl fluoride
PRM	parallel reaction monitoring
PSA	prostate specific antigen
PSM	peptide sequence match
Q	quadrupole
QqTOF	quadrupole-quadrupole-time-of-flight mass analyzer
Q-TOF	quadrupole – time-of-flight hybrid mass analyzer
Ras	“Rat sarcoma” protein; small GTPase, a cell signaling protein
RNA	ribonucleic acid
ROS	reactive oxygen species
RP	reverse phase
RPLC	reverse phase liquid chromatography
RPMI-1640	growth medium used in cell culture, name come from Roswell Park Memorial Institute
RPPA	reverse-phase protein array
RPS19	40S ribosomal protein S19
RT PCR	real-time polymerase chain reaction
RT	retention time
S. aureus	Staphylococcus aureus
SARS-CoV-2	severe acute respiratory syndrome-related coronavirus
sd	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SELDI	surface enhanced laser desorption/ionization
shRNA	short hairpin RNA
SILAC	stable isotope labeling by amino acids in cell culture
SP-A	surfactant protein A
Src	proto-oncogene tyrosine-protein kinase from Src protein family
SRM	selected reaction monitoring
SWATH	sequential windowed acquisition of all theoretical fragment ion
TBS/T	Tris-buffered saline with 0.1% Tween-20
TEZ buffer	Tris, EDTA and Zwittergent buffer
TFA	trifluoroacetic acid

TiMS TOF	trapped ion mobility spectrometry coupled to time-of-flight mass analyzer
TMT	tandem mass tags
TNC buffer	buffer containing Tris-Cl, NaCl, $CaCl_2$
TOF	time-of-flight mass analyzer
TOF/TOF	hybrid time-of-flight mass analyzer enabling tandem mass spectrometry analysis
TP53	cellular tumor antigen p53
TRAIL	tumor necrosis factor (TNF)-related apoptosis-inducing ligand
Turbo DECCS	transportable unit for research of biomarkers obtained and disposable exhaled condensate collection system
UFD1	ubiquitin recognition factor in ER-associated degradation protein 1
UPLC	ultra performance liquid chromatography
UTP11L	probable U3 small nucleolar RNA-associated protein 11
UV	ultraviolet
VEGF	vascular endothelial growth factor
WDR46	WD repeat-containing protein 46
Wnt	name created from the names Wingless and Int-1; signal transduction pathway
WT	wild-type

6 REFERENCES

3.5.1., R. D. C. T. (2018) ‘A Language and Environment for Statistical Computing’, *R Foundation for Statistical Computing*.

Ali, M. *et al.* (2018) ‘Global proteomics profiling improves drug sensitivity prediction: Results from a multi-omics, pan-cancer modeling approach’, *Bioinformatics*, 34(8), pp. 1353–1362. doi: 10.1093/bioinformatics/btx766.

Anastasi, E. *et al.* (2010) ‘HE4: A new potential early biomarker for the recurrence of ovarian cancer’, *Tumor Biology*, 31(2), pp. 113–119. doi: 10.1007/s13277-009-0015-y.

Azab, S., Osama, A. and Rafaat, M. (2012) ‘Does normalizing PSA after successful treatment of chronic prostatitis with high PSA value exclude prostatic biopsy?’, *Translational Andrology and Urology*, 1(3), pp. 148–152. doi: 10.3978/j.issn.2223-4683.2012.07.02.

Bass, J. J. *et al.* (2017) ‘An overview of technical considerations for Western blotting applications to physiological research’, *Scandinavian Journal of Medicine and Science in Sports*. doi: 10.1111/sms.12702.

Bekker-Jensen, D. B. *et al.* (2020) ‘A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients’, *Molecular and Cellular Proteomics*, 19(4), pp. 716–729. doi: 10.1074/mcp.TIR119.001906.

Blobel, G. A. *et al.* (1984) ‘Cytokeratins in normal lung and lung carcinomas’, *Virchows Archiv B Cell Pathology Including Molecular Pathology*, 45(1), pp. 407–429. doi: 10.1007/bf02889883.

Bloemen, K. *et al.* (2009) ‘Non-invasive bioarker sampling and analysis of the exhaled breath proteome’, *Proteomics - Clinical Applications*, 3(4), pp. 498–504. doi: 10.1002/prca.200800095.

Bloemen, K. *et al.* (2011) ‘A new approach to study exhaled proteins as potential biomarkers for asthma’, *Clinical and Experimental Allergy*, 41(3), pp. 346–356. doi: 10.1111/j.1365-2222.2010.03638.x.

Bowman, S. K. *et al.* (2006) 'The Drosophila NuMA Homolog Mud Regulates Spindle Orientation in Asymmetric Cell Division', *Developmental Cell*, 10(6), pp. 731–742. doi: 10.1016/j.devcel.2006.05.005.

Bredberg, A. *et al.* (2012) 'Exhaled endogenous particles contain lung proteins', *Clinical Chemistry*, 58(2), pp. 431–440. doi: 10.1373/clinchem.2011.169235.

Burnette, W. N. (1981) '"Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A', *Analytical Biochemistry*. doi: 10.1016/0003-2697(81)90281-5.

Byrum, S. *et al.* (2011) 'A quantitative proteomic analysis of FFPE melanoma', *Journal of Cutaneous Pathology*, 38(11), pp. 933–936. doi: 10.1111/j.1600-0560.2011.01761.x.

Cañas Montalvo, B. *et al.* (2006) 'Mass spectrometry technologies for proteomics', *Briefings in Functional Genomics and Proteomics*, 4(4), pp. 295–320. doi: 10.1093/bfgp/eli002.

Cha, Y. *et al.* (2018) 'Drug repurposing from the perspective of pharmaceutical companies', *British Journal of Pharmacology*, 175(2), pp. 168–180. doi: 10.1111/bph.13798.

Chen, D. *et al.* (2006) 'Disulfiram, a clinically used anti-alcoholism drug and copper-binding agent, induces apoptotic cell death in breast cancer cultures and xenografts via inhibition of the proteasome activity', *Cancer Research*, 66(21), pp. 10425–10433. doi: 10.1158/0008-5472.CAN-06-2126.

Chen, Jinliang *et al.* (2019) 'Epidermal Growth Factor in Exhaled Breath Condensate as Diagnostic Method for Non-Small Cell Lung Cancer', *Technology in cancer research & treatment*, 18, pp. 1–5. doi: 10.1177/1533033819872271.

Choe, L. *et al.* (2007) '8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease', *Proteomics*. doi: 10.1002/pmic.200700316.

Ciliberto, G., Mancini, R. and Paggi, M. G. (2020) ‘Drug repurposing against COVID-19: Focus on anticancer agents’, *Journal of Experimental and Clinical Cancer Research*. Journal of Experimental & Clinical Cancer Research, 39(1), pp. 1–9. doi: 10.1186/s13046-020-01590-2.

Cohen, G. H. *et al.* (1986) ‘Localization of discontinuous epitopes of herpes simplex virus glycoprotein D: use of a nondenaturing (“native” gel) system of polyacrylamide gel electrophoresis coupled with Western blotting.’, *Journal of Virology*, 60(1), pp. 157–166. doi: 10.1128/jvi.60.1.157-166.1986.

Collins, F. S. (2011) ‘Mining for therapeutic gold’, *Nature Reviews Drug Discovery*. Nature Publishing Group, 10(6), p. 397. doi: 10.1038/nrd3461.

Comisarow, M. B. and Marshall, A. G. (1974) ‘Fourier transform ion cyclotron resonance spectroscopy’, *Chemical Physics Letters*. doi: 10.1016/0009-2614(74)89137-2.

Coorssen, J. R. *et al.* (2002) ‘Quantitative femto- to attomole immunodetection of regulated secretory vesicle proteins critical to exocytosis’, *Analytical Biochemistry*, 307(1), pp. 54–62. doi: 10.1016/S0003-2697(02)00015-5.

Culurgioni, S. *et al.* (2011) ‘Inscuteable and NuMA proteins bind competitively to Leu-Gly-Asn repeat-enriched protein (LGN) during asymmetric cell divisions’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(52), pp. 20998–21003. doi: 10.1073/pnas.1113077108.

Cvek, B. (2012) ‘Nonprofit drugs as the salvation of the world’s healthcare systems: The case of Antabuse (disulfiram)’, *Drug Discovery Today*. Elsevier Ltd, 17(9–10), pp. 409–412. doi: 10.1016/j.drudis.2011.12.010.

Damia, G. and Broggini, M. (2004) ‘Improving the selectivity of cancer treatments by interfering with cell response pathways’, *European Journal of Cancer*, 40(17 SPEC. ISS.), pp. 2550–2559. doi: 10.1016/j.ejca.2004.07.020.

Dayon, L. *et al.* (2008) ‘Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags’, *Analytical Chemistry*, 80(8), pp. 2921–2931. doi: 10.1021/ac702422x.

Dewson, G. (2015) ‘Blue native PAGE and antibody gel shift to assess Bak and Bax conformation change and oligomerization’, *Cold Spring Harbor Protocols*, 2015(5), pp. 485–489. doi: 10.1101/pdb.prot086488.

Drabovich, A. P., Martínez-Morillo, E. and Diamandis, E. P. (2019) ‘Protein Biomarker Discovery’, *Proteomics for Biological Discovery*, pp. 63–88. doi: 10.1002/9781119081661.ch3.

Dufour, P. *et al.* (1993) ‘Sodium ditiocarb as adjuvant immunotherapy for high risk breast cancer: A randomized study’, *Biotherapy*. doi: 10.1007/BF01877380.

Duskova, K. and Vesely, S. (2015) ‘Prostate specific antigen. Current clinical application and future prospects’, *Biomedical Papers*, 159(1), pp. 18–26. doi: 10.5507/bp.2014.046.

Everley, P. A. *et al.* (2004) ‘Quantitative cancer proteomics: Stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research’, *Molecular and Cellular Proteomics*, 3(7), pp. 729–735. doi: 10.1074/mcp.M400021-MCP200.

Fedorchenko, K. U. *et al.* (2016) ‘Early diagnosis of lung cancer based on proteome analysis of exhaled breath condensate’, *Moscow University Chemistry Bulletin*, 71(2), pp. 134–139. doi: 10.3103/S0027131416020036.

Fleissig, Y. *et al.* (2010) ‘Comparative proteomic analysis of human oral fluids according to gender and age’, *Oral Diseases*, 16(8), pp. 831–838. doi: 10.1111/j.1601-0825.2010.01696.x.

Freedland, S. J. *et al.* (2003) ‘Defining the ideal cutpoint for determining PSA recurrence after radical prostatectomy’, *Urology*, 61(2), pp. 365–369. doi: 10.1016/S0090-4295(02)02268-9.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) ‘Regularization paths for generalized linear models via coordinate descent’, *Journal of Statistical Software*. doi: 10.18637/jss.v033.i01.

Fumagalli, M. *et al.* (2008) ‘Proteomic analysis of exhaled breath condensate from single patients with pulmonary emphysema associated to α1-antitrypsin deficiency’, *Journal of Proteomics*, 71(2), pp. 211–221. doi: 10.1016/j.jprot.2008.03.002.

Fumagalli, M. *et al.* (2012) ‘Profiling the proteome of exhaled breath condensate in healthy smokers and COPD patients by LC-MS/MS’, *International Journal of Molecular Sciences*, 13(11), pp. 13894–13910. doi: 10.3390/ijms131113894.

Gao, Y., Holland, R. D. and Yu, L. R. (2009) 'Quantitative proteomics for drug toxicity', *Briefings in Functional Genomics and Proteomics*, 8(2), pp. 158–166. doi: 10.1093/bfgp/elp006.

Gerber, S. A. *et al.* (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(12), pp. 6940–6945. doi: 10.1073/pnas.0832254100.

Gessner, C. *et al.* (2008) 'Presence of cytokeratins in exhaled breath condensate of mechanical ventilated patients', *Respiratory Medicine*, 102(2), pp. 299–306. doi: 10.1016/j.rmed.2007.08.012.

Ghosh, R., Gilda, J. E. and Gomes, A. V. (2014) 'The necessity of and strategies for improving confidence in the accuracy of western blots', *Expert Review of Proteomics*. doi: 10.1586/14789450.2014.939635.

Gianazza, E. *et al.* (2004) 'Increased keratin content detected by proteomic analysis of exhaled breath condensate from healthy persons who smoke', *American Journal of Medicine*, 117(1), pp. 51–54. doi: 10.1016/j.amjmed.2004.01.022.

Gillet, L. C. *et al.* (2012) 'Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis', *Molecular and Cellular Proteomics*, 11(6), pp. 1–17. doi: 10.1074/mcp.O111.016717.

Goldman, A., Harper, S. and Speicher, D. W. (2016) 'Detection of proteins on blot membranes', *Current Protocols in Protein Science*. doi: 10.1002/cpps.15.

Griese, M., Noss, J. and Von Bredow, C. (2002) 'Protein pattern of exhaled breath condensate and saliva', *Proteomics*, 2(6), pp. 690–696. doi: 10.1002/1615-9861(200206)2:6<690::AID-PROT690>3.0.CO;2-6.

Gygi, S. P. *et al.* (1999) 'Quantitative analysis of complex protein mixtures using isotope-coded affinity tags', *Nature Biotechnology*. doi: 10.1038/13690.

Hale, J. E. *et al.* (2003) 'Application of proteomics for discovery of protein biomarkers', *Briefings in Functional Genomics and Proteomics*, 2(3), pp. 185–193. doi: 10.1093/bfgp/2.3.185.

Holub, D. *et al.* (2019) ‘Mass Spectrometry Amyloid Typing Is Reproducible across Multiple Organ Sites’, *BioMed Research International*, 2019. doi: 10.1155/2019/3689091.

Horváth, I. *et al.* (2005) ‘Exhaled breath condensate: Methodological recommendations and unresolved questions’, *European Respiratory Journal*, 26(3), pp. 523–548. doi: 10.1183/09031936.05.00029705.

Huang, J. *et al.* (2018) ‘Repurposing psychiatric drugs as anti-cancer agents’, *Cancer Letters*. Elsevier Ltd, 419, pp. 257–265. doi: 10.1016/j.canlet.2018.01.058.

Ibarrola, N. *et al.* (2003) ‘A Proteomic Approach for Quantitation of Phosphorylation Using Stable Isotope Labeling in Cell Culture’, *Analytical Chemistry*, 75(22), pp. 6043–6049. doi: 10.1021/ac034931f.

Izaki, T. *et al.* (2006) ‘Two forms of human Inscuteable-related protein that links Par3 to the Pins homologues LGN and AGS3’, *Biochemical and Biophysical Research Communications*, 341(4), pp. 1001–1006. doi: 10.1016/j.bbrc.2006.01.050.

Ji, X. *et al.* (2018) ‘Forkhead box N1 inhibits the progression of non-small cell lung cancer and serves as a tumor suppressor’, *Oncology Letters*, 15(5), pp. 7221–7230. doi: 10.3892/ol.2018.8210.

Karas, M. and Hillenkamp, F. (1988) ‘Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons’, *Analytical Chemistry*. doi: 10.1021/ac00171a028.

Katoh, Masuko and Katoh, Masaru (2003) ‘Identification and characterization of human Inscuteable gene in silico.’, *International journal of molecular medicine*, 11(1), pp. 111–116. doi: 10.3892/ijmm.11.1.111.

Keppler, D. *et al.* (2011) ‘Novel expression of CST1 as candidate senescence marker’, *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 66 A(7), pp. 723–731. doi: 10.1093/gerona/glr033.

Kim, M. S. *et al.* (2014) ‘A draft map of the human proteome’, *Nature*, 509(7502), pp. 575–581. doi: 10.1038/nature13302.

Kondrat, R. W., McClusky, G. A. and Cooks, R. G. (1978) 'Multiple Reaction Monitoring in Mass Spectrometry/Mass Spectrometry for Direct Analysis of Complex Mixtures', *Analytical Chemistry*, 50(14), pp. 2017–2021. doi: 10.1021/ac50036a020.

Kononikhin, A. S. *et al.* (2017) 'Exhaled breath condensate analysis from intubated newborns by nano-HPLC coupled to high resolution MS', *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*. Elsevier B.V., 1047, pp. 97–105. doi: 10.1016/j.jchromb.2016.12.036.

Kononikhin, A. S. *et al.* (2019) 'Proteome profiling of the exhaled breath condensate after long-term spaceflights', *International Journal of Molecular Sciences*, 20(18). doi: 10.3390/ijms20184518.

Konstantinidi, E. M. *et al.* (2015) 'Exhaled Breath Condensate: Technical and Diagnostic Aspects', *Scientific World Journal*, 2015. doi: 10.1155/2015/435160.

Kurova, V. S. *et al.* (2009) 'Proteomics of exhaled breath: Methodological nuances and pitfalls', *Clinical Chemistry and Laboratory Medicine*, 47(6), pp. 706–712. doi: 10.1515/CCLM.2009.166.

Lacombe, M. *et al.* (2018) 'Proteomic characterization of human exhaled breath condensate', *Journal of Breath Research*. IOP Publishing, 12(2). doi: 10.1088/1752-7163/aa9e71.

Lahm, H. W. and Langen, H. (2000) 'Mass spectrometry: A tool for the identification of proteins separated by gels', *Electrophoresis*, 21(11), pp. 2105–2114. doi: 10.1002/1522-2683(20000601)21:11<2105::AID-ELPS2105>3.0.CO;2-M.

Lange, V. *et al.* (2008) 'Selected reaction monitoring for quantitative proteomics: A tutorial', *Molecular Systems Biology*, 4(222). doi: 10.1038/msb.2008.61.

Larsson, P. *et al.* (2012) 'Surfactant protein A and albumin in particles in exhaled air', *Respiratory Medicine*, 106(2), pp. 197–204. doi: 10.1016/j.rmed.2011.10.008.

Lazovic, J. *et al.* (2015) 'Nitroxoline induces apoptosis and slows glioma growth in vivo', *Neuro-Oncology*, 17(1), pp. 53–62. doi: 10.1093/neuonc/nou139.

Lee, P. Y., Saraygord-Afshari, N. and Low, T. Y. (2020) 'The evolution of two-dimensional gel electrophoresis - from proteomics to emerging alternative applications', *Journal of Chromatography A*. Elsevier B.V., 1615, p. 460763. doi: 10.1016/j.chroma.2019.460763.

Leong, S. *et al.* (2012) 'ITRAQ-based proteomic profiling of breast cancer cell response to doxorubicin and TRAIL', *Journal of Proteome Research*, 11(7), pp. 3561–3572. doi: 10.1021/pr2012335.

Leung, T. F. *et al.* (2005) 'Analysis of growth factors and inflammatory cytokines in exhaled breath condensate from asthmatic children', *International Archives of Allergy and Immunology*, 137(1), pp. 66–72. doi: 10.1159/000085106.

Li, J. *et al.* (2020) 'TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples', *Nature Methods*. Springer US, 17(April). doi: 10.1038/s41592-020-0781-4.

Li, Y. *et al.* (2015) 'Copper improves the anti-angiogenic activity of disulfiram through the EGFR/Src/VEGF pathway in gliomas', *Cancer Letters*, 369(1), pp. 86–96. doi: 10.1016/j.canlet.2015.07.029.

Liu, Y. and Yao, J. (2019) 'Research progress of cystatin SN in cancer', *OncoTargets and Therapy*, 12, pp. 3411–3419. doi: 10.2147/OTT.S194332.

Lomenick, B. *et al.* (2009) 'Target identification using drug affinity responsive target stability (DARTS)', *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), pp. 21984–21989. doi: 10.1073/pnas.0910040106.

Lomenick, B. *et al.* (2011) 'Target Identification Using Drug Affinity Responsive Target Stability (DARTS)', *Current Protocols in Chemical Biology*. doi: 10.1002/9780470559277.ch110180.

López-Sánchez, L. M. *et al.* (2017) 'Exhaled breath condensate biomarkers for the early diagnosis of lung cancer using proteomics', *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 313(4), pp. L664–L676. doi: 10.1152/ajplung.00119.2017.

MacKeigan, J. P. *et al.* (2003) 'Proteomic Profiling Drug-Induced Apoptosis in Non-Small Cell Lung Carcinoma: Identification of RS/DJ-1 and RhoGDIα', *Cancer Research*, 63(20), pp. 6928–6934.

Madda, R. *et al.* (2020) 'Proteomic profiling and identification of significant markers from high-grade osteosarcoma after cryotherapy and irradiation', *Scientific Reports*, 10(1), pp. 1–18. doi: 10.1038/s41598-019-56024-7.

Mahmood, T. and Yang, P. C. (2012) 'Western blot: Technique, theory, and trouble shooting', *North American Journal of Medical Sciences*, 4(9), pp. 429–434. doi: 10.4103/1947-2714.100998.

Makarov, A. (2000) 'Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis', *Analytical Chemistry*. doi: 10.1021/ac991131p.

Mamyrin, B. *et al.* (1973) 'The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution', *Soviet Journal of Experimental and Theoretical Physics*.

Mann, M., Hendrickson, R. C. and Pandey, A. (2001) 'Analysis of proteins and proteomes by mass spectrometry', *Annual Review of Biochemistry*. doi: 10.1146/annurev.biochem.70.1.437.

Mannelli, C. (2019) 'Tissue vs Liquid Biopsies for Cancer Detection: Ethical Issues', *Journal of Bioethical Inquiry*. Journal of Bioethical Inquiry, 16(4), pp. 551–557. doi: 10.1007/s11673-019-09944-y.

Mansfield, M. A. (1995) 'Rapid immunodetection on polyvinylidene fluoride membrane blots without blocking', *Analytical Biochemistry*. doi: 10.1006/abio.1995.1391.

Marx, V. (2013) 'Targeted proteomics', *Nature Methods*. Nature Publishing Group, 10(1), pp. 19–22. doi: 10.1038/nmeth.2285.

McDonald, W. H. and Yates, J. R. (2002) 'Shotgun proteomics and biomarker discovery', *Disease Markers*, 18(2), pp. 99–105. doi: 10.1155/2002/505397.

Medzihradszky, K. F. *et al.* (2001) 'Protein identification by in-gel digestion, high-performance liquid chromatography, and mass spectrometry: Peptide analysis by complementary ionization techniques', *Journal of the American Society for Mass Spectrometry*, 12(2), pp. 215–221. doi: 10.1016/S1044-0305(00)00214-2.

Mozdzanowski, J., Hembach, P. and Speicher, D. W. (1992) 'High yield electroblotting onto polyvinylidene difluoride membranes from polyacrylamide gels', *Electrophoresis*, 13(1), pp. 59–64. doi: 10.1002/elps.1150130112.

Muccilli, V. *et al.* (2015) 'Protein profile of exhaled breath condensate determined by high resolution mass spectrometry', *Journal of Pharmaceutical and Biomedical Analysis*. Elsevier B.V., 105, pp. 134–149. doi: 10.1016/j.jpba.2014.11.050.

Mutlu, G. M. *et al.* (2001) 'Collection and analysis of exhaled breath condensate in humans', *American Journal of Respiratory and Critical Care Medicine*, 164(5), pp. 731–737. doi: 10.1164/ajrccm.164.5.2101032.

Nadler, W. M. *et al.* (2017) 'MALDI versus ESI: The Impact of the Ion Source on Peptide Identification', *Journal of Proteome Research*, 16(3), pp. 1207–1215. doi: 10.1021/acs.jproteome.6b00805.

Nazarian, J. *et al.* (2008) 'Protein profiling of formalin fixed paraffin embedded tissue: Identification of potential biomarkers for pediatric brainstem glioma', *Proteomics - Clinical Applications*, 2(6), pp. 915–924. doi: 10.1002/prca.200780061.

Nechushtan, H. *et al.* (2015) 'A Phase IIb Trial Assessing the Addition of Disulfiram to Chemotherapy for the Treatment of Metastatic Non-Small Cell Lung Cancer', *The Oncologist*, 20(4), pp. 366–367. doi: 10.1634/theoncologist.2014-0424.

Nie, L. *et al.* (2016) 'An optimization of the LC-MS/MS workflow for deep proteome profiling on an Orbitrap Fusion', *Analytical Methods*, 8(2), pp. 425–434. doi: 10.1039/c5ay01900a.

Núñez-Naveira, L., Mariñas-Pardo, L. A. and Montero-Martínez, C. (2019) 'Mass Spectrometry Analysis of the Exhaled Breath Condensate and Proposal of Dermcidin and S100A9 as Possible Markers for Lung Cancer Prognosis', *Lung*. Springer US, 197(4), pp. 523–531. doi: 10.1007/s00408-019-00238-z.

Oda, Y. *et al.* (1999) 'Accurate quantitation of protein expression and site-specific phosphorylation', *Proceedings of the National Academy of Sciences of the United States of America*, 96(12), pp. 6591–6596. doi: 10.1073/pnas.96.12.6591.

Oh, S. S. *et al.* (2017) 'Extracellular cystatin SN and cathepsin B prevent cellular senescence by inhibiting abnormal glycogen accumulation', *Cell Death and Disease*. Nature Publishing Group, 8(4). doi: 10.1038/cddis.2017.153.

Ong, S. E. *et al.* (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(5), pp. 376–386. doi: 10.1074/mcp.M200025-MCP200.

Ong, S. E., Kratchmarova, I. and Mann, M. (2003) 'Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC)', *Journal of Proteome Research*, 2(2), pp. 173–181. doi: 10.1021/pr0255708.

Ong, S. E. and Mann, M. (2007) 'A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)', *Nature Protocols*, 1(6), pp. 2650–2660. doi: 10.1038/nprot.2006.427.

Ong, S. E., Mittler, G. and Mann, M. (2004) 'Identifying and quantifying in vivo methylation sites by heavy methyl SILAC', *Nature Methods*, 1(2), pp. 119–126. doi: 10.1038/nmeth715.

Ortega, M. A. *et al.* (2017) 'Using single-cell multiple omics approaches to resolve tumor heterogeneity', *Clinical and Translational Medicine*. Springer Berlin Heidelberg, 6(1), p. 46. doi: 10.1186/s40169-017-0177-y.

Ozdian, T. *et al.* (2017) 'Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells', *Journal of Proteomics*. doi: 10.1016/j.jprot.2017.05.005.

Park, Y. M. *et al.* (2018) 'Anti-cancer effects of disulfiram in head and neck squamous cell carcinoma via autophagic cell death', *PLoS ONE*, 13(9), pp. 1–12. doi: 10.1371/journal.pone.0203069.

Patton, W. F. (2002) 'Detection technologies in proteome analysis', *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 771(1–2), pp. 3–31. doi: 10.1016/S1570-0232(02)00043-0.

Paul, D. *et al.* (2016) 'Global proteomic profiling identifies etoposide chemoresistance markers in non-small cell lung carcinoma', *Journal of Proteomics*. Elsevier B.V., 138, pp. 95–105. doi: 10.1016/j.jprot.2016.02.008.

Paul, W. and Steinwedel, H. (1953) ‘Quadrupole mass filter’, *Z. Naturforsch. A*.

Pauly, F. *et al.* (2013) ‘Protein expression profiling of formalin-fixed paraffin-embedded tissue using recombinant antibody microarrays’, *Journal of Proteome Research*, 12(12), pp. 5943–5953. doi: 10.1021/pr4003245.

Pérès, S. *et al.* (2008) ‘A new method for 2D gel spot alignment: Application to the analysis of large sample sets in clinical proteomics’, *BMC Bioinformatics*, 9, pp. 1–9. doi: 10.1186/1471-2105-9-460.

Peterson, A. C. *et al.* (2012) ‘Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics’, *Molecular and Cellular Proteomics*, 11(11), pp. 1475–1488. doi: 10.1074/mcp.O112.020131.

Pushpakom, S. *et al.* (2018) ‘Drug repurposing: Progress, challenges and recommendations’, *Nature Reviews Drug Discovery*. Nature Publishing Group, 18(1), pp. 41–58. doi: 10.1038/nrd.2018.168.

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*. doi: 10.1021/ac026117i.

Righetti, P. G. and Candiano, G. (2011) ‘Recent advances in electrophoretic techniques for the characterization of protein biomolecules: A poker of aces’, *Journal of Chromatography A*. Elsevier B.V., 1218(49), pp. 8727–8737. doi: 10.1016/j.chroma.2011.04.011.

Rode, I. *et al.* (2015) ‘Foxn1 Protein Expression in the Developing, Aging, and Regenerating Thymus’, *The Journal of Immunology*, 195(12), pp. 5678–5687. doi: 10.4049/jimmunol.1502010.

Rosias, P. P. *et al.* (2008) ‘Biomarker reproducibility in exhaled breath condensate collected with different condensers’, *European Respiratory Journal*, 31(5), pp. 934–942. doi: 10.1183/09031936.00073207.

Ross, P. L. *et al.* (2004) ‘Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents’, *Molecular and Cellular Proteomics*, 3(12), pp. 1154–1169. doi: 10.1074/mcp.M400129-MCP200.

Safi, R. *et al.* (2014) 'Copper signaling axis as a target for prostate cancer therapeutics', *Cancer Research*, 74(20), pp. 5819–5831. doi: 10.1158/0008-5472.CAN-13-3527.

Scigelova, M. and Makarov, A. (2006) 'Orbitrap mass analyzer - Overview and applications in proteomics', *Proteomics*, 1(1-2 SUPPL.), pp. 16–21. doi: 10.1002/pmic.200600528.

Shahid, S. K. *et al.* (2002) 'Increased interleukin-4 and decreased interferon-γ in exhaled breath condensate of children with asthma', *American Journal of Respiratory and Critical Care Medicine*, 165(9), pp. 1290–1293. doi: 10.1164/rccm.2108082.

Shevchenko, A. *et al.* (1996) 'Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels', *Analytical Chemistry*, 68(5), pp. 850–858. doi: 10.1021/ac950914h.

Skalnikova, H. *et al.* (2011) 'Cancer drug-resistance and a look at specific proteins: Rho GDP-dissociation inhibitor 2, Y-box binding protein 1, and HSP70/90 organizing protein in proteomics clinical application', *Journal of Proteome Research*, 10(2), pp. 404–415. doi: 10.1021/pr100468w.

Skrott, Z. *et al.* (2017) 'Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4', *Nature*. doi: 10.1038/nature25016.

Spillier, Q. *et al.* (2019) 'Anti-alcohol abuse drug disulfiram inhibits human PHGDH via disruption of its active tetrameric form through a specific cysteine oxidation', *Scientific Reports*, 9(1), pp. 1–9. doi: 10.1038/s41598-019-41187-0.

Sun, C. *et al.* (2019) 'Proteomics of exhaled breath condensate in stable COPD and non-COPD controls using tandem mass tags (TMTs) quantitative mass spectrometry: A pilot study', *Journal of Proteomics*. Elsevier, 206(May), p. 103392. doi: 10.1016/j.jprot.2019.103392.

Tessitore, A. *et al.* (2013) 'Serum biomarkers identification by mass spectrometry in high-mortality tumors', *International Journal of Proteomics*, 2013. doi: 10.1155/2013/125858.

Thompson, A. *et al.* (2003) 'Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS', *Analytical Chemistry*, 75(8), pp. 1895–1904. doi: 10.1021/ac0262560.

Thompson, A. *et al.* (2019) 'TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set', *Analytical Chemistry*, 91(24), pp. 15941–15950. doi: 10.1021/acs.analchem.9b04474.

Towbin, H., Staehelin, T. and Gordon, J. (1979) 'Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.76.9.4350.

Tyleckova, J. *et al.* (2012) 'Cancer cell response to anthracyclines effects: Mysteries of the hidden proteins associated with these drugs', *International Journal of Molecular Sciences*, 13(12), pp. 15536–15564. doi: 10.3390/ijms131215536.

Ünlü, M., Morgan, M. E. and Minden, J. S. (1997) 'Difference gel electrophoresis: A single gel method for detecting changes in protein extracts', *Electrophoresis*, 18(11), pp. 2071–2077. doi: 10.1002/elps.1150181133.

Veschi, S. *et al.* (2020) 'Integrative proteomic and functional analyses provide novel insights into the action of the repurposed drug candidate nitroxoline in AsPC-1 cells', *Scientific Reports*, 10(1), pp. 1–12. doi: 10.1038/s41598-020-59492-4.

Vesterberg, O. (1993) 'A short history of electrophoretic methods', *Electrophoresis*, 14(1), pp. 1243–1249. doi: 10.1002/elps.11501401188.

Wang, Z. *et al.* (2020) '27-Plex Tandem Mass Tag Mass Spectrometry for Profiling Brain Proteome in Alzheimer's Disease', *Analytical Chemistry*, 92(10), pp. 7162–7170. doi: 10.1021/acs.analchem.0c00655.

Weisbrod, C. R. *et al.* (2012) 'Accurate peptide fragment mass analysis: Multiplexed peptide identification and quantification', *Journal of Proteome Research*, 11(3), pp. 1621–1632. doi: 10.1021/pr2008175.

Wiita, A. P. *et al.* (2013) 'Global cellular response to chemotherapy-induced apoptosis', *eLife*, 2, pp. 1–28. doi: 10.7554/elife.01236.

Wiley, W. C. and McLaren, I. H. (1955) 'Time-of-flight mass spectrometer with improved resolution', *Review of Scientific Instruments*. doi: 10.1063/1.1715212.

Wolf-Yadlin, A., Hu, A. and Noble, W. S. (2016) 'Technical advances in proteomics: New developments in data-independent acquisition', *F1000Research*, 5(0), pp. 1–12. doi: 10.12688/f1000research.7042.1.

Xu, H. *et al.* (2015) 'Proteomic profiling of paclitaxel treated cells identifies a novel mechanism of drug resistance mediated by PDCD4', *Journal of Proteome Research*, 14(6), pp. 2480–2491. doi: 10.1021/acs.jproteome.5b00004.

Yamashita, M. and Fenn, J. B. (1984) 'Electrospray ion source. Another variation on the free-jet theme', *Journal of Physical Chemistry*. doi: 10.1021/j150664a002.

Zha, J. *et al.* (2014) 'Disulfiram targeting lymphoid malignant cell lines via ROS-JNK activation as well as Nrf2 and NF-kB pathway inhibition', *Journal of Translational Medicine*, 12(1), pp. 1–9. doi: 10.1186/1479-5876-12-163.

Zhang, G. and Neubert, T. A. (2009) 'Use of stable isotope labeling by amino acids in cell culture (SILAC) for phosphotyrosine protein identification and quantitation.', *Methods in molecular biology (Clifton, N.J.)*. doi: 10.1007/978-1-60327-834-8_7.

Zhu, W., Smith, J. W. and Huang, C. M. (2010) 'Mass spectrometry-based label-free quantitative proteomics', *Journal of Biomedicine and Biotechnology*, 2010. doi: 10.1155/2010/840518.

Zhu, Y. *et al.* (2019) 'High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification', *Molecular Oncology*, 13(11), pp. 2305–2328. doi: 10.1002/1878-0261.12570.

Zieske, L. R. (2006) 'A perspective on the use of iTRAQ^{TM} reagent technology for protein complex and profiling studies', *Journal of Experimental Botany*, 57(7), pp. 1501–1508. doi: 10.1093/jxb/erj168.

Zong, C. *et al.* (2008) 'Two-dimensional electrophoresis-based characterization of post-translational modifications of mammalian 20S proteasome complexes', *Proteomics*. doi: 10.1002/pmic.200800387.

Zubarev, R. A. and Makarov, A. (2013) 'Orbitrap mass spectrometry', *Analytical Chemistry*, 85(11), pp. 5288–5296. doi: 10.1021/ac4001223.

7 BIBLIOGRAPHY

Articles related to PhD thesis are labeled *.

7.1. Original articles, reviews and utility models

***VÁCLAVKOVÁ J**, KOUŘILOVÁ P, VRBKOVÁ J, HOLUB D, HAJDÚCH M, DŽUBÁK P. Proteomic Analysis of Exhaled Breath Condensate Samples: High Reproducibility of Mass Spectrometric Measurements. Chem. Listy 2020, 114 (7), 470−479, online ISSN 1213-7103, IF: 0.39.

VÁCLAVKOVÁ J, OŽDIAN T, HAJDÚCH M, DŽUBÁK P. Tělní tekutiny jako zdroj proteomických biomarkerů různých onemocnění. Chem. Listy 2020, 114 (3), 209-215, online ISSN 1213-7103, IF: 0.39.

*ŠKROTT Z, MISTRÍK M, ANDERSEN K, FRIIS S, MAJERA D, GURSKÝ J, OŽDIAN T, BARTKOVA J, TURI Z, MOUDRÝ P, KRAUS M, MEDVEDÍKOVÁ M, **VÁCLAVKOVÁ J**, DŽUBÁK P, VROBEL I, POUCKOVA P, SEDLACEK J, MIKLOVICOVA A, KUTT A, MATTOVA J, DRIESSEN C, DOU Q, OLSEN J, HAJDÚCH M, CVEK B, DESHAIES R, BÁRTEK J. Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4, Nature, 2017, 552, 194-199, 0028-0836, IF: 41.577, PMID: 29211715.

*OŽDIAN T, HOLUB D, MACEČKOVÁ Z, VARANASI L, RYLOVÁ G, ŘEHULKA J, **VÁCLAVKOVÁ J**, SLAVÍK H, MOUDRÝ P, ZNOJEK P, STANKOVÁ J, DE SANCTIS J, HAJDÚCH M, DŽUBÁK P. Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells, Journal of Proteomics, 2017, 162, 73-85, 1874-3919, IF: 3.722, PMID: 28478306.

OŽDIAN T., HOLUB D, RYLOVÁ G, **VÁCLAVKOVÁ J**, HAJDÚCH M, DŽUBÁK P. Porovnání hmotnostně spektrometrických přístupů v proteomickém profilování léčiv. Chemagazín, 2016, 5, 8-11, RIV/61989592:15110/16:33162906.

KULTAN J., KOLEK V., DŽUBÁK P., **VÁCLAVKOVÁ J**, HAJDÚCH M, SZKORUPA M. Možnosti neinvazivní detekce karcinomu plic, Studia Pneumologica et Phthiseologica, 2018, 78, 69-75, 1213-810X.

JEDINÁKOVÁ P, HLAVÁČ J, **VÁCLAVKOVÁ J**, KONEČNÝ P, DŽUBÁK P, HAJDÚCH M. Fluorescent derivative for non-catalytic labeling of nucleic acid and peptide components (Jedináková), Priority Appl. PUV 2016-32558. Utility Model: CZ 30 136, Granted: 13. 12. 2016, Ownership: Palacky University, Olomouc, RIV/61989592:15110/16:33161235.

7.2. Oral and poster presentations

VÁCLAVKOVÁ J, OŽDIAN T, HOLUB D, DŽUBÁK P, RYLOVÁ G, KOLLAREDDY M, ŘEHULKA J, RADOVÁ L, HAJDÚCH M. Proteomický profil protinádorového účinku klinicky používaných rostlinných alkaloidů. In: IX. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 37, 2013.

VÁCLAVKOVÁ J, VRBKOVÁ J, HOLUB D, DŽUBÁK P, KOPŘIVA F, KULTAN J, KOLEK V, HAJDÚCH M. Proteomic Analysis of Exhaled Breath Condensates. In: XI. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, A11-A13, 2015.

VÁCLAVKOVÁ J, DŽUBÁK P, NEUBAUEROVÁ E, HOLUB D, HAJDÚCH M. Application of DARTS (Drug Affinity Responsive Target Stability) for identification of the ligand-binding part of the protein. In: 10th Central and Eastern European Proteomic Conference Book of Abstracts and Program, 135, 2016.

