UNIVERSITÀ DI PARMA

UNIVERSITY OF PARMA

Ph.D. IN BIOTECHNOLOGY AND LIFE SCIENCES

CYCLE XXXIV

Applications of mass spectrometry-based proteomics to characterize preclinical animal models

Coordinator:
Professor Marco Ventura

Tutor:
Professor Barbara Montanini

PhD Student: Maria Laura Faietti

Academic years 2018/2019 – 2020/2021

ABSTRACT

Proteomics, regarded as the comprehensive study of the gene expression at protein level at a particular time in different organs, tissues and cell types is a key enabling technology for the systems biology approach.

The aim of this Ph.D. work is to establish quantification and discovery approaches to measure and study proteins involved in the pulmonary system and associated models of disease to be implemented in drug discovery processes. Progression or establishment of a pulmonary disease, normal lung development and different pharmacodynamic preclinical models were taken into account to explore the applicability of mass spectrometry-based proteomics to identify new key molecular players -or to confirm the existing ones-, potential biomarkers and to help drug target prioritization. To achieve this aim, a comprehensive proteomic workflow, involving protein extraction protocols optimization, chromatographic and mass spectrometry methods and bioinformatic analysis setup, has been optimized. The whole work has been performed in collaboration with Chiesi Farmaceutici S.p.A.

Both bottom up targeted and discovery proteomics approaches, were considered precious tools to quantify target proteins through targeted experiments (MRM), and to lead to high-throughput discovery datasets, respectively. Both platforms were applied to answer to different key questions on pulmonary disease preclinical models.

Targeted proteomics allows to focus on specific proteins, based on proteotypic peptide detection. With the accurate selection of proteotypic peptides, it is possible to isolate and quantify a specific protein isoform, or to detect post-translational modifications on peptides without enrichment protocols. In this work, as proof of concept, three examples of targeted proteomics are described, applied to established biomarkers of fibrosis (fibrillar collagen and α-sma) and applied to pharmacodynamic study of a known HDACs inhibitor. These methods required and optimization of protein extraction protocol from lung, which has peculiar characteristics specific of the fibrotic tissue, and of protein enrichment.

Fibrillar collagen detection in fibrotic lung: a targeted method was developed to quantify collagen type I and collagen type III in rats treated with bleomycin with the focus on a time course analysis. A specific sample preparation was improved to enrich and digest extracellular matrix proteins. An MRM method was set up to investigate proteotypic peptides of the alpha chains of fibrillar collagens. The results confirm the knowledge about bleomycin-induced pulmonary fibrosis.

Detection of α-sma in fibrotic lung: a targeted method was developed to investigate this isoform of actin, selecting proteotypic peptides unique for α-sma different from other actin isoforms. MRM method was set

up and the results show that actin is not overexpressed at 28 days after bleomycin treatments, results supported by literature knowledge about the early expression in the bleomycin model.

Quantitation of acetylation on histone proteins after treatment with HDACs inhibitor: a new targeted method was developed for this aim. Acetylated peptides and unmodified peptides of Histone H3 and H4 were detected in the MRM method, after an enrichment protocol for histone proteins. The results show a reduction of Histone acetylation from 1 hour to 72 hours after drug administration, confirming the effect of the pan-HDACs inhibitors over time.

Discovery proteomics allows to identify and quantify thousands of proteins within a sample. Bioinformatic analysis leads to a deep understanding of the biological meaning and the processes involved. Two methods of quantitative high-throughput proteomics were studied and tested in this work.

Proteome profiling through TMT quantitative proteomics: the quantitative labeling technique was used to characterize two different experiments, (1) a bleomycin-induced pulmonary fibrosis experiment in rat at 28 days, testing Nintendanib, the approved drug for pulmonary fibrosis; and (2) the characterization of lung developmental stages studied in a rabbit model, from canalicular stage to alveolar stage. Quantitative proteomics were coupled with bioinformatic analysis, in order to understand the biological processes involved in the system.

Investigation of nintedanib mode of action through phosphoproteomics: As known, protein kinases and their substrates have been gaining increasing attention as therapeutic targets for the treatment of cancer and chronic inflammatory diseases. A specific sample preparation, based on enrichment of phosphopeptides, was employed to study the phosphoproteome of a multikinase inhibitor (Nintedanib), and three different bioinformatic approaches were applied to investigate the quantified phosphopeptides/proteins and to evaluate the mode of action of Nintedanib.

TABLE OF CONTENTS

Introduction

Proteome and Proteomics

In 1838 Jöns Jacob Berzelius defined molecules that have a fundamental role in supporting life as “proteins”, from the ancient Greek word “proteios”, meaning “in the first place”. Each protein is characterized by sequence, localization, interaction, post-translational modifications, and half-life. Proteins catalyse, regulate and coordinate biological processes in every living cell.

The proteome is the global protein content of a cell, a tissue or an organism. In 1996 Marc Wilkins used the term “proteomics” to describe the “protein complement of a genome”[1]. The proteome of eukaryotic cells is complex and constitutes about 50% of cell’s dry mass. The proteome network responds dynamically to stimuli, either external or internal, leading to define the functional state and role of the cell[2]. Protein levels not only depend on the corresponding mRNA, but also on translational control and regulation. Understanding the dynamic, the function and the structure of proteomes is a crucial challenge in biology.

Proteomics is the science that study the dynamic roles and effects of proteins in cells as well as in tissues, with the primary aim of generating biological insights. Different proteomics techniques have been developed over the years, and could be organized in:

- Conventional techniques, used for the analysis of few individual proteins, such as western blotting, ELISA or chromatography-based techniques;
- Advanced techniques, for high throughput and expression analysis of protein mixtures: protein microarray, gel-based approaches and mass spectrometry-based proteomics;
- Quantitative techniques based on tag labelling for complex samples (cells cultures, tissues, blood) , such as affinity tag (ICAT), stable isotope labelling (SILAC) and isobaric tag for relative and also absolute quantitation (iTRAQ/TMT)[3].

Mass spectrometry-based proteomics

Nowadays, powerful MS-based technologies grant identification and measurement of most proteins in order to collect understandings in the composition, control and function of a proteome.

Mass spectrometry is successfully used in proteomics because, measuring the mass to charge ratio (m/z), the molecular weight of proteins or peptides can be determined. Starting with the specific identification of each protein, mass-spectrometry-based techniques allow to quantify thousands of proteins across complex samples with a high degree of reproducibility.

Proteins can be analysed as intact entities with top-down proteomics; this technique is used to characterize the proteoforms of proteins, identifying all the modifications of the molecules. Instead, the preferred and the most widely used method to characterize a proteome is defined bottom-up proteomics and is used to identify and quantify the peptides derived from complex samples.

Bottom-up proteomics

In bottom-up proteomics experiments, a complex mixture of proteins is typically digested with lysin-C or more frequently with trypsin. This phase is followed by multidimensional high-performance liquid chromatography (HPLC), online coupled to the mass spectrometer, in order to separate the complex mixture of peptides.

A proteomic experiment starts from the lysis and denaturation of tissues or cell cultures to generate a protein extract. It is necessary to chemically control the reactive cysteines at the earliest possible stage to prevent mixtures of different cysteine modifications. Usually, the protection is done by reduction with dithiothreitol and alkylation with iodoacetamide. The proteolytic cleavage is most often performed with trypsin, a serine protease that specifically cleaves at the carboxylic side of lysine and arginine residues if these are not followed by proline[4].

The resulting peptides mixture is separated with reverse-phase chromatography and the solution is injected in the source of MS and ionized through electrospray ionization (ESI). Depending on the aim of the experiment different types of mass spectrometer can be used for the analysis (Figure 1).

Discovery proteomics is commonly achieved with data-dependent analysis (DDA) performed with quadrupole-orbitrap analysers and is aimed at reaching a global proteome characterization.

Targeted proteomics goal, instead, is to identify and quantify proteins of interest, that are predetermined and known, and the preferred analyser for this purpose is the triple quadrupole.

The choice of the mass analyser, or the proteomic method, is strictly related to the type of preclinical experiment performed and to the information to be obtained.

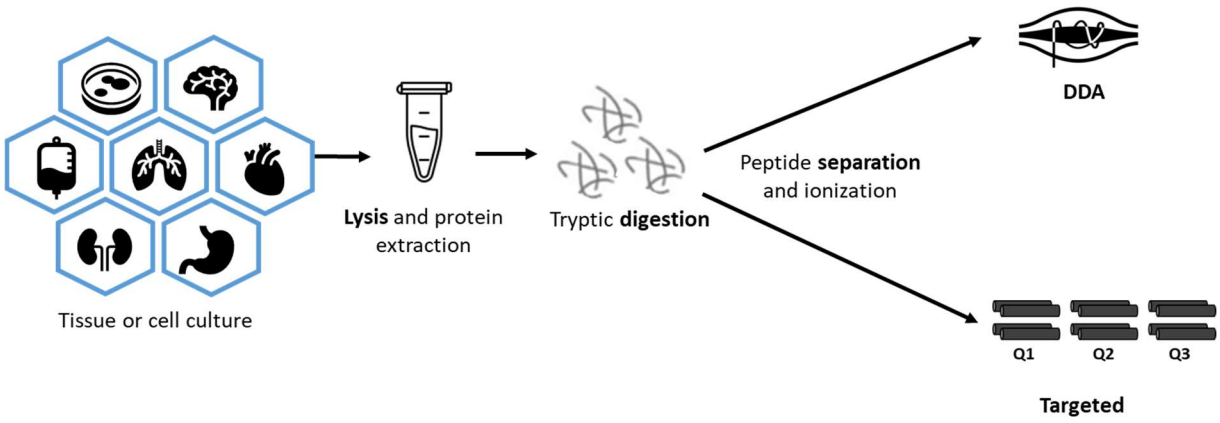


Figure 1: Bottom-up proteomics workflow

Discovery quantitative proteomics

A global discovery study to investigate proteins in complex mixture is usually performed with a discovery proteomics experiment, based on data-dependent acquisition methods (DDA). The mass analyser used is orbitrap: the peptides are recorded at the full-scan level (MS1), and then fragmented and isolated at MS2 (or

MS/MS) level. The mass spectrometer alternates the acquisition of one full-scan spectrum and about ten fragment-ion scans.

The Orbitrap mass analyser was invented by Makarov in 1999: following the injection, ions are trapped, orbit around a spindle-shaped electrode and oscillate along its axis with a frequency specific for their m/z; then the signal is Fourier transformed to yield high resolution mass spectra. This mass analyser owns the highest mass resolving power and the best mass accuracy among other mass analysers[5].

The high resolution of this type of instruments allow to identify with high confidence thousands of peptides, and consequently to identify and quantify thousands of proteins in complex mixtures within a single experiment. Diverse quantitative techniques have been developed during years, aimed to get increasingly resourceful, user-friendly, robust, and sensitive[6].

Label-free quantitation can be obtained by spectral counting, sum of MS/MS fragment ions, extraction of corresponding parent ion intensity from MS scans or a combination of the three. Consistency of the experimental design between samples is essential for this type of study. Variations could be due to the reproducibility of the LC system, spray stability, variation in chemical modification of amino acids introduced during sample preparation (such as methionine oxidation), and technical variance from trypsin digestion. The sample is simply digested and the quantitation is performed by specific software with detailed algorithms to normalize the quantitative data.