VÁCLAVKOVÁ J, DŽUBÁK P, NEUBAUEROVÁ E, HOLUB D, HAJDÚCH M. Application of DARTS (Drug Affinity Responsive Target Stability) for identification of the ligand-binding part of the protein. In: XII. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 36, 2016.

VÁCLAVKOVÁ J, HRUŠKA M, VRBKOVÁ J, KOPŘIVA F, LÁTALOVÁ V, HOLUB D, KULTAN J, KOLEK V, JAKUBEC P, HAJDÚCH M, DŽUBÁK P. Non-invasive lung cancer diagnostics using proteomic biomarkers in exhaled breath condensate. In: XIII. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 38-39, 2017.

VÁCLAVKOVÁ J, ŘEHULKA J, HOLUB D, HODOŇ J, URBAN M, GURSKÁ S, MEDVEDÍKOVÁ M, DŽUBÁK P, HAJDÚCH M. Affinity purification as a tool for identifying molecular targets of new triterpenic pyrazine compounds. In: XIV. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 28, 2018.

VÁCLAVKOVÁ J, VRBKOVÁ J, HOLUB D, DŽUBÁK P, KOPŘIVA F, GVOZDIAKOVÁ T, HAJDÚCH M. Proteomic analysis of exhaled breath condensates as a non-invasive diagnostics of pediatric asthma. In: Abstract Book IMTM REACTOR, 35, 2018.

VÁCLAVKOVÁ J, GVOZDIAKOVÁ T, VRBKOVÁ J, HOLUB D, DŽUBÁK P, KOPŘIVA F, LÁTALOVÁ V, KULTAN J, KOLEK V, JAKUBEC P, HAJDÚCH M. Proteomic Analysis of Exhaled Breath Condensates. In: Abstract Book IMTM REACTOR, 23, 2019.

VÁCLAVKOVÁ J, VRBKOVÁ J, DŽUBÁK P, HOLUB D, GVOZDIAKOVÁ T, KOPŘIVA F, LÁTALOVÁ V, KULTAN J, KOLEK V, JAKUBEC P, FISCHER O, HAJDÚCH M. Exhaled breath condensates proteomics as a tool to monitor condition of human lungs. In: XV. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 20, 2019.

VÁCLAVKOVÁ J, GVOZDIAKOVÁ T, VRBKOVÁ J, DŽUBÁK P, HOLUB D, KOPŘIVA F, LÁTALOVÁ V, KULTAN J, KOLEK V, HAJDÚCH M. Proteomic Analysis of Exhaled Breath Condensates. In: Book of Abstracts from the Eighth Annual Conference of the Czech Society for Mass Spectrometry, 41, 2019, ISBN 978-80-905045-9-2.

VÁCLAVKOVÁ J, DŽUBÁK P, GVOZDIAKOVÁ T, VRBKOVÁ J, HOLUB D, KOPŘIVA F, LÁTALOVÁ V, KULTAN J, KOLEK V, JAKUBEC P, NAKLÁDALOVÁ M, HAJDÚCH M. Proteomic Analysis of Exhaled Breath Condensates. In: XXI. Kongres České pneumologické a ftizeologické společnosti a Slovenskej pneumologickej a ftizeologickej spoločnosti: Abstrakta, 51 – 52, 2019, ISBN 978-80-7471-282-1.

7.3. Abstracts

OŽDIAN T, HOLUB D, RYLOVÁ G, **VÁCLAVKOVÁ J**, HAJDÚCH M, DŽUBÁK P. The comparison of mass spectrometry approaches in proteomic profiling of drug responses. In: 10th Central and Eastern European Proteomic Conference Book of Abstracts and Program, 135, 2016.

DŽUBÁK P, OŽDIAN T, HOLUB D, **VÁCLAVKOVÁ J**, STANKOVÁ J, RYLOVÁ G, ŘEHULKA J, DAS V, HAJDÚCH M. Strategies for molecular target identification. In: XIII. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 26, 2017.

OŽDIAN T, HOLUB D, MACEČKOVÁ Z, VARANASI L, RYLOVÁ G, ŘEHULKA J, **VÁCLAVKOVÁ J**, SLAVÍK H, MOUDRÝ P, ZNOJEK P, STANKOVÁ J, DE SANCTIS J, HAJDÚCH M, DŽUBÁK P. Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells. In: XIII. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 15, 2017.

URBAN M, HODOŇ J, SOURAL M, **VÁCLAVKOVÁ J,** HOLUB D, DŽUBÁK P, ŠAREK J, ŘEHULKA J, HAJDÚCH M. Semisynthetic triterpenoid heterocycles - synthesis, cytotoxicity, and studies of their mechanism of action. In: XIII. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 9-10, 2017.

HODOŇ J, **VÁCLAVKOVÁ J**, ŘEHULKA J, HOLUB D, DŽUBÁK P, HAJDÚCH M, URBAN M. Cytotoxic Triterpenes and Study of their Mechanism of Action. In: XIII. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 41, 2017.

URBAN M, HODOŇ J, BORKOVÁ L, POKORNÝ J, KRAJČOVIČOVÁ S, LIŠKOVÁ B, SOURAL M, **VÁCLAVKOVÁ J**, STANKOVÁ J, HOLUB D, DŽUBÁK P, HAJDÚCH M, ŠAREK J, ŘEHULKA J. Cytotoxic triterpenes and triterpenic conjugates used for studies of mechanisms of action. In: XIV. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 26, 2018.

DŽUBÁK P, **VÁCLAVKOVÁ J**, HRUŠKA M, VRBKOVÁ J, SLAVÍK H, SROVNAL J, HOLUB D, KOPŘIVA F, GVOZDIAKOVÁ T, KULTAN J, JAKUBEC P, KOLEK V, HAJDÚCH M. Non-invasive lung cancer diagnostics using proteomic biomarkers in exhaled breath condensate. In: XIV. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 33, 2018.

HRUŠKA M, VARANASI L, HOLUB D, **VÁCLAVKOVÁ J**, AGRAWAL K, VOJTA P, VOLLER J, DŽUBÁK P, HAJDÚCH M. CLAIRE: A system for detection of protein variants from tandem mass spectra. In: XIV. Diagnostic, Predictive and Experimental Oncology Days: Abstract Book, 35, 2018.

MISTRÍK M, ŠKROTT Z, HAJDÚCH M, FRIIS S, DŽUBÁK P, GURSKÝ J, MAJERA D, OŽDIAN T, **VÁCLAVKOVÁ J**, MICHALOVÁ M, POUCKOVA P, CVEK B, ANDERSEN K, BÁRTEK J. Alcohol-aversion drug disulfiram targets cancer via p97 segregase adaptor NPL4. In: Cancer Research, 78, 13, LB-264, AACR Annual Meeting 2018; April 14-18, 2018; Chicago, IL; American Association for Cancer Research, doi 10.1158/1538-7445.AM2018-LB-264; Online ISSN: 1538-7445.

OŽDIAN T, HOLUB D, MACEČKOVÁ Z, VARANASI L, RYLOVÁ G, ŘEHULKA J, **VÁCLAVKOVÁ J**, SLAVÍK H, MOUDRÝ P, ZNOJEK P, STANKOVÁ J, DE SANCTIS J, HAJDÚCH M, DŽUBÁK P. Cellular response to oxaliplatin treatment shows altered DNA damage, nucleolar and ribosomal stresses as main altered pathways. In: Cancer Research, 78, 13, 455, AACR Annual Meeting 2018; April 14-18, 2018; Chicago, IL; American Association for Cancer Research, doi 10.1158/1538-7445.AM2018-455; Online ISSN: 1538-7445.

OŽDIAN T, HOLUB D, MACEČKOVÁ Z, VARANASI L, RYLOVÁ G, ŘEHULKA J, **VÁCLAVKOVÁ J**, SLAVÍK H, DŽUBÁK P, HAJDÚCH M. Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells. In: Book of Abstracts from the Eighth Annual Conference of the Czech Society for Mass Spectrometry, 43, 2019, ISBN 978-80-905045-9-2.

KOPŘIVA F, **VÁCLAVKOVÁ J**, GVOZDIAKOVÁ T, DŽUBÁK P, LÁTALOVÁ V. Je stanovení proteinů v kondenzátu vydechovaného vzduchu další cestou k monitorování tajuplného zánětu u asthma bronchiale? In: XXI. Kongres České pneumologické a ftizeologické společnosti a Slovenskej pneumologickej a ftizeologickej spoločnosti: Abstrakta, 51 – 52, ISBN 978-80-7471-282-1.

8 APPENDIX – FULL-TEXT PUBLICATIONS RELATED TO THE THESIS

8.1. Proteomic Analysis of Exhaled Breath Condensate Samples: High Reproducibility of Mass Spectrometric Measurements

VÁCLAVKOVÁ J, KOUŘILOVÁ P, VRBKOVÁ J, HOLUB D, HAJDÚCH M, DŽUBÁK P. Proteomic Analysis of Exhaled Breath Condensate Samples: High Reproducibility of Mass Spectrometric Measurements. Chem. Listy 2020, 114 (7), 470−479, online ISSN 1213-7103, IF: 0.39.

PROTEOMIC ANALYSIS OF EXHALED BREATH CONDENSATE SAMPLES: HIGH REPRODUCIBILITY OF MASS SPECTROMETRIC MEASUREMENTS

Jana Václavková, Pavla Kouřilová, Jana Vrbková, Dušan Holub, Marián Hajdúch, and Petr Džubák

Laboratory of Experimental Medicine, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc, Czech Republic. petr.dzubak@upol.cz

Keywords: Exhaled breath condensate, mass spectrometry, proteomics, biomarkers, lung disease

Introduction

Exhaled breath condensate (EBC) sampling is a promising, easy and non-invasive method to obtain samples from human lungs. EBC was successfully taken and analyzed even in case of preschool children[1,2]. EBC collection could be performed by various devices including some lab-made or commercially available ones such as EcoScreen, TurboDECCS, RTube, and Anacon glass condenser. In this study, the TurboDECCS device was used. It consists of a portable Turbo Unit (Turbo is an acronym for Transportable Unit for Research of Biomarkers Obtained) and a Disposable Exhaled Condensate Collection System (DECCS). DECCS is equipped with a mouthpiece, a non-return valve, a tube, and a collecting cell inserted in an electrical cooling system of the Peltier type. After 10 minutes of quiet breathing, 1–3 mL of EBC could be collected from adults[3]. Approx. 99.9% of the EBC sample is constituted by condensed water vapor[3], while the rest contains various volatile and nonvolatile molecules from the lower respiratory tract[4]. Most biomarkers are highly diluted, close to the lower assay sensitivity limits of the methods[3].

Proteins in exhaled breath condensates were previously detected in subnanogram amounts by high-resolution 2D electrophoresis[5]. Cytokines and growth factors were identified in children with asthma using enzymatic immunoassays[6-7] or multiplex immunoassays[2]. The first mass spectrometry-based analyses included gel-based electrophoretic separation followed by mass spectrometric identification of excised spots[8-10]. This approach led to a few keratin identifications in healthy controls using an HPLC-MS / Q-TOF mass spectrometer[8], as well as in mechanically ventilated patients using MALDI-TOF MS[9] and in patients with pulmonary emphysema using the HPLC-MS / ESI-ION TRAP model LCQ-Advantage[10]. The authors of these studies faced a problem of a low protein concentration in EBC samples and therefore some of them used pooled samples[8,10].

As the low protein concentration represents the main problem in the study of the EBC proteome[11], the following methods to concentrate the proteins were studied: protein precipitation by trichloroacetic acid or pyrogallol red, precipitation on POROS-20-R2-RP [poly(styrene-*co*-divinylbenzene)] beads, EBC delipidation and precipitation with methanol and chloroform, protein precipitation with sodium deoxycholate and trichloroacetic acid, ultramembrane centrifugation on various membranes, solid phase extraction with ZipTipC18, freeze-drying, vacuum concentration and reversed-phase chromatography[11-13]. The precipitation with pyrogallol red showed high protein recovery but this dye is not compatible with MS protein analysis. Protein capture on POROS-20-R2-RP beads was selected by some authors as the best method for EBC samples protein concentration, while lyophilization was found to be highly irreproducible[11]. However, another group found the lyophilization as the finest and most rapid method of protein concentration with high protein recovery[12]. The reversed-phase chromatography concentration was also chosen as a suitable method[13]. In a study by Bloemen's group, the relative quantity of one of the proteins, calgranulin B, was found to be approximately 20-fold lower, as compared to the most abundant protein, cytokeratin 10. This was considered as an evidence that using a sample concentration method and gel-free approach improve the sensitivity[11].

The next step to complete the characterization of the EBC proteome was the analysis of pooled EBC samples from nine healthy subjects using the high-resolution Orbitrap-based mass spectrometry (HRMS), nanoESI ionization, combined with electrophoretic separation for protein identification. This approach led to the identification of 167 unique gene products[14]. Another study using electrophoretic separation and high resolution Orbitrap-based mass spectrometry has led to identification of 229 unique proteins in two pooled samples (each consisted of 10 individual EBC samples) and 153 proteins common to both pools[15]. A direct in-solution sample preparation and HPLC analysis in conjunction with Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry (FT-ICR MS) of EBC samples collected from 24 mechanically ventilated newborns were able to detect 119 proteins using label-free semi-quantitative approach[16].

Another increase of the number of protein identifications came with lung cancer and COPD studies. More than 300 proteins were identified in the samples that were processed by the gel-free method and analyzed by high-resolution HPLC FT-ICR MS[17]. In another study on early diagnosis of lung cancer, more samples were used, thereby increasing the coverage of the EBC proteome. EBCs were collected from 192 individuals: 49 controls, 49 smokers, 46 COPD, and 48 lung cancer patients; here, the nanoESI QqTOF mass spectrometry analysis revealed 348 different proteins[18]. Both studies suggested potential biomarkers of early-stage lung cancer[17,18]. However, higher number of proteins in these studies should be influenced by the fact that more proteins were identified in EBC from lung can-

cer patients, as compared to healthy controls and COPD patients[18]. In another approach[19], a total of 257 proteins were identified in the high resolution and quantitative Orbitrap-based mass spectrometry study. This study was performed with individual samples (not sample pools), comprising 19 COPD patients and 19 healthy controls. It revealed 24 proteins that were differentially expressed in COPD patients[19].

Thus, assays based on mass spectrometry have already solved successfully the problem of low protein concentrations in EBC samples. However, these studies are aimed at complete characterization of the EBC proteome or disease biomarkers discovery. Most studies focused on complete proteome analysis used pooled samples, which increased the number of identifications but lost information about the interindividual variability of the samples. In the present study, the measurements of the EBC samples in biological and technical replicates are introduced for the first time, comparing the protein profile between individual samples of healthy subjects and describing the reproducibility between three technical replicates of a single sample.

Materials and methods

Study cohort

The samples were taken from 3 individuals from one family. These individuals included two sisters of age 8 and 11 and their father (45).

EBC collection

The exhaled breath condensates were collected from the study participants using the Turbo 14 Turbo DECCS System (Medivac, Italy). The EBC was being collected for a period of 10 min without wearing a nose clip. Samples were taken at three time points in the interval of two hours. After collection, samples were immediately frozen and stored at –80 °C until analysis.

Sample preparation

The exhaled breath condensates were freeze-dried using a SpeedVac (Thermo Fisher Scientific). Proteins were solubilized in the TEZ buffer (10 mM Tris, pH 8.0; 1 mM EDTA; 0.002% Zwittergent 3-16), cleared at 14,000 g, at 20 °C for 10 min and the supernatants were transferred to 1.5 mL Eppendorf tubes. Heat-mediated antigen retrieval was performed at 98 °C for 90 min. Proteins were denatured by sonication in a water bath for 30 minutes at room temperature and digested with trypsin at 37 °C overnight. Trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega) was added at a final concentration of 0.5 µg per sample. Proteins were reduced with dithiothreitol (10 µM final concentration) and peptides were purified using Stage tips[20]. The purified samples were concentrated in vacuum concentrator (Eppendorf) and dissolved in 50 µL solution of 1% acetonitrile, 0.05% trifluoroacetic acid (TFA).

Mass spectrometric analysis

Mass spectrometric analysis was performed in 3 technical replicates using the Orbitrap Fusion™ mass spectrometer with the Proxeon EASY-Spray ionization source (Thermo Fisher Scientific) coupled to the Dionex UltiMate 3000 RSLC Nanoliquid chromatograph. The peptides were loaded on the Acclaim PepMap™ 100 nano trap column (nanoViper C18 100 µm i.d. × 2 cm, 5 µm particle size, 100 Å pore size; Thermo Fisher Scientific), loading solvent was 1% acetonitrile with 0.05% TFA in water and sample loading volume was 10 µL using a partial-loop injection mode. The trap column was directly connected to analytical EASY-Spray column PepMap™ RSLC C18 (75 µm × 15 cm, 3 µm particle size, 100 Å pore size; Thermo Fisher Scientific) heated to 35 °C. The peptides were separated for 95 minutes; the starting organic phase gradient was set to 2% (0 to 5 min), then it increased gradually to 35% (5 to 65 min), then to 90% (65 to 73 min), and finally dropped quickly back to 2%. As the aqueous and organic mobile phases, water and acetonitrile were used, respectively, both with 0.1% formic acid. Concerning MS analysis, the static positive ion spray voltage was set to 2000 V, ion transfer tube temperature was 200 °C, master scan was performed in Orbitrap, positive mode, FTMS resolution was set to 120,000 and the m/z mass range for scanning of precursor ions was 400–1500. Cycle time data dependent mode was chosen, time between master scans was set to 3 s. Ion fragmentation was performed by the collision-induced dissociation (CID). The MS2 scan was performed in an ion trap, the rapid scan rate was set, the collision energy was 30% and the activation time was 10 ms. The profile data type was recorded in both master scan and MS2.

Data processing

Protein search was performed in beta version of Proteome Discoverer™ 2.5 Software (Thermo Fisher Scientific). The processing workflow includes a search against the complete human UniProtKB database, which included both Swiss-Prot reviewed and TrEMBL computationally analyzed but unreviewed proteins (https://www.uniprot.org/uniprot/, downloaded in January 2020). Sequest HT search engine was used where oxidation, *N*-terminal acetylation, methionine loss and, finally, methionine loss in combination with *N*-terminal acetylation were set as dynamic modifications. The Fixed Value PSM Validator was used. The consensus workflow included retention time alignment, only unique peptides were used for quantification, total peptide amount normalization was performed, no scaling was done, protein relative quantities were calculated using the Top 3 average method, protein ratio calculation was performed using protein abundance based method, low abundance resampling mode was chosen for missing values imputation and ANOVA

(individual proteins) hypothesis test was used. The results were filtered only for high confidence peptides.

Statistical analysis

A peptide/protein was considered to be present in a sample, if it had been detected at least in one of the technical replicates of this sample. For each sample and each peptide/protein, abundance was calculated as the log2 value of the median of the three values (from Proteome Discoverer) from technical replicates of this sample. Statistical analysis was performed using the R program, ver. 3.5.2 (cit.[21]). Heatmaps were created with the Heatmap function (from the ComplexHeatmap library).

Results and discussion

We have evaluated 3 samples from 3 persons, namely, two children (two daughters, D1 and D2) and an adult (father, F); each sample was measured in three technical replicates, i.e. 27 samples altogether. These samples were evaluated by Proteome Discoverer software in one search. In our study, in addition to protein analysis, analyses were also performed at the peptide level. We have identified 2797 proteins across all samples and 2263 of them were quantified. The proteins were determined on the basis of the identification of 5006 peptides, of which 4189 were quantified. The number of proteins identified in our study was higher than that in previously published studies. Even if the pooled samples were used, 167 proteins were identified in a pool of nine healthy subjects[14], 229 proteins in two pooled samples from 10 individuals[15], and 124 proteins in two pooled samples from 6 individuals where endogenous particles were collected in exhaled air[22]. The number of 119 proteins were identified using individual samples in a cohort of 24 intubated newborns, where each sample was measured in triplicate[16]. Up to date, the highest protein count was 348 different proteins identified in the search for biomarkers of lung cancer and COPD in the entire group of 192 individuals[18].

We have detected 1056 ± 354 proteins per technical replicate. Three technical replicates were combined into one sample. A protein was considered to be detected in a sample, if it was detected in at least one of the technical replicates of the sample. The overlap of technical replicates was high; in spite of this, however, the number of detected proteins increased to 1198 ± 454 per sample, as compared to a single measurement (the number of proteins detected in individual replicates and samples are shown in Table I). The number of proteins in the sample from the adult (1603 ± 607) was higher than that from the daughters (996 ± 140). In the literature, there is still no evidence of differences in protein counts between children and adults. However, a comparison between children and adult samples is not the aim of this paper, and the group under study is too small to make any conclusion based on this difference. Instead, our intention was to show the reproducibility of biological and technical replicates of the samples, to present the advantages of introducing these replicates and to characterize samples of exhaled breath condensate taken in one family.

The overlap between biological replicates (individual samples) of each subject was relatively high and contained 644, 753 and 818 proteins for particular samples (Figure 1A). Of those proteins, which overlapped between biological replicates, 358 proteins were common to all the three individuals. Peptide analysis of peptide overlap showed results similar to protein analysis. However, as expected, all the numbers increased. The overlap between biological replications also remained relatively high, in particular 955, 1147, and 1606 peptides, 532 of which were common

Table I
Numbers of proteins detected in samples and technical replicates of samples. Samples were taken from three persons: father (F) and two daughters (D1 and D2). Three samples were taken from each person (the time delay between samples was about one hour), i.e. there were nine samples in the analysis. Each of the nine samples was measured in three technical replicates. A protein was considered to be detected in a sample if it was detected in at least one of the technical replicates of the sample

Sample	Number of detected proteins in the sample (biological replicate)	Number of detected proteins in the technical replicate (1^{st}| 2^{nd}| 3^{rd})
D1_01	1196	1052|1068|1069
D1_02	884	767| 794| 787
D1_03	888	753| 778| 803
D2_01	903	725| 772| 804
D2_02	956	793| 859| 764
D2_03	1148	986| 972|1006
F_01	2307	1869|1871|1900
F_02	1434	1276|1204|1213
F_03	1067	993| 949| 930

to all the three individuals (Figure 1B).

We performed a functional protein enrichment analysis common to all samples of all the three family members. Common proteins included 358 proteins that were submitted to STRING functional protein association networks analysis. Cornification, keratinocyte differentiation, skin development, keratinization, epidermis development, intermediate filament cytoskeleton organization, epithelial cell differentiation, intermediate filament organization, tissue development and epithelium development were among biological processes with false discovery rate (FDR) below 0.001. However, the same key proteins, especially keratins, take part in all these processes. Historically, the keratins were the first proteins identified in EBC samples[8] and were usually referred to as the most abundant proteins in EBC samples of healthy controls[8,11,12,14,15,23]. Other processes with FDR < 0.001, such as neutrophil degranulation, myeloid leukocyte activation, leukocyte-mediated immunity, cell activation, programmed cell death, regulated exocytosis, immune effector process, leukocyte activation, immune response, secretion by cell and immune system process, involved various proteins from interaction network. Using Gene Ontology, immune system processes have been reported to be the main biological processes of

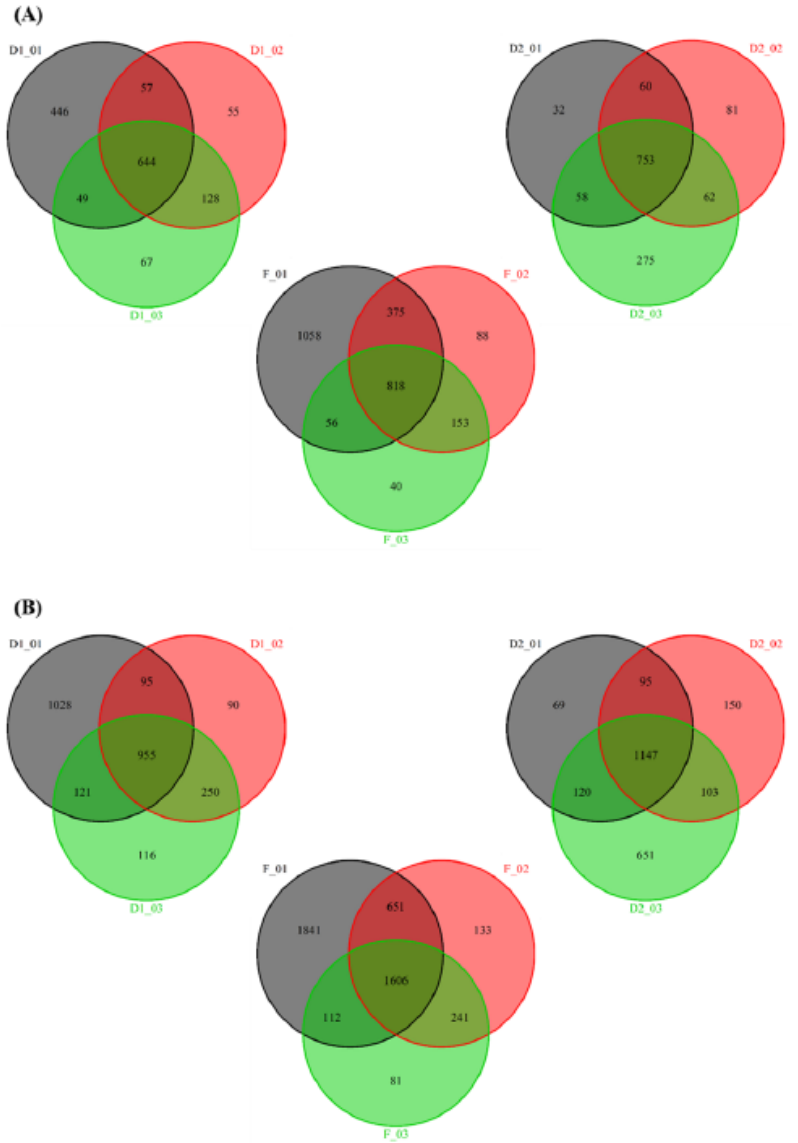


Fig. 1. The number of (A) proteins and (B) peptides detected in each sample and the overlap between biological replicates of each person expressed by Venn diagrams

proteins in EBC of intubated newborns[16] and were also overrepresented in EBC compared to lung proteome in healthy controls. This finding was considered as an evidence that EBC was a relevant matrix to study the major physiological functions of the respiratory tract, particularly mucosal secretion, innate and adaptive antimicrobial defense mechanisms and clearance of inhaled particles[15].

Heatmaps were used to express abundancies and clustering of proteins (Figure 2A) and peptides (Figure 2B) in biological replicates of individual family members. Only quantified proteins were used to construct the heatmap. Proteins that were detected but not quantified across all samples were excluded because we needed an abundance to draw a heatmap, and it was useless to include values that were all imputed, despite the fact that the imputed values were very low. We assessed whether or not the samples of the individuals were clustered together, so the heatmap was constructed for all the nine samples (Figure 3A). Daughters' samples were clustered together, except for the first biological replicate from D1 (D1_01). It was a sample with a higher number of proteins identified than that of the other samples of daughters and was clustered together with father's ones. Since the samples were taken from the members of the same family, it could be assumed

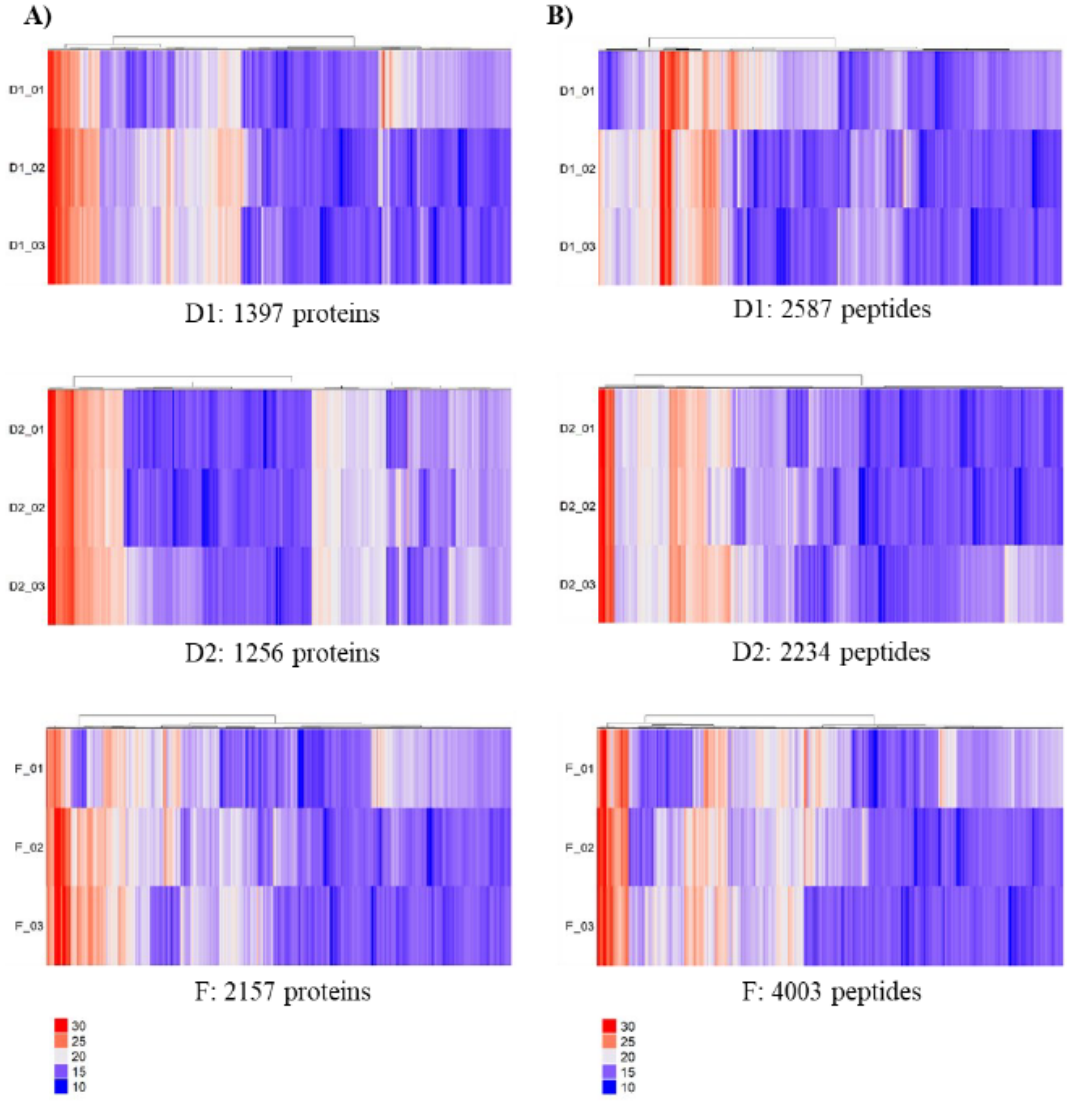


Fig. 2. **Heatmaps of abundance. (A) proteins and (B) peptides in three samples of each person.** In each person-heatmap there are proteins or peptides detected and quantified in at least one sample of that person

that a high similarity will be observed between the samples. Father's samples generally showed higher protein numbers than daughters' ones. This may be caused by the fact that the father is an adult with a higher lung volume and active lung surface; this, as a result, may lead to an increase in protein concentration at the same time. The number of proteins can have a big impact on such a situation. The same situation was observed in the heatmap at the peptide level (Figure 3B).

We have evaluated the distance of the abundancies of the three samples for each person and each protein. The mean value of protein abundancies and the difference of the three values of abundancies from the mean were calculated for each protein separately. This was done for each sample (biological replicate) of all three individuals. The following four intervals were set: ± sd (mean – sd; mean + sd), ± 2 × sd (mean – 2 × sd; mean + 2 × sd), ± 10% (0.9 ×mean; 1.1 × mean), and ± 25% (0.75 × mean; 1.25 × mean) and it was assessed whether the values for individual proteins were in or out of the interval (Figure 4). This analysis was also performed on peptide level. For peptide analysis, we used only fully quantified peptides, which were quantified in all technical and all biological samples (biological replicates) from the given individual. The number of this fully quantified peptides was 687 and 800 for daughters' samples (D1 and D2) and 1128 for father's samples (Figure 5). In both protein and peptide analyses, the difference of one standard deviation from the mean appeared to be too strict criterion because, for each of the proteins and peptides, at least one abundance value lay outside of the ± sd interval (denoted as OUT). In the case of double the standard deviation (± 2 × sd), all the proteins

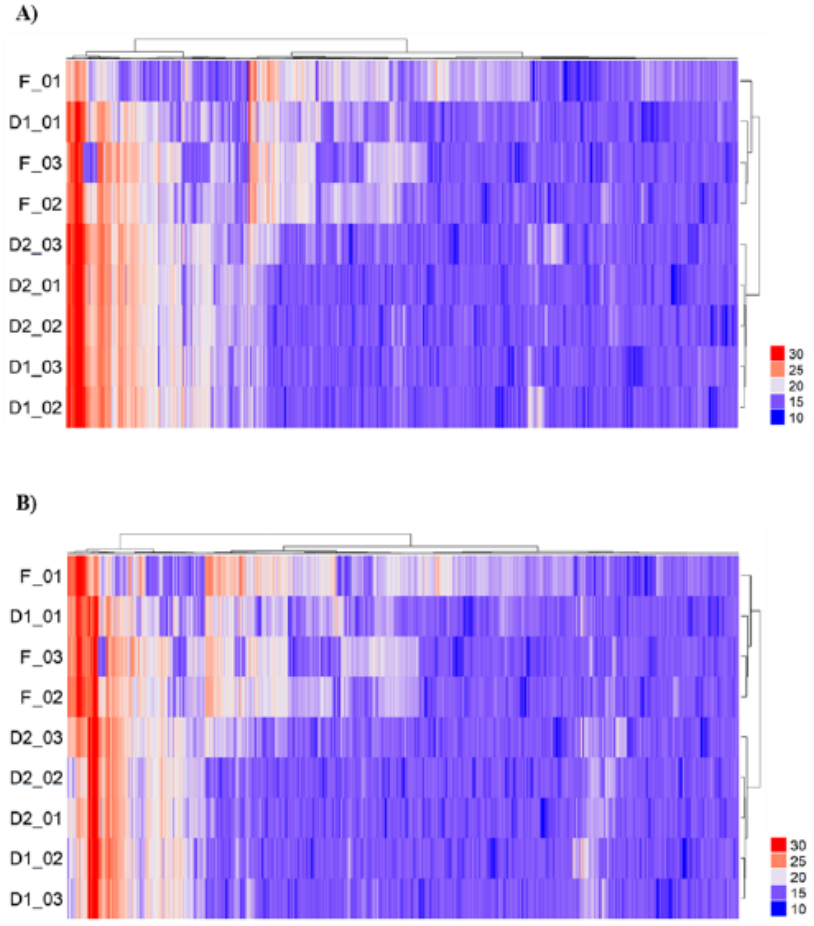
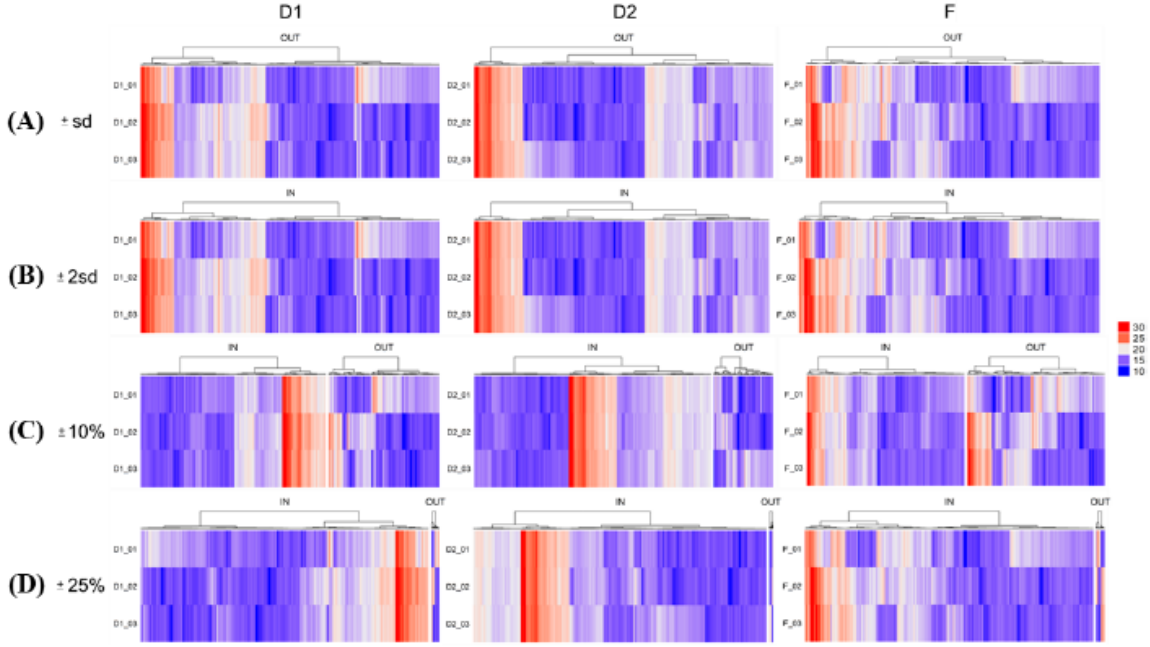


Fig. 3. **Heatmap expressing the abundance of all proteins and peptides across nine samples from individual family members.** (A) Heatmap with all 2263 quantified proteins and (B) heatmap with all 4189 quantified peptides. The presented proteins and peptides are those detected in at least one of the 9 samples. The columns represent individual proteins (or peptides), the rows represent the clustering of EBC samples of family members. Daughters' samples are clustered together except for the first replicate of D1, which contained the highest number of proteins and formed a cluster with father's samples

and peptides lay in this interval (denoted as IN). Analysis using the percent distance from the mean showed that 10% was also too strict criterion. However, more than half of the proteins and peptides belonged to this interval (IN), i.e. the distance of all the three abundances from the mean was less than 10%. When we analyzed proteins the distance of which from the mean was more than 25%, there were only a few tens of proteins or only 2 peptides lying outside this interval.

Then we tested whether these distances were due to variations in mass spectrometric measurements or rather to biological variability. We evaluated the maximum distance from the mean for technical replicates of the commercial bovine serum albumin standard (Pierce™ BSA Protein Digest Standard, LC/MS Grade, Thermo Fisher Scientific) which is routinely used to check the quality of mass spectrometry measurements by our group. The 500 pmol of lyophilized digested bovine serum albumin (BSA) was dissolved with 500 µL of 0.1% formic acid in water, and then 15 µL of this solution was further diluted with 285 µL of 0.1% formic acid in water to obtain 50 fmol/µL BSA. An amount of 100 fmol BSA was injected for each MS measurement. For an analysis of the technical replicates, we calculated the mean of log2 abundance of 3 technical

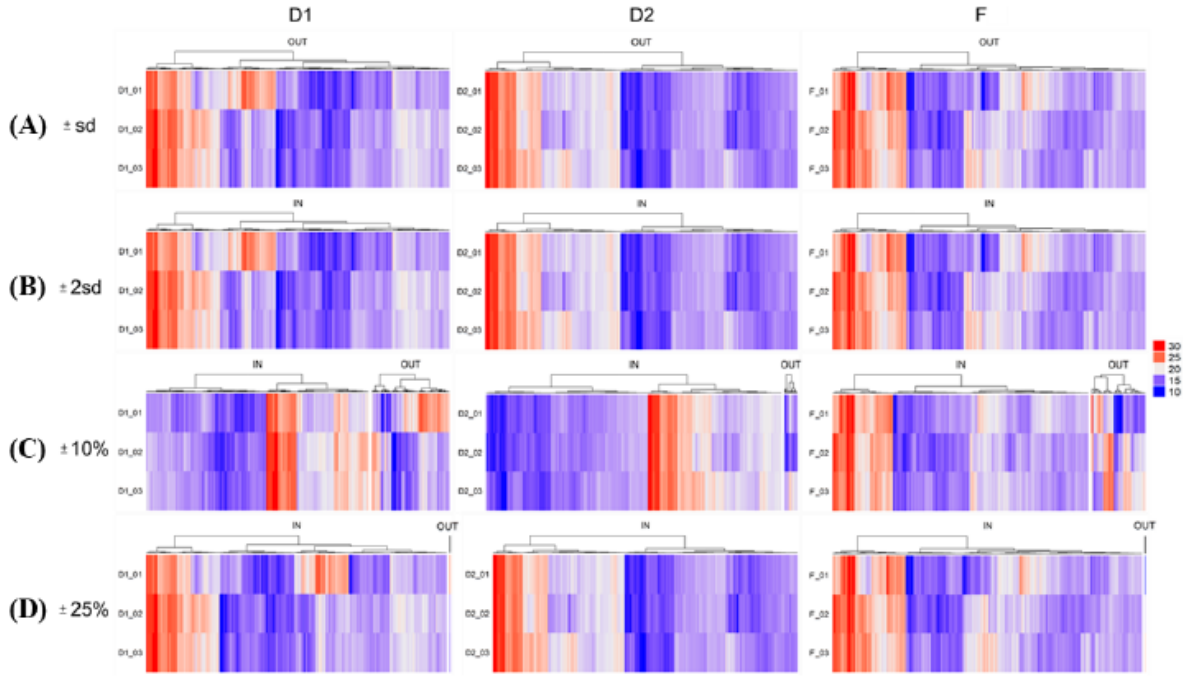


(E)

	D1		D2		F	
	IN	OUT	IN	OUT	IN	OUT
±sd	0	1397	0	1256	0	2157
±2sd	1397	0	1256	0	2157	0
±10%	871	526	1002	254	1152	1005
±25%	1357	40	1240	16	2094	63
total	1397		1256		2157	

Fig. 4. **The distance of the abundancies of the three samples analyzed for each person and each protein to evaluate the variability of the measured values.** Only proteins that have been quantified in at least one sample of individual person are included in the heatmaps. Thus, the total number of proteins is lower than reported before. Four types of intervals were determined and the proteins were divided into two groups: IN = the abundance of all three samples lies in the respective interval; OUT = the abundance of at least one of the three samples is outside this interval. Heatmaps (A – D) expressing proteins that lie in or out of intervals. Selected intervals were: **(A)** ± sd (mean – sd; mean + sd), **(B)** ± 2 × sd (mean – 2 × sd; mean + 2 × sd), **(C)** ± 10% (0.9 × mean; 1.1 × mean), **(D)** ± 25% (0.75 × mean; 1.25 × mean). The table **(E)** lists the number of proteins which lie in and out of the respective intervals. Mean = mean of the abundancies of the protein of the three samples of each person; sd = standard deviation of the mean

replicates for each peptide separately. The difference from the mean was calculated for all three values. The absolute value of the largest difference was chosen and expressed as a percentage of the mean value. This was performed for all fully quantified peptides of all replicates. For the BSA standard, all identified peptides were fully quantified. The analysis has shown that all maximum distances from the mean were less than 8.11%. Based on these results, it is clear that there is some deviation between the measurements of the technical replicates of the standard which is significantly lower than in the case of biological replicates. However, the BSA standard is a commercial mixture of peptides with defined concentrations of the components, which includes only the peptides of this single protein, so that the deviation can be expected to be lower for the standard than for complex biological samples. To test this hypothesis, we examined the maximum distance from the mean in the case of technical replicates of a single person sample. This was done using the same calculations as for the BSA. Calculations were performed for all fully quantified peptides of all samples of all subjects and the values thus obtained were further compared. In the case of tech-



(E)

	D1		D2		F	
	IN	OUT	IN	OUT	IN	OUT
±sd	0	687	0	800	0	1128
±2sd	687	0	800	0	1128	0
±10%	509	178	765	35	931	197
±25%	685	2	800	0	1126	2
total	687		800		1128	

Fig. 5. **The distance of the abundancies of the three samples analyzed for each person and each peptide.** For each individual, peptides quantified in all three technical replicates and in all samples of the person were included, i.e. only peptides that were quantified in 9 replicates (technical and biological) of the individuals, no values imputed by Proteome Discoverer software were included. Four types of intervals were set and the peptides were divided into two groups: IN = the abundance of all three samples lies in the respective interval; OUT = the abundance of at least one of the three samples is outside this interval. Heatmaps (A – D) expressing peptides that lies in or out of intervals. Selected intervals were: (A) ± sd (mean – sd; mean + sd), (B) ± 2 × sd (mean – 2 × sd; mean + 2 × sd), (C) ± 10% (0.9 × mean; 1.1 × mean), (D) ± 25% (0.75 × mean; 1.25 × mean). Table (E) lists the number of peptides that lie within and out the respective intervals. Mean = mean of the abundancies of the peptide in the three samples of each person; sd = standard deviation of the mean

Table II
The maximum distances from the mean abundance value of the three technical replicates (expressed in %) for individual samples

Sample	Max. distance (%) from the mean
D1_01	12.66
D1_02	19.69
D1_03	17.61
D2_01	13.13
D2_02	10.33
D2_03	8.31
F_01	9.04
F_02	13.39
F_03	13.44

nical replicates of human samples, the maximum distances from the mean received for each sample are given in Table II. All maximum distances from the mean were smaller than 19.69%. These results have shown that there is a greater variability between biological replicates of a given person than in the case of technical replicates of a single EBC sample. Thus, the difference in the protein or peptide abundancies which are distant less than 10% from the mean are much more due to the deviation of the mass spectrometry measurements than to the biological variability of the samples. These results also showed that, when mass spectrometry data are evaluated by label-free quantification, the change in protein abundancies below 20% of the mean value is due to the method variability and should not be considered as a change in protein expression.

Conclusion

In this work, a method of protein analysis in exhaled breath condensate based on high resolution mass spectrometry is described. At the same time, it is shown that this analysis is sensitive and reproducible, and the variability of the measurements is acceptable for further interpretation of the results and able to identify hundreds of unique proteins. This method therefore represents a suitable tool for other biologically oriented research work, especially for the search for biomarkers of lung diseases.

This work was supported by Ministry of Health of the Czech Republic, grant nr. 16-32302A. All rights reserved.

REFERENCES

1. Horváth I. a 33 coauthors: Eur. Respir. J. *26*, 523 (2005).
2. Rosias P. P., Robroeks C. M., van de Kant K. D., Rijkers G. T., Zimmermann L. J., van Schayck C. P., Heynens J. W., Jöbsis Q., Dompeling E: Pediatr Allergy Immunol. *21*, e235 (2010).
3. Konstantinidi E. M., Lappas A. S., Tzortzi A. S., Behrakis P. K.: Sci. World J. *2015*, 435160.
4. Mutlu G. M., Garey K. W., Robbins R. A., Danziger L. H., Rubinstein I.: Am. J. Respir. Crit. Care Med. *164*, 731 (2001).
5. Griese M., Noss J., von Bredow C.: Proteomics *2*, 690 (2002).
6. Shahid S. K., Kharitonov S. A., Wilson N. M., Bush A., Barnes P. J.: Am. J. Respir. Crit. Care Med. *165*, 1290 (2002).
7. Leung T. F., Wong G. W., Ko F. W., Li C. Y., Yung E., Lam C. W., Fok T. F.: Int. Arch. Allergy Immunol. *137*, 66 (2005).
8. Gianazza E., Allegra L., Bucchioni E., Eberini I., Puglisi L., Blasi F., Terzano C., Wait R., Sirtori C. R.: Am. J. Med. *117*, 51 (2004).
9. Gessner C., Dihazi H., Brettschneider S., Hammerschmidt S., Kuhn H., Eschrich K., Keller T., Engelmann L., Sack U., Wirtz H.: Respir. Med. *102*, 299 (2008).
10. Fumagalli M., Dolcini L., Sala A., Stolk J., Fregonese L., Ferrari F., Viglio S., Luisetti M., Iadarola P.: J. Proteomics *71*, 211 (2008).
11. Bloemen K., Hooyberghs J., Desager K., Witters E., Schoeters G.: Proteomics Clin. Appl. *3*, 498 (2009).
12. Kurova V. S., Anaev E. C., Kononikhin A. S., Fedorchenko K. Y., Popov I. A., Kalupov T. L., Bratanov D. O., Nikolaev E. N., Varfolomeev S. D.: Clin. Chem. Lab. Med. *47*, 706 (2009).
13. Núñez-Naveira L., Mariñas-Pardo L. A., Montero-Martínez C.: Lung *197*, 523 (2019).
14. Muccilli V., Saletti R., Cunsolo V., Ho J., Gili E., Conte E., Sichili S., Vancheri C., Foti S.: J. Pharm. Biomed. Anal. *105*, 134 (2015).
15. Lacombe M., Marie-Desvergne C., Combes F., Kraut A., Bruley C., Vandenbrouck Y., Chamel Mossuz V., Couté Y., Brun V.: J. Breath Res. *12*, 021001 (2018).
16. Kononikhin A. S. a 13 spoluautorů: J. Chromatogr. B *1047*, 97 (2017).
17. Fedorchenko K. U. a 11 coauthors: Mosc. Univ. Chem. Bull. *71*, 134 (2016).
18. López-Sánchez L. M., Jurado-Gámez B., Feu-Collado N., Valverde A., Cañas A., Fernández-Rueda J. L., Aranda E., Rodríguez-Ariza A.: Am. J. Physiol. Lung Cell. Mol. Physiol. *313*, L664 (2017).
19. Sun C., Zhou T., Xie G., Fu S., Gao L., Liao J., Wu Y., Wang G.: J. Proteomics *206*, 103392 (2019).
20. Rappsilber J., Ishihama Y., Mann M.: Anal. Chem. *75*, 663 (2003).
21. https://www.R-project.org/, staženo 20. 12. 2018.
22. Bredberg A., Gobom J., Almstrand A. C., Larsson P., Blennow K., Olin A. C., Mirgorodskaya E.: Clin. Chem. *58*, 431 (2012).
23. Fumagalli M. a 11 coauthors: Int. J. Mol. Sci. *13*, 13894 (2012).

Abstract

J. Václavková, P. Kouřilová, J. Vrbková, D. Holub, M. Hajdúch, and P. Džubák *(Laboratory of Experimental Medicine, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc):* **Proteomic Analysis of Exhaled Breath Condensate Samples: High Reproducibility of Mass Spectrometric Measurements**

Exhaled breath condensate (EBC) is considered to be a rich source of biomarkers that can provide valuable information for early diagnosis of respiratory and systemic diseases. However, EBC consists mainly of water vapor and the concentration of analytes in its condensed sample is very low, close to the detection limit of sensitive methods such as mass spectrometry. Here, we have developed a mass spectrometry-based approach to characterize the EBC proteome. We compared biological and technical replicates of the measured samples, compared the protein profile between individual samples of healthy individuals and described the reproducibility between the three technical replicates of the sample taken. We have found that repeated measurements of samples are reproducible and can increase the number of proteins identified. The analyses led to the detection of 2797 unique proteins in all samples. Based on statistical analysis, it was shown that (i) the 10% difference from the mean is due more to the deviations of the mass spectrometry measurements than to the biological variability of the samples and (ii) the change in protein amount of 20% of the average in the case of EBC is due to the method variability and should not be considered as a significant change in protein expression using label-free mass spectrometry techniques.

8.2. Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4.

ŠKROTT Z, MISTRÍK M, ANDERSEN K, FRIIS S, MAJERA D, GURSKÝ J, OŽDIAN T, BARTKOVA J, TURI Z, MOUDRÝ P, KRAUS M, MEDVEDÍKOVÁ M, **VÁCLAVKOVÁ J**, DŽUBÁK P, VROBEL I, POUCKOVA P, SEDLACEK J, MIKLOVICOVA A, KUTT A, MATTOVA J, DRIESSEN C, DOU Q, OLSEN J, HAJDÚCH M, CVEK B, DESHAIES R, BÁRTEK J. Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4, Nature, 2017, 552, 194-199, 0028-0836, IF: 40.137, PMID: 29211715.

ARTICLE

doi:10.1038/nature25016

Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4

Zdenek Skrott[1]*, Martin Mistrik[1]*, Klaus Kaae Andersen[2], Søren Friis[2], Dusana Majera[1], Jan Gursky[1], Tomas Ozdian[1], Jirina Bartkova[2,3], Zsofia Turi[1], Pavel Moudry[1], Marianne Kraus[4], Martina Michalova[1], Jana Vaclavkova[1], Petr Dzubak[1], Ivo Vrobel[1], Pavla Pouckova[5], Jindrich Sedlacek[6], Andrea Miklovicova[7], Anne Kutt[2], Jing Li[8], Jana Mattova[5], Christoph Driessen[4], Q. Ping Dou[9,10], Jørgen Olsen[2], Marian Hajduch[1], Boris Cvek[6]†, Raymond J. Deshaies[8,11]† & Jiri Bartek[2,3]

Cancer incidence is rising and this global challenge is further exacerbated by tumour resistance to available medicines. A promising approach to meet the need for improved cancer treatment is drug repurposing. Here we highlight the potential for repurposing disulfiram (also known by the trade name Antabuse), an old alcohol-aversion drug that has been shown to be effective against diverse cancer types in preclinical studies. Our nationwide epidemiological study reveals that patients who continuously used disulfiram have a lower risk of death from cancer compared to those who stopped using the drug at their diagnosis. Moreover, we identify the ditiocarb–copper complex as the metabolite of disulfiram that is responsible for its anti-cancer effects, and provide methods to detect preferential accumulation of the complex in tumours and candidate biomarkers to analyse its effect on cells and tissues. Finally, our functional and biophysical analyses reveal the molecular target of disulfiram's tumour-suppressing effects as NPL4, an adaptor of p97 (also known as VCP) segregase, which is essential for the turnover of proteins involved in multiple regulatory and stress-response pathways in cells.

Despite advances in the understanding of cancer biology, malignant diseases have a high global toll. Furthermore, the increasing average human life expectancy is predicted to have demographic consequences, including an increase in the incidence of cancer. The high cancer-associated morbidity and mortality highlight the need for innovative treatments. Given the high costs, failure rate and long testing periods of developing new medicines, using drugs that are approved for the treatment of diverse diseases as candidate anti-cancer therapeutics represents a faster and cheaper alternative[1], benefitting from available clinically suitable formulations and evidence of tolerability in patients. Among promising cancer-killing drugs[2] is disulfiram (tetraethylthiuram disulfide, DSF), a drug that has been used for over six decades as a treatment for alcohol dependence[3], with well-established pharmacokinetics, safety and tolerance at the US Food and Drug Administration (FDA)-recommended dosage[4]. In the body, DSF is metabolized to ditiocarb (diethyldithiocarbamate, DTC) and other metabolites, some of which inhibit liver aldehyde dehydrogenase[5]. Because DSF showed anti-cancer activity in preclinical models[3,6–9] and because adjuvant DTC was used to treat high-risk breast cancer in a clinical trial[10], DSF emerges as a candidate for drug repurposing in oncology. Additional advantages of DSF include a broad spectrum of malignancies sensitive to DSF, and its ability to also target the stem-like, tumour-initiating cells[11]. Although the mechanism of DSF's anti-cancer activity remains unclear and it has been suggested that the drug inhibits proteasome activity[6,12], it has been shown that DSF chelates bivalent metals and forms complexes with copper (Cu), which enhances its anti-tumour activity[6,13]. In addition to the lack of a well-defined mechanism of action in cancer cells, the main obstacles for DSF repurposing have been: (i) uncertainty about the active metabolite(s) of DSF *in vivo*; (ii) the lack of assays to measure these active derivative(s) in tumours; (iii) missing biomarker(s) to monitor the impact of DSF in tumours and tissues; (iv) the lack of insights into the preferential toxicity towards cancer cells compared to normal tissues; and (v) the absence of a specific molecular target that could explain the potent anti-tumour activity of DSF. Here, we combine experimental approaches and epidemiology to address the important characteristics of DSF in relation to cancer, pursuing the goal of repurposing DSF for cancer therapy. We identify the active metabolite of DSF, and provide biological validation and mechanistic insights, including the discovery of a biologically attractive protein that has previously not been considered as the target for the anti-cancer activity of DSF.

Epidemiological analyses of DSF and cancer

The relative lack of cancer-related clinical trials with DSF[10,14] prompted us to explore whether DSF use might reduce cancer mortality at a population level. Using the Danish nationwide demographic and health registries, we estimated hazard ratios of cancer-specific mortality associated with DSF use among patients with cancer for the first time during 2000–2013 (see Methods, Table 1 and Extended Data Fig. 1a). DSF users were categorized as (i) previous users, who were patients that were prescribed DSF for alcohol dependency only before their cancer diagnosis or (ii) continuing users, who were patients that were prescribed DSF both before and after diagnosis. As expected from the increase in cancer risk and the deleterious effect on prognosis[15] caused by alcohol abuse, cancer-specific mortality was higher among previous DSF users than among patients with cancer who had never

[1]Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic. [2]Danish Cancer Society Research Center, DK-2100 Copenhagen, Denmark. [3]Science for Life Laboratory, Division of Genome Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. [4]Kantonsspital St Gallen, Department Oncology/Hematology, St Gallen, Switzerland. [5]Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University, 120 00 Prague 2, Czech Republic. [6]Department of Cell Biology & Genetics, Palacky University, Olomouc, Czech Republic. [7]Psychiatric Hospital, 785 01 Šternberk, Czech Republic. [8]Division of Biology and Biological Engineering, Caltech, Pasadena, California 91125, USA. [9]Barbara Ann Karmanos Cancer Institute and Department of Oncology, School of Medicine, Wayne State University, Detroit, Michigan, USA. [10]School of Basic Medical Sciences, Affiliated Tumor Hospital of Guangzhou Medical University, Guangzhou 511436, China. [11]Howard Hughes Medical Institute, Caltech, Pasadena, California 91125, USA. †Present addresses: Olomouc University Social Health Institute, Palacky University, Olomouc, Czech Republic (B.C.); Amgen, Thousand Oaks, California 91320, USA (R.J.D.). *These authors contributed equally to this work.

Table 1 | Cancer-specific mortality associated with DSF use among Danish patients with cancer

	Overall				Localized stage				Non-localized stage				Unknown stage			
Cancer type	Number*	HR	95% CI	*P* value	Number*	HR	95% CI	*P* value	Number*	HR	95% CI	*P* value	Number*	HR	95% CI	*P* value
Any cancer†																
Previous users	3,038	1.00			1,429	1.00			1,054	1.00			555	1.00		
Continuing users	1,177	0.66	0.58–0.76	0.000	602	0.69	0.64–0.74	0.000	355	0.71	0.59–0.87	0.001	220	0.65	0.57–0.75	0.000
No prescriptions	236,950	0.68	0.64–0.73	0.000	113,354	0.59	0.57–0.61	0.000	73,933	0.80	0.73–0.88	0.000	49,663	0.66	0.62–0.71	0.000

Hazard ratios (HR) and 95% confidence intervals (CI) comparing continuing and previous users of DSF, relative to the time of their cancer diagnosis. For DSF exposure categories, statistics and clinical stages, see Methods.
*Number of patients included.
†Except cancers of the liver and kidney.

used DSF. Notably, we also found reduced cancer-specific mortality for cancer overall (Table 1), as well as for cancers of the colon, prostate and breast among continuing users compared to previous DSF users (Extended Data Fig. 1a). Stratification by clinical stage (Table 1) revealed reduced cancer-specific mortality with continuing use of DSF even among patients with metastatic disease. Although it is not possible to draw conclusions about causality, these findings supported the hypothesis that DSF may exert anti-cancer effects among patients suffering from common cancers, prompting us to perform pre-clinical analyses.

Anti-tumour activity of the DTC–copper complex

Because DSF anti-cancer activity has been suggested to be copper-dependent[6,13], we compared groups of mice injected with human MDA-MB-231 cancer cells, fed with a (i) normal diet; (ii) normal diet plus copper gluconate (CuGlu); (iii) normal diet plus DSF; or (iv) normal diet plus DSF and CuGlu (DSF/CuGlu); and tumour volume was measured over time (Fig. 1a and Extended Data Fig. 1b, c). Compared to matched controls, tumour volume in DSF- and DSF/CuGlu-treated groups at 32 days (at DSF doses equivalent to those used by alcoholics) were suppressed by 57% and 77%, respectively ($P = 0.0038$ in favour of the DSF/CuGlu treatment versus DSF alone). These results validate previous *in vitro*[6,11,13] and *in vivo*[6–9,13,16] studies, which indicated that DSF is an efficient anti-cancer agent and that copper potentiates its activity. As the reactive metabolite DTC forms complexes with metals, particularly copper[17], we argued that a DTC–copper complex (bis (diethyldithiocarbamate)–copper (CuET)) forms *in vivo* (Extended Data Fig. 1d), providing the anti-cancer metabolite. To test this hypothesis, we developed a high-resolution approach based on high-performance liquid chromatography–mass spectrometry to measure CuET in tissues, and readily detected CuET after a single oral dose of DSF (Extended Data Fig. 1e, f). Extracts from plasma, liver, brain and MDA-MB-231-xenografted tumours contained CuET in samples from mice treated for five days with DSF or DSF/CuGlu. The CuET levels in plasma and liver were slightly higher after DSF/CuGlu treatment compared to DSF alone. Notably, the CuET levels in the tumour specimens were almost an order of magnitude higher compared to corresponding levels in liver and brain tissues from the same animals (Fig. 1b), suggesting preferential accumulation of CuET in tumours. Importantly, we also confirmed formation of CuET in humans undergoing DSF treatment for alcoholism (Fig. 1c).

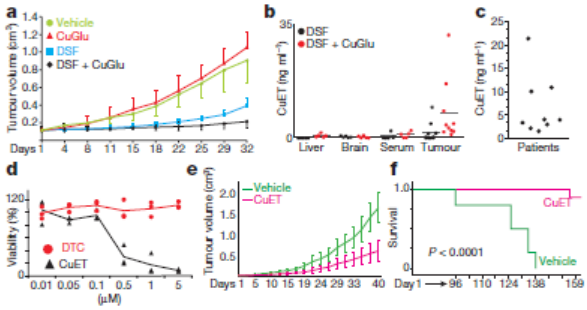


Figure 1 | Tumour-suppressing effects of DSF and CuET. a, Effects of orally administered DSF and CuGlu on subcutaneous growth of MDA-MB-231 tumours in mice. $n = 8$ mice per group. **b**, CuET levels in mouse tumours and tissues. $n = 5$ tissues, $n = 10$ tumours. **c**, CuET levels in human plasma after DSF treatment ($n = 9$ patients). **d**, Toxicity of DTC and CuET in MDA-MB-231 cells after 24 h treatment. $n = 3$ experiments. **e**, Effect of CuET on subcutaneous growth of MDA-MB-231 tumours in mice. $n = 20$ tumours. **f**, Survival of CuET- versus vehicle-treated mice with implanted AMO-1 xenografts. $n = 10$ animals per group. *P* value from a log-rank test. Data are mean ± s.d. (**a**, **e**) or mean (**b**) linked means with individual values (**d**) or individual values (**c**).

Next, we synthesized CuET and performed comparative cell culture and animal studies. Short-term (24-h) assays and long-term (colony-forming assay, CFA) assays consistently showed higher cytotoxicity of CuET than of the primary DSF metabolite DTC in various cancer cell lines (Fig. 1d and Extended Data Fig. 1g). The half-maximal lethal dose (LD_{50}) values of CuET in CFA experiments were ≤100 nM in three out of three tested breast cancer cell lines and similar potency was observed among cell lines derived from human lung, colon and prostate tumours (Extended Data Fig. 2a). These data were corroborated by tetrazolium dye ((2,3-bis-(2-methoxy-4-nitro-5-sulfophenyl)-2h-tetrazolium-5-carboxanilide) (XTT))-based 48-h cytotoxicity tests on a wider panel of cell types (Extended Data Fig. 2b). Unexpectedly, only the most sensitive cell lines (for example, AMO-1, Capan1) showed markers of apoptosis[18], which included annexin V and activated caspases, whereas in most cell lines, for example, MDA-MB-231 and U2OS cells, CuET induced apoptosis-independent cell death (Extended Data Fig. 2c–f).

Direct therapeutic effects of CuET *in vivo* were then investigated using the MDA-MB-231 breast cancer (Fig. 1e) and AMO-1 myeloma (Fig. 1f) xenograft models treated intraperitoneally with a CuET–albumin formulation, with which the anti-tumour activity and good tolerability of this DSF metabolite was confirmed (Extended Data Fig. 1h, i).

CuET inhibits p97-dependent protein degradation

Next, we investigated the interaction between CuET and cellular protein degradation, one of the suggested explanations for anti-tumour effects of DSF[6,12]. We confirmed that CuET induces phenotypic features shared with proteasome inhibitors, such as MG132 or bortezomib (BTZ), including accumulation of poly-ubiquitylated (poly-Ub) proteins (Fig. 2a and Extended Data Fig. 3a), rapid deubiquitylation of histone H2A (uH2A)[19] (Extended Data Fig. 3b) and accumulation of ubiquitylated proteins in the cytoplasm [19] (Extended Data Fig. 3c). Furthermore, TNF (also known as TNFα)-induced degradation of IκBα (ref. 20) was blocked after 1-h treatment with CuET or BTZ (Fig. 2b). Finally, CuET inhibited degradation of Ub(G76V)–GFP (an ubiquitin-fusion degradation substrate)[21] in a dose-dependent manner (Fig. 2c). However, although these data confirmed a defect in protein degradation, CuET had no effect on the CT-like, C-like or T-like activity of the 20S proteasome[22] (Extended Data Fig. 3d, e). This was further corroborated by the lack of a stabilizing effect of CuET on p53 tumour suppressor protein in dicoumarol-treated cells, in which

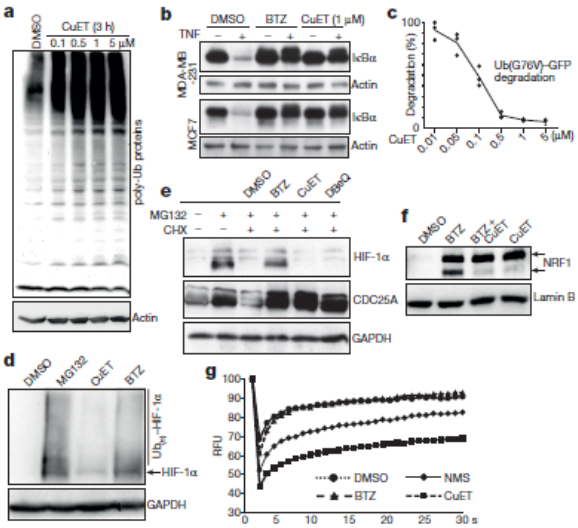


Figure 2 | CuET inhibits p97 segregase-dependent protein degradation. **a**, CuET causes accumulation of poly-ubiquitylated proteins in MCF7 cells. **b**, TNF-induced IκBα degradation is compromised after 1-h treatment with CuET or BTZ. **c**, Dose-dependent inhibition of Ub(G76V)–GFP degradation by CuET. HeLa cells were treated for 3 h. $n = 3$ experiments. **d**, HIF-1α levels after 2-h treatments with MG132 (5 μM), CuET (1 μM), BTZ (1 μM) in HeLa cells. **e**, Differential effect of BTZ (1 μM), CuET (1 μM) and DBeQ (10 μM) on CDC25A versus HIF-1α in MG132-pretreated (4 h, 5 μM), cycloheximide (CHX, 1 h, 50 μg ml^{-1})-exposed HeLa cells. **f**, BTZ (8 h, 1 μM) induces NRF1 120-kDa (top arrow) and 110-kDa (bottom arrow) forms; whereas CuET (8 h, 0.5 μM) only induced the non-cleaved 120-kDa form in NIH3T3 cells. **g**, FRAP quantification in U2OS Ub–GFP cells: slower mobility of accumulated cytoplasmic GFP–Ub after a 2-h pre-treatment with NMS873 (10 μM), CuET (1 μM) or BTZ (1 μM). **a**, **b**, **d**–**g**, Data are representative of two independent biological experiments. Data are linked means and individual values (**c**) and relative mean signal of the bleached region from 12 cells per treatment (**g**).

p53 turnover depends on the core 20S proteasome independently of ubiquitin[23,24]. In contrast to CuET, treatment with the 20S proteasome inhibitor BTZ stabilized p53 irrespective of dicoumarol (Extended Data Fig. 3f), indicating that 20S proteasome-dependent protein turnover remains operational with CuET treatment. Furthermore, CuET failed to inhibit 26S proteasome activity (Extended Data Fig. 3g), which was inferred from RPN11-dependent deubiquitylation[25]. Collectively, these results suggest that CuET stabilizes ubiquitylated proteins by blocking a step upstream of the proteasome.

Next we considered p97-dependent processing of poly-Ub proteins, as this pathway operates upstream of the proteasome and its malfunction resembles phenotypes of proteasome inhibition[26]. Unlike BTZ or MG132, CuET induced only modest accumulation (a small subfraction) of HIF-1α (Fig. 2d), consistent with reported modest accumulation of HIF-1α after knockdown of p97 compared to cells with inhibited proteasomes[27]. Next, we pre-treated cells with MG132, followed by wash-off and 1-h cycloheximide (an inhibitor of translation) treatment combined with BTZ, CuET or DBeQ (a direct inhibitor of p97 ATPase activity)[28]. All tested inhibitors prevented degradation of CDC25A (a known p97 target)[29], whereas degradation of the mostly p97-independent target, that is, most of HIF-1α[27], was inhibited only by BTZ (Fig. 2e). Furthermore, consistent with cleavage of the 120-kDa species of the endoplasmic reticulum-tethered transcription factor NRF1 into an active 110-kDa form being a p97-dependent process[30], appearance of the cleaved NRF1 form was inhibited by both CuET and NMS873 (another p97 ATPase inhibitor) (Fig. 2f and Extended Data Fig. 4a, b). These results suggest that the p97 pathway is compromised in cells treated with CuET.

Next, we asked whether CuET impairs the p97 segregase activity that extracts poly-Ub proteins from cellular structures, such as the endoplasmic reticulum, Golgi apparatus or chromatin for subsequent proteasomal degradation[31]. Using fluorescence recovery after photobleaching (FRAP) to investigate the mobility of accumulated poly-Ub proteins, we found that whereas GFP–ubiquitin in DMSO- or BTZ-treated cells diffused rapidly into bleached areas, this diffusion was slower after treatment with CuET or NMS873 (Fig. 2g and Extended Data Fig. 4c). This suggests that after treatment with CuET or NMS873 at least a subset of the accumulated poly-Ub proteins remains immobile, probably embedded into cellular structures. Consistently, upon detergent pre-extraction of mobile proteins, we observed greater immunofluorescence signals of extraction-resistant poly-Ub(K48) proteins (destined for proteasomal degradation) in NMS873- and CuET-treated cells compared to BTZ- or DMSO-treated controls (Extended Data Fig. 4d). Western blot analysis of endoplasmic reticulum-rich microsomal fractions also revealed enrichment of poly-Ub proteins after CuET and NMS873 treatment (Extended Data Fig. 4e). Malfunction of p97 segregase is furthermore associated with a cellular unfolded protein response (UPR)[32]. We confirmed UPR in cells treated with CuET or NMS873 by detecting increased markers of UPR induction, including the spliced form of XBP1s, ATF4 and phosphorylated (p-) eIF2α[33] (Extended Data Fig. 4f).

These studies are also of clinical relevance, because inhibition of p97 was suggested as an alternative treatment strategy for myeloma patients who had relapsed after therapy with BTZ (also known by the trade name Velcade)[34] or carfilzomib (CFZ)[35]. Thus, we performed cytotoxicity tests with CuET on a panel of BTZ- or CFZ-adapted and non-adapted human cell lines or on cells derived from samples of patients with myeloma before therapy and with BTZ therapy. All pairs of adapted and non-adapted cells showed similar sensitivity to CuET treatment, in contrast to BTZ (Extended Data Fig. 5a–d). These results suggest that treatment with DSF (best combined with copper) or CuET might become a feasible therapeutical option for patients with relapsed, BTZ-resistant multiple myeloma.

CuET binds and immobilizes NPL4

To elucidate how CuET inhibits the p97 pathway, we first used an assay of p97 ATPase activity[28]. In contrast to treatment with NMS873, CuET had no effect on p97 ATPase activity (Extended Data Fig. 6a). Because NPL4 and UFD1 proteins are key components of the p97 segregase[31], we examined whether CuET might target the pathway through these cofactors. Ectopic overexpression of NPL4–GFP, but not UFD1–GFP or p97–GFP, reduced CuET cytotoxicity, suggesting that NPL4 is a candidate target of CuET (Fig. 3a and Extended Data Fig. 6b). An analogous 'rescue effect' of ectopic NPL4–GFP was apparent from the reduction in accumulation of poly-Ub proteins caused by CuET (Extended Data Fig. 6c).

As shown by live-cell imaging, 2–3-h exposure to CuET induced prominent nuclear clustering of NPL4–GFP, but not of UFD1–GFP or p97–GFP (Fig. 3b). Within 2–3 h, most of cellular NPL4–GFP became immobilized in nuclear clusters and also in cytoplasmic areas, as shown by FRAP (Fig. 3c). CuET-induced immobilization of endogenous NPL4 was confirmed by accumulation, which was detectable by western blot, in the detergent-insoluble fractions from various cell lines (Fig. 3d) and by immunofluorescence on pre-extracted cells (Extended Data Fig. 6d). Notably, the immobilization of NPL4 was also detectable in pre-extracted sections of cryopreserved tumours from mice treated with DSF or DSF and CuGlu, thus providing a potential biomarker of CuET activity towards the p97 pathway *in vivo* (Fig. 3e).

NPL4 is an attractive candidate for CuET binding, because this protein contains two zinc finger domains: a C-terminal NZF (NPL4–zinc finger) and a putative zinc finger–NPL4[36], which bind bivalent metals and metal complexes that might chemically resemble CuET[37].

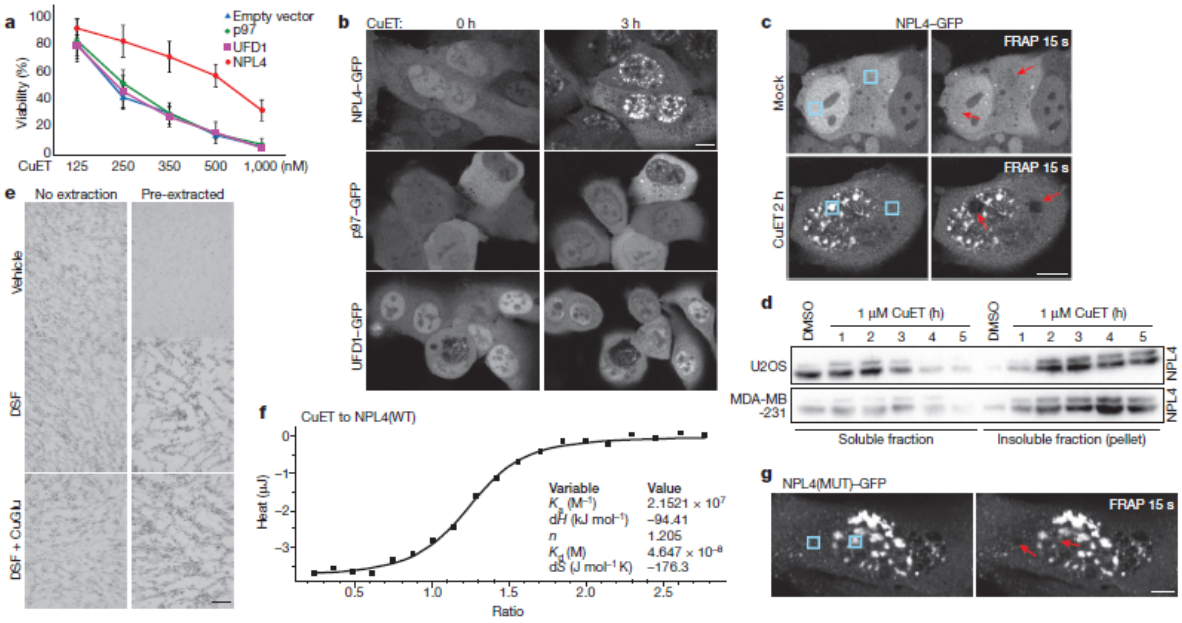


Figure 3 | CuET binds to and immobilizes NPL4. a, Ectopic NPL4–GFP, but not p97–GFP or UFD1–GFP rescues CuET toxicity in U2OS cells treated for 24 h. $n = 3$ experiments. Data are mean ± s.d. **b**, CuET (1 μM) induces intranuclear clustering of NPL4–GFP, but not p97–GFP or UFD1–GFP. **c**, CuET-induced (1 μM) immobilization of NPL4–GFP (FRAP) in U2OS cells treated for 2 h. Blue boxes, areas before bleaching; arrows, after bleaching. **d**, NPL4 enrichment in Triton X-100-insoluble fractions after CuET (1 μM) treatment. **e**, Immunohistochemistry demonstrates the non-extractable NPL4 in MDA-MB-231 tumours from mice treated with DSF or DSF and CuGlu. **f**, ICT shows that CuET binds to purified NPL4(WT). **g**, Spontaneous intranuclear clustering and immobilization of NPL4(MUT)–GFP using FRAP in U2OS cells. Blue boxes, areas before bleaching; arrows, after bleaching. Scale bars, 10 μm (**b**, **c**, **g**) or 50 μm (**e**). **b**–**g**, Data are representative of two independent experiments.

Using isothermal calorimetry analysis (ITC)[38], we observed a standard dose–response-dependent binding curve (Fig. 3f) compatible with one binding site for CuET on wild-type NPL4 (NPL4(WT)), and a K_d in nanomolar concentrations for the NPL4–CuET interaction. Next, we used mutagenesis to test whether the putative ZF–NPL4 domain has any role in the potential NPL4–CuET interaction. The putative zinc finger domain was preferred, because an endogenous larger form of NPL4 that lacks the C-terminal NZF sequence exists in human cells. This larger NPL4 form is detectable as an upper band on western blots (Fig. 3d) and it is immobilized after CuET treatment, suggesting that the C-terminal NZF is not necessary for the interaction with CuET. No ITC interaction was found with a NPL4 mutant (NPL4(MUT)) (Extended Data Fig. 6f) in which both histidines and both cysteines in the putative zinc finger domain were substituted by alanines (Extended Data Fig. 6e). We used drug affinity responsive target stability (DARTS) as another, independent approach, which is based on altered protease susceptibility of target proteins upon drug binding[39]. Consistently, exposure to CuET caused a differential pronase-dependent proteolysis pattern of NPL4(WT) but not NPL4(MUT) (Extended Data Fig. 6g). These results indicate that NPL4 is directly targeted by CuET and an intact putative zinc finger domain of NPL4 is essential for this interaction.