Another group of quantitative methods are stable isotope labelling strategies. These approaches can be divided in metabolic and chemical labelling. The metabolic labelling (SILAC) has the advantage to be incorporated in the samples at the beginning of experiment, during cell growth. The chemical labelling (iTRAQ or TMT), instead, is incorporated after protein digestion: this strategy is more convenient for clinical or preclinical samples.

TMT approach is the labeling with tandem mass tags, which consist of an amine-reactive, a balance and a reporter group. The reporter group is released upon fragmentation during MS/MS, and the intensity is used to calculate relative peptide amounts between samples [7]. The incorporation of differential stable isotopes ($_{15}$N, $_{13}$C) in the reporter and mass balancing regions of the tag has improved the multiplexing capability of TMT reagents (6-plex, 10-plex and 16-plex sets) for relative quantitative profiling of multiple conditions[8].

With this labeling strategy, digested peptides from multiple samples are first labeled in parallel with different tag variants (TMT). They are then mixed and analyzed using reversed phase high performance liquid chromatograph (HPLC) coupled with a mass spectrometer capable of MS/MS analysis.

TMT makes it possible to multiplex the analysis, to efficiently use the instrument time and exert further controls for technical variation.

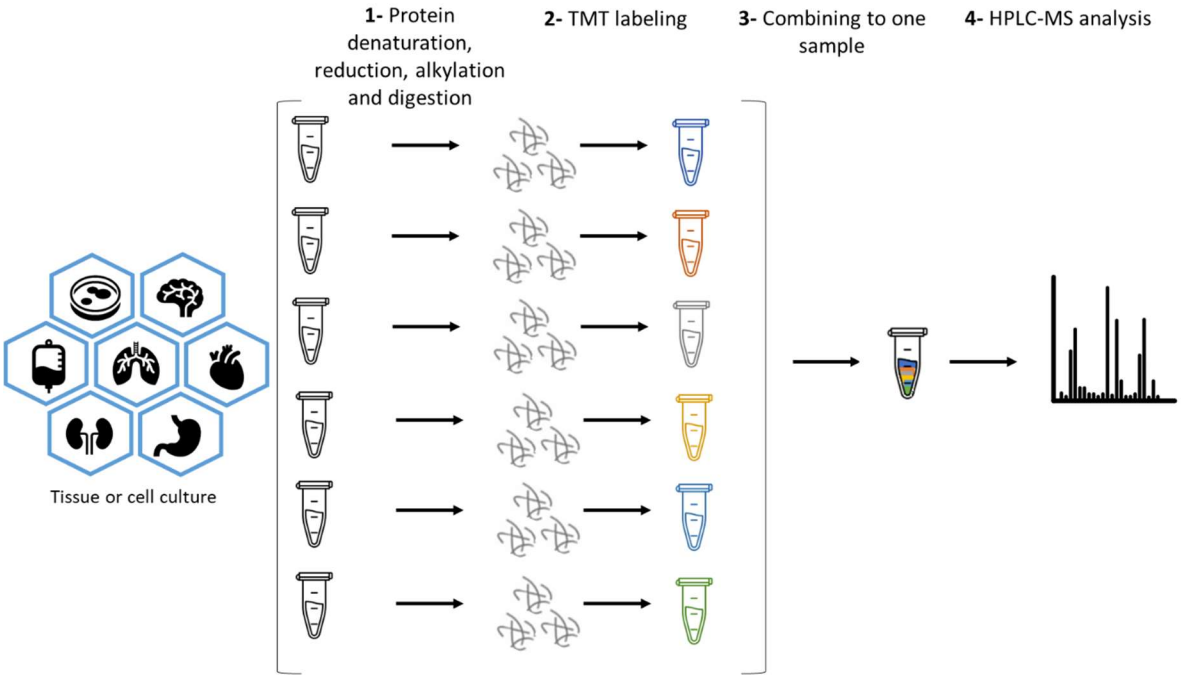


Figure 2: TMT labeling workflow

Targeted analysis

By using pre-existing information about proteins of interest, it is possible to perform a targeted analysis. After chromatographic separation, characteristic peptides, termed proteotypic peptides, are selectively isolated, and fragmented. The analytical method is usually set on a triple quadrupole instrument.

Among the multiple modes of operation of triple quadrupole instruments emerges the Multiple Reaction Monitoring, or MRM mode. This analysis allows to confirm the presence of specific proteins in a biological sample: peptides belonging to the protein of interest and having a unique sequence are selected and investigated as surrogates for the parent protein. The chosen surrogate peptide is found using mass analyzers with high resolution, like orbitrap.

The targeted analysis is possible because the first quadrupole selects the ion of the targeted peptide, which is subsequently fragmented in a collision cell. Specified fragment ions of the peptides are then detected by the third quadrupole (Figure 3). Only the specified transitions, i.e. the m/z relative to the fragment ions, are recorder, so other product ions are not detected[9].

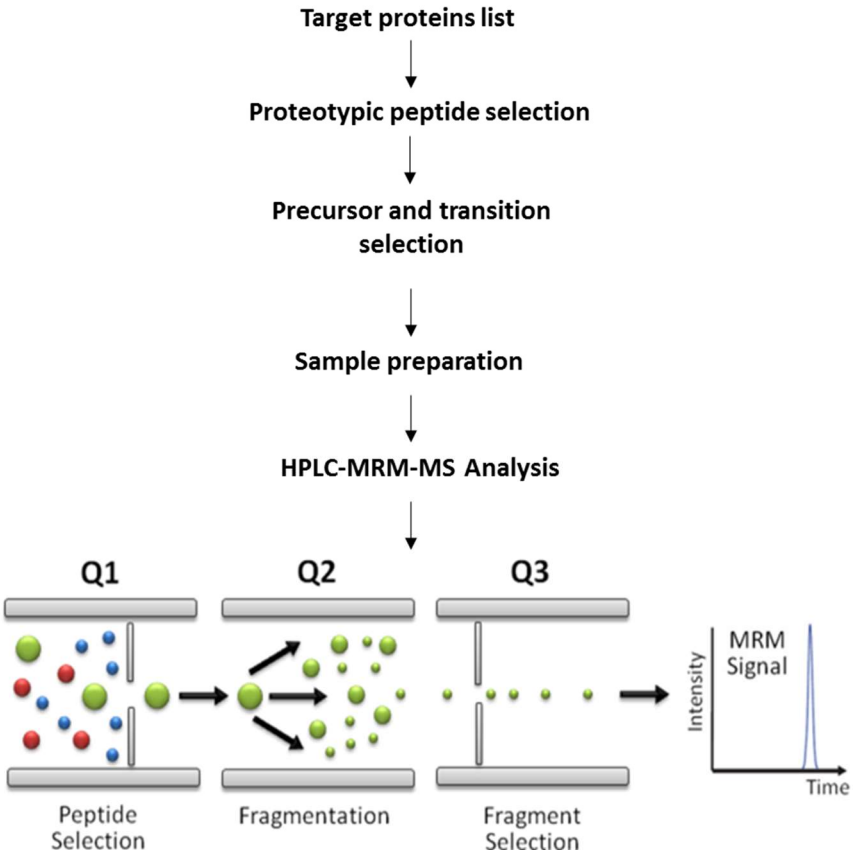


Figure 3:Workflow for the design of an MRM experiment

A limit of this analysis, dependent on the low resolution of quadrupole analyzers, is the possible interference of nominally isobaric peptides, which can exist in complex biological mixtures.

The type of peptide fragment ions depends on various factors, like CID (collision-induced dissociation) and metastable decay. With low-energy CID, fragmentation of peptides occurs at the peptide amide bond, obtaining two types of fragment: one preserves the N-terminus of the peptide (*b* ions), the other conserves the C-terminus (*y* ions) (Figure 4).

This fragmentation, unique for every peptide, allows the identification of proteins in a biological sample, using *ad hoc* search programs, based on protein databases.

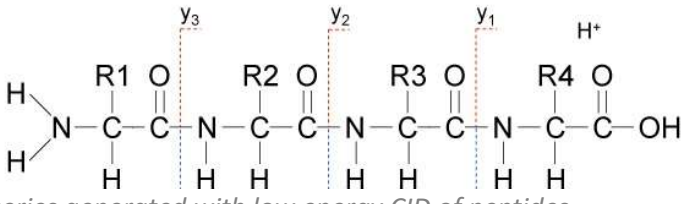


Figure 4: Principal ion series generated with low energy CID of peptides.

Post-translational modification characterization

Mass-spectrometry-based proteomics is widely used to analyse post-translational modifications (PTMs), because of the high sensitivity of the instrument. The difference in mass of the modification is detected and measured with a single amino acid resolution, through the peptide fragments.

To improve the identification of post-translational modifications, new workflows for sample preparation have been developed. The key step is the enrichment of the proteins or the peptides of interest[2].

The most studied PTMs are phosphorylation, ubiquitylation, acetylation and glycosylation. Each of these modifications regulates signalling events or interactions, essential for living cells. Modifications can also lead to changes in enzymatic activity or alterations in protein conformation, localization or lifetime (i.e. ubiquitylation).

The study of post-translational modifications in proteomes is leading to a deeper knowledge of cell behaviours and to new biological insights.

Bioinformatic analysis of proteomic data

Advances in mass spectrometry-based proteomics have generated large volumes of data, but a key challenge lies in data analysis, in order to generate new biological insight: discoveries about the regulation of protein functions and, consequently, about cellular mechanisms can help to understand how the proteome is perturbated in specific diseases and expand the knowledge about mechanisms of establishment and development of human diseases.

Following the rapid improvements in sample preparation methods and mass spectrometry techniques for proteomics, downstream bioinformatic analyses have been developed that allow to understand regulatory mechanisms, cellular behaviour, and relationships between proteins.

Bioinformatic analysis of proteomic data starts with the determination of the peptide sequences in order to reconstruct the sequence of the proteins, then moves on to the quantification of the identified proteins, and finally allows to identify biological pathways and protein networks.

Different tools are available to perform identification and quantification of proteins in complex biological samples, among which the most known are Proteome Discoverer and MaxQuant[10,11].

Despite their differences, the procedure to identify and quantify proteins from MS spectra with these bioinformatic tools always starts with the search against a fragmentation spectra database[12].

This target database is established from the in-silico digestion of all protein sequences in the animal model of interest. Then a peptide spectrum match (PSM) score is calculated for each fragmentation spectrum of the

target database against the query peptide. The *in-silico* digested peptide that has the highest PSM score is used as a candidate for the query peptide.

The key for a reliable database searching is choosing appropriate input parameters. The ppm and Dalton tolerances (for precursor mass and fragment mass, respectively) are important parameters, which depend on both mass spectrometer and analyser used. Other essential parameters are the minimum and maximum established length of the peptide, the proteolytic enzyme used and the maximum allowed number of missed cleavages. Not least is searching for post-translational modifications, identifiable by the specific and unique mass of the chemical group linked to an amino acid residue.

Once peptide identification is completed, the next step is protein inference, i.e. the reconstruction of the original proteins from peptide sequences. Longer peptides and unique peptides are very informative in this step due to their specificity. To increase the consistency of the inference, it is important to verify that protein abundance is derived from at least two unique peptides.