Notably, ectopically expressed NPL4(MUT)–GFP formed immobile nuclear clusters spontaneously in untreated cells, reminiscent of events seen in cells upon CuET treatment (Fig. 3c, g). Moreover, unlike ectopic NPL4(WT)–GFP, ectopically expressed NPL4(MUT)–GFP not only did not render cells resistant to CuET but also was toxic to the acceptor cells (Extended Data Fig. 6h). We also confirmed that multiple CuET-induced cellular phenotypes were mimicked by the ectopic NPL4(MUT)–GFP model, including accumulation of poly-Ub proteins and UPR activation (Extended Data Fig. 6i).

NPL4 aggregates trigger a heat-shock response

Although the nuclear NPL4 clusters occupied DAPI-unlabelled areas of chromatin (Extended Data Fig. 6d) co-localization with DAPI-excluded structures, such as nucleoli and nuclear speckles, were not found (Extended Data Fig. 7a). In late-G2 cells, NPL4 clusters were evidently excluded from the partially condensed chromatin (Extended Data Fig. 7b), suggesting that the NPL4 aggregates exclude chromatin rather than accumulating in specific nuclear areas. Both the nuclear clusters and the immobilized cytoplasmic NPL4 co-localized with poly-Ub proteins (confirmed by anti-Ub(K48) and FK2 antibodies), small ubiquitin-like modifiers (SUMOs) (only in nuclei) and with TDP43 protein[40] (Fig. 4a and Extended Data Fig. 7d), which are all features typical of aggregated defective proteins[41]. The same co-localization patterns were observed for spontaneous clusters formed by NPL4(MUT)–GFP showing that NPL4 aggregation is sufficient for the induction of these phenotypes even without CuET treatment (Extended Data Fig. 7c, e). Blockade of cellular ubiquitylation with a chemical inhibitor (MLN7243) of the E1 ubiquitin-activating enzyme failed to prevent either NPL4–GFP nuclear aggregation or cytoplasmic immobilization (Extended Data Fig. 7d), excluding the immobilization of NPL4 via recognition of ubiquitylated and SUMOylated substrates, but rather suggesting that immobilized NPL4 attracts ubiquitylated proteins or proteins that subsequently become ubiquitylated and/or SUMOylated. A key protein commonly associated with intracellular protein aggregates is HSP70, a chaperone implicated in aggregate processing[42]. Indeed, pre-extracted cells showed co-localization of HSP70 with both CuET-induced NPL4(WT)–GFP and spontaneous NPL4(MUT)–GFP aggregates (Fig. 4b, c). Both the CuET-induced NPL4(WT) aggregates and spontaneous NPL4(MUT) aggregates also co-localized with p97 (Extended Data Fig. 7f, g), as is particularly evident after pre-extraction. In the NPL4–GFP model, the amount of p97 immunoreactivity within

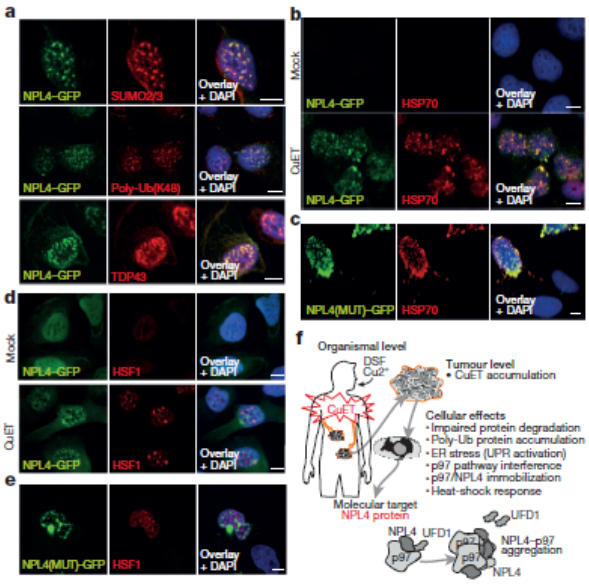


Figure 4 | NPL4 protein aggregation triggers HSR. a, NPL4–GFP co-localizes with SUMO2/3, poly-Ub(K48) and TDP43 in U2OS cells. Cells were treated with 1 μM CuET for 3 h and pre-extracted. **b**, NPL4–GFP co-localizes with HSP70 in mock- and CuET-treated U2OS cells. Cells were treated with 1 μM CuET for 3 h and pre-extracted. **c**, NPL4(MUT)–GFP co-localizes with HSP70 in U2OS cells after pre-extraction. **d**, CuET-induced HSF1 stress bodies. NPL4–GFP U2OS cells were treated with 1 μM CuET for 3 h. **e**, HSF1 stress bodies in U2OS cells expressing NPL4(MUT)–GFP. **f**, Model of DSF anti-cancer activity in patients. Scale bars, 10 μm. **a**–**e**, Data are representative of two independent experiments.

the NPL4–GFP clusters correlated with the GFP signal intensity, suggesting that p97 is immobilized via its interaction with NPL4. The other NPL4-binding partner, UFD1, was almost undetectable in detergent-insoluble pellets of CuET-treated or NPL4(MUT)–GFP-expressing cells despite clear p97 immobilization (Extended Data Fig. 8a, b), suggesting that UFD1 cannot bind to, or becomes only loosely attached to, the aggregated NPL4–p97 complex. Notably, non-extractable cellular p97 is detectable after CuET treatment (Extended Data Fig. 8c), including in stained tumour sections from mice treated with DSF or DSF and CuGlu, providing an additional candidate marker for CuET activity *in vivo* (Extended Data Fig. 8d).

Because aggregation of misfolded or damaged proteins triggers cellular heat-shock response (HSR) through an HSF1-dependent mechanism[43], we confirmed that CuET treatment indeed triggered a robust HSR accompanied by characteristic HSF1 nuclear stress foci (Fig. 4d) that were also detectable in cells spontaneously aggregating NPL4(MUT)–GFP (Fig. 4e). HSR markers, including accumulation of heat-shock proteins and a phosphorylation shift in HSF1, were detectable by western blot (Extended Data Fig. 8e, f).

Discussion

Our results help to explain the anti-cancer activity of the alcohol-abuse drug disulfiram. We propose a model for DSF cytotoxic activity, featuring rapid conversion of DSF into CuET, which accumulates in tumours. After entering cells, CuET binds NPL4 and induces its aggregation, consequently disabling the vital p97–NPL4–UFD1 pathway and inducing a complex cellular phenotype leading to cell death (Fig. 4f). Supporting CuET as the active metabolite is the correlation of CuET concentrations (active in the nanomolar range) with the biological effects and functional impact on the targeted pathway(s) *in vivo*. In addition, CuET is the only known metabolite of DSF containing copper ions, a metal that enhances the anti-tumour effects of DSF; it is unlikely that another DSF metabolite could represent the major anti-cancer agent as levels of non-CuET metabolites should be lowered by copper addition. We also present a method for CuET detection in tissues and plasma, as well as data suggesting that preferential accumulation of CuET in tumours may contribute to cancer cell toxicity, consistent with the high therapeutic tolerability of DSF[3], as documented even after years of daily administration at doses comparable to those we used in our mouse experiments. Considering the numerous studies on DSF and diverse opinions about the potential target of its anti-cancer effects[44], identification of NPL4, a key component of the p97–NPL4–UFD1 segregase complex, as the molecular target of CuET is surprising. The CuET–NPL4 interaction leads to rapid formation of protein aggregates and immobilization of this otherwise very mobile multifunctional protein complex, resulting in a severe phenotype, induction of HSR and eventually cell death. While additional potential targets of CuET cannot be excluded, the malfunction of the p97 pathway due to the NPL4–p97 aggregate formation explains the major cell phenotypes and the consequent cell death. Our work also reconciles the controversial studies[6,12], suggesting that the proteasome is the DSF target, by demonstrating that neither 20S nor 26S proteasome, but the processing of ubiquitylated proteins by the NPL4-dependent segregase, is targeted by CuET. Our results support the notion that the p97–NPL4 pathway is a promising therapeutic target in oncology[45,46]. Indeed, reports on p97 overabundance correlating with progression and metastasis of carcinomas of the breast, colon and prostate[47–49] are consistent with our present nationwide epidemiological analysis, which revealed an association between continued use of DSF and favourable prognosis, an intriguing finding that should be investigated further, particularly given the currently limited therapeutic options for patients with metastatic cancer. From a broader perspective, our study illustrates the potential of multifaceted approaches to drug repurposing, providing novel mechanistic insights, identification of new cancer-relevant targets and encouragement for further clinical trials, here with DSF, an old, safe and public domain drug[4] that might help to save lives of patients with cancer worldwide.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October 2015; accepted 8 November 2017.
Published online 6 December 2017.

1. Collins, F. S. Mining for therapeutic gold. *Nat. Rev. Drug Discov.* **10,** 397 (2011).
2. Pantziarka, P. *et al.* The repurposing drugs in oncology (ReDO) project. *Ecancermedicalscience* **8,** 442 (2014).
3. Iljin, K. *et al.* High-throughput cell-based screening of 4910 known drugs and drug-like small molecules identifies disulfiram as an inhibitor of prostate cancer cell growth. *Clin. Cancer Res.* **15,** 6070–6078 (2009).
4. Cvek, B. Nonprofit drugs as the salvation of the world's healthcare systems: the case of Antabuse (disulfiram). *Drug Discov. Today* **17,** 409–412 (2012).
5. Shen, M. L., Johnson, K. L., Mays, D. C., Lipsky, J. J. & Naylor, S. Determination of *in vivo* adducts of disulfiram with mitochondrial aldehyde dehydrogenase. *Biochem. Pharmacol.* **61,** 537–545 (2001).
6. Chen, D., Cui, Q. C., Yang, H. & Dou, Q. P. Disulfiram, a clinically used anti-alcoholism drug and copper-binding agent, induces apoptotic cell death in breast cancer cultures and xenografts via inhibition of the proteasome activity. *Cancer Res.* **66,** 10425–10433 (2006).
7. Zha, J. *et al.* Disulfiram targeting lymphoid malignant cell lines via ROS–JNK activation as well as Nrf2 and NF-κB pathway inhibition. *J. Transl. Med.* **12,** 163 (2014).
8. Safi, R. *et al.* Copper signaling axis as a target for prostate cancer therapeutics. *Cancer Res.* **74,** 5819–5831 (2014).
9. Liu, P. *et al.* Liposome encapsulated disulfiram inhibits NFκB pathway and targets breast cancer stem cells *in vitro* and *in vivo*. *Oncotarget* **5,** 7471–7485 (2014).
10. Dufour, P. *et al.* Sodium dithiocarb as adjuvant immunotherapy for high risk breast cancer: a randomized study. *Biotherapy* **6,** 9–12 (1993).
11. Yip, N. C. *et al.* Disulfiram modulated ROS–MAPK and NFκB pathways and targeted breast cancer cells with cancer stem cell-like properties. *Br. J. Cancer* **104,** 1564–1574 (2011).

12. Lövborg, H. *et al.* Inhibition of proteasome activity, nuclear factor-κB translocation and cell survival by the antialcoholism drug disulfiram. *Int. J. Cancer* **118,** 1577–1580 (2006).
13. Allensworth, J. L. *et al.* Disulfiram (DSF) acts as a copper ionophore to induce copper-dependent oxidative stress and mediate anti-tumor efficacy in inflammatory breast cancer. *Mol. Oncol.* **9,** 1155–1168 (2015).
14. Nechushtan, H. *et al.* A phase IIb trial assessing the addition of disulfiram to chemotherapy for the treatment of metastatic non-small cell lung cancer. *Oncologist* **20,** 366–367 (2015).
15. Jin, M. *et al.* Alcohol drinking and all cancer mortality: a meta-analysis. *Ann. Oncol.* **24,** 807–816 (2013).
16. Li, Y. *et al.* Copper improves the anti-angiogenic activity of disulfiram through the EGFR/Src/VEGF pathway in gliomas. *Cancer Lett.* **369,** 86–96 (2015).
17. Suzuki, Y. *et al.* The origin of an EPR signal observed in dithiocarbamate-loaded tissues. Copper(ii)-dithiocarbamate complexes account for the narrow hyperfine lines. *Biochim. Biophys. Acta* **1335,** 242–245 (1997).
18. Kepp, O., Galluzzi, L., Lipinski, M., Yuan, J. & Kroemer, G. Cell death assays for drug discovery. *Nat. Rev. Drug Discov.* **10,** 221–237 (2011).
19. Doil, C. *et al.* RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* **136,** 435–446 (2009).
20. Li, J. M., Wu, H., Zhang, W., Blackburn, M. R. & Jin, J. The p97–UFD1L–NPL4 protein complex mediates cytokine-induced IκBα proteolysis. *Mol. Cell. Biol.* **34,** 335–347 (2014).
21. Chou, T. F. & Deshaies, R. J. Quantitative cell-based protein degradation assays to identify and classify drugs that target the ubiquitin–proteasome system. *J. Biol. Chem.* **286,** 16546–16554 (2011).
22. Kisselev, A. F. & Goldberg, A. L. Monitoring activity and inhibition of 26S proteasomes with fluorogenic peptide substrates. *Methods Enzymol.* **398,** 364–378 (2005).
23. Asher, G., Lotem, J., Cohen, B., Sachs, L. & Shaul, Y. Regulation of p53 stability and p53-dependent apoptosis by NADH quinone oxidoreductase 1. *Proc. Natl Acad. Sci. USA* **98,** 1188–1193 (2001).
24. Asher, G., Tsvetkov, P., Kahana, C. & Shaul, Y. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev.* **19,** 316–321 (2005).
25. Verma, R. *et al.* Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298,** 611–615 (2002).
26. Dai, R. M. & Li, C. C. Valosin-containing protein is a multi-ubiquitin chain-targeting factor required in ubiquitin–proteasome degradation. *Nat. Cell Biol.* **3,** 740–744 (2001).
27. Alexandru, G. *et al.* UBXD7 binds multiple ubiquitin ligases and implicates p97 in HIF1α turnover. *Cell* **134,** 804–816 (2008).
28. Chou, T. F. *et al.* Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways. *Proc. Natl Acad. Sci. USA* **108,** 4834–4839 (2011).
29. Riemer, A. *et al.* The p97–Ufd1–Npl4 ATPase complex ensures robustness of the G2/M checkpoint by facilitating CDC25A degradation. *Cell Cycle* **13,** 919–927 (2014).
30. Radhakrishnan, S. K., den Besten, W. & Deshaies, R. J. p97-dependent retrotranslocation and proteolytic processing govern formation of active Nrf1 upon proteasome inhibition. *eLife* **3,** e01856 (2014).
31. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat. Cell Biol.* **14,** 117–123 (2012).
32. Magnaghi, P. *et al.* Covalent and allosteric inhibitors of the ATPase VCP/p97 induce cancer cell death. *Nat. Chem. Biol.* **9,** 548–556 (2013).
33. Samali, A., Fitzgerald, U., Deegan, S. & Gupta, S. Methods for monitoring endoplasmic reticulum stress and the unfolded protein response. *Int. J. Cell Biol.* **2010,** 830307 (2010).
34. Auner, H. W. *et al.* Combined inhibition of p97 and the proteasome causes lethal disruption of the secretory apparatus in multiple myeloma cells. *PLoS ONE* **8,** e74415 (2013).
35. Soriano, G. P. *et al.* Proteasome inhibitor-adapted myeloma cells are largely independent from proteasome activity and show complex proteomic changes, in particular in redox and energy metabolism. *Leukemia* **30,** 2198–2207 (2016).
36. Lass, A., McConnell, E., Fleck, K., Palamarchuk, A. & Wójcik, C. Analysis of Npl4 deletion mutants in mammalian cells unravels new Ufd1-interacting motifs and suggests a regulatory role of Npl4 in ERAD. *Exp. Cell Res.* **314,** 2715–2723 (2008).
37. Voráčková, I., Suchanová, S., Ulbrich, P., Diehl, W. E. & Ruml, T. Purification of proteins containing zinc finger domains using immobilized metal ion affinity chromatography. *Protein Expr. Purif.* **79,** 88–95 (2011).
38. Holdgate, G. *et al.* Biophysical methods in drug discovery from small molecule to pharmaceutical. *Methods Mol. Biol.* **1008,** 327–355 (2013).
39. Lomenick, B. *et al.* Target identification using drug affinity responsive target stability (DARTS). *Proc. Natl Acad. Sci. USA* **106,** 21984–21989 (2009).
40. Becker, L. A. *et al.* Therapeutic reduction of ataxin-2 extends lifespan and reduces pathology in TDP-43 mice. *Nature* **544,** 367–371 (2017).
41. Guo, L. *et al.* A cellular system that degrades misfolded proteins and protects against neurodegeneration. *Mol. Cell* **55,** 15–30 (2014).
42. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82,** 323–355 (2013).
43. Dai, C. & Sampson, S. B. HSF1: guardian of proteostasis in cancer. *Trends Cell Biol.* **26,** 17–28 (2016).
44. Cvek, B. Targeting malignancies with disulfiram (Antabuse): multidrug resistance, angiogenesis, and proteasome. *Curr. Cancer Drug Targets* **11,** 332–337 (2011).
45. Deshaies, R. J. Proteotoxic crisis, the ubiquitin–proteasome system, and cancer therapy. *BMC Biol.* **12,** 94 (2014).
46. Anderson, D. J. *et al.* Targeting the AAA ATPase p97 as an approach to treat cancer through disruption of protein homeostasis. *Cancer Cell* **28,** 653–665 (2015).
47. Cui, Y. *et al.* High expression of valosin-containing protein predicts poor prognosis in patients with breast carcinoma. *Tumour Biol.* **36,** 9919–9927 (2015).
48. Yamamoto, S. *et al.* Expression of valosin-containing protein in colorectal carcinomas as a predictor for disease recurrence and prognosis. *Clin. Cancer Res.* **10,** 651–657 (2004).
49. Tsujimoto, Y. *et al.* Elevated expression of valosin-containing protein (p97) is associated with poor prognosis of prostate cancer. *Clin. Cancer Res.* **10,** 3007–3012 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Škvor, M. Zadinová, J. Večerka and D. Doležal for help with animal experiments, Jana Vrbkova for statistical analysis, D. Fridecky and T. Adam for help with HPLC, I. Protivankova and M. Grønvig Nielsen for technical assistance. This work was supported by grants from the Kellner Family Foundation, Czech National Program of Sustainability, Grant Agency of the Czech Republic, MEYS CR project Czech-BioImaging, the Czech Health Research Council, of the Danish Cancer Society, the Danish National Research Foundation (project CARD), the Danish Council for Independent Research, the Novo Nordisk Foundation, the Czech Cancer League, the Swedish Research Council, Cancerfonden of Sweden, the European Commission (EATRIS), the Czech Ministry of Education, youth and sports (OPVKCZ), Cancer Research Czech Republic and the Howard Hughes Medical Institute.

Author Contributions Z.S., M.Mis., B.C., R.J.D. and J.Barte. conceived the study. Z.S. and M.Mis. performed most biochemical and microscopy experiments and wrote the manuscript. D.M. established the expression cell lines and performed most cytotoxicity tests. T.O., P.D. and I.V. performed the HPLC experiments. K.K.A., S.F. and J.O. performed the epidemiological analyses. J.Bartk. performed the immunohistochemical analyses. J.V. and P.D. performed DARTS experiments. P.M. performed cell death analyses. Z.T. performed cytotoxicity tests and heat-shock response analyses. A.K. performed cytotoxicity tests. A.M. designed and performed phlebotomies of patients treated with Antabuse. M.Mic. performed the ITC. J.G. performed FACS analyses, cell death assays and cell sorting. J.S. performed 20S proteasome assays. J.L. performed 26S proteasome assays. M.K. and C.D. performed the cytotoxicity experiments on myeloid- and patient-derived cell lines. P.P., J.M. and M.H. performed mouse experiments. J.Barte., B.C., Q.P.D. and R.J.D. helped to design the experiments, interpreted the data and wrote/edited the manuscript. All authors approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.Barte. (jb@cancer.dk), B.C. (cvekb@seznam.cz) and R.J.D. (deshaies@caltech.edu).

Reviewer Information *Nature* thanks P. Brossart and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

The experiments were not randomized.

Epidemiological analyses and access to health registers. We conducted a population-based cohort study by combining Danish nationwide demographic and health registers. This study was approved by the Danish Data Protection Agency and Statistics Denmark's Scientific Board. As the epidemiological study was based solely on register data and did not involve any contact with patients, no ethical approval was required from the Danish Scientific Ethical Committee[50]. The cohort consisted of all Danes aged 35–85 years with a first-time diagnosis of cancer between January 2000 and December 2013. Because DSF (Antabuse) is a relative contra-indication among individuals with liver or kidney diseases, we excluded patients with cancers of the liver or kidney from the cohort. Cohort members were categorized according to use of DSF into two main groups: (i) those who filled at least one prescription of DSF within five years before the cancer diagnosis, but did not fill DSF prescription(s) during the first year after the diagnosis (previous users), that is, individuals suffering from alcoholism but taking DSF only before their cancer diagnosis; and (ii) those who used DSF before their cancer diagnosis and also filled one or more DSF prescriptions during the first year after the cancer diagnosis (continuing users), that is, individuals who underwent DSF therapy both before and after the cancer diagnosis. We also defined a category of patients with cancer who did not fill prescription(s) for DSF either before or after ($\leq$1 year) the cancer diagnosis (never users). In the main analyses, we calculated hazard ratios and 95% confidence intervals estimating cancer-specific mortality among continuing DSF users compared to previous DSF users based on a Cox model regressing on both propensity scores and disulfiram use. By including propensity scores in the regression, we used demographic data and comorbid conditions/diagnostic codes as well as prescription data for selected concomitant drugs, to balance baseline characteristics of previous and continuing users of DSF and to adjust estimated hazard ratios of cancer-specific mortality associated with DSF use[51]. The patients with cancer were followed from one year after the diagnosis until death, migration or end of study (31 December 2014). The propensity scores thus estimate the probability of being treated with DSF in the exposure window 0–1 year after the cancer diagnoses conditional on the following other covariates in the calculation of propensity scores using logistic regression: gender, age at diagnosis, calendar time, highest achieved education and disposable income; medical histories of diabetes mellitus, chronic obstructive pulmonary disease, ischaemic heart disease, congestive heart failure, cerebrovascular disease, atrial fibrillation or atrial flutter, dementia and ulcer disease; and use of non-steroidal anti-inflammatory drugs (NSAIDs; including aspirin), non-aspirin antithrombotic agents (anticoagulants), statins, antihypertensive medication, other cardiovascular drugs, anti-diabetics and psychotropic drugs. In the Cox model, the propensity score is further included as a restricted cubic spline to model possible nonlinearities, in addition to the categorical disulfiram use, which is the variable of interest. Analyses were run for cancer overall and for breast, prostate and colon cancer, separately. Furthermore, all analyses were stratified by stage (localized, non-localized or unknown). Statistical significance of DSF use was evaluated by likelihood ratio tests. We used the software R for statistical computing[52].

***In vivo* tumour experiments.** The human breast cancer cell line MDA-MB-231 was injected (10^7 cells transplanted subcutaneously) to grow tumours in athymic NU/NU female mice (AnLab Ltd) with a body weight of 23.6–26.9 g, aged 12 weeks. Inclusion criteria were: female, appropriate age and weight (15–30 g). Exclusion criteria were: tumour size must not exceed 20 mm (volume 4,000 mm^3) in any direction in an adult mouse, the tumour mass should not proceed to the point where it significantly interferes with normal bodily functions, or causes pain or distress owing to its location, persistent self-induced trauma, rapid or progressive weight loss of more than 25%, for seven days. In none of the experiments were these approved ethical limits exceeded. After the tumours grew to 0.114–0.117 cm^3 on average, mice were randomly divided into four groups, each of eight mice, and treated as follows: (i) normal diet; (ii) normal diet plus oral administration of 0.15 mg kg^{-1} copper gluconate (CuGlu); (iii) normal diet plus oral administration of 50 mg kg^{-1} DSF; (iv) normal diet plus oral administration of 0.15 mg kg^{-1} CuGlu and 50 mg kg^{-1} DSF. Administration of compounds was carried out as a blinded experiment (all information about the expected outputs and the nature of used compounds were kept from the animal technicians). CuGlu was administered each day in the morning (08:00) and DSF each day in the evening (19:00) to mimic a clinical trial combining DSF and CuGlu in treatment of tumours involving the liver (NCT00742911). After treatment began, mice were weighed and their tumours measured twice per week. At day 32, mice were euthanized, and the tumours were removed and frozen at −80 °C. The experiment was evaluated by comparing growth curves of tumours in the experimental groups with those in controls. The rates of tumour growth inhibition (TGI) were calculated by the formula $TGI = (1 - V_{treated}/V_{control}$ where $V_{treated}$ is the mean of tumour volumes in the treated group and $V_{control}$ is the mean of tumour volumes in the control group). Mean tumour volume values at specific time intervals were statistically evaluated. To test directly the effect of CuET, we used MDA-MB-231 and AMO-1 models. MDA-MB-231 was injected (5×10^6 cells were transplanted subcutaneously) to grow tumours in SCID mice (ENVIGO) aged 10 weeks (±2 weeks). AMO-1 xenografts were expanded in SCID mice. Each group consisted of 10 animals, each bearing two tumours. CuET was formulated in bovine serum albumin solution (1%) to a final concentration of 1 mg ml^{-1}. CuET was applied intraperitoneally with a schedule of five days on and two days off. All aspects of the animal study met the accepted criteria for the care and experimental use of laboratory animals, and protocols were approved by the Animal Research Committee of the 1st Faculty of Medicine Charles University in Prague and Ethical Committee of Faculty of Medicine and Dentistry, Palacky University in Olomouc. For HPLC–MS and immunohistochemistry analysis, we used MDA-MB-231 xenografted mice treated with the same DSF and DSF plus copper gluconate regime as described for the anti-cancer activity assessment with the notable difference that mice were euthanized after five days of treatment.

HPLC–MS analysis of CuET. The HR-MRM analysis was performed on a HPLC-ESI-QTOF system consisting of HPLC chromatograph Thermo UltiMate 3000 with AB Sciex TripleTOF 5600+ mass spectrometer, using the DuoSpray ESI source operated at an ion source voltage of 5,500 V, ion source gas flow rates of 40 units, curtain gas flow rate of 30 units, declustering potential of 100 V and temperature 400 °C. Data were acquired in product ion mode with two parent masses (358.9 and 360.9) for analysis of CuET. Chromatographic separation was done by PTFE column, which was especially designed for analysis of strong metal chelators filled by C18 sorbent (IntellMed, IM_301). Analysis was performed at room temperature and with a flow rate of 1,500 µl min^{-1} with isocratic chromatography. Mobile phase consisted of HPLC grade acetone (Lachner) 99.9%, HPLC water (Merck Millipore) 0.1% and 0.03% HPLC formic acid (Sigma-Aldrich). Acquired mass spectra were evaluated in software PeakView 1.2. Extracted ion chromatograms of transitions 88.0 and 116.0 (common for both parent masses) with a 0.1 mass tolerance were Gaussian smoothened with width of two points. Peak area was then recorded and recalculated to ng ml^{-1} according to the calibration curve.

Sample preparation for HPLC–MS analysis. Liquid nitrogen-frozen biological samples were cut into small pieces using a scalpel. Samples (30–100 mg) were immediately processed by homogenization in 100% acetone in a ratio of 1:10 sample: acetone (for plasma or serum the ratio was 1:4). Homogenization was done in a table-top homogenizer (Retsch MM301) placed in a cold room (4 °C) in 2-ml Eppendorf tubes with 2 glass balls (5 mm) for 1 min at 30 Hz. The tube was immediately centrifuged at 4 °C, 20,000*g* for 2 min. Supernatant was decanted into a new 1.5-ml Eppendorf tube and immediately centrifuged for 30 min using a small table-top centrifuge (BioSan FVL-2400N) placed inside a −80 °C freezer. Supernatant was quickly decanted into a glass HPLC vial and kept at −80 °C for no longer than 6 h. Just before the HPLC analysis, the vial was placed into a pre-cooled (4 °C) LC-sample rack and immediately analysed. To enable an approximate quantification of analysed CuET, a calibration curve was prepared. Various amounts of CuET were spiked in plasma, frozen in liquid nitrogen, and placed at −80 °C to mimic sample processing. Standards were then processed as the samples described above. To measure circulating CuET concentrations, mice were given a single oral dose of DSF (50 mg kg^{-1}) and euthanized at different time points. Serum was collected and frozen for analysis.

Blood collection from humans for HPLC–MS analysis of CuET. Blood samples were collected before and 3 h after oral application of DSF (Antabuse, 400 mg) dissolved in water. Phlebotomy needles were specific for metal analysis—Sarstedt Safety Kanule 21*G* × 1½ inches, 85.1162.600. Collection tubes were specific for metal analysis —Sarstedt, S-Monovette 7.5 ml LH, 01.1604.400. Immediately after blood collection samples were centrifuged in a pre-cooled centrifuge (4 °C at 1,300*g* for 10 min). After centrifugation, tubes were placed on ice and the plasma fraction was immediately aliquoted into the 1.5-ml Eppendorf tubes with approximately 500 µl per tube. The tubes with plasma were immediately frozen on dry ice and later stored in −80 °C. Blood samples were collected from volunteers who gave informed consent and were undergoing Antabuse therapy because of alcohol abuse. Participants were four males (aged 34, 38, 41, 60 years) and five females (aged 37, 56, 46, 59, 63 years). All individuals were freshly diagnosed for alcohol-use disorder and were scheduled for Antabuse therapy. Blood samples were collected before and after the first use of Antabuse. All relevant ethical regulations were followed for the study. The study, including the collection of blood samples, was approved by the Ethical Committee of Faculty of Medicine and Dentistry, Palacky University in Olomouc.

Cell lines. Cell lines were cultured in appropriate growth medium supplemented with 10% fetal bovine serum (FBS) and penicillin–streptomycin; and maintained in a humidified, 5% CO_2 atmosphere at 37 °C. Cell lines cultured in DMEM medium were: HCT116 (ATCC), DU145 (ECACC), PC3 (ECACC), T47D (NCI60),

HS578T (NCI60), MCF7 (ECACC), MDA-MB-231 (ATCC), U2OS (ECACC), HeLa (ATCC), NIH-3T3 (ATCC), CAPAN-1 (ATCC), A253 (ATCC), FaDu (ATCC), h-TERT-RPE1 (ATCC), HeLa-Ub(G76V)-GFP-ODD-Luc[21]. Cell lines cultured in RPMI1640 medium were: NCI-H358 (ATCC), NCI-H52 (ATCC), HCT-15 (ATCC), AMO-1 (ATCC), MM-1S (ATCC), ARH77 (ATCC), RPMI8226 (ATCC), OVCAR-3 (NCI60), CCRF-CEM (ATCC), K562 (ATCC), 786-0 (NCI60). Cell lines cultured in EMEM medium were: U87-MG (ATCC), SiHA (ATCC). Cell line A549 (ATCC) was cultured in F12K medium, HT29 (ATCC) in McCoy's medium and LAPC4 (provided by Z. Culig, University of Innsbruck) in IMDM medium supplemented with metribolone R1881 (Sigma-Aldrich). RWPE-1 (ATCC) cells were cultured in a keratinocyte serum-free medium supplemented with bovine pituitary extract and human recombinant epidermal growth factor (Thermo Fisher Scientific). BTZ- and CFZ-resistant multiple myeloma cell lines were previously described in ref. 35. Cell lines were tested for mycoplasma contamination and authenticated by STR method. None of the cell lines used in this study is listed in the database of commonly misidentified cell lines maintained by ICLAC.

Stable cell line construction. For construction of all stably transfected cell lines we used the U2OS cell line (ECACC). For U2OS Ub–GFP, we used the commercial Ub–GFP EGFP-C1 vector (Addgene); for U2OS NPL4–GFP, we used the commercial NPLOC4–GFP pCMV6-AC-GFP vector (Origene); for U2OS p97–GFP, we used the commercial VCP–GFP pCMV6-AC-GFP vector (Origene); and for U2OS UFD1–GFP, we used the commercial UFD1L–GFP pCMV6-AC-GFP vector (Origene). Cells were transfected using Promega FugeneHD according to the manufacturer's instructions. Cells were further cultured in the appropriate antibiotics (geneticin, 400 $\mu g\ ml^{-1}$). Medium with geneticin was replaced every 2–3 days until the population of resistant cells was fully established. Cells were further refined by sorting for cells expressing GFP (BD FACS Aria). For preparation of inducible NPL4(MUT)–GFP cells, U2OS cells were transfected with a pcDNA6/TR plasmid (Invitrogen, V1025-20) using the FugeneHD transfection reagent (Promega, E2311) according to the manufacturer's protocol. To generate a cell line that stably expressed the Tet repressor, U2OS cells were cultured in selective medium with blasticidin (10 $\mu g\ ml^{-1}$) for 10 days. Blasticidin-resistant colonies were picked, expanded and screened for clones that exhibited the lowest basal levels and highest inducible levels of expression. Next, the most suitable clones were transfected with the PCDNA4/TO expression vector encoding the mutated NPL4–GFP protein using the Fugene transfection reagent. Cells were cultured in medium with zeocin (500 $\mu g\ ml^{-1}$) to select clones that contain pcDNA4/TO-mutated NPL4–GFP. The NPL4(MUT)–GFP-encoding plasmid was obtained from Generi Biotech. To induce expression of protein, cells were incubated with doxycycline (Sigma-Aldrich) 1 $\mu g\ ml^{-1}$ for 16–48 h.

Colony-formation assay. Cells were seeded into six-well plates at 100–300 cells per well (depending on the cell line). The next day, cells were treated with compounds as indicated in the specific experiments and kept in culture for 7–14 days. Colonies were visualized by crystal violet and counted.

XTT assay. Cells were plated at a density of 10,000 per well in a 96-well plate. The next day, cells were treated as indicated. After 24 h, an XTT assay was performed according to the manufacturer's instructions (Applichem). XTT solution was added to the medium and incubated for 30–60 min, and then the dye intensity was measured at the 475 nm wavelength using a spectrometer (TECAN, Infinite M200PRO). Results are shown as mean ± s.d. from three independent experiments, each performed in triplicate. For LD_{50} analysis across the panel of cell lines listed in Extended Data Fig. 2b, cell lines were treated with various doses (at least five doses) for 48 h. LD50 values were calculated using Graphpad Prism software based on survival curves from at least two independent experiments.

Annexin V staining. Cell cultures were treated as indicated and collected by trypsinization. Initial culture medium and washing buffer were collected to include detached cells. Cells were centrifuged (250*g*, 5 min) and re-suspended in a staining buffer (140 mM NaCl, 4 mM KCl, 0.75 mM $MgCl_2$, 10 mM HEPES) containing 2.5 mM $CaCl_2$, Annexin-V–APC (1:20, BD Biosciences) and 2.5 $\mu g\ ml^{-1}$ 7-AAD (BD Biosciences) for 15 min on ice in the dark. Samples were analysed by flow cytometry using BD FACSVerse (BD Biosciences) and at least 10,000 events were acquired per sample. Collected data were processed using BD FACSSuite (BD Biosciences) and exported into Microsoft Excel.

Caspases 3/7 assay. Activity of caspase-3 and -7 was quantified by cleavage of fluorogenic substrate CellEvent Caspase-3/7 Green Detection Reagent (Thermo Fisher Scientific). In brief, samples prepared in the same staining buffer as described for annexin V staining above, supplemented with 2% FBS, 0.5 μM CellEvent Caspase-3/7 Green Detection Reagent and incubated for 45 min at room temperature in the dark. Subsequently, 0.5 $\mu g\ ml^{-1}$ DAPI was added and samples were analysed by flow cytometry using BD FACSVerse (BD Biosciences) and at least 10,000 events were acquired per sample. Collected data were processed using BD FACSSuite (BD Biosciences) and exported into Microsoft Excel.

Viability assay of multiple myeloma cells. The CellTiter 96 MTS-assay (Promega) was used according to the manufacturer's instructions to determine the cell viability of BTZ (Janssen Cilag), CFZ and CuEt in cell lines and the absorbance of the formazan product was measured in 96-well microplates at 492 nm. The assay measures dehydrogenase enzyme activity found in metabolically active cells.

For patient cells, the more sensitive luminescent CellTiterGlo assay (Promega) was used to determine cell viability, measured by ATP production of metabolically active cells. The primary myeloma cell samples were obtained after written informed consent and approval by the independent ethics review board (St Gallen ethics committee—Ethikkommission Ostschweiz), in accordance with ICH-GCP and local regulations. Malignant plasma cells were retrieved by PBMC isolation from a patient with multiple myeloma progressing under BTZ-containing therapy, based on IMWG criteria (BTZ-resistant) and an untreated patient with multiple myeloma (BTZ-sensitive). The purity of the cell samples was >80% myeloma cells, as assessed by morphology.

Immunoblotting and antibodies. Equal amounts of cell lysates were separated by SDS–PAGE on hand-cast or precast tris–glycine gradient (4–20%) gels (Life Technologies), and then transferred onto a nitrocellulose membrane. The membrane was blocked with 5% bovine milk in Tris-buffered saline containing 0.1% Tween-20 for 1 h at room temperature, and then incubated overnight at 4 °C or for 1 h at room temperature, with one of the following primary antibodies (all antibodies were used in the system under study (assay and species) according to the instructions of the manufacturer): anti-ubiquitin (1:1,000; Cell Signaling, 3933), anti-H2A, acidic patch (1:1,000; Merck Millipore, 07-146), anti-monoubiquityl-H2A (1:1,000; Merck Millipore, clone E6C5), anti-IκBα (1:500; Santa Cruz Biotechnology, sc-371), anti-p53 (1:500; Santa Cruz Biotechnology, clone DO-1), anti-HIF-1α (1:1,000; BD Biosciences, 610958), anti-CDC25A (1:500; Santa Cruz Biotechnology, clone DCS-120), anti-NRF1 (1:1,000; Cell Signaling, clone D5B10), anti-VCP (1:2,000; Abcam, ab11433), anti-VCP (1:1,000; Novus Bio, NBP100-1557), anti-NPLOC4 (1:1,000; Novus Bio, NBP1-82166), anti-ubiquitin lys48-specific (1:1,000; Merck Millipore, clone Apu2), anti-β-actin (1:2,000; Santa Cruz Biotechnology, sc-1616; or 1:500, Santa Cruz Biotechnology, sc-87778), anti-GAPDH (1:1,000,GeneTex, clone 1D4), anti-lamin B (1:1,000; Santa Cruz Biotechnology, sc-6217), anti-calnexin (1:500; Santa Cruz Biotechnology, sc-11397), anti-α-tubulin (1:500; Santa Cruz Biotechnology, sc-5286), anti-XBP1 (1:500; Santa Cruz Biotechnology, sc-7160), UFD1 (1:500; Abcam, ab155003), cleaved PARP1 (1:500; Cell Signaling, 9544), p-eIF2α (1:500; Cell Signaling, 3597), ATF4 (1:500; Merck Millipore, ABE387), HSP90 (1.500; Enzo, ADI-SPA-810), HSP70 (1:500; Enzo, ADI-SPA-830), HSF1 (1:500; Cell Signaling, 4356), p-HSP27 (1:1,000; Abcam, 155987), HSP27 (1:1,000; Abcam, 109376) followed by detection by secondary antibodies: goat anti-mouse IgG–HRP (GE Healthcare), goat anti-rabbit (GE Healthcare), donkey anti-goat IgG–HRP (Santa Cruz Biotechnology, sc-2020). Bound secondary antibodies were visualized by ELC detection reagent (Thermo Fisher Scientific) and images were recorded by imaging system equipped with CCD camera (ChemiDoc, Bio-Rad) operated by Image Laboratory software or developed on film (Amersham).

Immunofluorescence staining. Cells were grown in 24-well plates with a 0.170-mm glass bottom (In Vitro Scientific). Where indicated, the cells were pre-extracted before fixation with pre-extraction buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 1.5 mM $MgCl_2$, 300 mM sucrose, 0.5% Triton X-100, 1 mM DTT, 5 $\mu g\ ml^{-1}$ leupeptin, 2 $\mu g\ ml^{-1}$ aprotinin, 0.1 mM PMSF) for 20 min at 4 °C, washed by PBS and then immediately fixed with 4% formaldehyde for 15 min at room temperature. Cells were stained with primary antibodies: anti-ubiquitylated conjugated mouse FK2 antibody (1:500; Enzo, BML-PW8810), anti-VCP (1:500; Abcam; ab11433), anti-NPL4 (1:500; Novus Bio, NBP1-82166), HSP70 (1:100; Enzo, ADI-SPA-830), HSF1 (1:500; Cell Signaling, 4356), anti-ubiquitin lys48-specific (1:500; Merck Millipore, clone Apu2), SUMO2/3 (1:500; Abcam, ab3742), TDP43 (1:300; Proteintech, 10782-2-AP) and appropriate Alexa Fluor 488- and 568-conjugated secondary antibodies (Invitrogen, 1:1,000). Cytochrome *c* was stained using an Alexa Fluor 555-conjugated mouse anti-cytochrome *c* antibody according the manufacturer's protocol (BD Pharmingen, 558700).

Microscopy, FRAP and image analysis. Samples were analysed using a Zeiss Axioimager Z.1 platform equipped with the Elyra PS.1 super-resolution module for structured illumination (SIM) and the LSM780 module for CLSM. High resolution images were acquired in super-resolution mode using a Zeiss Pln Apo100×/1.46 oil objective (total magnification, 1,600×) with appropriate oil (Immersol 518F). SR-SIM setup involved five rotations and five phases for each image layer and up to seven *z*-stacks (101 nm) were acquired per image. The CLSM setup for FRAP and life cells acquisition had a c-Apo 40×/1.2 W water immersion objective. Bleaching of regions of interest (ROIs) was performed using an Argon 488 nm laser. Lower resolution images of fixed samples were acquired using a Plan Apo 63×/1.4 oil objective (total magnification 1,008×). FRAP and image acquisitions were performed using Zeiss Zen 11 software. For FRAP, internal Zen's 'Bleach'

and 'Regions' modules were used. Data from FRAP analysis involving multiple bleached ROIs were exported into Microsoft Excel and plotted. Basic processing of acquired images, such as contrast and brightness settings, was done in Adobe Photoshop on images exported as TIFFs. Quantitative microscopy-based cytometry of the immunofluorescence-stained samples was performed using an automatic inverted fluorescence microscope BX71 (Olympus) using ScanR Acquisition software (Olympus) and analysed with ScanR Analysis software (Olympus).

Cell fractionation for Triton-X100 insoluble pellets. Cells were treated as indicated, washed in cold PBS and lysed in lysis buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 2 mM $MgCl_2$, 10% glycerol, 0.5% Triton X-100, protease inhibitor cocktail by Roche) for 10 min gently agitating at 4 °C. Then, cells were scraped into Eppendorf tubes and kept for another 10 min on ice with intermittent vortexing. After that, the lysate was centrifuged at 20,000*g* for 10 min at 4 °C. The insoluble fraction and supernatant were separately re-suspended in 1× LSB buffer.

Isolation of microsomal fraction. After the desired treatment in cell culture, cells were washed with cold PBS and lysed (250 mM sucrose, 20 mM HEPES pH 7.4, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, protease inhibitor cocktail). Lysates were homogenized by Potter-Elvejhem PTFE homogenizer and kept on ice for 20 min. The homogenates were subjected to serial centrifugation steps (720*g* and 10,000*g* for 5 min each, and 100,000*g* for 1 h). Pellets and supernatants from the last ultracentrifugation step were resuspended in the 1× LSB buffer and used for western blot analysis.

Immunoperoxidase staining of pre-extracted tissue sections. Frozen sections (4–5 μm thick) from xenograft-grown, cryopreserved tumour tissues were cut on a cryostat and placed on commercial adhesion slides (SuperFrost Plus, Menzel, Germany) and air-dried for 2 h at room temperature. The dried sections were carefully covered with the cold extraction buffer: 50 mM Tris-HCL (pH 7.5), 150 mM NaCl, 1 mM $MgCl_2$, 5% glycerol, 1 mM DTT, 1% Triton X-100, 1% IGEPAL, protease inhibitor cocktail (Phos Stop Easy pack, 04906837001, Roche) or cold PBS (controls) and incubated in a cold room for 20 min. Pre-extracted and control PBS-treated sections were gently washed three times in cold PBS, and fixed in 4% paraformaldehyde fixative for 15 min, followed by another three washes in PBS. Washed sections were then subject to a sensitive immunoperoxidase staining protocol, using the primary rabbit monoclonal antibody against VCP antibody (EPR3307(2)) (1:10,000; ab109240, Abcam) and rabbit polyclonal antibody against NPLOC4 (1:500; NBP1-82166, Novus Biologicals) and Vectastain Elite kit as secondary reagents (Vector Laboratories, USA), followed by a nickel-sulfate-enhanced diaminobenzidine reaction without nuclear counterstaining, mounted and microscopically evaluated and representative images documented by an experienced oncopathologist.

Isothermal titration calorimetry (ITC). Experiments were performed at 25 °C with a Nano ITC Low Volume (TA Instruments) and analysed by Nano Analyze Software v.2.3.6. During all measurements, injections of 2.5 μl of ligand (16 μM) were titrated into 250 μl protein (2 μM) with time intervals of 300 s, a stirring speed of 250 r.p.m. All ITC experiments were conducted with degassed buffered solutions 20 mM HEPES buffer pH 7.3, in the presence of 1% DMSO. Purified GST–NPL4(WT) and GST–NPL4(MUT) proteins were used in ITC experiment.

Drug affinity responsive target stability (DARTS). DARTS was performed according to a modified published protocol[38]. Purified GST–NPL4(WT) and GST–NPL4(MUT) proteins were diluted by 100 mM phosphate buffer, pH 7.4 to final concentration of 0.03 μg μl^{-1}. The proteins were treated with CuET (final concentration of 5 μM; dissolved in DMSO) for 1 h and equal amounts of DMSO were added to the solutions, which served as control samples. Pronase (Sigma-Aldrich) was dissolved in TNC buffer (50 mM Tris-Cl, 50 mM NaCl, 10 mM $CaCl_2$, pH 7.5). The 0.025 μg of pronase was added to 50 μl of protein solution and incubated for 1 h at 37 °C. Samples without pronase served as the non-digested controls. The pronase reaction was stopped by addition of 5× SDS loading buffer; the samples were boiled at 95 °C for 15 min and loaded on SDS–PAGE gels. After SDS–PAGE, gels were silver-stained and scanned on a GS-800 Calibrated Densitometer (Bio-Rad) or used for western blot analysis.

20S proteasome activity. To measure proteasome activity in cell extracts, cell lines were seeded in 100-mm Petri dishes at a density of 3×10^6 cells per dish. After 24 h, cells were washed twice with 2 ml of ice-cold PBS and scraped in to 1,000 μl ice-cold PBS. The cells were then isolated and suspended in buffer (50 mM HEPES (pH 7.5), 150 mM NaCl, 1% Triton X-100 and 0.1 μM PMSF) and then centrifuged at 15,000 r.p.m. for 15 min at 4 °C. The cell lysates (10 μg) were incubated with 20 μM of substrates for measurement of chymotrypsin-like, trypsin-like and caspase-like activities (Suc-LLVT-AMC, Ac-RLR-AMC and Z-LLE-AMC (Boston Biochem)) in 90 μl of assay buffer (30 mM Tris-HCl, 0.035% sodium dodecylsulfate (pH 7.4)) in the presence CuET (1 μM and 5 μM) and BTZ (1 μM) for the investigation of proteasome inhibition; BTZ or an equivalent volume of solvent (DMSO) was used as a control. After 2 h of incubation at 37 °C, inhibition of proteasome activity was measured by the release of hydrolysed free AMC groups by fluorimeter at 380/460 nm (TECAN, Infinite M200PRO). To measure proteasome activity in live cells, the cells were seeded in 24-well plate at a density of 0.2×10^6 cells per well. Cell lines were treated with CuET (1 μM and 5 μM), vehicle control or 1 μM BTZ for 1 h. After incubation, cells were twice washed with 0.5 ml of 1× ice-cold PBS and scraped into 100 μl ice-cold lysis buffer and then centrifuged at 15,000 r.p.m. for 15 min at 4 °C. Subsequently, the cell extract (10 μg) was incubated with 20 μM substrates to measure chymotrypsin-like, trypsin-like and caspase-like activities in assay buffer (30 mM Tris-HCl (pH 7.4)). After 2 h of incubation at 37 °C, inhibition of proteasome enzymatic activities was measured by the release of hydrolysed free AMC as described above.

Ub(G76V)–GFP degradation. HeLa Ub(G76V)–GFP-ODD-Luc cells expressing Ub(G76V)–GFP were seeded at a density of 10^4 cells per well in 96-well plates. The next day, cells were treated with 4 μM MG132 for 3 h. After that, the medium was discarded and cells were washed twice with PBS and then incubated with the tested compound in the presence of 30 μg ml^{-1} cycloheximide for another 3 h. The GFP signal was acquired using an ImageXpress automated microscope. For each well, four images were taken (corresponding to 200–250 cells). Cells were analysed every 30 min during 3 h of treatment. Normalized GFP signal intensity was calculated using the following formula: (test compound − background)/(basal GFP signal intensity × background) where 'test compound' is defined as the mean GFP signal intensity of Ub(G76V)–GFP-expressing cells treated with the test compound. 'Background' is defined as the background GFP signal intensity of HeLa cells. 'Basal GFP signal intensity' is defined as mean GFP signal intensity of Ub(G76V)–GFP-expressing cells treated with DMSO. The degradation rate constant (k) was obtained from the slope of the linear range of plotting ln(normalized GFP signal intensity) versus time ranging from 90 to 180 min. The percentage of remaining k for each compound is calculated using the following formula (test compound/DMSO control) × 100.

p97 ATPase activity assay. P97 ATPase assay was performed as described previously[28]. A total of 250 nM of p97 protein was diluted in assay buffer (50 mM Tris-HCl pH 7.4, 20 mM $MgCl_2$, 0.5 mM DTT). Test compounds were added in DMSO (final concentration of DMSO was 5%). After 10 min of incubation, the reaction was started with ATP (100 μM final concentration) followed by a 1-h incubation at room temperature. The reaction was stopped by adding Biomol green solution (Enzo) and free phosphate was measured according to the manufacturer's instructions. Results are expressed as the percentage of activity of the control (a well containing only DMSO).

26S proteasome activity. The RPN11 assay is described in PubChem (AID588493). In brief, a synthetic fluorescently labelled substrate, Ub4pepOG, was used to measure RPN11 activity. Fluorescence polarization assay was performed in a low-volume 384-well solid black plate in the presence of (i) 5 μl of the compound 1,10-phenanthroline or CuEt in 3% DMSO or 3% DMSO control; (ii) 5 μl of BioMol 26S proteasome; and (iii) 5 μl of substrate (15 nM Ub4pepOG). Fluorescence polarization is measured using a plate reader with excitation of 480 nm and emission of 520 nm filter set. The activity was normalized to DMSO control and fit to a dose–response curve.

Protein expression and purification. All proteins were expressed in *E. coli* BL21 (DE3) cells (Novagen). p97-His (pET28a vector) and Ufd1-His (pET28a vector) expression were induced by 1 mM IPTG (Life Technologies) at an OD_{600} of 0.6 for 10 h at 22 °C. NPL4(WT) and NPL4(MUT) (pGEX-2TK) were induced by 0.4 mM IPTG at an OD_{600} of 0.8 overnight at 16 °C. For p97 and UFD1, the bacterial pellet was suspended in buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 2.5 mM $MgCl_2$, 20 mM imidazole, 5% glycerol) and lysed by sonication and centrifuged (14,000*g* for 20 min). Proteins were purified by Ni-NTA chromatography (Qiagen) according to the manufacturer's instructions. For p97, the protein was further purified by gel filtration (Superdex 200, GE Healthcare). For GST–NPL4(WT) and GST–NPL4(MUT), the bacterial pellet was suspended in phosphate buffer (PBS, 0.1% Triton X-100, 300 mM NaCl) and lysed by sonication and centrifuged (14,000*g* for 10 min). Proteins were purified by glutathione sepharose 4B (Life Technologies) according the manufacturer's protocol. The proteins were further purified by gel filtration (Superdex 200, GE Healthcare).

Chemicals. CuET was prepared by direct synthesis from water solutions of diethyldithiocarbamate sodium salt and copper(II) chloride as described previously[53]. CuET for *in vivo* experiments was prepared equally with a slight modification. The reaction between diethyldithiocarbamate sodium salt and copper(II) chloride was performed in a sterile 1% aqueous solution of bovine serum albumin. The resulting solution was used directly. The following chemicals were purchased from commercial vendors: tetraethylthiuram disulfide (disulfiram, DSF) (Sigma-Aldrich), sodium diethyldithiocarbamate trihydrate (Sigma-Aldrich), copper D-gluconate (Sigma-Aldrich), BTZ (Velcade, Janssen-Cilag International N.V.), MG132 (Sigma-Aldrich), DBeQ (Sigma-Aldrich), NMS873 (Abmole), cycloheximide (Sigma-Aldrich), dicoumarol (Sigma-Aldrich), 1,10-phenanthroline (Sigma) and MLN7243 (Active Biochem).

Statistical analyses and reproducibility. For the epidemiological study, we calculated hazard ratios and 95% confidence intervals estimating cancer-specific mortality, based on a Cox model regressing of both propensity scores and disulfiram use, balancing baseline characteristics of previous and continuing users of DSF and adjusting estimated hazard ratios of cancer-specific mortality associated with DSF use[51]. The propensity score estimates were conditional on multiple covariates, based on using logistic regression (see 'Epidemiological analyses and access to health registers' for specifics of cohorts and covariates). In the Cox model, the propensity score is further included as a restricted cubic spline to model possible nonlinearities, in addition to the categorical disulfiram use as the variable of interest. Statistical significance of DSF use was evaluated by likelihood ratio tests, using the software R for statistical computing[52].

For evaluation of the animal studies, STATISTICA software, v.12 (StatSoft) was used to estimate sample size. For a power of 80%, the level of significance set at 5%, 4 groups and RMSSE = 0.8, seven mice per group were estimated. For usage of non-parametrical statistical methods, the number of eight mice per group was finally planned. The differences between tumour volumes were statistically analysed by non-parametrical Kruskal–Wallis test, not requiring any assumptions of normality and homoscedascity. To test the effect of CuET treatment on survival of AMO-1-xenografted mice, a Kaplan–Meier graph and log-rank statistical test were used. For other experiments, the statistics, such as number of repetitions, centre value and error bars, are specified in figure legends.

Data availability. Most data generated or analysed during this study are included in the article and its Supplementary Information. Uncropped images of all gels and blots can be found in Supplementary Fig. 1. Source Data for all graphs are provided in the online version of the paper. Additional datasets generated during and/or analysed during the current study and relevant information are available from the corresponding authors upon reasonable request.

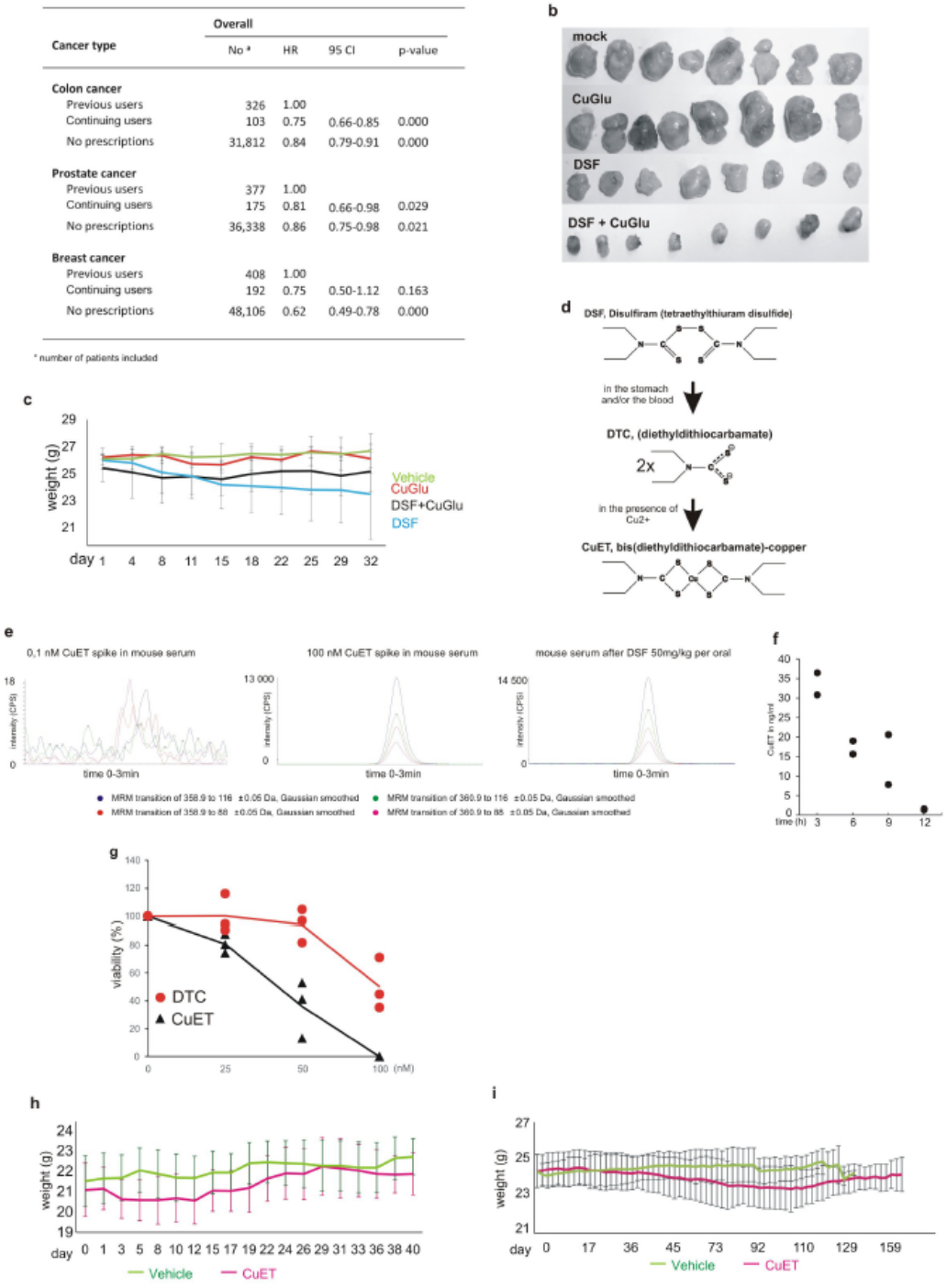
50. Thygesen, L. C., Daasnes, C., Thaulow, I. & Brønnum-Hansen, H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand. J. Public Health* **39** (Suppl), 12–16 (2011).
51. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55 (1983).
52. R Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing* https://www.R-project.org/ R v.3.2.3 (2015-12-10) (R Foundation for Statistical Computing, 2016).
53. Cvek, B., Milacic, V., Taraba, J. & Dou, Q. P. Ni(II), Cu(II), and Zn(II) diethyldithiocarbamate complexes show various activities against the proteasome in breast cancer cells. *J. Med. Chem.* **51,** 6256–6258 (2008).

Extended data.

a

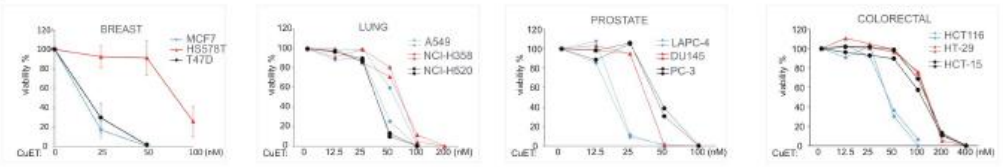
Cancer type	Overall No *	HR	95 CI	p-value
Colon cancer				
Previous users	326	1.00		
Continuing users	103	0.75	0.66-0.85	0.000
No prescriptions	31,812	0.84	0.79-0.91	0.000
Prostate cancer				
Previous users	377	1.00		
Continuing users	175	0.81	0.66-0.98	0.029
No prescriptions	36,338	0.86	0.75-0.98	0.021
Breast cancer				
Previous users	408	1.00		
Continuing users	192	0.75	0.50-1.12	0.163
No prescriptions	48,106	0.62	0.49-0.78	0.000

* number of patients included



Extended Data Figure 1 | Epidemiological and pre-clinical data of the anti-cancer effects of DSF. a, Summary of hazard ratios (HR) and 95% confidence intervals (CI) for cancer-specific mortality among Danish patients with cancer, comparing continuing and previous users of DSF for selected types of cancer (for statistical analysis and definitions of DSF exposure categories, see Methods). **b**, Photographs of subcutaneously growing human MDA-MB-231 tumours extracted from mice at day 32. **c**, Time-course diagram of mouse weight. $n = 8$ animals per group. **d**, Model of CuET formation during metabolic processing of orally administered DSF in the human body. **e**, Examples of mass-spectrometry spectra of CuET expressed as peaks of 4 MRM transitions in mouse serum after CuET spikes, compared to orally applied DSF (50 mg kg^{-1}). Data are representative of two independent experiments. **f**, Pharmacokinetic analysis of CuET levels in mouse serum after orally applied DSF (50 mg kg^{-1}). $n = 2$ animals per time point. **g**, Effect of DTC and CuET on MDA-MB-231 cells analysed by colony formation assay. $n = 3$ independent experiments. **h**, Time-course diagram of weight in CuET- and vehicle-treated mice. $n = 10$ animals per group. **i**, Extended time-course diagram of weight in CuET- and vehicle-treated mice. $n = 10$ animals per group. Data are mean ± s.d. (**c**, **h**, **i**) or linked means (**g**).

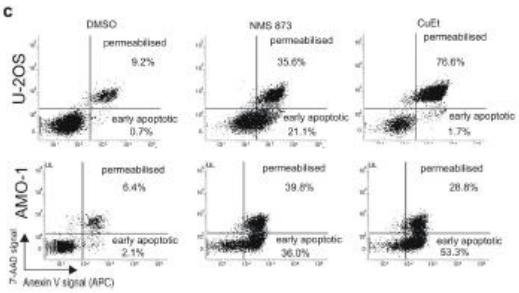
a



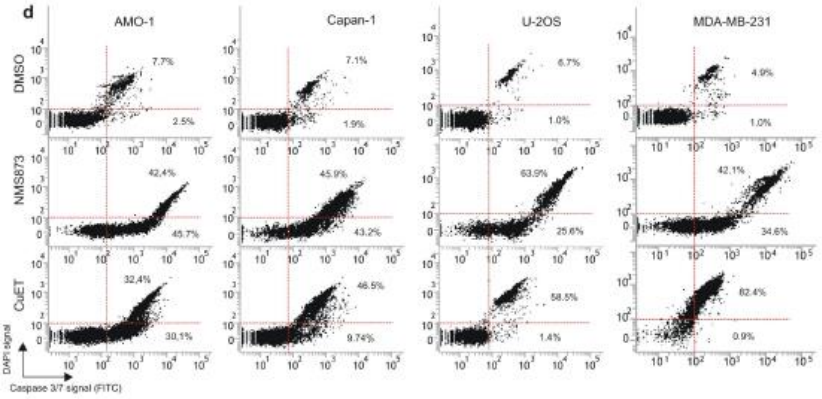
b

Cell line	Type	IC50 (nM)	
AMO-1	Plasmacytoma	78,44	73,39
CAPAN-1	Pancreatic adenocarcinoma	83,51	55,68
MM1S	Multiple myeloma	65,22	73,73
CCRF-CEM	Acute lymphoblastic leukemia	78,6	70,81
OVCAR3	Ovarian adenocarcinoma	79,7	85,43
A549	Lung carcinoma	99,92	90,8
HCT116	Colorectal carcinoma	103,6	146,5
A253	Salivary gland carcinoma	143,8	284,4
786-O	Renal cell adenocarcinoma	304,6	165
U-2OS	Osteosarcoma	286,8	244,9
FaDU	Squamous cell carcinoma	287,6	371,7
hTERT-RPE-1	Normal retinal epithelial cells	538,3	237,1
K562	Chronic myelogenous leukemia	339,8	322,8
MDA-MB-231	Breast adenocarcinoma	508,9	417,7
U87-MG	Glioblastoma	760,7	631,6
SiHA	Squamous cell carcinoma	697,9	856,2
RWPE-1	Normal prostate epithelial cells	4673	1951

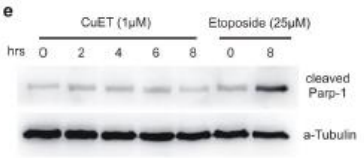
c



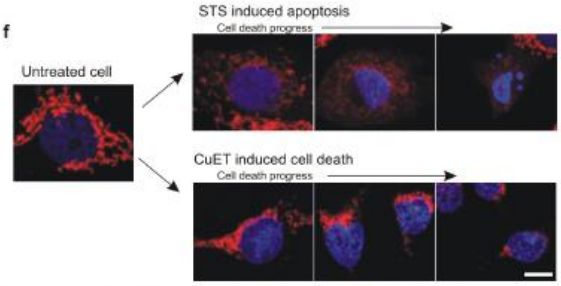
d



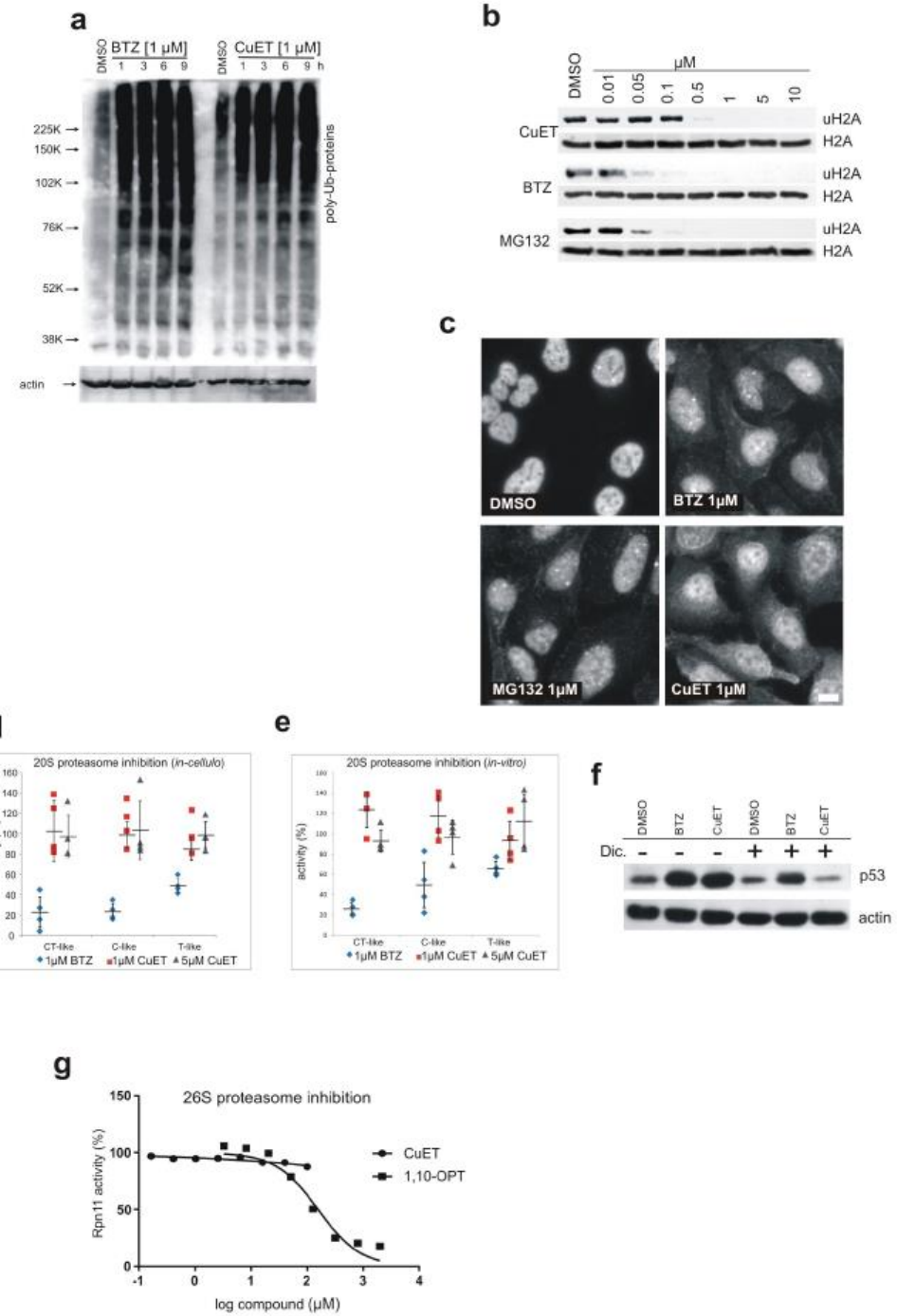
e



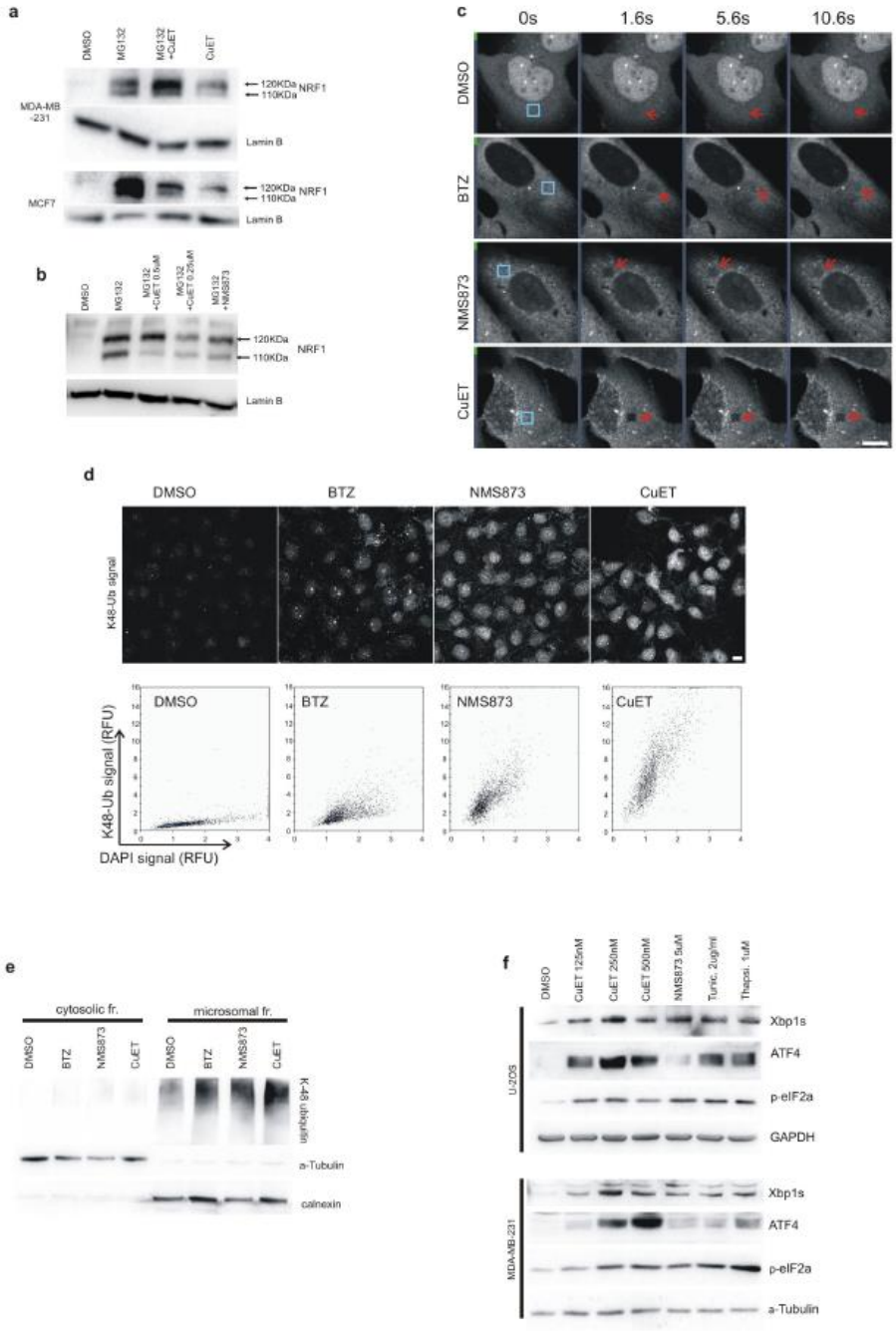
f



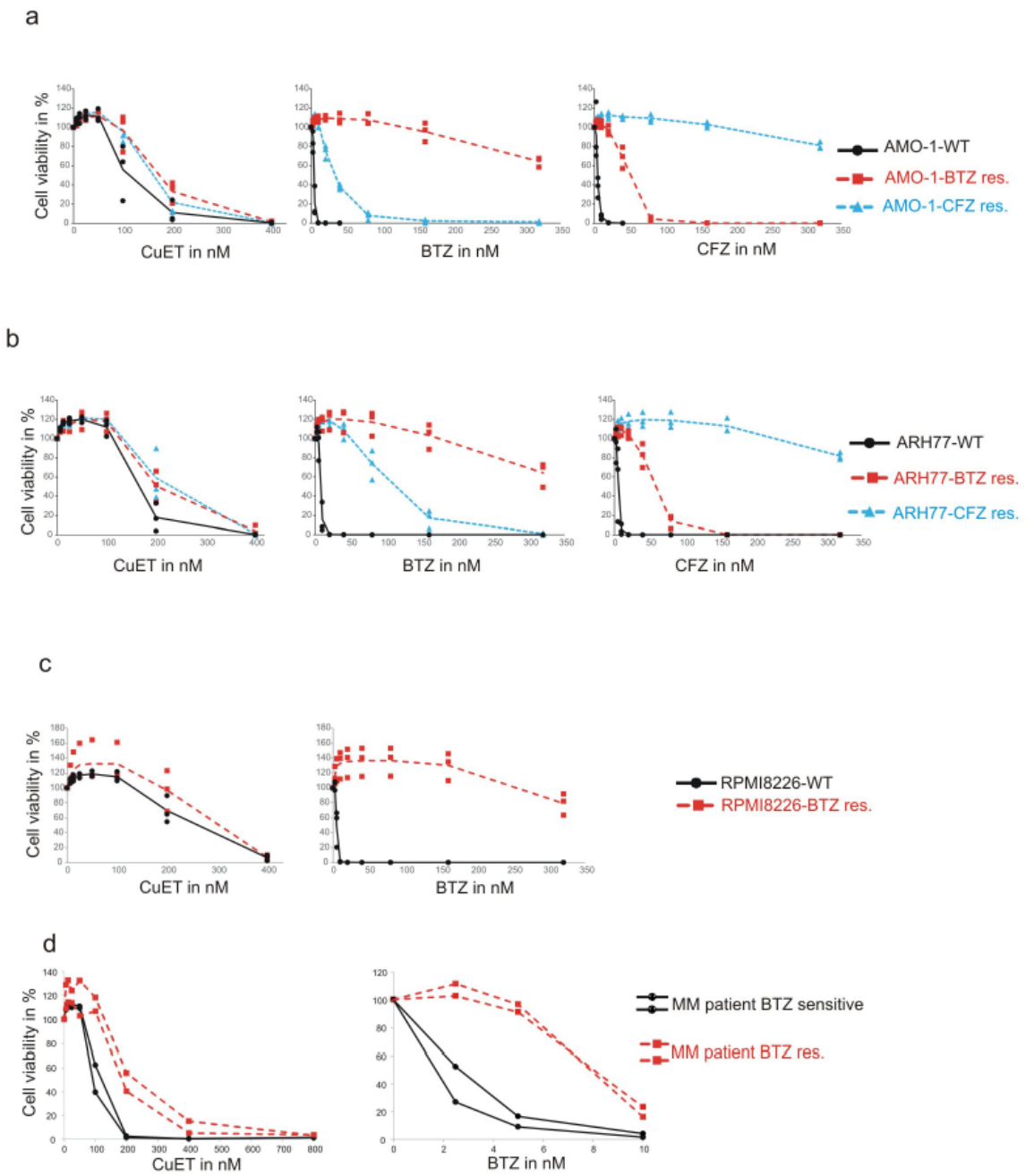
Extended Data Figure 2 | CuET is the major anti-cancer metabolite of DSF. a, CuET cytotoxicity measured by a colony-formation assay in human cell lines derived from breast, lung, colon and prostate carcinomas. Data are mean ± s.d. of three independent experiments (breast) or presented individually for two independent biological experiments for each cell line (lung, colon and prostate). **b**, IC_{50} values from two independent biological experiments documenting differential CuET-induced cytotoxicity across a panel of cancer and non-cancerous cell lines (48 h treatment). **c**, Analysis of annexin V signal in AMO-1 cells exposed to toxic doses of NMS873 (5 μM, 16 h) or CuET (100 nM, 16 h) and in U2OS cell exposed to toxic doses of NMS873 (10 μM, 16 h) or CuET (1 μM, 16 h). **d**, Analysis of caspase 3/7 activity in selected cell lines after apoptosis induction by NMS873 (AMO-1: 6 h, 5 μM; Capan1: 16 h, 10 μM; U2OS: 16 h, 10 μM; MDA-MB-231: 24 h, 10 μM) or CuET (AMO-1: 16 h, 100 nM; Capan1: 16 h, 250 nM; U2OS: 16 h, 1 μM; MDA-MB-231: 24 h, 1 μM). **e**, Absence of cleaved PARP1 after a toxic dose of CuET in U2OS cells, compared to etoposide treatment as a positive control. **f**, Analysis of cytochrome *c* (in red) release from mitochondria in U2OS cells during cell death induced by the positive control staurosporin (STS, 1 μM) compared to cell death induced by CuET (1 μM). Blue, DAPI. Scale bar, 10μm. **c–f**, Data are representative of two independent biological experiments.