After protein identification, bioinformatic tools perform the quantification step. Experimental quantification methods belong to two categories: labelled methods and label-free methods, depending on the sample preparation protocol used. Specific algorithms integrate peak area and normalize protein abundances.

Normalization is a necessary step in data correction, helpful for removing any non-biological variation and obtaining consistent results, as well as for alignment between samples prior to downstream analysis. Several types of normalization methods have been developed based on different statistical assumptions[13].

Once the protein abundance data have gone through filtering, cleaning and normalization, further statistical analysis and downstream studies are performed.

The most common statistical analysis consists in examining whether there are significant changes in protein levels between two different conditions. This is typically performed with a t-test between the protein abundances observed in two different groups. Many tools help in this analysis, such as Proteome Discoverer, which has statistical analyses integrated in the data processing workflow, or Perseus, a platform to interpret protein quantification[14] that has specific functions to perform different statistical analyses (t-test, ANOVA, Principal component analysis). Based on selected cut off values for fold change and p-value (or adjusted p-value), is possible to define differentially expressed proteins. Several clustering methods (hierarchical clustering, K mean, WGCNA) are also suited to investigate changes in protein levels, mainly when several samples are analysed together.

Starting from lists of differentially expressed proteins or of proteins deriving from clusters of interest, it is possible to go deeper into the understanding of biological issues. Enrichment analysis categorizes genes or proteins that are overrepresented in the original protein list into predefined gene sets of interest, such as

pathways or gene ontology classes. The advantage of performing enrichment analysis on proteomic data over transcriptomic data is that the systematic proteome profiling can lead to a deeper understanding of protein-protein interactions and post-translational modifications, and consequently of biological processes and signalling networks.

The aim of studying signalling pathways is interpreting the relationships between enzymes and their substrates. Particularly interesting are the physical interactions between kinases and their targets, because they are essential signalling molecules, frequently associated with diseases such as cancers or degenerative diseases. In phosphoproteomics, the information about this PTM can help predict the direction of the nodes and edges in the phosphorylation network.

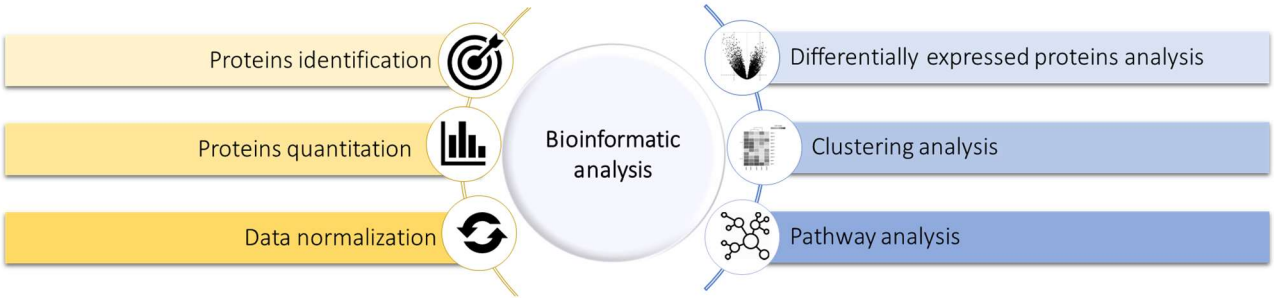


Figure 5: Basics for Bioinformatic analysis of proteomic data

Preclinical research and respiratory diseases

Pharmaceuticals are released following a protracted development procedure. Despite comprehensive preclinical evaluation, there is still a chance that a therapeutic candidate will cause side effects or have insufficient efficacy in clinical trials. If a medication candidate fails in the later stages of the clinical trial, the cost in terms of both time and money is considerable. Pharmaceutical companies are concerned about improving the drug development process and develop innovative technology in the preclinical stage.[15]

Preclinical research has long used model organisms to formulate hypotheses about the mechanisms of human diseases and to test treatments. However, it is important to know the complex ways in which model organisms are similar to humans and, crucially, the ways in which they differ.

Several preclinical models have been developed to investigate diseases as well as to investigate drug mechanisms of action. Drug discovery and early drug development require pharmacokinetics/pharmacodynamics (PK/PD) studies that are performed in preclinical animal models to identify safe and effective doses and dosing regimens[16]. Since there are considerable differences between animal models and humans (regarding for example genetics, physiology and immunology), a detailed understanding of these features and their effects on the function of the whole organism is necessary to provide reliable information about relevant disease mechanisms, potential drug targets, drug modes of action and possible biomarkers, which can be used for further clinical investigations[17].

A biomarker is a trait that may be measured and assessed objectively as a sign of normal or pathogenic biological processes, or pharmacologic reactions to a therapeutic treatment. There are many distinct sorts of biomarkers, such as disease biomarkers, target engagement or mechanism related biomarkers. Because biomarkers related to target engagement or physiological response to a treatment are generally comparable in animals and humans, biomarkers play a significant role in translating early preclinical research in model systems to the clinical stage.[15]

To evaluate potential biomarkers and the effects of drug candidates on animal models, preclinical investigations in the discovery stage involve a variety of biochemical tests. The majority of drug candidates and biologics have protein targets that are involved in cellular pathways and networks and are physically and functionally associated with a variety of other proteins and cellular components. Furthermore, the hundreds of different cell types that make up the organs create various physical and functional contexts in which proteins exist and on which drugs may act with desired or undesirable outcomes, such as post-translational modification or protein activation through specific proteolytic cleavages.

Given this complexity, it seems only reasonable to use proteomics in preclinical research to better understand the impact of drug candidates on their protein targets, to generate reliable biomarkers and to shed light on the cellular pathways that result in the intricate phenotype.

Respiratory diseases are pathological conditions affecting human respiration. Chronic Respiratory diseases (CRDs) such as cystic fibrosis, asthma and chronic obstructive pulmonary disease (COPD) are important causes of morbidity and mortality both for adults and children. Interstitial lung disease (ILD) defines a group of lung diseases, such as pulmonary fibrosis, that affect the interstitium and cause scaring and fibrosis. Currently, no specific effective therapies are available for these leading respiratory diseases and, as a result, there is a great need for novel approaches in therapy and research, as well as for relevant preclinical models[18].

Many of these issues could be resolved by proteomics. With innate and adaptive immunity, extracellular matrix/interstitium, resident and recruited leucocytes, and an epithelium that is constantly exposed to the external environment, the lung offers an interesting and complicated field for proteomic investigations. The relevance of proteomics in lung cancer has received a lot of attention recently, although proteomics is a rapidly expanding science, useful for many other diseases.[19]

Aim of the work

Proteomics, regarded as the comprehensive study of the gene expression at protein level at a particular time in different organs, tissues and cell types is a key enabling technology for the systems biology approach.

The goal of this thesis work is to create several workflows that will aid preclinical studies in pharmaceutical research by utilizing mass spectrometry-based proteomics to discover, identify, quantify, and study proteins involved in the pulmonary system as well as disease biomarkers. Progression or establishment of a pulmonary disease, normal lung development and different pharmacodynamic preclinical models were taken into account to explore the applicability of mass spectrometry-based proteomics to identify key molecular players, potential biomarkers and to help drug target prioritization.

Different methods and technologies known in the proteomic field were tested to evaluate the best protocol choice for specific experiments in preclinical research. The workflows developed started from the sample preparation coming up to the data analysis in order to provide a complete scenario of proteome to understand biological meaning of the studies. Each step of the workflow has to be optimized or adapted to the specific tissue or biological context.

Bottom up proteomics is a precious tool to quantify target proteins, through targeted experiments (MRM), but also through discovery proteomics experiments, leading to high-throughput datasets.

Targeted proteomics allows to focus on specific proteins, based on proteotypic peptide detection.
With the accurate selection of proteotypic peptides, it is possible to isolate and quantify a specific protein isoform, or to detect post-translational modifications on peptides without enrichment protocols. In this work, targeted proteomics was applied to three case studies, with the specific aim to quantify fibrillar collagen and actin in fibrotic lungs, and to monitor the target engagement of acetylated histones in a pharmacokinetic experiment.

Discovery proteomics allows to identify and quantify thousands of proteins within a tissue and joined with bioinformatic analysis leads to a deep understanding of the biological meaning and the processes involved. Quantitative proteomics was applied and studied in this thesis work with the specific aim of characterize the proteome profile of developing lung in a rabbit model.

Moreover, in this thesis work, a chapter was dedicated to phosphoproteomic analysis. As known, protein kinases and their substrates have been gaining increasing attention as therapeutic targets for the treatment of cancer and chronic inflammatory diseases. For this reason, a specific sample preparation, based on enrichment of phosphopeptides, was employed to study the phosphoproteome of a multikinase inhibitor, and three different bioinformatic approaches were applied to investigate the quantified phosphopeptides/proteins.

Target proteomics to study biomarkers of pulmonary fibrosis

Background

Pulmonary fibrosis and the bleomycin model

Interstitial lung disease (ILD) comprises many parenchymal lung disorders characterized by varying degrees of inflammation and fibrosis, as well as of morbidity and mortality.

Idiopathic pulmonary fibrosis (IPF) is the most lethal amongst the interstitial lung diseases and is characterized by progressive fibrosis commonly leading to end-stage lung disease, respiratory failure and a fatal outcome[18].

IPF and many of these fibrotic ILDs lack effective therapy despite the recent approval of two drugs (Nintedanib and Pirfenidone) to reduce progression in IPF patients. Because of the pathology complexity, the use of animal models that reproduce key known features of the disease is necessary. Different animal models have been settled to investigate key mechanisms underlying pathogenesis of pulmonary fibrosis, via exposure to FITC, silica, radiation, or bleomycin. While no animal model can reproduce all manifestations of the human disease, several are available to address selected features of IPF. Among the first to be settled and widely used is the bleomycin model, which is the best-characterized and most extensively used animal model due to its ability to reproduce many aspects of IPF and other fibrotic ILDs, its good reproducibility and ease of induction.[20] Studies based on the bleomycin model have identified specific cellular and molecular mechanisms recognized as important in IPF pathogenesis, as well as new potential therapeutic targets for this disease.

Bleomycins are a family of complex glycosylated peptides with anti-neoplastic properties, originally isolated from the strain of *Streptomyces verticillus*[21]. Bleomycin-induced toxicity occurs predominantly in the organs that lack the bleomycin-inactivating enzyme, the bleomycin hydrolase, which are lungs, skin, and mucous membranes.

Different bleomycin-induced fibrosis time course experiments in rodents have revealed specific aspects of pulmonary fibrosis, which starts with an acute inflammatory phase followed by a fibrotic phase, with a continuous increase of collagens over time. Comparison of human IPF-patient derived samples with bleomycin-induced rat lung fibrosis samples have shown conserved biological processes, such as epithelial cell injury, basement membrane damage and interstitial and intra-alveolar fibrosis with dense collagen deposition.[22]

The progression of lung injury and fibrosis in different bleomycin animal models is quite similar, even changing species or administration. The lung injury progresses through three stages:

- Acute inflammatory response, characterized by infiltration of inflammatory cells with activation and elaboration of a multitude of cytokines, chemokines and plasma proteins, due to damage to the alveolar epithelium combined with vascular leak.
- Transition phase from inflammation to active fibrosis, characterized by increase in fibroproliferation and connective tissue disorders, initiation of fibroblast-myofibroblast transition and gene expression of interstitial collagen, which starts to accumulate.
- Chronic fibrosis stage: intra alveolar and septal fibrosis becomes morphologically manifest. In this phase myofibroblast population grows, as well as the deposition of extracellular matrix. The increased deposition of lung collagen usually peaks at about 28 days post-bleomycin treatment in rodent models [22,23].