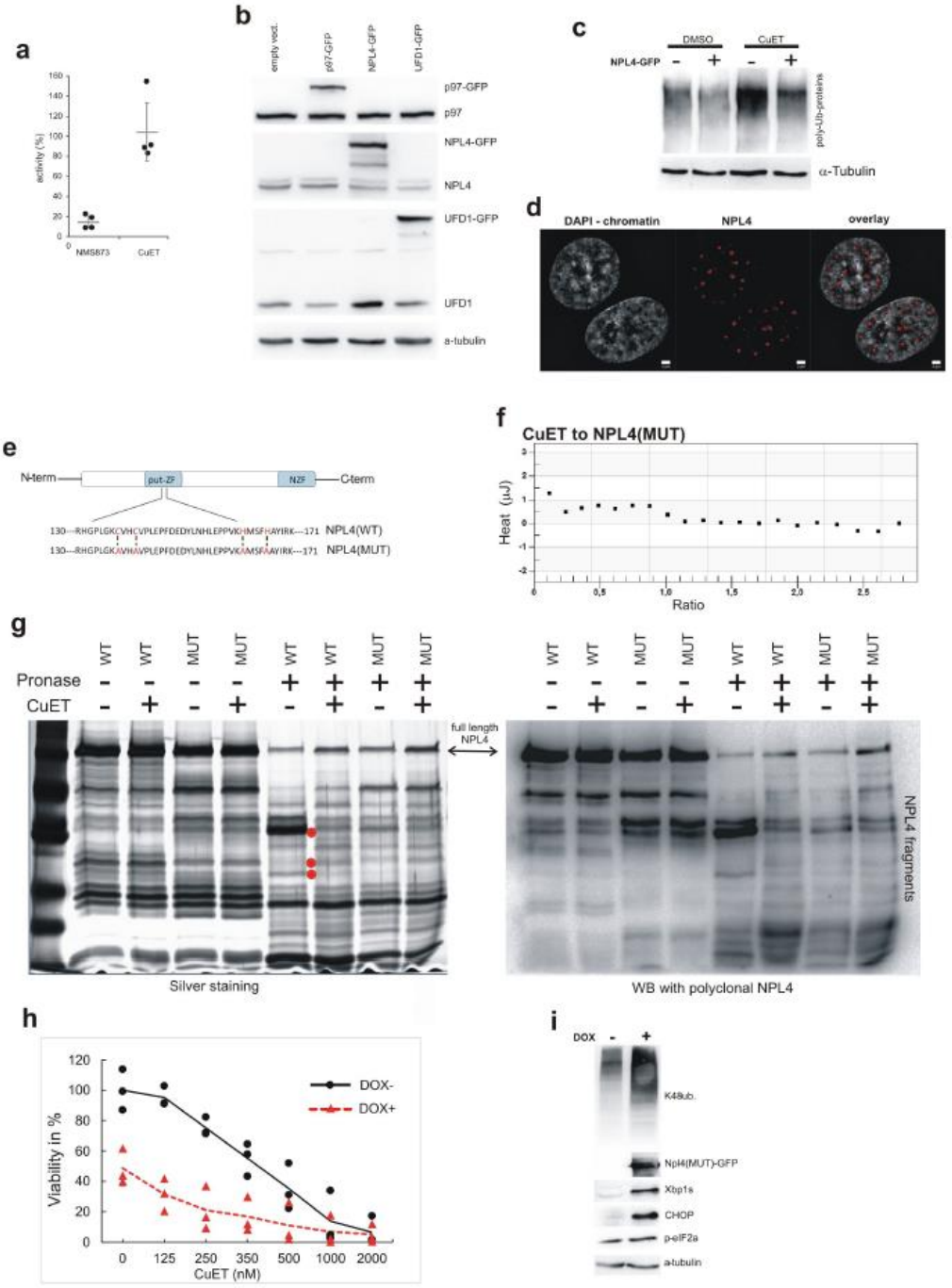
Extended Data Figure 3 | CuET-induced proteasome inhibition-like response is not due to proteasome inhibition. a, Kinetics of poly-Ub protein accumulation in U2OS cells treated with CuET or the proteasome inhibitor BTZ. **b**, CuET treatment (1.5 h) induces rapid deubiquitylation of ubiquitylated histone H2A (uH2A) similarly to proteasome inhibitors BTZ or MG132 in U2OS cells. **c**, CuET treatment (1.5 h) induces rapid cytoplasmic accumulation of poly-ubiquitylated proteins (FK2 antibody staining) in U2OS cells, similar to BTZ and MG132 treatment. Scale bar, 10μm. **d**, **e**, 20S proteasome activity is not inhibited by CuET as examined in live MDA-MB-231 cells (**d**) or in lysates from MDA-MB-231 cells (**e**). Data are mean ± s.d. of four independent experiments. **f**, CuET treatment (1 μM, 6 h) does not cause accumulation of p53 in the presence of dicoumarol (300 μM) in MCF7 cells. **g**, *In vitro* 26S proteasome function measured as RPN11 deubiquitylation activity, is not inhibited by CuET; 1,10-phenanthroline (1,10-OPT) served as a positive control. Data are representative of two (**a**–**c**, **f**) or three (**g**) independent experiments.



Extended Data Figure 4 | CuET inhibits the p97 pathway and induces cellular UPR. a, MG132-treated cells (5 μM, 6 h) accumulate both forms of NRF1 (120-kDa and 110-kDa bands, top and bottom arrows, respectively), whereas CuET-treated cells (1 μM, 6 h) accumulate only the non-cleaved 120-kDa form. **b**, Inhibition of the NRF1 cleavage process (appearance of the lower band) by CuET and NMS873 (a p97 inhibitor; 5 μM) in mouse NIH3T3 cells co-treated with the proteasome inhibitor MG132 (5 μM for 6 h). **c**, Time-course example images from a FRAP experiment, for which the quantitative analysis is shown in Fig. 2g (U2OS cells, blue boxes mark areas before bleaching, arrows after bleaching). **d**, U2OS cells pre-extracted with Triton X-100 and stained for poly-Ub(K48). The antibody signal intensities for cells treated with DMSO, BTZ (1 μM), NMS873 (10 μM) and CuET (1 μM) are analysed by microscopy-based cytometry and plotted below. **e**, Western blot analysis of accumulated poly-Ub proteins in the ultracentrifugation-separated microsomal fraction from U2OS cells treated with mock, CuET (1 μM), NMS873 (10 μM) or BTZ (1 μM) for 3 h. **f**, UPR in U2OS and MDA-MB-231 cell lines induced by 6-h treatment with CuET (various concentrations) or positive controls (5 μM NMS873, 2 μg ml^{-1} tunicamycin, 1 μM thapsigargin) is shown by increased levels of XBP1s, ATF4 and p-eIF2α. **a**–**f**, Data are representative of two independent experiments. All scale bars, 10μm.

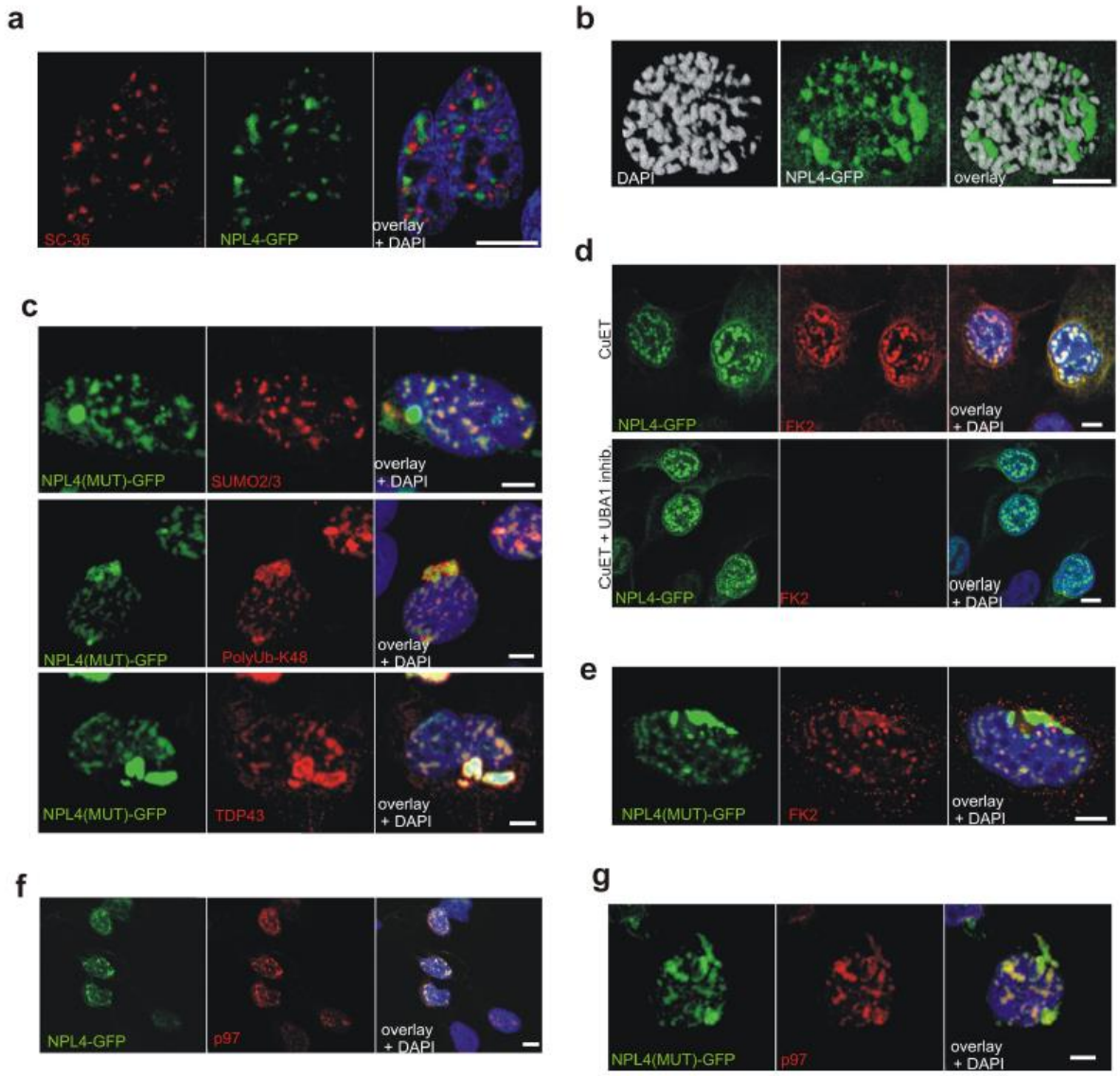


Extended Data Figure 5 | CuET kills BTZ-resistant cells. a, BTZ-adapted (BTZres), CFZ-adapted (CFZres) and non-adapted AMO-1 human myeloma cells are equally sensitive to treatment with CuET. **b**, BTZ-adapted, CFZ-adapted and non-adapted ARH77 human plasmocytoma cells are equally sensitive to treatment with CuET. **c**, BTZ-adapted and non-adapted RPMI8226 human myeloma cells are equally sensitive to treatment with CuET. **d**, Human myeloma cells derived from a patient with BTZ-resistant myeloma show CuET sensitivity comparable to myeloma cells derived from a patient with BTZ-sensitive myeloma. Data are means linked of three independent experiments (**a**–**c**) or data are from two independent experiments (**d**).

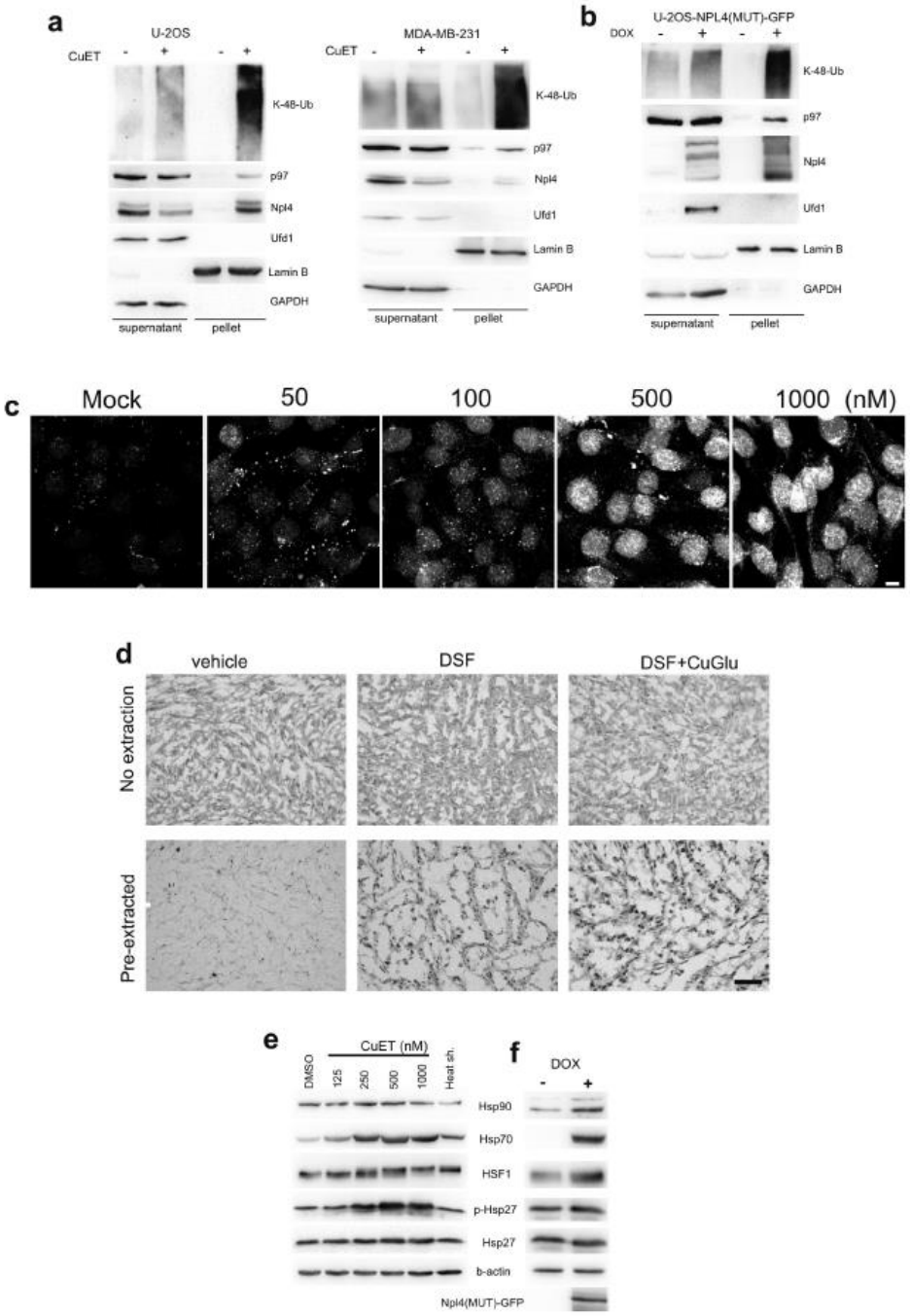


Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | CuET targets NPL4, causing immobilization and nuclear clustering of NPL4. a, CuET (1 μM) does not inhibit ATPase activity of p97. NMS873 (5 μM) was used as a positive control. Data are mean ± s.d. from four independent experiments. **b**, Western blotting analysis showing levels of ectopic p97–GFP, NPL4–GFP and UFD1–GFP in stable U2OS-derived cell lines used for the CuET-treatment rescue and cluster formation experiments. **c**, Ectopic expression of NPL4–GFP alleviates CuET-induced (125 nM, 4 h) accumulation of poly-Ub proteins in U2OS cells. **d**, Distribution of NPL4 nuclear clusters relative to chromatin in cells treated with CuET (1 μM, 2 h). Scale bars, 2 μm. **e**, Schematic representation of site-directed mutagenesis within the amino acid sequence of the putative zinc finger domain of NPL4. **f**, ITC curve showing the lack of CuET binding to purified NPL4(MUT) protein. **g**, DARTS analysis of recombinant NPL4 proteins shows that differential pronase-mediated proteolysis after CuET addition is apparent for NPL4(WT) but not for NPL4(MUT); detected by either silver-stained SDS–PAGE (the most prominent differential bands are marked by red dots) or by blotting with an anti-NPL4 polyclonal antibody. **h**, Viability of cells expressing doxycycline-inducible NPL4(MUT)–GFP, treated with CuET for 48 h. Data are from three independent experiments, means are linked. **i**, Accumulation of K48-ubiquitinated proteins and activation of UPR in cells expressing the doxycycline-inducible NPL4(MUT)–GFP. **b**–**d**, **f**, **g**, **i**, Data are representative of two independent experiments.



Extended Data Figure 7 | Immobilized NPL4 forms insoluble protein aggregates. a, NPL4–GFP aggregates induced by CuET treatment (1 μM, 3 h) do not co-localize with nuclear speckles (stained by SC-35 antibody) or nucleoli (visible as a $DAPI^-$ nuclear signal). **b**, NPL4–GFP nuclear aggregates induced by CuET (1 μM, 3 h) are excluded from chromatin in early prometaphase U2OS cells. **c**, Co-localization of spontaneous NPL4(MUT)–GFP aggregates with SUMO2/3, poly-UB(K48) and TDP43 in pre-extracted U2OS cells. **d**, NPL4–GFP aggregates are formed independently of ubiquitylation, as shown in CuET-treated (1 μM, 3 h) cells pre-treated with a chemical UBA1 inhibitor (MLN7243, 10 μM, 1 h). The lack of cellular FK2 staining of ubiquitylated proteins validates the efficacy of the MLN7243 inhibitor. **e**, Co-localization of FK2 signal with the spontaneous NPL4(MUT)–GFP aggregates in pre-extracted U2OS cells. **f**, Analysis of p97 in CuET-induced (1 μM, 3 h) NPL4–GFP aggregates in pre-extracted U2OS cells. **g**, Analysis of p97 in spontaneous NPL4(MUT)–GFP aggregates in pre-extracted U2OS cells. **a**–**g**, Data are representative of two independent biological experiments. All scale bars, 10μm.



Extended Data Figure 8 | NPL4 aggregation immobilizes the p97 binding partner and induces a global cellular HSR. a, Immobilization of selected proteins in Triton X-100-resistant pellet fractions of CuET-treated (1 μM, 3 h) U2OS cells. **b**, Immobilization of selected proteins in Triton X-100-resistant pellet fractions from U2OS cells expressing doxycycline-inducible NPL4(MUT)–GFP (48 h after induction). **c**, CuET dose-dependent immobilization of p97 in Triton X-100 pre-extracted MDA-MB-231 cells (3 h). Scale bar, 10μm. **d**, Immunohistochemical staining showing non-extractable p97 in MDA-MB-231 xenografts from mice treated with DSF or DSF and CuGlu, compared to vehicle. Scale bar, 50μm. **e**, HSR after CuET (8 h treatment) is shown by various HSR markers detected by western blotting of U2OS cell extracts. **f**, HSR markers in U2OS cells expressing doxycycline-inducible NPL4(MUT)–GFP (24 h after induction). **a–f**, Data are representative of two independent biological experiments.

8.3. Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells.

OŽDIAN T, HOLUB D, MACEČKOVÁ Z, VARANASI L, RYLOVÁ G, ŘEHULKA J, **VÁCLAVKOVÁ J**, SLAVÍK H, MOUDRÝ P, ZNOJEK P, STANKOVÁ J, DE SANCTIS J, HAJDÚCH M, DŽUBÁK P. Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells, Journal of Proteomics, 2017, 162, 73-85, 1874-3919, IF: 3.867, PMID: 28478306.

Journal of Proteomics 162 (2017) 73–85

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Proteomic profiling reveals DNA damage, nucleolar and ribosomal stress are the main responses to oxaliplatin treatment in cancer cells

Tomas Ozdian [a], Dusan Holub [a], Zuzana Maceckova [a], Lakshman Varanasi [a], Gabriela Rylova [a], Jiri Rehulka [a], Jana Vaclavkova [a], Hanus Slavik [a], Pavel Moudry [b], Pawel Znojek [a], Jarmila Stankova [a], Juan Bautista de Sanctis [c], Marian Hajduch [a], Petr Dzubak [a,*]

[a] Laboratory of Experimental Medicine, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University and University Hospital in Olomouc, Olomouc, Czech Republic
[b] Laboratory of Genome Integrity, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic
[c] Institute of Immunology, National Center of Clinical Immunology and FOCIS Center of Excellence, Faculty of Medicine, Universidad Central de Venezuela, Caracas, Venezuela

ARTICLE INFO

Article history:
Received 18 September 2016
Received in revised form 26 April 2017
Accepted 2 May 2017
Available online 3 May 2017

Keywords:
Oxaliplatin
Proteomics
LC-MS
Nucleolar stress
Ribosomal stress

ABSTRACT

Oxaliplatin is widely used to treat colorectal cancer in both palliative and adjuvant settings. It is also being tested for use in treating hematological, esophageal, biliary tract, pancreatic, gastric, and hepatocellular cancers. Despite its routine clinical use, little is known about the responses it induces in cancer cells. Therefore the whole-cell proteomics study was conducted to characterize the cellular response induced by oxaliplatin. Chemosensitive CCRF-CEM cells were treated with oxaliplatin at 29.3 μM ($5 \times IC_{50}$) for 240 min (half-time to caspase activation). The proteomes of un−/treated cells were then compared by high-resolution mass spectrometry, revealing 4049 proteins expressed over 3 biological replicates. Among these proteins, 76 were significantly downregulated and 31 significantly upregulated in at least two replicates. In agreement with the DNA-damaging effects of platinum drugs, proteins involved in DNA damage responses were present in both the upregulated and downregulated groups. The downregulated proteins were divided into three subgroups; i) centrosomal proteins, ii) RNA processing and iii) ribosomal proteins, which indicates nucleolar and ribosomal stress. In conclusion, our data supported by further validation experiments indicate the initial cellular response to oxaliplatin is the activation of DNA damage response, which in turn or in parallel triggers nucleolar and ribosomal stress.

Biological significance: We have performed a whole-cell proteomic study of cellular response to oxaliplatin treatment, which is the drug predominantly used in the treatment of colorectal cancer. Compared to its predecessors, cisplatin and carboplatin, there is only a small fraction of studies dedicated to oxaliplatin. From those studies, most of them are focused on modification of treatment regimens or study of oxaliplatin in new cancer diagnoses. Cellular response hasn't been studied deeply and to our best knowledge, this is the first whole-cell proteomics study focused exclusively to this important topic, which can help to understand molecular mechanisms of action.

1. Introduction

Oxaliplatin ([(1R,2R)-cyclohexane-1,2-diamine](ethanedioato-O,O′)platinum(II); L-OHP) is a member of the platinum drug family. Members of this family are derived from cisplatin, which can be regarded as the first generation platinum drug. Cisplatin has some serious side effects that limit its clinical utility, particularly nephrotoxicity [1]. The second generation platinum agents (of which carboplatin is the most well-known) were introduced to avoid these drawbacks. While these drugs have fewer side-effects than cisplatin, they suffer from cross-resistance with the first generation agent [2]. To overcome this issue, the newer and safer platinum drug L-OHP was synthesized in 1980 and approved by the FDA in 2005 [3]. L-OHP differs from cisplatin in that the amine groups of the latter are replaced with a diaminocyclohexane moiety. In plasma, L-OHP rapidly undergoes non-enzymatic transformations to form various reactive species. Most of these are pharmacologically inactive, but some are dichloro platinum complexes that can enter cells and exert cytotoxic effects.

L-OHP is widely used to treat colorectal, but also gastric and pancreatic metastatic cancer [4,5], and is being evaluated for the treatment of hematological, ovarian, breast, biliary tract, hepatocellular and non-small cell lung cancers [6,7]. It is generally applied in combination with other therapeutic agents. For example, in the FOLFOX regimen it is combined with 5-fluorouracil and folinic acid; when applied as a first-line therapy, this combination yields response rates of up to 50% in colorectal cancer, compared to 15% for 5-fluorouracil

* Corresponding author at: Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc, Hnevotinska 1333/5, Olomouc 775 15, Czech Republic.
E-mail address: petr.dzubak@upol.cz (P. Dzubak).

http://dx.doi.org/10.1016/j.jprot.2017.05.005
1874-3919/

alone [8]. Similarly, in the GEMOX regimen it is applied in combination with gemcitabine to treat pancreatic cancer and non-Hodgkin lymphomas [7,8]. It can also be combined with a range of other agents, including targeted biologics such as bevacizumab [9] or rituximab [7]. The use of targeted therapy and monoclonal antibody drugs with conventional drugs like L-OHP should be based on rational and scientific combinations. It is therefore valuable to study and understand molecular mechanisms of conventional chemotherapeutic agents which can help to design new synergistic combinations for clinical trials and avoid rapid development of drug resistant disease.

The most serious side-effect of L-OHP treatment is peripheral neuropathy, and most of the research on its use has focused on this side-effect [10] and the development of existing and novel treatment combinations. Only a few studies have investigated L-OHP resistance [11] or DNA damage induced by L-OHP treatment [12] and the potential resulting ionic [13] and/or oxidative stress [14]. L-OHP binds to nucleophiles, DNA, RNA and proteins, and forms intra-strand adducts between guanine and/or adenine residues, thus disrupting replication and transcription. Nucleotide excision repair pathway is believed to be main DNA damage repair system induced by L-OHP treatment. However, it has also been reported to induce epigenetic changes, detoxification pathways, cell death, and changes in NF-kB signaling [15].

Cellular response to L-OHP can be elucidated by very powerful tools, such as systematic proteomic analysis. Previously, we have conducted complex proteomic studies based on 2D electrophoresis profiling to characterize the responses of cancer cells to anthracyclines (doxorubicin, daunorubicin and mitoxantrone), revealing that exposure to these drugs changes the expression and activity of proteins involved in metabolic processes, transport proteins, and immune system proteins, among others [16]. A similar workflow based on 2D electrophoresis has also been used by us to elucidate the mechanisms responsible for resistance to the kinase inhibitor bohemine [17]. This proteomic approach is thus a valuable tool for elucidation of cellular response to single drugs. Proteomic profilling of single drugs cellular response can be valuable tool in comparison of drug effects in vitro.

In recent years there have been several important advances in the methods and technologies available for use in proteomic analysis; as a result, we have replaced our electrophoresis-based approach with a more straightforward alternative based on high resolution mass spectrometry (Orbitrap) that offers increased sensitivity and a greater ability to detect and quantify changes in the expression of less abundant proteins. Here, we describe this new approach and its application in an analysis of the cellular effects of L-OHP treatment. Our aim in conducting this work was both to obtain new knowledge concerning the response to the drug and (perhaps more importantly) to establish a generally applicable set of conditions for mechanism of action testing and -omics profiling of cytotoxic drugs. Our method involves treating a cancer cell line (in this work, the highly chemosensitive CCRF-CEM line was used as a model of L-OHP use in hematological cancers) with the agent of interest at 5 times its IC_{50} and for half the time required for caspase activation. The time to caspase activation is a measure of the time elapsed from exposure of cells to cytotoxic drug to induction of cell death; the aim of stopping the experiment at half the time to caspase activation is to minimize the risk of identifying late and non-specific changes in protein expression and processing induced by apoptotic machinery and the proteome more generally.

2. Material and methods

2.1. Chemicals and reagents

Oxaliplatin (Eloxatin) was obtained as a 5 mg/ml stock solution from Sanofi. If not mentioned otherwise, all chemicals were obtained from Sigma. Fetal calf serum (FCS) was purchased from Pan Biotech (Germany). The MagicRed Caspase detection kit for caspases 3 and 7 was purchased from Immunochemistry Technologies, USA. RPMI media lacking arginine and lysine was obtained from Biowest, USA. HPLC/MS grade water was produced using a Merck Millipore Milli-Q Direct-8 purifier with an LC-PAK polisher (Merck), and HPLC/MS grade acetonitrile was purchased from J. T. Baker, USA. Proteomic grade trypsin was purchased from Promega. Benzonase and Luminata Forte were obtained from Merck. Pierce 660 nm Protein Assay kits for protein concentration measurements with Ionic Detergent Compatibility Reagent were purchased from ThermoFisher Scientific, USA. Pertex rapid drying medium was used for slide mounting (Histolab, Sweden). The following antibodies were used: Anti-UTP11 (AbCam), Anti-WDR46 (Sigma), Anti-TRF1 (Santa Cruz), Anti-RPS19 (Santa Cruz), Anti-p53 (Cell Signaling), Anti-Fibrilarin (Sigma) and Anti-β-actin (Sigma), horseradish peroxidase conjugated anti-rabbit (Sigma), horseradish peroxidase conjugated anti-mouse (Sigma), Alexa-488 anti-goat (Life technologies) and Alexa 488 anti-rabbit (Invitrogen). ELISA kits were purchased from Thermo Fisher (Ferritin) and RnD Systems (Apolipoprotein A1).

2.2. Cell culture, analysis of IC50 and time to apoptosis

The T-lymphoblastic leukemia-derived cell line CCRF-CEM was purchased from the American Tissue Culture Collection (ATCC) and cultivated according to the recommended procedure in RPMI media supplemented with 20% FCS, 100 IU/ml penicillin and 10 μg/ml streptomycin. Inhibition of tumor growth/survival, measured in terms of the 50% inhibition constant (IC_{50}), was determined by the MTT assay as described previously [18]. Time to apoptosis was determined using the Magic Red assay (Immunochemistry Technologies). For proteomic analyses, CCRF-CEM cells (10^6/ml) were treated with a 5 × IC 50 concentration of L-OHP, which approximately corresponds to the IC 50 value when compared to cell count used in cytotoxic MTT assay [18]. Assay was performed according to the manufacturer's instructions, with fluorescence analysis based on excitation at 590 nm and emission at 630 nm.

2.3. SILAC labeling and treatment with drugs

Three biological replicates of the proteomic analysis were performed. CCRF-CEM cells were grown in complete RPMI media lacking arginine and lysine or media containing either $^{12}C_6$ (light) or $^{13}C_6$ (heavy) labeled arginine and lysine. In all cases, the cells were cultured for five passages to ensure that the "light" or "heavy" amino acids were completely incorporated. The completeness of the labeling was verified by mass spectrometry (data not shown). Light cell lines were diluted to a concentration of 10^6 cells per ml and treated with a 5 × IC_{50} concentration of L-OHP for half the time required to induce apoptosis ($TA_{1/2}$). After treatment, light and heavy cells were mixed in a 1:1 ratio, and the mixed cell suspension was washed twice in ice-cold PBS with protease inhibitors and once with ice-cold PBS. The washed cells were lysed in lysis buffer containing 20 mM Tris-HCl, 7 M Urea, 10 mM DTT, 1% Triton X-100 and 0.5% SDS, and 2 U benzonase, and transferred to 1.5 ml microfuge tubes. The lysate was then incubated on ice for 5 min, after which two extra units of benzonase were added and the resulting mixture was incubated for an additional 5 min. Lysates were cleared at 16,000 g at 4 °C for 10 min and the supernatants were transferred into 1.5 ml Eppendorf tubes and stored at − 80 °C.

2.4. Sample preparation for LC-MS/MS

The 100 μl of whole cell lysates was resolved by preparative SDS-PAGE (MiniPrep Cell, BioRad), and the resulting gels were cut into 20 slices corresponding to protein fractions separated by molecular weight. The gel slices were then dehydrated with acetonitrile, after which the proteins were reduced with 50 mM Tris(2-carboxyethyl)phosphine at

90 °C for 10 min and then alkylated with 50 mM iodoacetamide for 1 h in the dark. The alkylated samples were washed three times with water and acetonitrile successively, and finally washed with 50% acetonitrile. After washing, the samples were solubilized and trypsinized in trypsin buffer (6.25 ng/μl trypsin, 50 mM 4-ethylmorpholine, 10% v/v acetonitrile, pH 8.3) overnight at 37 °C. The resulting supernatant was transferred into a new Eppendorf tube and peptides were extracted from the gel pieces, first with 80% acetonitrile containing 0.1% TFA, then with 0.1% TFA in water, and finally with 50% acetonitrile. The extracts were then pooled and dried in a SpeedVac (Eppendorf). Dried samples were reconstituted in 5 μl of 80% ACN with 0.1% TFA, diluted with 145 μl of 0.1% TFA, and purified using a C-18 MacroTrap column (Michrom Bioresources, USA). The purified samples were then re-dried in the SpeedVac and resuspended in 20 μl of 5% acetonitrile with 0.1% formic acid for LC-MS/MS analysis.

2.5. LC/MS

Mass spectrometric analysis was performed on an Orbitrap Elite (Thermo) instrument fitted with a Proxeon Easy-Spray ionization source, coupled to an Ultimate 3000 RSLCnano chromatograph. One microliter of sample was loaded on a PepMap 100 (75 μm × 2 cm, 3 μm, 100 Å pore size) desalting column (Thermo) "in-line" with a PepMap RSLC (75 μm × 15 cm, 3 μm, 100 Å pore size) analytical column (Thermo) heated to 35 °C. The peptides were subsequently separated on the analytical column by ramping the organic phase from 5% to 35% over a total run time of 150 min. The aqueous and organic mobile phases were, respectively, 0.1% formic acid in water and acetonitrile containing 0.1% formic acid. The FTMS resolution was set to 120,000 and precursor ions were scanned across an m/z range of 300.0–1950.0. The twenty most intense ions were selected in the linear ion-trap for fragmentation by collision (CID) in the Orbitrap. A collision energy of 35 eV was applied throughout.

2.6. Data analysis

Peak picking and peptide searching were performed with MaxQuant v.1.3.0.5 [19] using the SwissProt human database (downloaded 4/4/2013). All fractions of a single sample were searched together. Carbamidomethylation of cysteines was selected as a fixed modification, with oxidation of methionines and protein N-terminal acetylation as variable modifications. The minimal peptide and razor peptide count was set to 1 and the peptide length to 6. The mass tolerance was set to 20 ppm for parent ions and 0.5 Da for fragments. The FDR rate for peptides and proteins was set to 1%. Arginine and lysine were specified as special amino acids to filter for labeled amino acids. Razor peptides were used for quantification, with unmodified counterpart peptides being discarded; re-quantification and iBAQ were permitted. Search results were processed in Perseus 1.4.0.6, which was used to remove decoy search results and contaminant items. Significance was evaluated using the significance B test, using normalized ratios of replicates and their respective intensities to capture the effect of abundance on significance. Proteins that had significance B scores below 0.05 in at least two replicates were selected for further study.

Venn diagrams were created in Perseus and Venn Diagram Plotter 1.5 (http://omics.pnl.gov). Pearson coefficients were calculated in Microsoft Excel by plotting the $\log_2$ of each replicate's L/H ratio against the $\log_2$ of the average L/H ratio for proteins that were quantified in all replicates.

Bioinformatics analyses of the B significant protein list were performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.7, http://david.abcc.ncifcrf.gov/, downloaded 13/1/2016) [20]. Results were visualized using DAVID's Functional Annotation Clustering method. Protein-protein interactions of B significantly changed proteins were analysed using the STRING database (http://string-db.org, downloaded 13/1/2016) [21]. The graphical output of STRING was divided into clusters, with a cluster being defined as a close group of at least 7 proteins that are closer to one-another than to proteins outside the cluster.

2.7. Validation of proteomic data

2.7.1. Immunoblotting

CCRF-CEM cells were-treated with L-OHP as before. Cells were harvested and centrifuged at 90g for 5 min at 4 °C and then washed twice in PBS with 5 mM sodium pyrophosphate, 1 mM sodium orthovanadate, 5 mM sodium fluoride and 1 mM PMSF. After washing, the cells were lysed with SDS lysis buffer (62.5 mM Tris-HCl, pH 6.8, 10% glycerol, 2% SDS, 1% mercaptoethanol, 0.5% bromophenol blue), heated for 10 min at 95 °C, and incubated with 2 units of benzonase for one hour at room temperature to prepare a whole cell lysate. Protein content was measured using the Pierce 660 nm Protein Assay containing the Ionic Detergent Compatibility Reagent. Samples were loaded to SDS-PAGE with uniform loading 10 μg of total protein and were resolved by 12% T gels. The separated proteins were then blotted onto a nitrocellulose membrane (0.2 μm pore size, Bio-Rad) using the TransBlot Turbo semi-dry system (Bio-Rad), after which the membrane was blocked for 1 h with 5% non-fat dry milk in TBS/T (Tris-buffered saliva with 0.1% Tween-20) and incubated with appropriate antibodies in 5% v/v BSA in TBS/T overnight, at 4 °C, with agitation. After incubation, the membranes were washed with TBS/T, incubated with a peroxidase-labeled secondary antibody, and visualized using the Luminata Forte peroxidase substrate. Chemiluminiscence was recorded using a HCD camera (Li-Cor Odyssey FC). Band intensities were normalized against those of the corresponding actin band, and averages of three biological replicates were processed into bar graphs. The significance of observed changes in protein abundance was evaluated using the t-test, with a p-value threshold of 0.05.

2.8. ELISA assay

Cell lysates were prepared in the same way as for immunoblotting. Proteins from approximately 10 million cells were precipitated with ethanol and centrifuged for 10 min at 2000g, after which the supernatant was removed and the pellet was reconstituted in water. The total protein content was then measured using the bicinchoninic acid assay. Additionally, ELISA assays for ferritin and apolipoprotein A1 were performed using the kits specified above, according to the manufacturer's instructions.

2.9. Immunoprecipitation

For preparation of lysates for immunoprecipitation, cells were washed three times in PBS and lysed in TNE buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5% NP-40) supplemented with cOmplete and PhosSTOP tablets (Roche). After 30 min incubation on ice, lysates were cleared by centrifugation. Appropriate antibodies were pre-conjugated to Dynabeads M-280 Sheep anti Rabbit IgG (Novex) for 1 h at 4 °C and cleared lysates incubated with beads and antibodies for 2 h at 4 °C. Immunoglobulin-antigen complexes were washed extensively before elution in 20 μl 2× Laemmli sample buffer before SDS-PAGE. Protein interaction was detected by immunoblot as described above.

2.10. Reverse transcription PCR

The washed cell pellets were resuspended in 1 ml of TRI Reagent (Molecular Research Centre, Cincinnati, OH, USA) and stored at −20 °C until RNA purification. Total RNA was extracted according to manufacturer's instructions. RNA concentration and purity were assessed by a Nanodrop ND 1000 (ThermoScientific, Wilmington, DE, USA). For reverse transcription, we pre-incubated 3 μg of total RNA with 0.3 μg of Random Primers

(Promega, Madison, WI, USA) at 70 °C for 5 min, and immediately placed the mixture on ice. Then 6 μl of RevertAid 5× RT buffer (Fermentas, Vilnius, Lithuania), 3 μl of 10 mM deoxyribonucletotide triphosphates (dNTPs), 0.75 μl of 40 U/μl RNAsin ribonuclease inhibitor (Promega, Madison, WI, USA), and DEPC treated water (Ambion, Austin, TX, USA) up to a final volume of 30 μl were added, and the mixture was incubated for 5 min at room temperature. During the final step, 150 U of RevertAid Moloney Murine Leukemia Virus reverse transcriptase (Fermentas, Vilnius, Lithuania) were added to each tube and the samples were incubated at room temperature for 10 min. Finally, samples were incubated at 42 °C for 60 min and then 70 °C for 10 min. Prepared cDNA were stored at −20 °C until qPCR.

TaqMan technology was used for gene expression analyses for each sample and marker (CCNB1, UTP11L, WDR46, DDX56, RPS19, TP53, APOA1, FTH1, CDKN1A and ACTB). Reaction mixtures were prepared using 1 μl 20× TaqMan Gene Expression Assay (ThermoScientific, Wilmington, DE, USA), 10 μl 2× LightCycler 480 Probes Master (Roche, Basil, Switzerland), 8 μl DEPC treated water (Ambion, Austin, TX, USA) and 1 μl cDNA. Reactions were performed on LightCycler 480 II/96 (Roche, Basil, Switzerland) in two replicates according to manufacturer's instructions.

Relative quantification was used for data evaluation. ACTB was used for endogenous normalization of Ct values for each marker. Fold differences were calculated from ΔΔCt values as treated samples versus controls. Results are average values from three biological replicates with standard deviations.

2.11. Nucleolar staining

The CCRF-CEM cell line was treated in the same way as for proteomic analysis. After treatment, the cells were spun onto microscopy slides using a Cytospin centrifuge (Sakura, Japan) at 500 rpm for 5 min. Spun cells were allowed to dry, fixed with 4% paraformaldehyde in PBS, and stained with toluidine blue as described by Smetana [22]. After staining, the slides were mounted with Pertex (Histolab, Sweden). Three replicates were observed using a Zeiss Axio microscope with a Zeiss AxioCamERc 5 s digital camera at 400× magnification. In each sample, 100 nucleoli were counted and classified visually as either (a) compact nucleoli, (b) ring nucleoli, or (c) micronucleoli as described by Smetana et al. [22].

2.12. Immunofluorescence

Cells were spun using a Cytospin centrifuge as described above, fixed with 4% paraformaldehyde in PBS for 15 min at room temperature, and permeabilized with 0.5% TritonX-100 in PBS for 15 min. The cells were then washed three times in PBS with 0.1% Tween 20 and incubated overnight at 4 °C with primary rabbit or goat antibodies against fibrillarin, DDX56, WDR46, RPS9 and UT11L in 5% FCS. The samples were then washed three times in PBS with 0.1% Tween 20 and incubated with the corresponding anti-rabbit/goat secondary antibodies conjugated to Alexa Fluor 488 in 5% FCS for 1 h at room temperature. Coverslips were stained with DAPI for nuclei counterstaining and mounted in mounting medium. The localization of the proteins of interest was examined using confocal spinning disc fluorescent microscopy (Carl-Zeiss Observer Z1) at 1000× magnification. Image quantification was calculated as number of spots per nuclei in Columbus software version 2.71 (Perkin-Elmer). The number of nuclei entered into the analysis ranged from 900 to 30,000 per sample. Obtained data were analysed with General Linear Model (GLM) and significance was calculated with post hoc Tukey's HSD test in Statistica 12 (StatSoft).

Data analysed with General Linear Model (GLM) and significance was calculated with post hoc Tukey HSD test.

2.13. Fluorescent in-situ hybridization of 47S pre-rRNA

Fluorescent in-situ hybridization of 47S pre-rRNA has been performed according to [23]. Briefly, cells were treated, transferred to microscope glass and fixed as in immunofluorescence experiment. After being rinsed in PBS, cells were permeabilized for 18 h in 70% ethanol at 4 °C. After two washes in 2 × SSC containing 10% formamide, hybridization was performed in the dark at 37 °C for 5 h in a buffer containing 10% formamide, 2 × SSC, 0.5 mg/ml tRNA, 10% dextran sulfate, 250 μg/ml BSA, 10 mM ribonucleoside vanadyl complexes and 0.5 ng/μl of ETS1-1399 probe (5′-cgctagagaaggcttttctc-3′) conjugated to Cy5 (which hybridizes to the 47S pre-rRNA between 01/A′ and A0 cleavage sites). After two washes at 37 °C with 2 × SSC containing 10% formamide, the cover slips were rinsed in PBS and DNA was counterstained with DAPI, mounted and examined as in immunofluorescence.

2.14. Flow cytometry

Flow cytometry analysis was performed using standard protocol described in Borková et al. [24]. Briefly, CCRF-CEM cell line at concentration 2.5. 10^5 cells/ml were seeded in 6-well plates, incubated for 24 h and treated with L-OHP concentrations 1 × IC_{50} and 5 × IC_{50} in time intervals 1, 2, 4, 8, 12, 24 and 36 h in three replicates. Non-treated control for each analysis and time point was done in one replicate, except analysis of proteosynthesis, where negative control was performed together with positive control, 5 μM puromycin, for 24 h.

After treatment, samples for cell cycle analysis were washed with cold PBS and fixed with 70% ethanol and stored overnight at −20 °C. For cell cycle analysis, fixed cells were split into two halves. First half was washed in hypotonic citrate buffer, treated with RNAse (50 μg/mL) and stained with propidium iodide. Second half was used for $pH3^{Ser10}$ antibody (Sigma) labeling and subsequent flow cytometry analysis of mitotic cells.

Samples for DNA synthesis analysis were processed as following. Before harvesting, 10 μM 5-bromo-2-deoxyuridine (BrDU), was added to the cells for pulse-labeling for 30 min. Cells were fixed with ice-cold 70% ethanol and stored overnight. Before the analysis, cells were washed with PBS, and resuspended in 2 M HCl for 30 min at room temperature to denature their DNA. Following neutralization with 0.1 M $Na_2B_4O_7$ (Borax), cells were washed with PBS containing 0.5% Tween-20 and 1% BSA. Staining with primary anti-BrDU antibody for 30 min at room temperature in the dark followed. Cells were than washed with PBS and stained with secondary FITC antibody. Cells were then washed with PBS again and incubated with propidium iodide (0.1 mg/ml) and RNAse A (0.5 mg/ml) for 1 h at room temperature in the dark.

For analysis of RNA synthesis, pulse-labeling with 1 mM 5-bromo uridine (BrU) for 30 min followed the time of treatment. The cells were then fixed in 1% buffered paraformaldehyde with 0.05% of NP-40 in room temperature for 15 min, and then stored at 4 °C overnight. Cells were then washed in PBS with 1% glycine, washed in PBS again, and stained by primary anti-BrDU antibody cross-reacting to BrU for 30 min at room temperature in the dark. After another washing step in PBS cells were stained by secondary FITC antibody. Following the staining, cells were washed with PBS and fixed with 1% PBS-buffered paraformaldehyde with 0.05% of NP-40 for 1 h. Cells were then treated as in DNA analysis, washed with PBS, incubated with propidium iodide and RNAse A for 1 h at room temperature in the dark.

Analysis of protein synthesis was performed using commercial Thermo Click-It kit [25]. Samples were after incubation washed in PBS, transferred into methionine-free RMPI 1640 medium for 1 h to deplete intracellular deposits of methionine. Then, L-azidohomoalanine was added for another 1 h. Samples were then washed with PBS, fixed with 4% paraformaldehyde, permeabilized with 0.25% Triton-X100 in PBS and washed in 3% BSA in PBS. Detection was performed using Click-It Cell reaction buffer kit.

All flow cytometry analyses were analysed on FACS Calibur (Becton Dickinson) flow cytometer equipped with 488 nm single beam laser.

3. Results

The IC_{50} value of L-OHP was determined to be 5.9 μM by the 3-day MTT cytotoxicity assay. In all subsequent experiments, the CCRF-CEM cells were treated with a concentration of L-OHP equal to 5 times the calculated IC_{50}, i.e. 29.3 μM. To determine appropriate timings for sample preparation and to ensure that only early events leading later to apoptosis were included in the analysis, we analysed caspase 3 activity in L-OHP-treated samples using the Magic Red assay kit. In these assays, the time to onset of apoptosis (TA) was defined as the time required for the fluorescence signal of treated cells to become 10% greater than that for untreated control cells. On the basis of these assays, the TA (i.e. the time to caspase activation) was determined to be approximately 480 min, so in all subsequent experiments the CCRF-CEM cells were treated with 5 × IC_{50} L-OHP and then lysed 240 min later, at TA_{50}.

MS analysis of SILAC-labeled whole cell lysates resulted in the identification of 4164, 4087 and 3896 proteins across three replicates, with 3171, 3369 and 3069 of these proteins being quantified, respectively. The three replicates had 3260 identified and 2356 quantified proteins in common. The Pearson coefficients for the three replicates, which reflect the quality of the quantification, were 0.94, 0.88 and 0.84 (Fig. 1). A complete list of the identified and quantified proteins is presented in Supplementary Table 1 and the corresponding raw data have been deposited with the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [26], using the dataset identifier PXD003543.

We defined a protein as having been significantly up or downregulated if it had a significance B p-value below 0.05 in at least two replicates (Supplementary Table 2). Bioinformatic analyses of these proteins were performed using the functional annotation clustering method implemented in DAVID. This method assigned high scores to ribosomal and nucleolar clusters (Table 1). A second, complementary, analysis based on protein-protein interactions was performed using STRING. This revealed four distinct clusters (Fig. 2) corresponding to proteins related to the centromere and G2/M checkpoint (Fig. 2A), secreted proteins (Fig. 2B), ribosomal (Fig. 2C) and nucleolar proteins (Fig. 2D). Proteins from clusters A, C and D are also involved in responses to cellular damage and stress [27].

For a next step, we have selected proteins, which were the most significantly changed in each cluster. Change in abundance of each of selected proteins, CCNB1, UTP11L, WDR46, DDX56, RPS19, TP53 was determined using immunoblotting. For determination of change in abundance of the remaining two proteins, APOA1 and FTH1, commercially available ELISA kit was used. The change in abundance of all those proteins was confirmed - the immunoblot and ELISA results shared the same trend of change, although the numbers weren't exactly the same (Fig. 3). As another orthogonal method, we have used RT-PCR for those proteins. However, the only protein which mRNA changed in expression was CCNB1 with L-OHP/control ratio 0.69 ± 0.14, other tested proteins (UTP11L, WDR46, DDX56, RPS19, TP53, APOA1 and FTH1) haven't changed in mRNA expression.

The first cluster from STRING analysis is cluster of proteins involved in G2/M stop or in centromere assembly. The representing protein of this group, CCNB1 showed decrease in abundance in both protein and mRNA level. To evaluate possible stop in G2/M, or more generally changes in cell cycle after L-OHP treatment, a flow cytometry analysis was performed with multiple time points - 1, 2, 4, 8, 12, 24 and 36 h and with two concentrations – 1 × IC_{50} and 5 × IC_{50} (Fig. 4). The reason for two concentrations is simple - standardised flow cytometry protocol uses lower concentration of cells - 1 million cells in 4 ml, thus 1 × IC_{50} represents cell adjusted concentration and 5 × IC_{50} volume adjusted concentration of L-OHP. The changes of cell cycle after treatment are following: In 1 × IC_{50} is coming about decrease in G0/G1 and increase in S. The G2/M phase remains more less constant; however, phosphorylation of histone H3 at serine 10, a marker of mitosis, is decreasing

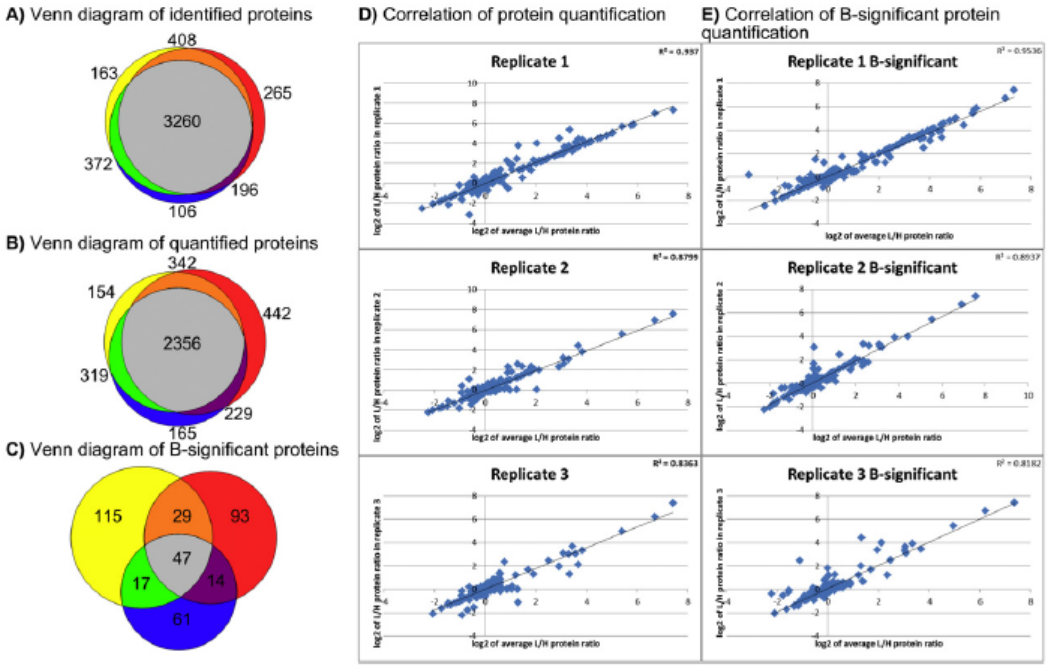


Fig. 1. Characterization of the proteomic dataset. Venn diagram of all identified proteins (A) and quantified proteins (B). Results shown in yellow, red, and blue correspond to replicates 1, 2, and 3, respectively. Correlation of protein quantifications in individual replicates relative to the average (C) and correlations of B-significant protein quantifications relative to the average (D).

Table 1
DAVID annotation analysis of B-significant proteins. For simplicity, only annotation clusters with enrichment scores above 10 are shown, along with the six most highly scored processes for each cluster. Annotation cluster 1 includes ribosomal proteins, cluster 2 includes proteins involved in ribosome biogenesis and rRNA processing, and cluster 3 includes nucleolar proteins.

Category	Term	Count	%	P value
Annotation cluster 1	*Enrichment score: 42.00*			
KEGG_PATHWAY	hsa03010:Ribosome	38	36.89	6.54E-58
GOTERM_BP_FAT	GO:0006414 – translational elongation	38	36.89	1.26E-57
GOTERM_MF_FAT	GO:0003735 – structural constituent of ribosome	41	39.81	7.40E-56
SP_PIR_KEYWORDS	Ribosomal protein	41	39.81	1.10E-54
SP_PIR_KEYWORDS	Ribonucleoprotein	43	41.75	9.59E-51
SP_PIR_KEYWORDS	Ribosome	31	30.10	1.14E-50
Annotation Cluster 2	*Enrichment score: 12.26*			
GOTERM_BP_FAT	GO:0042254 – ribosome biogenesis	19	18.45	7.37E-20
GOTERM_BP_FAT	GO:0022613 – ribonucleoprotein complex biogenesis	19	18.45	9.22E-17
GOTERM_BP_FAT	GO:0006364 – rRNA processing	15	14.56	7.94E-16
GOTERM_BP_FAT	GO:0016072 – rRNA metabolic process	15	14.56	1.48E-15
GOTERM_BP_FAT	GO:0034470 – ncRNA processing	15	14.56	1.76E-11
GOTERM_BP_FAT	GO:0034660 – ncRNA metabolic process	16	15.53	2.27E-11
Annotation Cluster 3	*Enrichment score: 11.78*			
GOTERM_CC_FAT	GO:0005730 – nucleolus	33	32.04	1.32E-17
GOTERM_CC_FAT	GO:0031981 – nuclear lumen	40	38.83	1.03E-13
GOTERM_CC_FAT	GO:0070013 – intracellular organelle lumen	44	42.72	1.21E-13
GOTERM_CC_FAT	GO:0043233 – organelle lumen	44	42.72	2.70E-13
GOTERM_CC_FAT	GO:0031974 – membrane-enclosed lumen	44	42.72	5.36E-13
SP_PIR_KEYWORDS	Nucleus	38	36.89	8.83E-04

over the time. Thus more cells are blocked in G2 and not enter to mitosis. In 5 × IC_{50} is distribution of cell cycle stages similar to control. The most visible changes are slight increase of G0/G1, slight decrease of S and G2/M and, mainly, increase in sub G1 fraction, which is fraction of apoptotic cells (Fig. 4).

Other clusters were validated in the following way: Cluster of secreted proteins has been validated using ELISA kit only. Influence of L-OHP to ribosomes has been validated using immunofluorescence microscopy (IFM) for RPS19 showing its translocation from nucleus to the whole cell and disappearing of positive foci (Fig. 5; with L-OHP/control ratio of spot numbers 0.09, p-value < 0.0001). Further validation of ribosomal stress has been performed by 47S pre-rRNA FISH, where the pure nucleolar concentration of 47 s pre-rRNA changed after L-OHP treatment to whole cellular localization (Fig. 6; ratio of spot numbers 0.05, p-value < 0.0001). Cluster of nucleolar proteins has been validated using IFM for regulated proteins UTP11L with L-OHP/control ratio of spot number 155.32 (p-value < 0.0001) showing disintegration of nucleoli (Fig. 5) and DDX56 (ratio 0.01, p-value < 0.0001) as well as for known nucleolar markers FBL (ratio 0.09, p-value < 0.0001) and NCL (ratio 1.76 showing nucleolar decomposition, p-value < 0.0001; Fig. 5). In all cases, there is a visible trend of nucleolar disintegration. Nucleolar morphology was further examined using light microscopy with toluidine blue staining with a shift towards metabolically inactive micronucleoli (12% to 69%, p-value < 0.0001 Fig. 6).

The most often discussed effect of L-OHP, the DNA damage, is not included in any of bioinformatic result for this study. However, according to literature mining of proteins changed in abundance, there are several proteins involved in DNA damage response (Table 2). The DNA damage caused by L-OHP was observed using IFM of H2AX with increase of protein phosphorylation at serine 139 (Supplementary Fig. 1) without change in total protein expression (Supplementary Table 1).

The common downstream result of all three stress pathways, DNA damage, nucleolar and ribosomal stress, is activation of p53. In the proteomic result, there were significant upregulation of TP53 with ratio 1.73 in only one of three replicates and this protein was not quantified in remaining two (Supplementary Table 1). Western blot verification showed upregulation of this protein with lower ratio 1.21. However, immunoprecipitation experiment using enrichment of the HDM2 and RPL5, the known activators of p53 in those stresses, haven't show any increase in this pathway (Supplementary Fig. 2).

The next step was confirmation of obtained results in other cell lines. For this purpose, we have selected osteosarcoma cell line U2OS and colorectal cancer cell line HCT116. Those cell lines were treated with 5× of their specific IC_{50} (34.5 μM for U2OS and 53.5 μM for HCT116) for uniform time 4 h corresponding to $TA_{1/2}$ of CCRF-CEM. Those cells were examined for levels of UTP11L, WDR46 and DDX56 proteins, for nucleolar stress by monitoring FBL and NCL and 47S pre-rRNA FISH for ribosomal stress. Whereas U2OS showed similar trend in behaviour with CCRF-CEM based on IFM or FISH, the trend in HCT116 was just opposite to CCRF-CEM based on FISH or immunoblot DDX56 (Fig. 7).

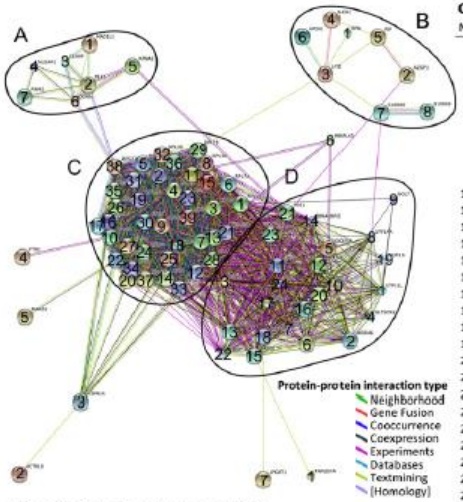
4. Discussion

The IC_{50} concentration of L-OHP was determined to 5.9 μM. We have used concentration 5 × IC_{50} of L-OHP, which is similar concentration of drug as in reported blood concentration of L-OHP, which is 4.15 μM [28]. Last but not least, advance in proteomic methods allowed us to use lower concentration than in previous study [16], where concentration 10 × IC_{50} was necessary to observe visible results, whereas 5 × IC_{50} haven't caused any changes observable at 2D electrophoresis (Tylečková, unpublished results). The advance in instrumentation and implementation of Orbitrap-based proteomics led to increase of significantly changed proteins. In this study we have identified 107 differently regulated proteins compared to Tylečková with an average of 31 differently regulated protein spots [16]. Such an increase allowed us to utilize modern bioinformatic tools such as DAVID or STRING. Mainly STRING analysis (Fig. 2) showed totally four clusters of interacting proteins.

Some of the proteins remained unclustered in the STRING analysis (FAM207A, ACTR10, HSPA1A, FTH1, MARS2, MRPL43 and LPCAT1). The relevant information in the literature is available only for proteins HSPA1A, FTH1 and LPCAT1. HSPA1A functions as a protein chaperone, but it is also involved in anti-inflammatory and anti-oxidant processes [29] and DNA damage responses [30]. FTH1 is involved in innate immune reactions and is overexpressed upon lipopolysaccharide stimulation [31], anoxia or oxidative stress [32], among other things. FTH1 (ferritin) has been considered as a vehicle for platinum drugs; it is possible that such drugs may form ferritin conjugates and thus induce its overexpression. LPCAT1 is reported to have anti-inflammatory properties [33]. The change in the levels of FTH1 was verified by ELISA (Fig. 3).

The first cluster from STRING analysis, cluster A (Fig. 2) contained proteins related to the centromere and G2/M cell cycle phase. L-OHP has been reported to arrest mitosis at the G2/M checkpoint [34]. This is consistent with our observation that L-OHP treatment caused the downregulation of cyclin B1 (CCNB1) and PLK1 [36], as well as proteins interacting both with centromeres and microtubules such as NUSAP1 [37], MAD1L1 [38], HOOK3 [39], CEP41 [40] and CENPF [41] (Table 2). The downregulation of CCNB1 following L-OHP treatment has been observed previously [42] and was verified by immunoblotting (Fig. 3) and mRNA expression. Arrest in G2/M was confirmed by flow cytometric analysis of cellular cycle and decrease phosphorylation of serine 10 of histone H3 in both analysed concentrations.

Protein cluster B (Fig. 2B) consisted of proteins secreted extracellularly, all of which were upregulated and many of them seems to associate with histogenetic origin of CCRF-CEM cells (T lymphocytes/lymphoblasts). LYZ is involved in innate immunity and ameliorates oxidative stress [43], and has previously been reported to be elevated after cisplatin treatment [44]. SPN is a one of the major glycoproteins in T-



A Proteins related to centromere or mitosis

Nr.	Protein names	Gene	Ratio L/H
1	Mitotic spindle assembly checkpoint protein MAD1	MAD1L1	0.58
2	Serine/threonine-protein kinase PLK1	PLK1	0.74
3	Centromere protein F	CENPF	0.82
4	Nucleolar and spindle-associated protein 1	NUSAP1	0.85
5	Importin subunit alpha	KPNA2	0.86
6	G2/mitotic-specific cyclin-B1	CCNB1	0.87
7	Ribonucleoside-diphosphate reductase subunit M2	RRM2	1.11

B Secretory proteins

Nr.	Protein names	Gene	Ratio L/H
1	Leukosialin	SPN	3.27
2	Zinc-alpha-2-glycoprotein	AZGP1	3.97
3	Lysozyme C	LYZ	5.49
4	Lipocalin-1	LCN1	5.70
5	Prolactin-inducible protein	PIP	8.46
6	Apolipoprotein A-I	APOA1	9.30
7	Protein S100-A9	S100A9	10.45
8	Protein S100-A8	S100A8	12.61

Unclustered proteins

Nr.	Protein names	Gene	Ratio L/H
1	Protein FAM207A	FAM207A	0.68
2	Actin-related protein 10	ACTR10	0.74
3	Heat shock 70 kDa protein 1A/1B	HSPA1A	1.13
4	Ferritin	FTH1	1.22
5	Methionine--tRNA ligase, mitochondrial	MARS2	3.79
6	39S ribosomal protein L43, mitochondrial	MRPL43	4.38
7	Lysophosphatidylcholine acyltransferase 1	LPCAT1	10.96

C Ribosomal proteins

Nr.	Protein names	Gene	Ratio L/H
1	60S ribosomal protein L7-like 1	RPL7L1	0.43
2	60S ribosomal protein L35a	RPL35A	0.86
3	60S ribosomal protein L15	RPL15	0.86
4	60S ribosomal protein L13	RPL13	0.86
5	60S ribosomal protein L34	RPL34	0.86
6	60S ribosomal protein L7a	RPL7A	0.86
7	60S ribosomal protein L27a	RPL27A	0.87
8	60S ribosomal protein L36	RPL36	0.87
9	60S ribosomal protein L6	RPL6	0.87
10	60S ribosomal protein L19	RPL19	0.87
11	60S ribosomal protein L18a	RPL18A	0.88
12	40S ribosomal protein S8	RPS8	0.88
13	60S ribosomal protein L7	RPL7	0.88
14	60S ribosomal protein L4	RPL4	0.88
15	40S ribosomal protein S5	RPS5	0.88
16	40S ribosomal protein S16	RPS16	0.88
17	40S ribosomal protein S4, X isoform	RPS4X	0.88
18	60S ribosomal protein L10a	RPL10A	0.89
19	60S ribosomal protein L18	RPL18	0.89
20	40S ribosomal protein S23	RPS23	0.89
21	60S ribosomal protein L23a	RPL23A	0.89
22	60S ribosomal protein L13a	RPL13A	0.89
23	40S ribosomal protein S24	RPS24	0.89
24	Ribosomal protein L14 variant	RPL14	0.90
25	60S ribosomal protein L8	RPL8	0.90
26	40S ribosomal protein S3a	RPS3A	0.90
27	40S ribosomal protein S11	RPS11	0.90
28	60S ribosomal protein L3	RPL3	0.90
29	60S ribosomal protein L21	RPL21	0.90
30	40S ribosomal protein S6	RPS6	0.90
31	60S ribosomal protein L32	RPL32	0.90
32	60S ribosomal protein L35	RPL35	0.90
33	40S ribosomal protein S13	RPS13	0.90
34	40S ribosomal protein S3	RPS3	0.91
35	60S ribosomal protein L10	RPL10	0.91
36	60S ribosomal protein L12	RPL12	0.92
37	60S ribosomal protein L26	RPL26	0.96
38	Ubiquitin-40S ribosomal protein S27a	RPS27A	1.13
39	60S ribosomal protein L27	RPL27	1.20

D Nucleolar proteins

Nr.	Protein names	Gene	Ratio L/H
1	Probable U3 small nucleolar RNA-associated protein 11	UTP11L	0.24
2	WD repeat-containing protein 46	WDR46	0.28
3	Ribosome biogenesis protein NSA2 homolog	NSA2	0.31
4	Glioma tumor suppressor candidate region gene 2 protein	GLTSCR2	0.36
5	Probable ATP-dependent RNA helicase DDX56	DDX56	0.53
6	DDB1- and CUL4-associated factor 13	DCAF13	0.54
7	Ribosomal L1 domain-containing protein 1	RSL1D1	0.58
8	U3 small nucleolar RNA-associated protein 14 homolog A	UTP14A	0.63
9	Nucleolar protein 7	NOL7	0.66
10	U3 small nucleolar RNA-associated protein 6 homolog	UTP6	0.69
11	Guanine nucleotide-binding protein-like 3	GNL3	0.73
12	ATP-dependent RNA helicase DDX54	DDX54	0.76
13	Protein RRP5 homolog	PDCD11	0.76
14	Probable rRNA-processing protein EBP2	EBNA1BP2	0.76
15	RRP12-like protein	RRP12	0.78
16	Probable ATP-dependent RNA helicase DDX27	DDX27	0.80
17	Ribosome biogenesis protein BRX1 homolog	BRIX1	0.80
18	Pumilio domain-containing protein KIAA0020	KIAA0020	0.81
19	Nucleolar protein 16	NOP16	0.82
20	RNA-binding protein 28	RBM28	0.83
21	Pescadillo homolog	PES1	0.87
22	Myb-binding protein 1A	MYBBP1A	0.87
23	Nucleolar GTP-binding protein 1	GTPBP4	0.90
24	Ribosome production factor 2 homolog	RPF2	1.10

Fig. 2. STRING analysis of B-significant proteins. The yellow cluster (A) includes centromere proteins and proteins involved in G2/M arrest. The purple cluster (B) includes common secretory proteins. The blue cluster (C) includes ribosomal proteins. The red cluster (D) includes nucleolar proteins. The proteins in the clusters are listed in the corresponding tables.

cells; it is involved in cellular adhesion and its upregulation activates the p53 protein, which can lead to cell cycle arrest [45]. In addition, p53 plays a central role in DNA damage responses, as discussed below. Lipocalin 1 (LCN1) is a transporter of small hydrophobic molecules [46], and the increase in its expression is probably linked to L-OHP detoxification. According to a previous report [47], LCN1 is a potential biomarker of cisplatin-induced nephrotoxicity. AZGP1 is involved in lipid metabolism and has been identified as a promising biomarker of different types of carcinomas. It effects the cell cycle by downregulating the cyclin dependent kinase cdc2 and slows the G2/M transition [48]. PIP, which was also upregulated in our experiments, is a potential cancer biomarker whose abundance increases in breast and prostate cancers and which binds to AZGP1 [49]. Variation of APOA1 level can be linked to poor prognosis in cancer, both without [50] and with cisplatin treatment [51]. APOA1 reportedly has anti-inflammatory, anti-apoptotic, and (of particular relevance during L-OHP treatment), antioxidant activity. As such, it is reasonable to suppose that the increase in APOA1 expression following L-OHP treatment is due to the high oxidative stress induced by the drug, so an increase in APOA1 expression can be regarded as a biomarker for effective L-OHP therapy [52]. The interaction between APOA1 and its cellular receptor, ABCA1, triggers several signaling events. These include the activation of the Cdc42 protein (which leads to cytoskeletal reorganization) and changes in the ability of other oncoproteins, including Ras and EGFR, to induce cellular transformation. In addition, binding partners of APOA1 such as APOL1 are involved in autophagy. Identifying the mechanisms that modulate APOA1 gene expression could lead to a deeper understanding of L-OHP's mechanism of action,

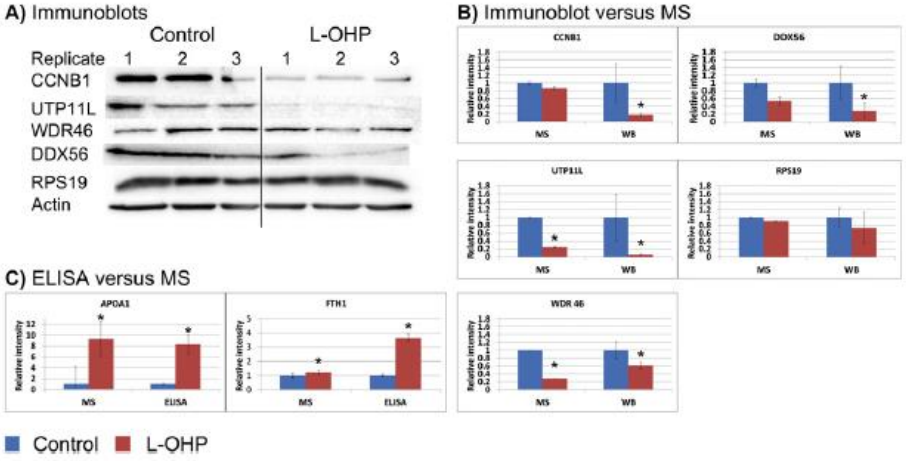


Fig. 3. Verification of the up/down-regulation of the proteins TERF1, UTP11L, WDR46, DDX56, RPS1, APOA1 and FTH1. The results of triplicate Western blot analyses for L-OHP-treated and control cells are shown in A, with the corresponding normalized base intensities in B. The results for the proteins APOA1 and FTH1 were further verified by ELISA. The relative intensities for these proteins in the treated and untreated cells are compared to the relative MS intensities in C.

and APOA1 expression could be monitored as a biomarker for treatment response. Two other proteins that were identified as being significantly upregulated, S100A8 and S100A9, are known to be involved in acute inflammatory responses and the induction of apoptosis via the release of nitric oxide and reactive oxygen species [53]. The differential overexpression of APOA1 following L-OHP treatment was verified using ELISA (Fig. 3).

We observed statistically significant downregulation of several ribosomal proteins in cluster C (Fig. 2C). This is a sign of ribosomal stress, which causes the shutdown of ribosome biosynthesis [27]. However, ribosomal proteins were not regulated uniformly: some ribosomal proteins were downregulated non-significantly (e.g. RPS19, whose downregulation was verified by immunoblotting and microscopy - Figs. 3 and 5) and others exhibited no detectable change in expression (e.g. RPL5, RPL11; Supplementary Table 1). These differences in ribosomal protein expression may be connected to the roles of the ribosomal proteins in nucleolar stability, rRNA synthesis or p53 activation [54]. Decrease in ribosome content was further confirmed by decrease of 47S pre-rRNA level and its translocation from nucleoli to nucleoplasm and cytoplasm. Contrary to decrease of ribosome synthesis, the proteosynthesis decreased very slowly indicating full functionality of remaining ribosomes (Fig. 4). The L-OHP treatment is reported to affect ribosomes [55], and the ribosomal pathway is involved in the development of L-OHP resistance [56].

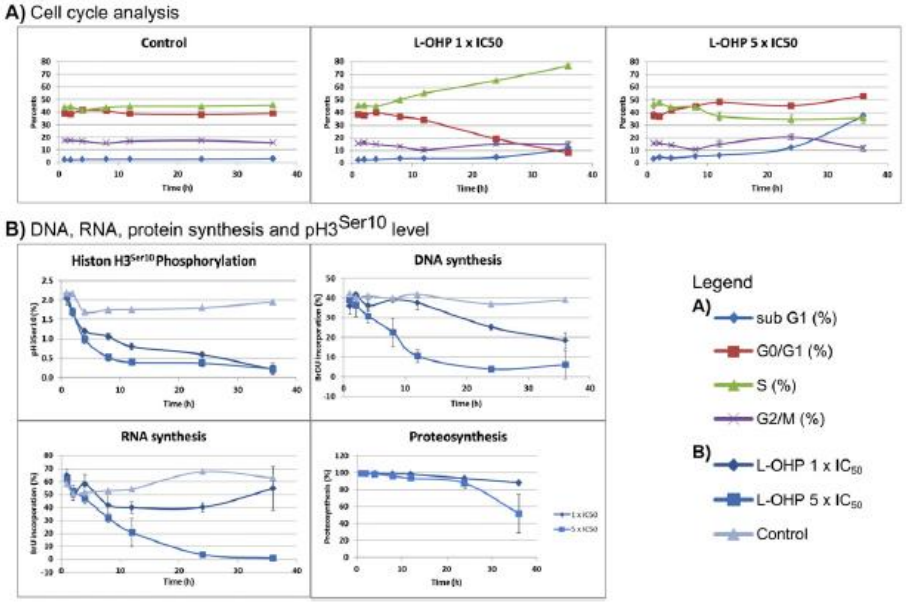
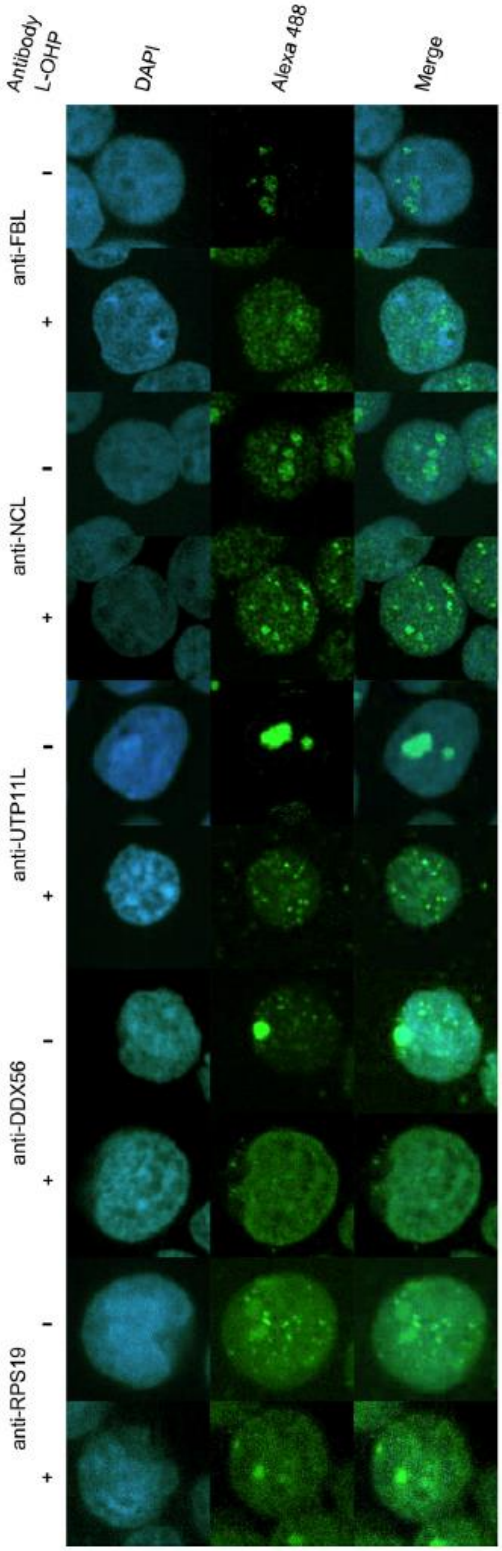


Fig. 4. Flow cytometric analysis of cellular response to L-OHP. This analysis is split into cell cycle analysis (A) showing cell cycle to be influenced mainly at lower L-OHP concentration and analysis of mitosis rate by phosphorylation of histone H3 at Ser10, DNA, RNA and proteosynthesis (B).



Correspondingly to down-regulation or ribosome biosynthesis, the concentration and distribution of proteins localized in nucleolus shown in Fig. 2D may indicate nucleolar stress, which reportedly causes the shrinkage or disruption of nucleoli under stress conditions [57] and has been reported to occur after L-OHP treatment [58,59]. We were able to reproduce these results during the early stages of the L-OHP response (Fig. 6) and observed a clear shift towards metabolically inactive micronucleoli (12% to 69%, Fig. 6D). This phenomenon is reflected in our proteomic data, which show the downregulation of proteins involved in the small (UTP11L, UTP6, UTP14A, DCAF13, PNO1 and WDR46 [60,61]) and large (NSA2, PES1, DDX27, EBNA1BP2 (EBP2), RBM28, RPF2 and EFTUD1 [62–66]) ribosomal subunit processomes, RNA processing proteins (NOP16, DDX proteins [67,68]), nucleolar stress sensors (MYBBP1 [69]), and PES1 interactors (BRIX1, RRP12 [62]). We selected three proteins from this set - UTP11L, WDR46 and DDX56 - for immunoblot verification because of their high fold changes (Fig. 2); in all three cases, immunoblotting confirmed their downregulation (Fig. 3) and change of nucleolar localization was confirmed by IFM (Fig. 5). Nucleolar desintegration was also observed using FBL and NCL as nucleolar markers and observed on FISH monitoring of 47S pre-rRNA, where decrease of rRNA followed the same trend as was observed in toluidine blue staining (Fig. 6).

In addition to the proteins assigned to the clusters shown in Fig. 2, we found that several of the proteins identified in the proteomic analysis have previously been reported to be involved in DNA damage responses (Table 2). DNA damage is known to trigger nucleolar and ribosomal stress [27] and the formation of DNA-platinum crosslinks is well established as the main effect of platinum drugs [6]. In our experiments with L-OHP treatment, we observed the upregulation of the DNA damage response proteins TERF2 [70] and RRM2 [71], as well as the downregulation of other DNA damage response proteins such as NHEJ1 [72], NUSAP1 [73], GLTSCR2 [74], and KIAA0020 [75]. The DNA damage was confirmed by the increase of H2AX phosphorylation at Ser139 (Supplementary Fig. 1).

Ribosomal and nucleolar stress are closely linked because nucleolus is the place of ribosome biogenesis and this biogenesis is thus dependent on nucleoli in good condition. Nucleolar stress is triggered after DNA damage by inhibition of RNA polymerase I [27]. Our experiments revealed no changes in the expression of RNA polymerase I upon L-OHP treatment (Supplementary Table 1). However, nucleolar stress responses can also be activated via TOPBP1 [76]. While our analysis identified TOPBP1, we were not able to quantify this protein (Supplementary Table 1). The actual pathway of DNA-damage-activated nucleolar stress caused by L-OHP is not elucidated yet.