Collagen content and α-smooth muscle actin (αSMA) are the most commonly used biomarkers of fibrosis. The development of fibrosis and collagen deposition in bleomycin models can be assessed histologically and biochemically by measurement of lung hydroxyproline content as surrogate to estimate total lung collagen content. However, hydroxyproline assay is not specific for collagen types, i.e. fibril-forming or network-forming.

Collagen, the main biomarker of fibrosis

During the 1980s, it was discovered that collagen increased in biopsy samples obtained from patients affected by lung fibrosis compared to control patients. In particular, it was found an increase in collagen type I over the other types of fibrillar collagen. These studies laid the foundations for approaching fibrotic pathologies, pointing to the increase in this type of collagen, the less elastic one, as a cause of the abnormalities in alveolar architecture.[24]

With fibronectin, elastin and other extracellular matrix proteins, collagen creates a conglomerate of fibrillar proteins: in case of pulmonary fibrosis these proteins are overexpressed and lead to changes in the normal architecture of lung tissue. The excessive production and accumulation of collagen in fibrosis do not only depend on an increased biosynthesis, but also on a decreased degradation of this polypeptide. Intact fibrillar collagen is cleaved by matrix metalloproteinases (MMPs), then the collagen fragments are ingested by macrophages and undergo lysosomal degradation. In fibrotic samples a defect in both intracellular and extracellular degradation pathways has been observed.[25]

Collagen is the primary insoluble fibrous protein in the extracellular matrix and in connective tissue. Collagens are a family of proteins with an amazing heterogeneity: up to now, 27 collagen types are known, encoded by over 40 different genes. The diversity of collagen types is related to their molecular assembly and structures, and also to tissue distribution and function.

Molecular hallmarks of collagen are the several repetitions of Gly-X-Y sequences, the high content of proline and hydroxyproline residues and the unique triple helical structure composed by three polypeptide chains, called α-chains. Each α-chain is composed by about 1400 amino acids, with a molecular weight of about 140 kDa.[26]

The triple helical structure is the most interesting feature of all collagen types: it is a well-designed structural motif in which three parallel polypeptide chains, left-handed, are wound together in a right-handed triple helix. The triple helix consists in three domains: the N-terminal and C-terminal non-triple helical domains and the central triple helical domain, which represents more than 95% of the molecule.

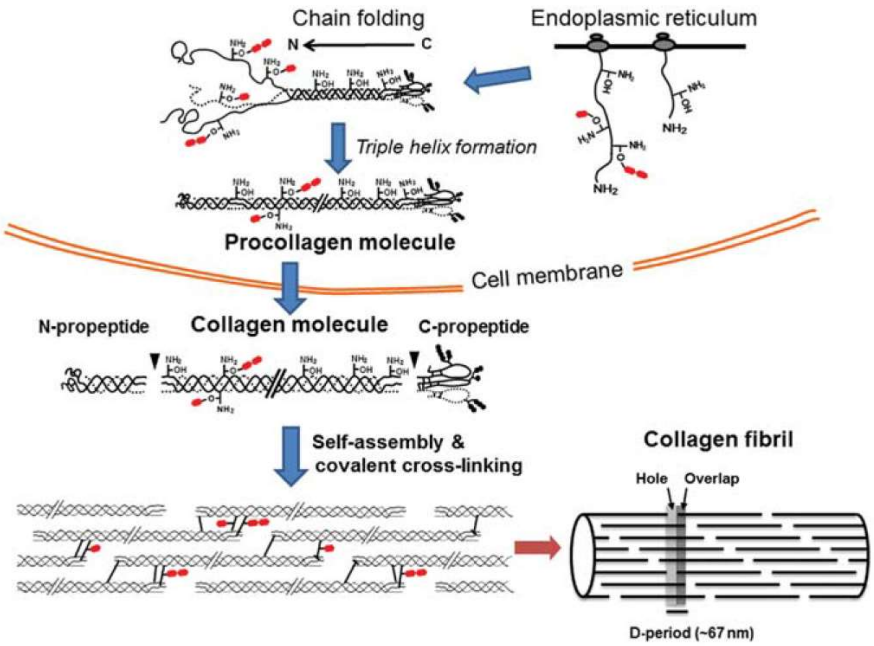


Figure 6: Scheme of collagen biosynthesis. From Yamauchi et al., 2012 [33]

Collagen is a secreted protein and its biosynthesis is composed by specific steps[27], also described in Figure 6:

- Synthesis and first PTMs: the synthesis of pro-α-chains occurs in the endoplasmic reticulum. Here, the polypeptides undergo different post translational modifications catalyzed by specific enzymes on specific lysine residues that are hydroxylated and glycosylated, in addition to proline residues that are hydroxylated.

- Association of α-chains: starting from C-terminus, two α1 and one α2 chains associate wrapping one with the other towards the N-terminus to form the procollagen molecules.
- Secretion into the extracellular space: the procollagen molecule is secreted, then in the extracellular space the N- and C-ends of procollagen are cleaved by ADAMTS, a type of metalloproteinases, and by BMP1/tolloid-like proteinases.
- Self-assembly and cross-linking: in the extracellular matrix collagen molecules spontaneously assemble in parallel and longitudinal fibrils. Covalent inter- and intra-molecular cross-linking occurs through an enzymatic oxidative deamination and through non-enzymatic condensation reactions. Collagen cross-linking and triple helical structure confer resistance to proteases.

In several recent studies, collagen expression was found to be increased in areas of fibrotic tissue[28].

The members of the collagen family have characteristic domains with a triple-helical structure formed by three collagen α chains twisted together. Specific features in the α chain structure enable the formation of the typical collagen triple helix. Among these features, glycine as every third residue and the presence of post-translationally hydroxylated proline residues are the most important for the formation of the macromolecular fibrillar structure[29].

Type I collagen is the most abundant collagen and represents the key structural element of several tissues. It is expressed in almost all connective tissues and it is the predominant component of the interstitial membrane. It is composed of two alpha chain 1 (col1a1) and one alpha chain 2 (col1a2)[30].

Type III collagen is secreted by fibroblasts and other mesenchymal cell types, consequently this collagen type is a major player in various inflammation-associated pathologies such as lung injury. It is composed of three alpha chain 1 (col3a1)[31].

Type III collagen together with type I collagen are the main constituents of the interstitial matrix.

αSMA, the fibroblast-to-myofibroblast transition biomarker

Actin protein family plays an important role in many cellular processes: in cell division, cell motility and in the creation of contractile forces. In vertebrates three main actin isoforms have been found: α-isoforms are present in skeletal muscle, in cardiac muscle and in smooth muscle; the β- and γ- isoforms are expressed both in non-muscle and muscle cells. The aminoacidic sequence is highly conserved both intra isoforms and inter species[32].

Measuring 7 nm in diameter, actin filaments are the thinnest apparatuses of the cytoskeleton and can be found assembled in two types of structures, forming through cross-links between actin filaments: bundles and networks. Bundles play a role in cell division and cytokinesis, in cell movement and in muscle contraction, when F-actin is joined to myosin. The actin network, as suggested by recent studies, serves as barrier localized beneath the plasma membrane[32].

The smooth muscle α 2 actin (Acta2) isoform is largely found in smooth muscle but is also expressed in other specialized cells such as pericytes and myofibroblasts, the latter being typical of wound healing. The contractile behavior of myofibroblasts is correlated with the expression of α-SMA, whose presence improves wound repair, contracting the edges of the wound[33].

Different isoforms exist in actin family. α-Smooth muscle actin is the actin isoform that predominates within vascular smooth-muscle cells and plays an important role in fibrogenesis. Smooth muscle actin shows high identity with other isoforms, ACTB and ACTG, which represent cytoplasmic actin isoforms highly expressed in lung homogenates (Figure 7). The differences regard 18 amino acids upon 377 total amino acids in the sequence. The ELISA assays are less sensitive in the recognition of single amino acid variants in respect with peptide mapping based on mass spectrometry.

```
P60711 ACTB_RAT             1   --MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA       58
P63259 ACTG_RAT             1   --MEEEIAALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA       58
P62738 ACTA_RAT             1   MCEEEDSTALVCDNGSGLCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA       60
                                    ::: :*** *****:*******************************************

P60711 ACTB_RAT            59   QSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREK       118
P63259 ACTG_RAT            59   QSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREK       118
P62738 ACTA_RAT            61   QSKRGILTLKYPIEHGIITNWDDMEKIWHHSFYNELRVAPEEHPTLLTEAPLNPKANREK       120
                                ****************:************:*************.***************

P60711 ACTB_RAT           119   MTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRL       178
P63259 ACTG_RAT           119   MTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRL       178
P62738 ACTA_RAT           121   MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL       180
                                **********.**********************:********.*************:**

P60711 ACTB_RAT           179   DLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEK       238
P63259 ACTG_RAT           179   DLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEK       238
P62738 ACTA_RAT           181   DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK       240
                                **********************.**********************:*************

P60711 ACTB_RAT           239   SYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTV       298
P63259 ACTG_RAT           239   SYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTV       298
P62738 ACTA_RAT           241   SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV       300
                                ********************:******:****.******:*******:*********.*

P60711 ACTB_RAT           299   LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIS       358
P63259 ACTG_RAT           299   LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIS       358
P62738 ACTA_RAT           301   LSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIS       360
                                ************************************************************

P60711 ACTB_RAT           359   KQEYDESGPSIVHRKCF                                            375
P63259 ACTG_RAT           359   KQEYDESGPSIVHRKCF                                            375
P62738 ACTA_RAT           361   KQEYDEAGPSIVHRKCF                                            377
                                ******:**********
```

Figure 7: Alignment of actin isoforms. The alignment was performed with Uniprot tool

α-SMA is the most used fibroblast-to-myofibroblast transition (FMT) marker, though this protein is expressed also in other cell types, like smooth muscle cells. Fibroblasts and myofibroblasts incorporate α-SMA into stress fiber augmenting the contractile and motile ability and leading to tissue remodeling. For this reason,

α-sma is considered a biological marker also for the epithelial to mesenchymal transition[34]. Epithelial to mesenchymal transition occurs during the healing process after injury and, with ongoing chronic inflammation conditions, the abnormal formation of myofibroblasts causes progressive fibrosis, leading to parenchymal lung damage by excessive ECM deposition.

Aim

The assessment of hydroxyproline (HYP) content in lung homogenate in preclinical models is the fastest and easiest way to estimate collagen content, because HYP is one of the key components of this complex protein[35]. A limitation of this assay regards the presence of hydroxyproline in all collagen types, not only in fibrillar collagens, and in other proteins, such as Elastin and Hypoxia-inducible factor[36].

Another limitation is the complex biological matrix derived from fibrotic lung. The increased accumulation of extracellular matrix leads the tissue to high resistance to proteases, probably due to the higher presence of crosslinker between collagens and other extracellular matrix proteins.

A specific analytic assay that helps the quantitation of fibrillar collagen in fibrotic lung was conceived and developed, because the hydroxyproline assay results does not represent the real fibrotic condition compared with histologic analysis. Collagen quantification is a clear and important readout of bleomycin and pharmacologic therapy effects in determining the fibrotic stage. For this reason, a targeted approach to selectively identify and quantify three different chains of fibrillar collagen was developed, focused on proteotypic peptides containing the HYP modification.

Moreover, quantitation of α-smooth muscle actin (αSMA) is often used as marker for the identification of contractile myofibroblast, the major contractile cells in fibrotic organs, and activated fibroblasts, the cells responsible for the exaggerated production of collagen in fibrosis[37].

Nevertheless, αSMA isoform shares high sequence similarity with other cytoskeletal isoforms, that coexist in most cell types. Therefore, the correct isoform identification is essential, when αSMA is employed as biomarker of fibrotic diseases.

To focus on the correct isoform a targeted assay was developed to identify and quantify α-SMA as biomarker in pulmonary fibrosis bleomycin-induced rat model. The final purpose was to select peptides to develop a target analysis specific for the isoform ACTA.

Results

Preliminary untargeted analysis to identify specific targeted peptides

The first step for the development of a correct targeted proteomic analysis is the collection of data from an untargeted proteomic approach.

The key point of the untargeted approach is to accurately identify proteotypic peptides for proteins of interest. Mass over charge of precursor peptides and of transitions (peptide fragments) have to be well defined at this stage.

Another key point is to collect the untargeted data with the same HPLC method that will be used for the targeted method, in order to have the same retention time during the chromatographic separation.

In this work, the preliminary untargeted analysis was performed through Q-Exactive instrument, which employs a quadrupole analyser followed by an Orbitrap.

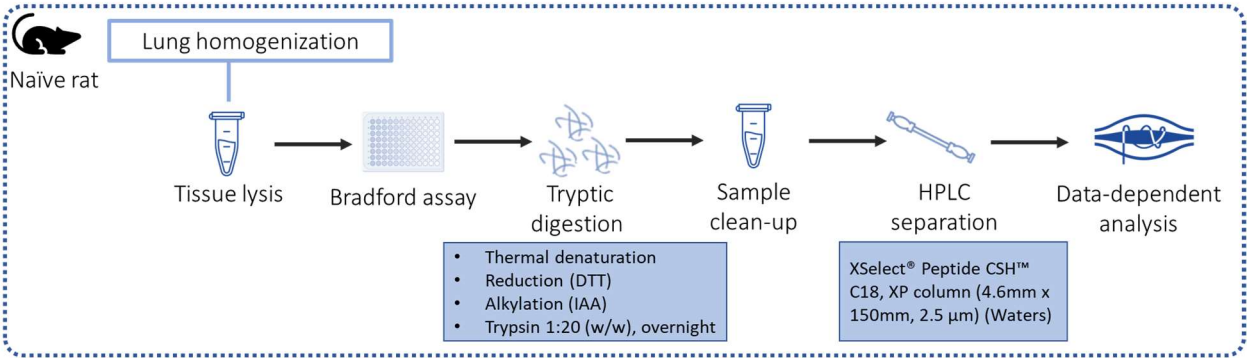


Figure 8: Preliminary untargeted analysis: sample preparation and analysis workflow

Lung homogenate from rat naïve samples were submitted to a shotgun proteomic analysis. The biological samples were quantified before the digestion processes. Heat denaturation, reduction and alkylation reactions were performed following the protocol showed in Figure 8 and the tryptic digestion was able to cleave the majority of the homogenate proteins. The aim of this step is to obtain a comprehensive peptide mapping of the entire biological sample and following this protocol proteins with a molecular weight from 5 kDa up to more than 250 kDa can be digested. Samples deriving from two independent biological replicates were pooled prior UHPLC-HRMS analysis, to improve peptide identification.

Peptide solutions were then submitted to UHPLC-MS analysis. The chromatographic separation occurred through a reverse-phase C18 column and the mass analyser is a Q-Exactive instrument. The samples were

analysed with a data-dependent method, to obtain full scan spectrum and MS/MS spectrum in a timely manner.

The raw data were analysed with Proteome discoverer to identify proteins suitable for the development of targeted methods, using the parameters described in Table 1.

Table 1: Parameters used in Proteome discoverer for preliminary untargeted data analysis

Enzyme name	Trypsin (full)
Max missed cleavage sites	3
Min. peptide length	5
Max. peptide length	144
Precursor mass tolerance	10 ppm
Fragment mass tolerance	0.02 Da
Dynamic modification	Oxidation / +15.995 Da (K,M,P)
Dynamic modification	Deamidated / +0.984 Da (N,Q)
N-terminal modification	Acetyl / +42.011 Da (N-terminus)
Static modification	Carbamidomethyl / +57.021 Da (C)

The untargeted analyses highlighted hundreds of proteins. Among these, several proteins suitable for the development of targeted methods were identified.

Therefore, the sample preparation here described, from lung homogenization to HPLC separation, was applied to all the targeted approaches described in the next paragraphs.

Quantitation of Fibrillar collagens in a pulmonary fibrosis rat model

Study design

The experimental design consisted of 5 time points (at days 7, 14, 21, 28 and 56) for both bleomycin (BLM)-treated animals and saline controls, plus one T_0 time-point (Figure 9).

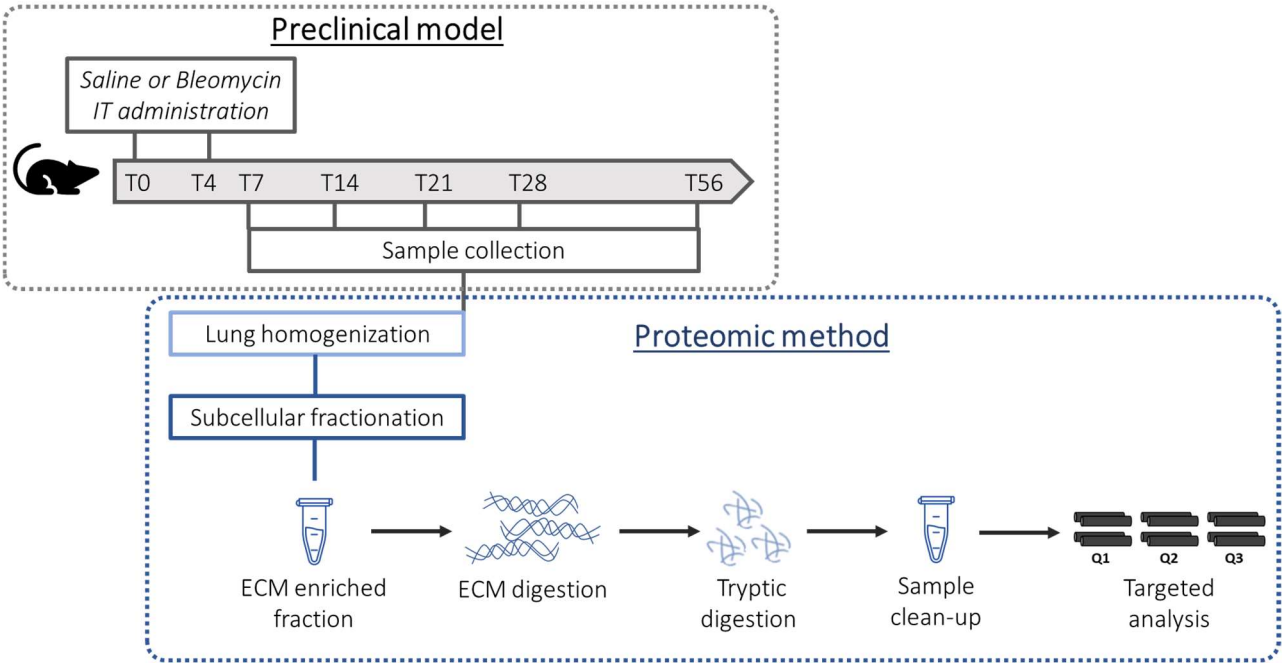


Figure 9: Experimental design and workflow for quantitation of fibrillar collagens in IPF bleomycin-induced rat model

The preclinical experiment used for IPF bleomycin-induced rat model is here briefly described: Bleomycin (2U/kg body-weight) was administered twice (at T_0 and T_4) intra-tracheally (IT) in adult rats. Three sample replicates were collected for each time point, with the exception of T_0 and T56 BLM, where only two replicates were available. Samples were homogenized in PBS.

Subcellular fractionation for ECM proteins enrichment

Pools of lung homogenates of replicates at each time point and treatment (V = vehicle; B = bleomycin) were collected. Each pool (T0, T7 V, T14 V, T21 V, T28 V, T56 V, T7 B, T14 B, T21 B, T28 B and T56 B) was fractionated with ProteoExtract® Subcellular Proteome Extraction Kit, as described in the experimental section.

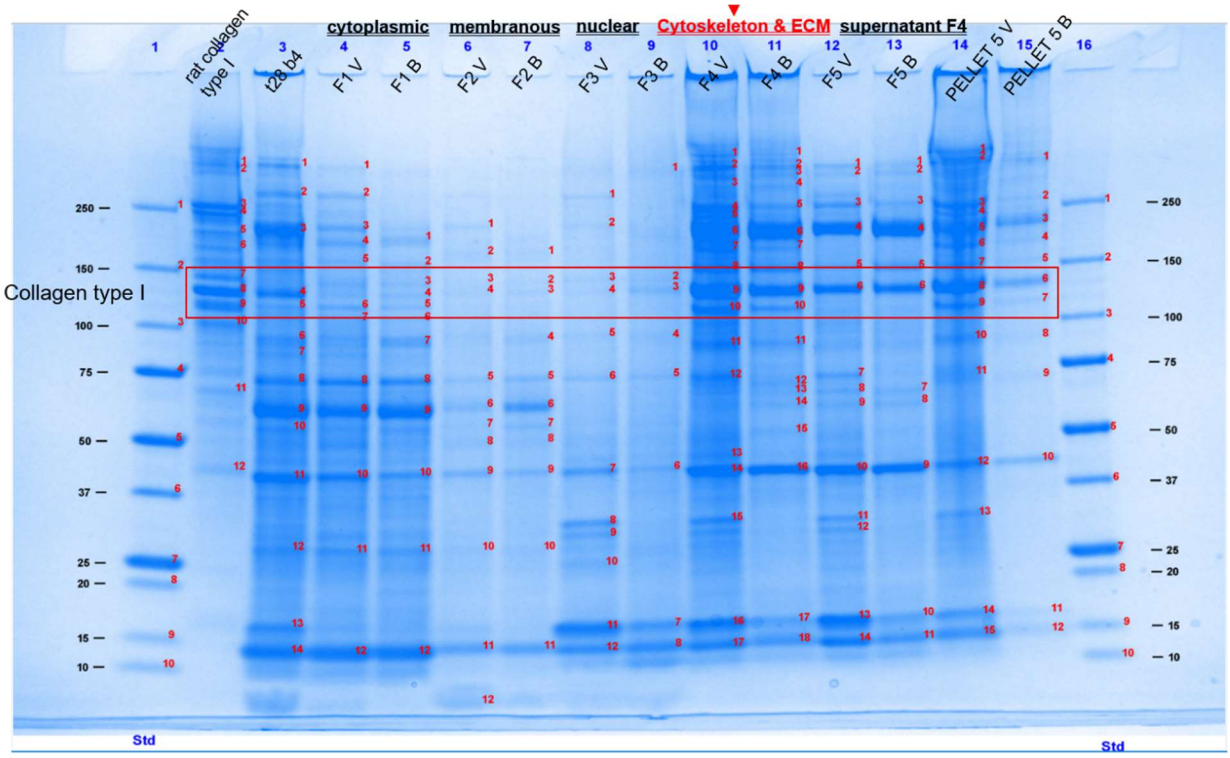


Figure 10: SDS-PAGE of lung homogenate fractionation.
Lane 1 and 16. Molecular weight marker Precision Plus Protein™ Dual Color Standards. Lane 3 show a total lung homogenate of a Bleomycin sample at T28. The red box underlines the bands relative to collagen, compared with the standard of collagen type I showed in lane 2. It is possible to appreciate the presence of collagen only in F4, and accordingly in F5 and Pellet 5.