All three stress responses (ribosomal, nucleolar and DNA damage) are known to activate the p53 pathway. The activation of this pathway by DNA damage responses is well described and reviewed elsewhere, and the main pathway of DNA damage-dependent p53 activation is through nucleolar disruption and activation of p53 as a consequence of the binding of MDM 2 to the free ribosomal subunits RPL5 and RPL11 [27]. Based on results from immunoprecipitation experiment, this pathway is not activated in CCRF-CEM after L-OHP treatment (Supplementary Fig. 2). Protein p53 is thus not participating in L-OHP induced nucleolar and ribosomal stress in CCRF-CEM cell line. However, there are other mechanisms of apoptosis activation driven by nucleolar stress occurring in cancer or organisms lacking a p53 protein homolog are reviewed by James et al. [77]. Cancer cells generally have unusually large nucleoli [78], and one of the proteins involved in the p53-independent nucleolar stress response, pescadillo (PES1), has been shown to be upregulated in $p53^{-/-}$ cells and in cancer. In keeping with our results,

Fig. 5. Immunofluorescent staining of FBL, NCL, UTP11L, DDX56 and RPS19 visualizes changes in their nucleolar localization after L-OHP treatment. The localization of the proteins of interest was examined using confocal spinning disc fluorescent microscopy at 1000× magnification.

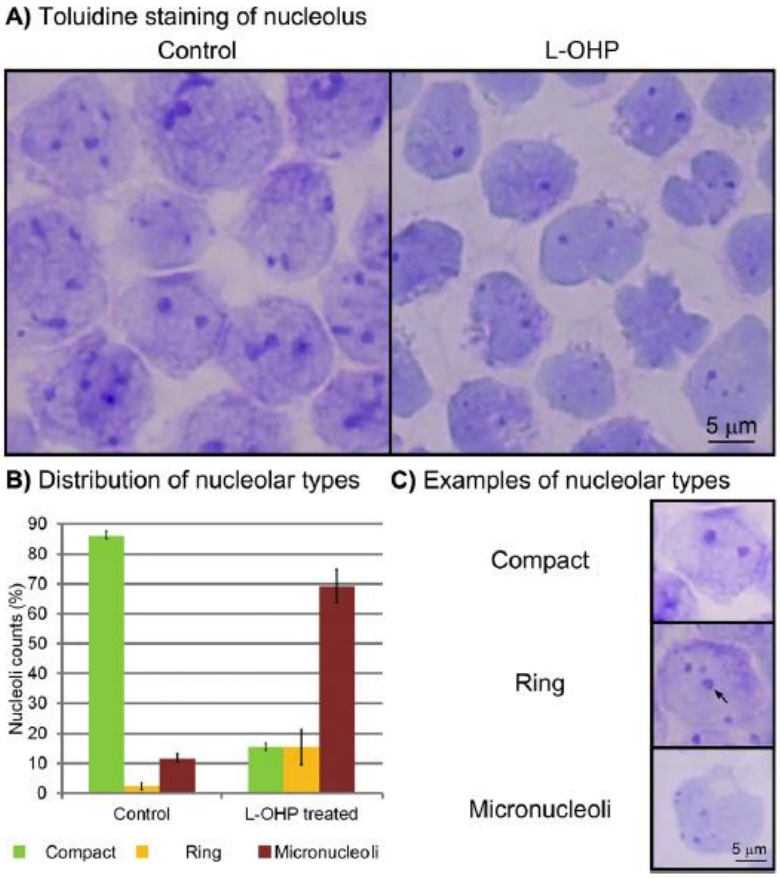


Fig. 6. Fluorescent in-situ hybridization of 47S pre-rRNA (A) and toluidine blue staining of control and L-OHP treated CCRF-CEM cells (B). The distribution of nucleolar types is shown in subfigure C and representative examples of each type are shown in subfigure D. Definitions of single nucleolar types are provided in [87–89].

downregulation of PES1 reportedly leads to cell cycle arrest and apoptosis [79] (Supplementary Table 2).

The activation of nucleolar and ribosomal stress by L-OHP has been already described previously [10,55,56,58] and is described in this paper in more detail. However, nucleolar stress proteins are not often reported in large-scale genomic or proteomic experiments of cellular response to drugs known to induce such a stress e.g. doxorubicin used to induce nucleolar stress in [80] but without its detection in [16,81]. The same situation is in the case of L-OHP, where nucleolar and ribosomal stress were detected in this paper, but in study of Briffa [82] comparing genomic, transcriptomic and proteomic responses of different colorectal cancers was not observed any signs of such stress responses.

Table 2
Proteins exhibiting significant changes in abundance that are involved in DNA damage responses and G2/M arrest. Proteins are shown along with their average fold change, intensity, significance B p-value, and references detailing their involvement in DNA damage responses.

Protein names	Gene names	Ratio L/H	Reference
Proteins involved in DNA damage response			
Glioma tumor suppressor candidate region gene 2 protein	GLTSCR2	0.36	Kim et al. [74]
Non-homologous end-joining factor 1	NHEJ1	0.55	Yano et al. [72]
Pumilio domain-containing protein KIAA0020	KIAA0020	0.81	Chang et al. [75]
Nucleolar and spindle-associated protein 1	NUSAP1	0.85	Kotian et al. [73]
Ribonucleoside-diphosphate reductase subunit M2	RRM2	1.11	Lu et al. [71]
Telomeric repeat-binding factor 2	TERF2	1.39	de Lange et al. [70]
Proteins involved in G2/M arrest			
Centrosomal protein of 41 kDa	CEP41	0.38	Lee et al. [40]
Protein Hook homolog 3	HOOK3	0.39	Ge et al. [39]
Mitotic spindle assembly checkpoint protein MAD1	MAD1L1	0.58	Kruse et al. [38]
Serine/threonine-protein kinase PLK1	PLK1	0.74	Lee et al. [36]
Centromere protein F	CENPF	0.82	Waters et al. [41]
G2/mitotic-specific cyclin-B1	CCNB1	0.87	Menon et al. [35]

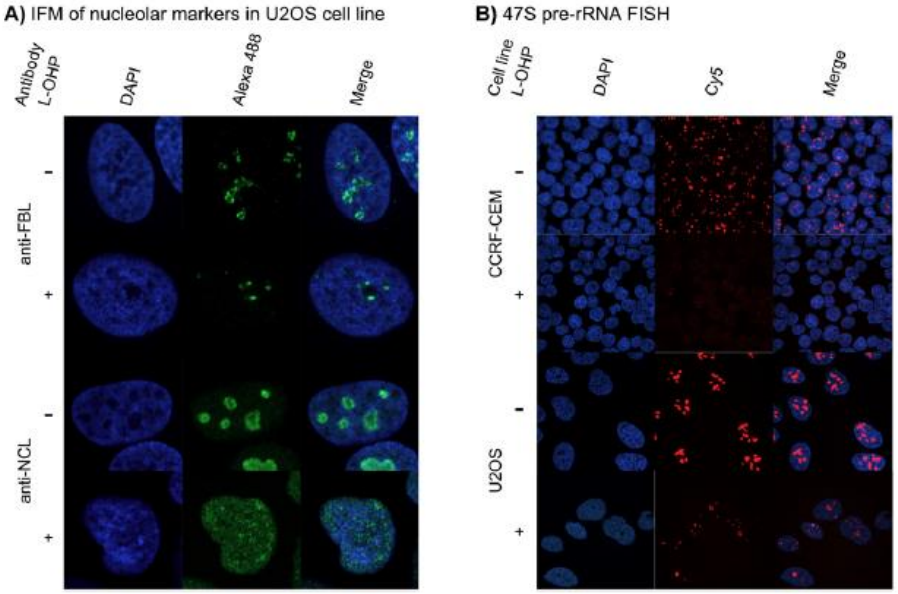


Fig. 7. Nucleolar stress induced by L-OHP in U2OS and HCT116 cell lines.

This is not in strict contradiction with results of this study, because mRNA of nucleolar and ribosomal proteins tested in this study has not changed in expression too and proteomic experiment in [82] was targeted to selected proteins only. However, to address the issue of nucleolar and ribosomal stress, two solid cancer cell lines, osteosarcoma derived U2OS and colorectal cancer derived HCT116, were examined. Whereas U2OS followed the CCRF-CEM in both parameters with shrinking nucleoli and reduction of 47S pre-rRNA, HCT116 cell line responded in opposite manner by increasing of 47S pre-rRNA and activation of protein DDX56 (Fig. 7). Those results show different mechanisms of cellular response in HCT116 and could be the promising background for future comparison of L-OHP cellular responses.

Nucleolar stress is a common stress response induced by a plethora of stressors, e.g. UV irradiation, heat, or chemical stress [83]. Induction of nucleolar stress may play an important role in cancer treatment. This can be illustrated by the wide range of anticancer drugs that cause such stress, including L-OHP, cisplatin [84], doxorubicin [80] and etoposide [85]. Indeed, some authors have suggested that nucleolar stress proteins may be viable targets for anti-cancer drug development [86] because drugs selectively targeting the nucleolus would be unlikely to cause major genotoxic stress and additional mutations. Prior to this work, the effects of L-OHP treatment on nucleolar stress had only received passing attention, notably in Jamieson's investigation into the use of microscopy to monitor nucleolar shrinkage [58]. This shrinkage correlates with observed level of neuropathy in rats or mice, and can be modulated by sequential dosage with paclitaxel and L-OHP.

Interestingly, extensive proteomic profiling of L-OHP treated CCRF-CEM cells did not reveal activation of apoptotic machinery, thus confirming validity of our approach based on treatment of cells with cytotoxic drugs for time corresponding to half-time to induction of apoptosis.

5. Conclusion

We have analysed the proteomic responses of CCRF-CEM cells to L-OHP treatment. The main cellular response to the drug was the activation of DNA damage responses, which triggered nucleolar stress and significantly impaired ribosome biogenesis. With adequate follow-up validation, proteins such as APOA1 and FTH1 that were significantly upregulated by L-OHP treatment could potentially be used as surrogate biomarkers of therapeutic responses to L-OHP.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2017.05.005.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We would like to thank Pavlina Kalabova for help with the verification of MS results. This work was supported by grants LF_2016_019, Ministry of School, Education and Youth (LO1304 and LM2015064) and Technological Agency of the Czech Republic (TE02000058).

MH and PD designed the experiments; TO, DH, GR and PD prepared samples for MS; TO and DH performed MS analysis and data processing; TO, JV and JS performed immunoblotting and staining verification; JBS performed ELISA assays; TO, ZM, JR and PZ performed IFM; ZM performed the FISH; PM performed immunoprecipitation; HS performed RT-PCR; TO, LV, MH and PD wrote the article.

References

[1] R.P. Miller, R.K. Tadagavadi, G. Ramesh, W.B. Reeves, Mechanisms of cisplatin nephrotoxicity, Toxins 2 (2010) 2490–2518, http://dx.doi.org/10.3390/toxins2112490.
[2] M.E. Gore, I. Fryatt, E. Wiltshaw, T. Dawson, B.A. Robinson, A.H. Calvert, Cisplatin/carboplatin cross-resistance in ovarian cancer, Br. J. Cancer 60 (1989) 767–769.
[3] Y. Kidani, M. Noji, T. Tashiro, Antitumor activity of platinum(II) complexes of 1,2-diamino-cyclohexane isomers, Gann 71 (1980) 637–643.
[4] S. Stintzing, Management of Colorectal Cancer, F1000Prime Rep, 6, 2014http://dx.doi.org/10.12703/P6-108.
[5] I. Garrido-Laguna, M. Hidalgo, Pancreatic cancer: from state-of-the-art treatments to promising novel therapies, Nat. Rev. Clin. Oncol. (2015)http://dx.doi.org/10.1038/nrclinonc.2015.53.
[6] S. Dilruba, G.V. Kalayda, Platinum-based drugs: past, present and future, Cancer Chemother. Pharmacol. 77 (2016) 1103–1124, http://dx.doi.org/10.1007/s00280-016-2976-z.
[7] N. Mounier, T. El Gnaoui, H. Tilly, D. Canioni, C. Sebban, R.-O. Casasnovas, R. Delarue, A. Sonet, P. Beaussart, T. Petrella, S. Castaigne, S. Bologna, G. Salles, A. Rahmouni, P.

Gaulard, C. Haioun, Rituximab plus gemcitabine and oxaliplatin in patients with refractory/relapsed diffuse large B-cell lymphoma who are not candidates for high-dose therapy. A phase II lymphoma study association trial, Haematologica 98 (2013) 1726–1731, http://dx.doi.org/10.3324/haematol.2013.090597.
[8] T. Alcindor, N. Beauger, Oxaliplatin: a review in the era of molecularly targeted therapy, Curr. Oncol. 18 (2011) 18–25.
[9] R.K. Mehmood, Review of cisplatin and oxaliplatin in current immunogenic and monoclonal antibody treatments, Oncol. Rev. 8 (2014) 256, http://dx.doi.org/10.4081/oncol.2014.256.
[10] M.J. McKeage, T. Hsu, D. Screnci, G. Haddad, B.C. Baguley, Nucleolar damage correlates with neurotoxicity induced by different platinum drugs, Br. J. Cancer 85 (2001) 1219–1225, http://dx.doi.org/10.1054/bjoc.2001.2024.
[11] M. Stukova, M.D. Hall, S.D. Tsotsoros, J.P. Madigan, N.P. Farrell, M.M. Gottesman, Reduced accumulation of platinum drugs is not observed in drug-resistant ovarian cancer cell lines derived from cisplatin-treated patients, J. Inorg. Biochem. 149 (2015) 45–48, http://dx.doi.org/10.1016/j.jinorgbio.2015.05.003.
[12] L. Li, R. Liu, F. Xu, Y. Zu, Z. Liu, Study on the interactions between anti-cancer drug oxaliplatin and DNA by atomic force microscopy, Micron. Oxf. Engl. 1993 (76) (2015) 46–51, http://dx.doi.org/10.1016/j.micron.2015.05.002.
[13] T. Ochiai, K. Nishimura, T. Watanabe, M. Kitajima, A. Nakatani, T. Sato, K. Kishine, S. Futagawa, S. Mashiko, I. Nagaoka, Mechanism underlying the transient increase of serum iron during FOLFOX/FOLFRI therapy, Mol. Clin. Oncol. 2 (2014) 968–972, http://dx.doi.org/10.3892/mco.2014.385.
[14] T.J. Preston, J.T. Henderson, G.P. McCallum, P.G. Wells, Base excision repair of reactive oxygen species-initiated 7,8-dihydro-8-oxo-2′-deoxyguanosine inhibits the cytotoxicity of platinum anticancer drugs, Mol. Cancer Ther. 8 (2009) 2015–2026, http://dx.doi.org/10.1158/1535-7163.MCT-08-0929.
[15] S. Ahmad, Platinum-DNA interactions and subsequent cellular processes controlling sensitivity to anticancer platinum complexes, Chem. Biodivers. 7 (2010) 543–566, http://dx.doi.org/10.1002/cbdv.200800340.
[16] J. Tyleckova, R. Hrabakova, K. Mairychova, P. Halada, L. Radova, P. Dzubak, M. Hajduch, S.J. Gadher, H. Kovarova, Cancer cell response to anthracyclines effects: mysteries of the hidden proteins associated with these drugs, Int. J. Mol. Sci. 13 (2012) 15536–15564, http://dx.doi.org/10.3390/ijms131215536.
[17] H. Skalnikova, J. Martinkova, R. Hrabakova, P. Halada, M. Dziechciarkova, M. Hajduch, S.J. Gadher, A. Hammar, D. Enetoft, A. Ekefjard, O. Forsstrom-Olsson, H. Kovarova, Cancer drug-resistance and a look at specific proteins: Rho GDP-dissociation inhibitor 2, Y-box binding protein 1, and HSP70/90 organizing protein in proteomics clinical application, J. Proteome Res. 10 (2011) 404–415, http://dx.doi.org/10.1021/pr100468w.
[18] M. Hajdúch, V. Mihál, J. Minarík, E. Faber, M. Safárová, E. Weigl, P. Antálek, Decreased in vitro chemosensitivity of tumour cells in patients suffering from malignant diseases with poor prognosis, Cytotechnology 19 (1996) 243–245.
[19] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372, http://dx.doi.org/10.1038/nbt.1511.
[20] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2008) 44–57, http://dx.doi.org/10.1038/nprot.2008.211.
[21] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, L.J. Jensen, STRING v9.1: protein-protein interaction networks, with increased coverage and integration, Nucleic Acids Res. 41 (2013) D808–D815, http://dx.doi.org/10.1093/nar/gks1094.
[22] K. Smetana, J. Karban, M. Trneny, To the nucleolar bodies (nucleoli) in cells of the lymphocytic lineage in patients suffering from B - chronic lymphocytic leukemia, Neoplasma 57 (2010) 495–500.
[23] C. Carron, M.-F. O'Donohue, V. Choesmel, M. Faubladier, P.-E. Gleizes, Analysis of two human pre-ribosomal factors, bystin and hTsr1, highlights differences in evolution of ribosome biogenesis between yeast and mammals, Nucleic Acids Res. 39 (2011) 280–291, http://dx.doi.org/10.1093/nar/gkq734.
[24] L. Borkova, S. Gurska, P. Dzubak, R. Burianova, M. Hajduch, J. Sarek, I. Popa, M. Urban, Lupane and 18α-oleanane derivatives substituted in the position 2, their cytotoxicity and influence on cancer cells, Eur. J. Med. Chem. 121 (2016) 120–131, http://dx.doi.org/10.1016/j.ejmech.2016.05.029.
[25] D.C. Dieterich, A.J. Link, J. Graumann, D.A. Tirrell, E.M. Schuman, Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT), Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 9482–9487, http://dx.doi.org/10.1073/pnas.0601637103.
[26] J.A. Vizcaíno, R.G. Côté, A. Csordas, J.A. Dianes, A. Fabregat, J.M. Foster, J. Griss, E. Alpi, M. Birim, J. Contell, G. O'Kelly, A. Schoenegger, D. Ovelleiro, Y. Pérez-Riverol, F. Reisinger, D. Ríos, R. Wang, H. Hermjakob, The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res. 41 (2013) D1063–D1069, http://dx.doi.org/10.1093/nar/gks1262.
[27] I. Grummt, The nucleolus—guardian of cellular homeostasis and genome integrity, Chromosoma 122 (2013) 487–497, http://dx.doi.org/10.1007/s00412-013-0430-0.
[28] H. Ito, H. Yamaguchi, A. Fujikawa, N. Tanaka, A. Furugen, K. Miyamori, N. Takahashi, J. Ogura, M. Kobayashi, T. Yamada, N. Mano, K. Iseki, A full validated hydrophilic interaction liquid chromatography-tandem mass spectrometric method for the quantification of oxaliplatin in human plasma ultrafiltrates, J. Pharm. Biomed. Anal. 71 (2012) 99–103, http://dx.doi.org/10.1016/j.jpba.2012.08.010.
[29] D.C. Henstridge, M. Whitham, M.A. Febbraio, Chaperoning to the metabolic party: the emerging therapeutic role of heat-shock proteins in obesity and type 2 diabetes, Mol. Metab. 3 (2014) 781–793, http://dx.doi.org/10.1016/j.molmet.2014.08.003.
[30] Y. Duan, S. Huang, J. Yang, P. Niu, Z. Gong, X. Liu, L. Xin, R.W. Currie, T. Wu, HspA1A facilitates DNA repair in human bronchial epithelial cells exposed to benzo[a]pyrene and interacts with casein kinase 2, Cell Stress Chaperones 19 (2014) 271–279, http://dx.doi.org/10.1007/s12192-013-0454-7.
[31] D.S.T. Ong, L. Wang, Y. Zhu, B. Ho, J.L. Ding, The response of ferritin to LPS and acute phase of pseudomonas infection, J. Endotoxin Res. 11 (2005) 267–280, http://dx.doi.org/10.1179/096805105X58698.
[32] D. Finazzi, P. Arosio, Biology of ferritin in mammals: an update on iron storage, oxidative damage and neurodegeneration, Arch. Toxicol. 88 (2014) 1787–1802, http://dx.doi.org/10.1007/s00204-014-1329-0.
[33] E. Stanca, G. Serviddio, F. Bellanti, G. Vendemiale, L. Siculella, A.M. Giudetti, Down-regulation of LPCAT expression increases platelet-activating factor level in cirrhotic rat liver: potential antiinflammatory effect of silybin, Biochim. Biophys. Acta 1832 (2013) 2019–2026, http://dx.doi.org/10.1016/j.bbadis.2013.07.005.
[34] S. William-Faltaos, D. Rouillard, P. Lechat, G. Bastian, Cell cycle arrest by oxaliplatin on cancer cells, Fundam. Clin. Pharmacol. 21 (2007) 165–172, http://dx.doi.org/10.1111/j.1472-8206.2007.00462.x.
[35] V.R. Menon, E.J. Peterson, K. Valerie, N.P. Farrell, L.F. Povirk, Ligand modulation of a dinuclear platinum compound leads to mechanistic differences in cell cycle progression and arrest, Biochem. Pharmacol. 86 (2013) 1708–1720, http://dx.doi.org/10.1016/j.bcp.2013.10.012.
[36] K.S. Lee, D.-Y. Oh, Y.H. Kang, J.-E. Park, Self-regulated mechanism of Plk1 localization to kinetochores: lessons from the Plk1-PBIP1 interaction, Cell Div 3 (2008) 4, http://dx.doi.org/10.1186/1747-1028-3-4.
[37] T. Raemaekers, K. Ribbeck, J. Beaudouin, W. Annaert, M. Van Camp, I. Stockmans, N. Smets, R. Bouillon, J. Ellenberg, G. Carmeliet, NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization, J. Cell Biol. 162 (2003) 1017–1029, http://dx.doi.org/10.1083/jcb.200302129.
[38] T. Kruse, M.S.Y. Larsen, G.G. Sedgwick, J.O. Sigurdsson, W. Streicher, J.V. Olsen, J. Nilsson, A direct role of Mad1 in the spindle assembly checkpoint beyond Mad2 kinetochore recruitment, EMBO Rep. 15 (2014) 282–290, http://dx.doi.org/10.1002/embr.201338101.
[39] X. Ge, C.L. Frank, F. Calderon de Anda, L.-H. Tsai, Hook3 interacts with PCM1 to regulate pericentriolar material assembly and the timing of neurogenesis, Neuron 65 (2010) 191–203, http://dx.doi.org/10.1016/j.neuron.2010.01.011.
[40] J.E. Lee, J.L. Silhavy, M.S. Zaki, J. Schroth, S.L. Bielas, S.E. Marsh, J. Olvera, F. Brancati, M. Iannicelli, K. Ikegami, A.M. Schlossman, B. Merriman, T. Attié-Bitach, C.V. Logan, I.A. Glass, A. Cluckey, C.M. Louie, J.H. Lee, H.R. Raynes, I. Rapin, I.P. Castroviejo, M. Setou, C. Barbot, E. Boltshauser, S.F. Nelson, F. Hildebrandt, C.A. Johnson, D.A. Doherty, E.M. Valente, J.G. Gleeson, CEP41 is mutated in Joubert syndrome and is required for tubulin glutamylation at the cilium, Nat. Genet. 44 (2012) 193–199, http://dx.doi.org/10.1038/ng.1078.
[41] A.M. Waters, R. Asfahani, P. Carroll, L. Bicknell, F. Lescai, A. Bright, E. Chanudet, A. Brooks, S. Christou-Savina, G. Osman, P. Walsh, C. Bacchelli, A. Chapgier, B. Vernay, D.M. Bader, C. Deshpande, M. O'Sullivan, L. Ocaka, H. Stanescu, H.S. Stewart, F. Hildebrandt, E. Otto, C.A. Johnson, K. Szymanska, N. Katsanis, E. Davis, R. Kleta, M. Hubank, S. Doxsey, A. Jackson, E. Stupka, M. Winey, P.L. Beales, The kinetochore protein, CENPF, is mutated in human ciliopathy and microcephaly phenotypes, J. Med. Genet. 52 (2015) 147–156, http://dx.doi.org/10.1136/jmedgenet-2014-102691.
[42] P. Noordhuis, A.C. Laan, K. van de Born, N. Losekoot, I. Kathmann, G.J. Peters, Oxaliplatin activity in selected and unselected human ovarian and colorectal cancer cell lines, Biochem. Pharmacol. 76 (2008) 53–61, http://dx.doi.org/10.1016/j.bcp.2008.04.007.
[43] H. Liu, F. Zheng, Q. Cao, B. Ren, L. Zhu, G. Striker, H. Vlassara, Amelioration of oxidant stress by the defensin lysozyme, Am. J. Physiol. Endocrinol. Metab. 290 (2006) E824–E832, http://dx.doi.org/10.1152/ajpendo.00349.2005.
[44] K. Pai, A. Sodhi, Effect of cisplatin, rIFN-Y, LPS and MDP on release of H2O2, O2- and lysozyme from human monocytes in vitro, Indian J. Exp. Biol. 29 (1991) 910–915.
[45] L. Kadaja, S. Laos, T. Maimets, Overexpression of leukocyte marker CD43 causes activation of the tumor suppressor proteins p53 and ARF, Oncogene 23 (2004) 2523–2530, http://dx.doi.org/10.1038/sj.onc.1207359.
[46] D.R. Flower, A.C. North, T.K. Attwood, Structure and sequence relationships in the lipocalins and related proteins, Protein Sci. Publ. Protein Soc. 2 (1993) 753–761, http://dx.doi.org/10.1002/pro.5560020507.
[47] J. Mishra, K. Mori, Q. Ma, C. Kelly, J. Barasch, P. Devarajan, Neutrophil gelatinase-associated lipocalin: a novel early urinary biomarker for cisplatin nephrotoxicity, Am. J. Nephrol. 24 (2004) 307–315, http://dx.doi.org/10.1159/000078452.
[48] M.I. Hassan, A. Waheed, S. Yadav, T.P. Singh, F. Ahmad, Zinc α2-glycoprotein: a multidisciplinary protein, Mol. Cancer Res. 6 (2008) 892–906, http://dx.doi.org/10.1158/1541-7786.MCR-07-2195.
[49] M.I. Hassan, A. Waheed, S. Yadav, T.P. Singh, F. Ahmad, Prolactin inducible protein in cancer, fertility and immunoregulation: structure, function and its clinical implications, Cell. Mol. Life Sci. 66 (2009) 447–459, http://dx.doi.org/10.1007/s00018-008-8463-x.
[50] Q. Quan, Q. Chen, P. Chen, L. Jiang, T. Li, H. Qiu, B. Zhang, Decreased apolipoprotein A-I level indicates poor prognosis in extranodal natural killer/T-cell lymphoma, nasal type, OncoTargets Ther. 9 (2016) 1281–1290, http://dx.doi.org/10.2147/OTT.S96549.
[51] T. Cheng, X. Dai, D.-L. Zhou, Y. Lv, L.-Y. Miao, Correlation of apolipoprotein A-I kinetics with survival and response to first-line platinum-based chemotherapy in advanced non-small cell lung cancer, Med. Oncol. Northwood Lond. Engl. 32 (2015) 407, http://dx.doi.org/10.1007/s12032-014-0407-8.
[52] M. Zamanian-Daryoush, J.A. DiDonato, Apolipoprotein A-I and cancer, Front. Pharmacol. 6 (2015)http://dx.doi.org/10.3389/fphar.2015.00265.
[53] M. Sattari, Y. Pazhang, M. Imani, Calprotectin induces cell death in human prostate cancer cell (LNCaP) through surviving protein alteration, Cell Biol. Int. 38 (2014) 1311–1320, http://dx.doi.org/10.1002/cbin.10328.

[54] E. Nicolas, P. Parisot, C. Pinto-Monteiro, R. de Walque, C. De Vleeschouwer, D.L.J. Lafontaine, Involvement of human ribosomal proteins in nucleolar structure and p53-dependent nucleolar stress, Nat. Commun. 7 (2016) 11390, http://dx.doi.org/10.1038/ncomms11390.
[55] K. Burger, B. Mühl, T. Harasim, M. Rohrmoser, A. Malamoussi, M. Orban, M. Kellner, A. Gruber-Eber, E. Kremmer, M. Hölzel, D. Eick, Chemotherapeutic drugs inhibit ribosome biogenesis at various levels, J. Biol. Chem. 285 (2010) 12416–12425, http://dx.doi.org/10.1074/jbc.M109.074211.
[56] G. Samimi, G. Manorek, R. Castel, J.K. Breaux, T.C. Cheng, C.C. Berry, G. Los, S.B. Howell, cDNA microarray-based identification of genes and pathways associated with oxaliplatin resistance, Cancer Chemother. Pharmacol. 55 (2005) 1–11, http://dx.doi.org/10.1007/s00280-004-0819-9.
[57] M. Horky, G. Wurzer, V. Kotala, M. Anton, B. Vojtesek, J. Vacha, J. Wesierska-Gadek, Segregation of nucleolar components coincides with caspase-3 activation in cisplatin-treated HeLa cells, J. Cell Sci. 114 (2001) 663–670.
[58] S.M.F. Jamieson, J. Liu, T. Hsu, B.C. Baguley, M.J. McKeage, Paclitaxel induces nucleolar enlargement in dorsal root ganglion neurons in vivo reducing oxaliplatin toxicity, Br. J. Cancer 88 (2003) 1942–1947, http://dx.doi.org/10.1038/sj.bjc.6601012.
[59] C.L. Renn, V.A. Carozzi, P. Rhee, D. Gallop, S.G. Dorsey, G. Cavaletti, Multimodal assessment of painful peripheral neuropathy induced by chronic oxaliplatin-based chemotherapy in mice, Mol. Pain 7 (2011) 29, http://dx.doi.org/10.1186/1744-8069-7-29.
[60] K. Wada, M. Sato, N. Araki, M. Kumeta, Y. Hirai, K. Takeyasu, K. Furukawa, T. Horigome, Dynamics of WD-repeat containing proteins in SSU processome components, Biochem. Cell Biol. 92 (2014) 191–199, http://dx.doi.org/10.1139/bcb-2014-0007.
[61] M. Kornprobst, M. Turk, N. Kellner, J. Cheng, D. Flemming, I. Koš-Braun, M. Koš, M. Thoms, O. Beminghausen, R. Beckmann, E. Hurt, Architecture of the 90S pre-ribosome: a structural view on the birth of the eukaryotic ribosome, Cell 166 (2016) 380–393, http://dx.doi.org/10.1016/j.cell.2016.06.014.
[62] M. Kellner, M. Rohrmoser, I. Forné, K. Voss, K. Burger, B. Mühl, A. Gruber-Eber, E. Kremmer, A. Imhof, D. Eick, DEAD-box helicase DDX27 regulates 3′ end formation of ribosomal 47S RNA and stably associates with the PeBoW-complex, Exp. Cell Res. 334 (2015) 146–159, http://dx.doi.org/10.1016/j.yexcr.2015.03.017.
[63] K. Shimoji, J. Jakovljevic, K. Tsuchihashi, Y. Umeki, K. Wan, S. Kawasaki, J. Talkish, J.L. Woolford, K. Mizuta, Ebp2 and Brx1 function cooperatively in 60S ribosomal subunit assembly in *Saccharomyces cerevisiae*, Nucleic Acids Res. 40 (2012) 4574–4588, http://dx.doi.org/10.1093/nar/gks057.
[64] J. Nousbeck, R. Spiegel, A. Ishida-Yamamoto, M. Indelman, A. Shani-Adir, N. Adir, E. Lipkin, S. Bercovici, D. Geiger, M.A. van Steensel, P.M. Steijlen, R. Bergman, A. Bindereif, M. Choder, S. Shalev, E. Sprecher, Alopecia, neurological defects, and endocrinopathy syndrome caused by decreased expression of RBM28, a nucleolar protein associated with ribosome biogenesis, Am. J. Hum. Genet. 82 (2008) 1114–1121, http://dx.doi.org/10.1016/j.ajhg.2008.03.014.
[65] K. Saito, Y. Iizuka, S. Ohta, S. Takahashi, K. Nakamura, H. Saya, K. Yoshida, Y. Kawakami, M. Toda, Functional analysis of a novel glioma antigen, EFTUD1, Neuro-Oncology 16 (2014) 1618–1629, http://dx.doi.org/10.1093/neuonc/nou132.
[66] S. Wu, B. Tutuncuoglu, K. Yan, H. Brown, Y. Zhang, D. Tan, M. Gamalinda, Y. Yuan, Z. Li, J. Jakovljevic, C. Ma, J. Lei, M.-Q. Dong, J.L. Woolford, N. Gao, Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes, Nature 534 (2016) 133–137, http://dx.doi.org/10.1038/nature17942.
[67] C. Zhang, C. Yin, L. Wang, S. Zhang, Y. Qian, J. Ma, Z. Zhang, Y. Xu, S. Liu, HSPC111 governs breast cancer growth by regulating ribosomal biogenesis, Mol. Cancer Res. 12 (2014) 583–594, http://dx.doi.org/10.1158/1541-7786.MCR-13-0168.
[68] Z. Xu, T.C. Hobman, The helicase activity of DDX56 is required for its role in assembly of infectious West Nile virus particles, Virology 433 (2012) 226–235, http://dx.doi.org/10.1016/j.virol.2012.08.011.
[69] T. Kumazawa, K. Nishimura, N. Katagiri, S. Hashimoto, Y. Hayashi, K. Kimura, Gradual reduction in rRNA transcription triggers p53 acetylation and apoptosis via MYBBP1A, Sci. Rep. 5 (2015)http://dx.doi.org/10.1038/srep10854.
[70] T. de Lange, How shelterin solves the telomere end-protection problem, Cold Spring Harb. Symp. Quant. Biol. 75 (2010) 167–177, http://dx.doi.org/10.1101/sqb.2010.75.017.
[71] A.-G. Lu, H. Feng, P.-X.-Z. Wang, D.-P. Han, X.-H. Chen, M.-H. Zheng, Emerging roles of the ribonucleotide reductase M2 in colorectal cancer and ultraviolet-induced DNA damage repair, World J. Gastroenterol. 18 (2012) 4704–4713, http://dx.doi.org/10.3748/wjg.v18.i34.4704.
[72] K. Yano, K. Morotomi-Yano, N. Adachi, H. Akiyama, Molecular mechanism of protein assembly on DNA double-strand breaks in the non-homologous end-joining pathway, J. Radiat. Res. 50 (2009) 97–108 (Tokyo).
[73] S. Kotian, T. Banerjee, A. Lockhart, K. Huang, U.V. Catalyurek, J.D. Parvin, NUSAP1 influences the DNA damage response by controlling BRCA1 protein levels, Cancer Biol. Ther. 15 (2014) 533–543, http://dx.doi.org/10.4161/cbt.28019.
[74] J.-Y. Kim, K.-O. Seok, Y.-J. Kim, W.K. Bae, S. Lee, J.-H. Park, Involvement of GLTSCR2 in the DNA damage response, Am. J. Pathol. 179 (2011) 1257–1264, http://dx.doi.org/10.1016/j.ajpath.2011.05.041.
[75] H.-Y. Chang, C.-C. Fan, P.-C. Chu, B.-E. Hong, H.J. Lee, M.-S. Chang, hPuf-A/KIAA0020 modulates PARP-1 cleavage upon genotoxic stress, Cancer Res. 71 (2011) 1126–1134, http://dx.doi.org/10.1158/0008-5472.CAN-10-1831.
[76] M. Sokka, K. Rilla, I. Miinalainen, H. Pospiech, J.E. Syväoja, High levels of TopBP1 induce ATR-dependent shut-down of rRNA transcription and nucleolar segregation, Nucleic Acids Res. 43 (2015) 4975–4989, http://dx.doi.org/10.1093/nar/gkv371.
[77] A. James, Y. Wang, H. Raje, R. Rosby, P. DiMario, Nucleolar stress with and without p53, Nucleus 5 (2014) 402–426, http://dx.doi.org/10.4161/nucl.32235.
[78] L. Montanaro, D. Treré, M. Derenzini, Nucleolus, ribosomes, and cancer, Am. J. Pathol. 173 (2008) 301–310, http://dx.doi.org/10.2353/ajpath.2008.070752.
[79] J. Li, L. Yu, H. Zhang, J. Wu, J. Yuan, X. Li, M. Li, Down-regulation of pescadillo inhibits proliferation and tumorigenicity of breast cancer cells, Cancer Sci. 100 (2009) 2255–2260, http://dx.doi.org/10.1111/j.1349-7006.2009.01325.x.
[80] T. Maehama, K. Kawahara, M. Nishio, A. Suzuki, K. Hanada, Nucleolar stress induces ubiquitination-independent proteasomal degradation of PICT1 protein, J. Biol. Chem. 289 (2014) 20802–20812, http://dx.doi.org/10.1074/jbc.M114.571893.
[81] E. Hammer, S. Bien, M.G. Salazar, L. Steil, C. Scharf, P. Hildebrandt, H.W.S. Schroeder, H.K. Kroemer, U. Völker, C.A. Ritter, Proteomic analysis of doxorubicin-induced changes in the proteome of HepG2cells combining 2-D DIGE and LC-MS/MS approaches, Proteomics 10 (2010) 99–114, http://dx.doi.org/10.1002/pmic.200800626.
[82] R. Briffa, I. Um, D. Faratian, Y. Zhou, A.K. Turnbull, S.P. Langdon, D.J. Harrison, Multiscale genomic, transcriptomic and proteomic analysis of colorectal cancer cell lines to identify novel biomarkers, PLoS One 10 (2015), e0144708. http://dx.doi.org/10.1371/journal.pone.0144708.
[83] S. Boulon, B.J. Westman, S. Hutten, F.-M. Boisvert, A.I. Lamond, The nucleolus under stress, Mol. Cell 40 (2010) 216–227, http://dx.doi.org/10.1016/j.molcel.2010.09.024.
[84] P. Jordan, M. Carmo-Fonseca, Cisplatin inhibits synthesis of ribosomal RNA in vivo, Nucleic Acids Res. 26 (1998) 2831–2836, http://dx.doi.org/10.1093/nar/26.12.2831.
[85] M. Pietrzak, S.C. Smith, J.T. Geralds, T. Hagg, C. Gomes, M. Hetman, Nucleolar disruption and apoptosis are distinct neuronal responses to etoposide-induced DNA damage, J. Neurochem. 117 (2011) 1033–1046, http://dx.doi.org/10.1111/j.1471-4159.2011.07279.x.
[86] J.E. Quin, J.R. Devlin, D. Cameron, K.M. Hannan, R.B. Pearson, R.D. Hannan, Targeting the nucleolus for cancer intervention, Biochim. Biophys. Acta 1842 (2014) 802–816, http://dx.doi.org/10.1016/j.bbadis.2013.12.009.
[87] K. Smetana, A. Vlastiborová, I. Iscenko, Studies on micronucleoli of immature human leukemic neutrophils, Neoplasma 20 (1973) 491–498.
[88] N.J. Lane, Spheroidal and ring nucleoli in amphibian oocytes. Patterns of uridine incorporation and fine structural features, J. Cell Biol. 35 (1967) 421–434.
[89] P. Kasík, J. Lejnar, M. Hrobon, K. Smetana, Nucleoli in human pseudostratified columnar epithelium cells. (Microscopic classification of nucleoli), Folia Biol. 23 (1977) 354–358 (Praha).