This fractionation kit allows to separate proteins belonging to cytosol (F1), membrane (F2), nucleus (F3) and extracellular matrix (F4), for a total of four fractions. The last fraction, including ECM and cytoskeleton proteins, has been further separated after centrifugation in supernatant (F5) and pellet (Pellet 5) to discriminate between soluble and insoluble collagen.

Fractions (F1 to F5 plus Pellet 5) corresponding to the pools T28 V and T28 B were visualized on an SDS-PAGE (Figure 10). A sample of purified rat collagen was also loaded on gel, as reference. Collagen was mainly evident in F4 and derived (F5 and Pellet F5) samples. From this experiment, is possible to deduce that collagen is present both in supernatant and in pellet. Therefore, fraction 4, corresponding to ECM and cytoskeleton proteins, was chosen for the identification of collagen with the targeted approach.

Collagen detection: Bottom up targeted proteomics

A total of 11 samples, corresponding to fraction 4 of each pool, were therefore processed.

Briefly, human MMP1 was added to samples to improve collagen peptide digestion, which proceeded overnight at 37°C. The MMP1 collagenase improves collagen cleavage by unwinding the triple helices, through a specific cutting site at three-quarter of the triple helix[38]. Lung homogenates were digested with trypsin after heat denaturation, reduction with dithiothreitol (DTT) and alkylation with iodoacetamide.

For each sample, 30 µg of digested peptides were desalted. The resulting eluates were dried prior chromatographic analysis, which was performed by high-pressure liquid chromatography –mass spectrometry (HPLC-MS).

Targeted method

The instrument used for the targeted analysis is a Q-TRAP 4000. The acquisition method developed is a Multiple Reaction Monitoring analysis (MRM). This method allows to analyse the proteotypic peptides (listed in Table 2) based on the transitions, which allows to identify the specific position of hydroxyproline within the peptides.

The triple quadrupole, acting as a mass analyser, selects the m/z of proteotypic peptide, then the second quadrupole fragments the peptide ion with collision energy, and the third quadrupole, selects and measures the transitions (peptide fragments).

Several proteotypic peptides for each collagen protein were selected based on position in the chain, on the intensity of m/z signal and based on the sequence coverage with MS/MS spectra, using high-resolution data obtained in the preliminary untargeted analysis.

The peptides selected for the targeted proteins are reported in Table 2, where the proline in lower case represents the hydroxyproline modification in the peptides.

Table 2: Proteotypic peptides of fibrillar type I and type III collagen

Chain	***Peptide Sequence***	***Residue***	***M/Z [Da]***	***Charge***	***Transition***	
Col1a1	GVVGLpGQR	948-956	449.76	2	y6: 643.35	y7: 742.41
Col1a1	GQAGVMGFpGpK	564-575	589.29	2	y7: 765.35	y10:992.48
Col1a1	GQAGVMGFpGPK	564-575	581.29	2	y7: 749.36	y9: 905.45
Col1a1	GFpGADGVAGPk	483-494	552.77	2	y10 2+: 450.72	y9:787.39

Col1a1	GFpGADGVAGPK	483-494	544.77	2	y10 2+: 442.72	y9:771.39
Col1a1	GLTGSpGSpGPDGK	528-541	629.80	2	y9: 843.37	y6:586.28
Col1a1	GFSGLDGAK	258-265	426.22	2	y7: 647.33	y6: 560.30
Col1a1	GFSGLDGAkGDTGPAGPK	258-274	549.94	3	y9: 799.39	y10: 943.48
Col1a2	GApGAIGApGPAGASGDR	682-698	755.86	2	y10: 900.41	b6: 483.25
Col1a2	SGHpGPVGPAGVR	1074-1086	401.88	3	b5: 452.18	y6: 556.32
Col1a2	GATGLpGVAGApGLpGPR	316-333	796.92	2	y13: 1193.62	y10: 924.48
Col1a2	GLpGEFGLpGPAGPR	580-594	727.38	2	y7: 667.34	y13: 1283.63
Col1a2	GLpGEFGLpGpAGPR	580-594	735.37	2	y7: 683.34	y13: 1299.62
Col1a2	GLVGEPGpAGSK	349-360	542.79	2	y7:629.32	y5: 459.25
Col1a2	GEAGNIGFpGPK	493-504	580.29	2	y6: 618.32	y4:414.23
Col1a2	GEAGNIGFpGPk	493-504	588.29	2	y6: 634.31	y4: 430.22
Col3a1	DGSSGHPGpIGpPGPR	1150-1165	506.24	3	y10: 976.52	y5: 539.29
Col3a1	GEAGSpGIpGPK	449-460	549.77	2	y7: 697.38	y4: 414.23
Col3a1	GEMGpAGIpGApGLLGAR	395-412	834.92	2	y10: 940.52	b8: 729.32
Col3a1	GEMGPAGIpGApGLLGAR	395-412	826.92	2	y7: 699.41	b11: 954.44
Col3a1	GGpGGpGLpGPAGK	596-609	583.79	2	y6: 542.29	b5: 342.14
Col3a1	GLAGPpGMpGPR	956-967	569.79	2	y8: 840.40	y7: 743.34
Col3a1	GLAGppGMpGPR	956-967	577.78	2	y8: 856.39	y7: 743.34
Col3a1	GPpGTAGTpGLR	683-694	556.79	2	y10: 958.49	y6: 616.34
Col3a1	GPVGPHGPpGK	1139-1149	339.18	3	y5: 471.25	b6: 545.28
Col1a3	GRpGLpGAAGAR	308-319	371.20	3	b5: 497.28	y7: 615.32

Type I and type III collagen detection in the time course study

The output of the analysis described in the previous paragraph is a chromatogram reporting only peaks relative to the transitions specified in the method. In the method we also specified two proteotypic peptides and transition of human MMP1 used in the sample preparation for collagen degradation. This enzyme was the best choice as internal standard, because it undergoes all the reaction steps as lung homogenate proteins. Furthermore, the enzyme is of human origin and therefore distinguishable from rat MMP1, because the sequence of the selected proteotypic peptide was different from the rat MMP1.

Protein	*Peptide sequence*	*m/z precursor*	*Transitions*	
hMMP1	IENYTPDLPR	609,31	y8: 975,48	y7: 861,44

For each peptide two transitions were selected: the most intense is considered the "ion quantifier", the less intense is called "ion qualifier". The area under the peak of the ion quantifier was then measured. The area value for each peptide of each sample was normalized as a ratio to the area of the proteotypic peptide of hMMP1. The resulting value is therefore normalized taking into account differences due to sample preparation, quantity of total peptide in the mixture or to instrumental variation.

The fold increase for each time point was obtained by dividing the normalized value of each bleomycin-treated sample with the normalized value of the corresponding vehicle-treated sample at each time point (Figure 11).

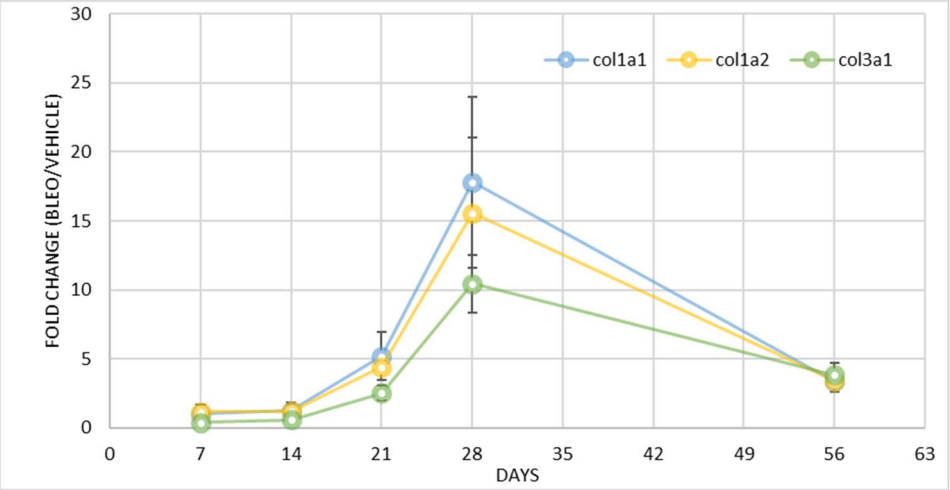


Figure 11: Collagen quantification: collagen type I (alpha chain 1, blue; alpha chain 2,yellow) and collagen type III (alpha chain 1, green) fold change (bleomycin/vehicle) in rat lung samples. The trend during time course is the same for all the alpha chain. Error bars represent standard deviation and are calculated on the values obtained from each proteotypic peptide.

The relative increase of all collagen types considered at time points 21 and 28 is statistically significant compared to time points 7 and 14. The decrease at time point 56 is statistically significant compared to time point 28, based on Dunn's multiple comparison test.

The elevated standard deviation at time point 21 and 28 highlight the heterogeneity of macromolecular structure of fibrillar collagens that reflects the heterogeneity of pulmonary fibrosis development.

Quantitation of alpha-SMA isoform in pulmonary fibrosis bleomycin-induced rat model

Selection of proteotypic peptides for ACTA

From the untargeted data, we identified two proteotypic peptides of α-smooth muscle actin: peptide 199 - 208 and peptide 98 - 115. These peptides differ from the other actin isoforms by only one amino acid: the 5th position Val instead of Thr in peptide 199 - 208 and the 8th position Thr instead of Val in peptide 98 – 115. The selected transitions (y6: m/z 676.36) confirmed the different amino acid in the peptide sequence.

In the Figure 12 the masses of transition and the MS/MS spectrum of peptide 199 - 208 revealed by untargeted analysis are shown.

Sequence: **GYSFVTTAER**, pI: **6.00141**

Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
G	1	58.02879	1130.54771	10
Y	2	221.09211	1073.52624	9
S	3	308.12414	910.46291	8
F	4	455.19256	823.43089	7
V	5	554.26097	676.36247	6
T	6	655.30865	577.29406	5
T	7	756.35633	476.24638	4
A	8	827.39344	375.19870	3
E	9	956.43603	304.16159	2
R	10	1112.53714	175.11900	1

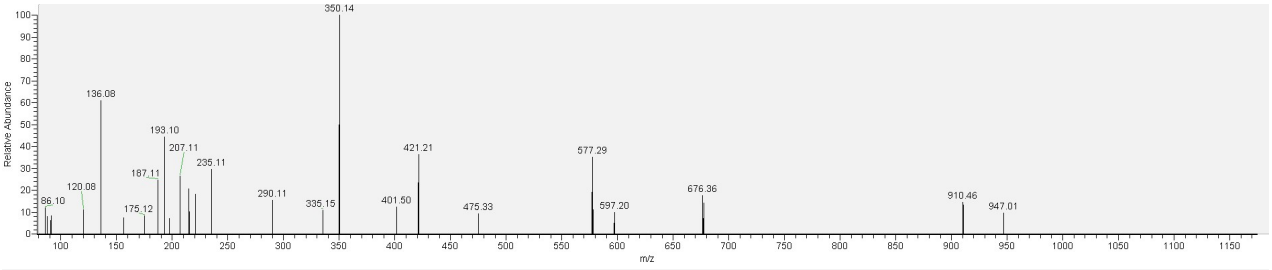


Figure 12: Fragment ion table and MS/MS spectrum of peptide 199-208

Peptide 98 - 115 allowed the identification of the different isoforms in the untargeted analysis. ACTA peptide is characterized by with a m/z of 652.68 instead of 652.02 for the other isoforms. ACTA peptide and ACTB/ACTG peptide are separated by chromatography with a retention-time difference of 90 seconds, and their MS/MS spectra differs for the transition y12, containing the characterizing amino acid. In Figure 13, the fragment ions of the two peptides are reported, while in panel B the two MS/MS spectra confirming the identification of peptide are shown.

A)

ACTA peptide

Sequence: **VAPEEHPTLLTEAPLNPK**, pI: **4.75193**

Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
V	1	100.07574	1956.04371	18
A	2	171.11285	1856.97529	17
P	3	268.16561	1785.93818	16
E	4	397.20821	1688.88542	15
E	5	526.25080	1559.84282	14
H	6	663.30971	1430.80023	13
P	7	760.36247	1293.74132	12
T	8	861.41015	1196.68856	11
L	9	974.49421	1095.64088	10
L	10	1087.57828	982.55681	9
T	11	1188.62596	869.47275	8
E	12	1317.66855	768.42507	7
A	13	1388.70566	639.38248	6
P	14	1485.75843	568.34537	5
L	15	1598.84249	471.29260	4
N	16	1712.88542	358.20854	3
P	17	1809.93818	244.16561	2
K	18	1938.03314	147.11285	1

Mass/Charge Table

	Mass	
	Mono	Avg
(M)	1955.03643	1956.22613
$(M+H)^{+}$	1956.04371	1957.23340
$(M+2H)^{2+}$	978.52551	979.12036
$(M+3H)^{3+}$	652.68612	653.08268
$(M+4H)^{4+}$	489.76642	490.06384

ACTB/ACTG peptide

Sequence: **VAPEEHPVLLTEAPLNPK**, pI: **4.75193**

Fragment Ion Table, monoisotopic masses

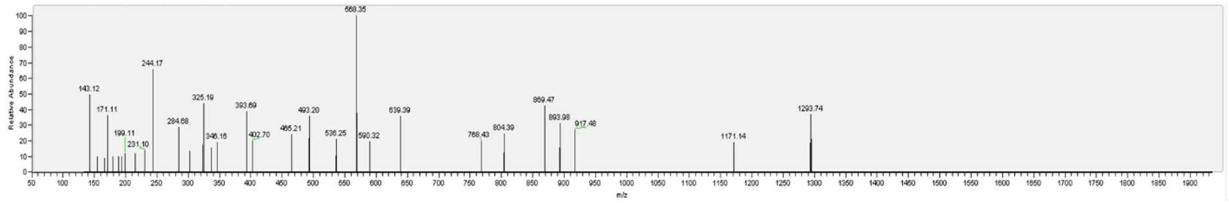
Seq	#	B	Y	# (+1)
V	1	100.07574	1954.06444	18
A	2	171.11285	1854.99603	17
P	3	268.16561	1783.95892	16
E	4	397.20821	1686.90615	15
E	5	526.25080	1557.86356	14
H	6	663.30971	1428.82097	13
P	7	760.36247	1291.76206	12
V	8	859.43089	1194.70929	11
L	9	972.51495	1095.64088	10
L	10	1085.59901	982.55681	9
T	11	1186.64669	869.47275	8
E	12	1315.68928	768.42507	7
A	13	1386.72640	639.38248	6
P	14	1483.77916	568.34537	5
L	15	1596.86323	471.29260	4
N	16	1710.90615	358.20854	3
P	17	1807.95892	244.16561	2
K	18	1936.05388	147.11285	1

Mass/Charge Table

	Mass	
	Mono	Avg
(M)	1953.05717	1954.25361
$(M+H)^{+}$	1954.06444	1955.26088
$(M+2H)^{2+}$	977.53588	978.13410
$(M+3H)^{3+}$	652.02636	652.42517
$(M+4H)^{4+}$	489.27160	489.57071

B)

MS/MS spectrum of precursor 652.68 m/z



MS/MS spectrum of precursor 652.02 m/z

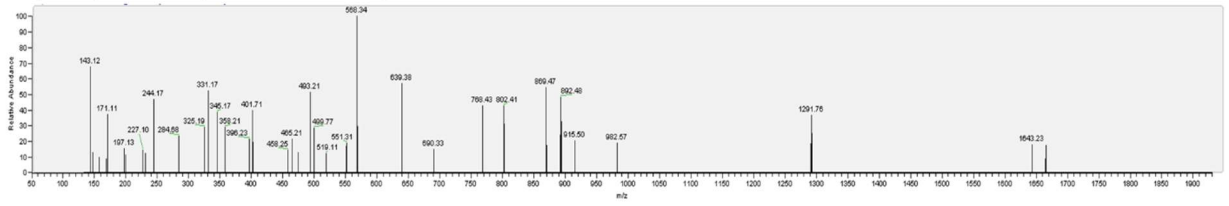


Figure 13: Fragment ion table and MS/MS spectra of peptide 98-115

Therefore, the proteotypic peptides selected for targeted analysis of Actin, aortic smooth muscle (P62738) are shown in Table 3.

Table 3: Proteotypic peptides selected for MRM analysis of alpha-smooth muscle actin

Annotated Sequence	*Residue*	*m/z [Da]*	*Charge*	*Transition*	
GYSFVTTAER	199-208	565,77	2	Y5: 577,29	Y6: 676,36
VAPEEHPTLLTEAPLNPK	98-115	652,68	3	Y5: 568,34	Y8: 869,47

Study design and targeted proteomics for actin detection

The targeted method described was applied to different preclinical experiments. In this thesis work, the data from an IPF bleomycin-induced experiment in rat are reported.

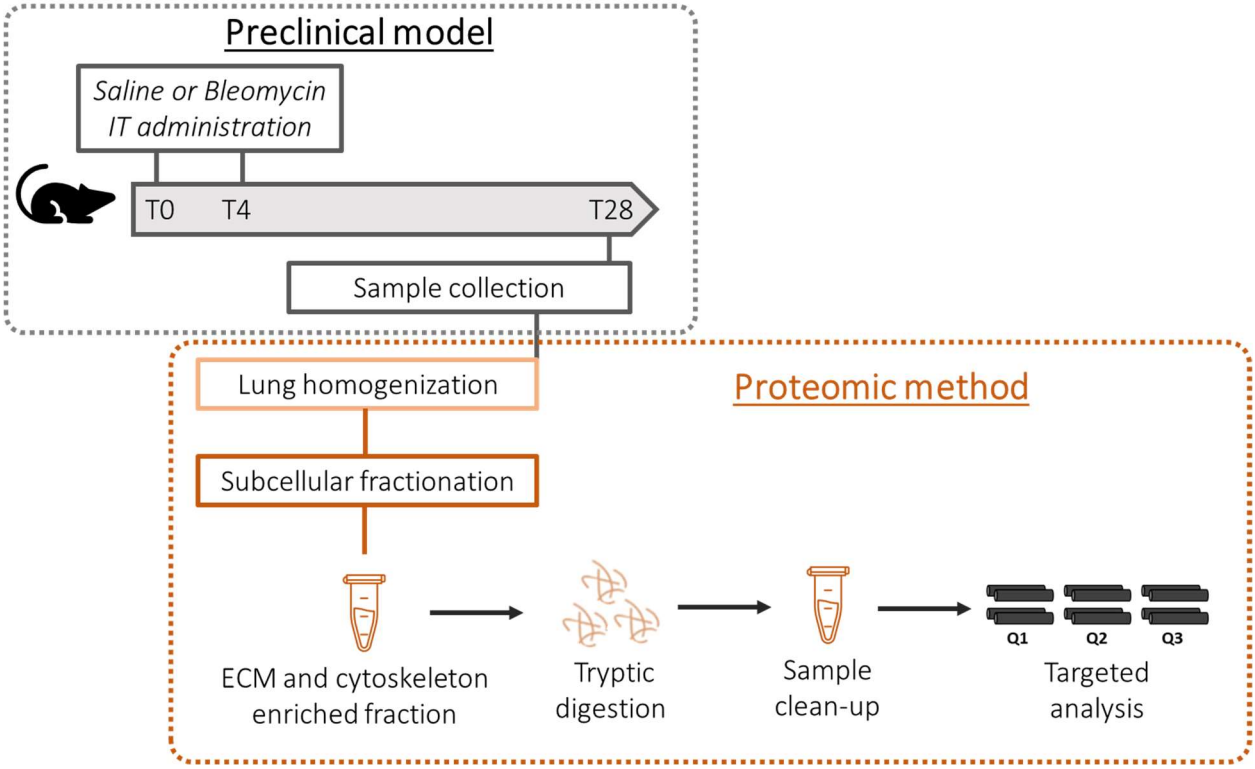


Figure 14: Experimental design and workflow for quantitation of a-sma in IPF bleomycin-induced rat model

The experimental design consisted of 10 bleomycin (BLM)-treated and 10 saline-control rats sacrificed at day 28 after the first administration of bleomycin.

Lung samples were homogenized and the samples were pooled based on the treatment group.

Actin is a cytoskeletal protein; therefore, the subcellular fractionation was used in this workflow as well. Fraction 4, enriched in ECM and cytoskeletal proteins, was digested with trypsin overnight. hMMP1 was added before the digestion in order to have an internal standard for normalization.

The targeted method was developed for the Q-trap 4000, coupled with HPLC. The chromatographic conditions were set as described in the experimental section.

The peptides were detected in lung homogenates digested and analyzed for relative quantitation, normalizing the bleomycin treatment on vehicle values.

α-Smooth muscle actin expression did not show a significant variation 28 days after bleomycin administration (Figure 15) and did not correlate with the collagen expression, that was instead overexpressed in bleomycin upon vehicle (data obtained with the method described in paragraph *Collagen detection: Bottom up targeted proteomics).*

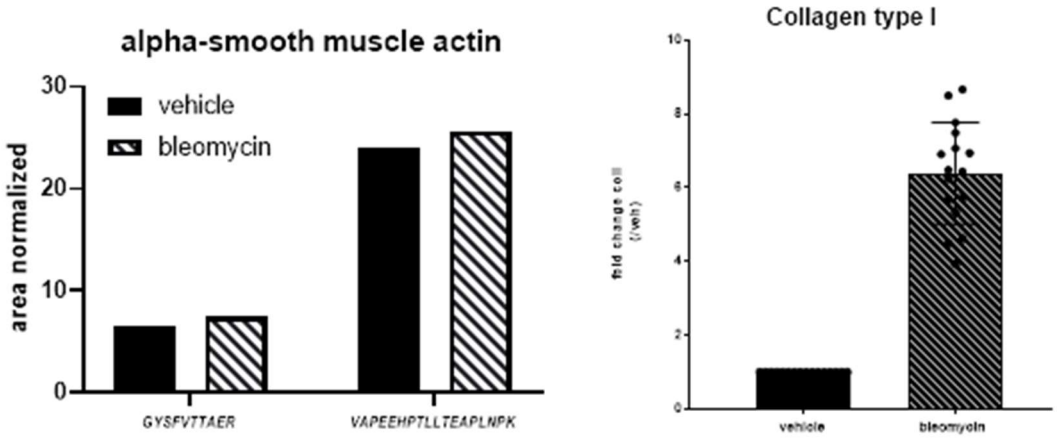


Figure 15: Results of actin quantitative analysis based on the MRM method, on the left, and result of collagen type I quantitation, on the right.

Discussion

Fibrillar collagen detection

Optimization of the targeted proteomic method

The first study on collagen characterized by MS was published in 1970 [39]. The number of articles on collagen and MS has gradually increased since then.

One of the most important developments was the invention of electrospray ionization, which made it possible to study (large) biomolecules such as peptides and proteins.

Collagen has been analyzed in a widespread variety of liquid, such as serum or urine, and solid biomaterials, like bones and tissues with different mass spectrometry applications, from bottom-up proteomics, to targeted proteomics, to MALDI imaging[40,41].

Enrichment of the sample for collagen or extracellular matrix fraction leads to a higher number of identified peptides, and therefore to a better identification of different collagen types.

There are several methods to enrich samples for collagen, of which two are especially suitable for tissues. The first described is decellularization, the second is based on tissue fractionation.

Naba et al.[42] showed the capability of the second enrichment method to extract collagen and extracellular matrix and identify several alpha chains in lung and colon tissues. This method is based on the fractionation of proteins from different cell compartments: cytoplasm, nucleus, membrane, and cytoskeleton plus ECM.

Independently from the different kit used in this thesis work, the aim of extracting ECM proteins from lung tissue was achieved. The depletion of cytosolic and nuclear components simplified the biological samples and allowed a successful collagen quantification.

Moreover, in this thesis work, the protocol was improved with the employment of specific protease that unwind the supramolecular structure of fibrillar collagens.

For proteomics applications proteolytic cleavages obtained with specific proteases, such as trypsin or endoproteinase AspN, are commonly used in order to obtain distinctive fragments of the protein of interest. However, the insolubility of native collagen, due to its triple-helical conformation, makes interstitial collagens resistant to most proteases used for proteomic applications.

Nimptsch and colleagues tested a bacterial collagenase to digest native collagen[41]. The use of bacterial collagenases can be regarded as a straightforward approach because these enzymes fragment native form of triple helices. The limitation of this approach with bacterial collagenases is that the collagen is fragmented into small unspecific peptides, leaving no information concerning the collagen type.

In vertebrates, the degradation of collagens is an integral part of many biological processes (embryogenesis, organ morphogenesis, tissue remodeling, angiogenesis and wound healing), and is managed by the matrix metalloproteinase family (MMPs).

The action of these enzymes is critical for the initiation of collagen denaturation, because collagens are cleaved into 3/4 and 1/4 fragments by the MMPs. Once the collagens are unwind, the common protease are able to digest the alpha chains of collagens[38].

These enzymes act at neutral pH, so we decided to unwind collagen in its native conformation, after ECM enrichment.

All these sample preparation steps, optimized for tissue collagen analysis, allowed to perform a bottom-up targeted proteomic approach to detect and quantify collagen type I and collagen type III in lung homogenate.

Collagen quantification in bleomycin-induced pulmonary fibrosis

The bleomycin-induced murine lung fibrosis model is commonly used to study mechanisms of lung fibrosis and to test potential therapeutic interventions. Bleomycin induced lung injury, and the pathological response, leads to damage to the alveolar epithelium, leakage of fluid and plasma proteins into the alveolar space, inflammatory infiltrates, and fibrotic accumulation.

Commonly, the resultant accumulation of collagen in the lung is measured both by histological and biochemical techniques, most notably via accumulation of hydroxyproline, used as a surrogate for whole lung collagen content[23].

The proteomic method developed in this thesis work offers an alternative to hydroxyproline assay and to ELISA assay to monitor the abundance of collagen in fibrotic samples.

The proteotypic peptides selected for the MRM assay cover the mature sequence of three different triple helices: col1a1, col1a2 and col1a3. Moreover, the position of hydroxyproline in the peptides is monitored, making the proteomic assay accurate and specific for fibrillar mature collagens.

A bleomycin-induced rat lung fibrosis time course model, described by Bauer et al.[22], reported the genome-wide expression changes in the model as well as the histological characterization and the hydroxyproline content. The results showed an increase of hydroxyproline content from day 14 to day 56 after bleomycin administration, in accordance with the results obtained with the targeted method. This study is one of the few time course studies performed on rat models.

Instead, many time course studies were described for mice, and the collagen content was determined with the hydroxyproline assay by Liu et al.[43], showing an increase from day 14 to day 28 in the mouse model.

Another interesting study, taking in account the problem of the insoluble part of extracellular matrix, tested a combination of dynamic proteomics (mice labeled with heavy water) and tissue decellularization to quantify changes in ECM fractional synthesis associated with progression of fibrotic disease *in vivo* in the

mouse. The newly synthetized collagen (col1a1) increased from week one to week three after bleomycin administration[44].

Therefore, the results obtained with the targeted proteomic method described in this thesis work confirmed the trend of bleomycin-induced IPF murine model. The increment of fibrillar collagen type 1 and type 3 starts from the day 21 with a maximum at day 28, then a decremental trend is observed until day 56.

In a 2020 review[45], Yanagihara outlined the three stages of bleomycin-induced pulmonary fibrosis. After administration of bleomycin, an acute inflammation generally lasts for seven days, followed by fibrogenic changes with extracellular matrix (ECM) deposition from 28 up to 35 days. These fibrogenic changes with ECM deposition resolve after 35 days.

The time course rat model studied aligned with the scheme described by Yanagihara, highlighting the role of collagen deposition during extracellular matrix deposition, from day 21 to day 28.

Due to the many literature confirmations, this targeted method is validated and can be used to study samples from pulmonary fibrosis rat model, supporting the preclinic research and the drug discovery for IPF.

Detection of actin with targeted method

Myofibroblasts produce α-smooth muscle actin (α-SMA) and have contractile and secretory activities that are important for tissue architectural control. α-SMA is a marker for a subset of activated fibrogenic cells, which are essential tissue fibrogenesis effector cells.

The presence of contractile myofibroblasts in lung interstitium may promote retraction of parenchymal tissue, resulting in alveolar collapse and the honeycombing characteristic of the lungs of IPF patients[46].

The high percentage of sequence identity between actin isoforms represents the major difficulty in measuring abundances of actin isoforms extracted from various tissues with an immunological assay. Each isoform is remarkably similar to any other isoform, with only few specific variations in amino acid sequence[47].

Other methods have been developed to identify different isoforms of actin, for example for the accurate mass determination of the α-skeletal and α-cardiac actin in cardiac tissue[48]. This method involves purification of muscle actin and subsequent liquid chromatography separation of the intact proteins coupled with electrospray ionization Fourier transform ion cyclotron resonance (LC/FTICR-MS) mass spectrometry.

The cardiac alpha-actin have been measured with a MRM method also for extraocular muscles in skeletal muscle diseases[49].

The targeted method developed here to identify and quantify the α-smooth muscle actin isoform in lung homogenate, as well as in other complex biological tissues, helps to quantify this specific isoform, relevant for the fibrotic diseases.

The sensitivity of the targeted method based on MRM experiment was the key feature to set-up an assay focused only on α-smooth muscle actin.

The results obtained in the case study described in this thesis work, concerning the abundance of α-SMA in lung homogenates derived from a rat model of bleomycin-induced pulmonary fibrosis after 4 weeks from bleomycin administration, showed no difference between the control (vehicle) and the bleomycin group.

Moreover, the α-SMA expression level did not correlate with the fibrillar collagen deposition monitored during the bleomycin-induced pulmonary fibrosis.

The results obtained by Sun at al.[37] suggest that α-SMA expression is quite inconstant in collagen-producing cells across different models of tissue fibrosis. Referring back to Yanagihara's review[45], it is important to note that, at the earliest stage of bleomycin-induced pulmonary fibrosis, the accumulation of α-smooth muscle actin (α-SMA) positive cells in fibroblastic foci derives from overexpression of IL-1β and consequent neutrophilic inflammation.

Therefore, α-SMA expression should be monitored at the early stage of bleomycin-induced fibrosis, in order to evaluate the inflammation and the fibroblast to myofibroblast transition.

Target engagement in a pharmacodynamic experiment

Background

Pharmacodynamics in preclinical model

An important tool in the discovery and development of new drugs in the pharmaceutical industry is the characterization of the relationship between the pharmacokinetics (PK, concentration vs. time) and pharmacodynamics (PD, effect vs. time). The main objective of PK/PD models is to identify promising safe compounds and effective doses and dosing regimens[16].

Evaluating the effective role of a drug in vivo is essential to estimate the efficacy of drug binding to its target. These types of experiments are known as target engagement experiments.

Mass spectrometry is a sensitive tool to perform target engagement assays, very useful specially to monitor the effect of drugs that inhibit enzymes responsible for post-translational modifications.

Post-translational modification analysis: focus on acetylation

N-Acetylation is a post-translational modification (PTM) that can occur at either the N-terminal amino group or the ε-NH_2 of lysine residues. It is a central PTM involved in the regulation of key cellular processes including transcription, protein–protein interactions, and protein stability[50]. In particular, the acetylation of lysines in histones has been widely and thoroughly studied[51].

Acetylation of lysine is a reversible PTM, which neutralizes the positive charge of this amino acid, changing protein function and physical-chemical characteristics. Acetylation has a crucial role in the regulation of gene expression through the modification of core histone tails by histone acetyltransferases (HATs) or histone deacetylases (HDACs), enzymes that remove the acetyl groups from both histones and other proteins. [52]. In general, acetylation of histones, made by HATs, leads to a more relaxed chromatin structure, allowing transcriptional activation, instead HDACs act as transcription repressors, leading to histone deacetylation, and consequently promoting chromatin condensation.

HDACs are important drug targets in cancer and in other degenerative diseases, because of their involvement in cell-growth and cell-death regulation such as in regulation of gene expression[53]. Although cancer is one of the most studied pathology connected to epigenetic modifications, some data link epigenomic alterations to variable disease phenotypes, which include fibroproliferative lung pathologies such as IPF. Inhibitors of HDAC (HDACi), molecules that block HDAC activity, were tested in preclinical studies with positive results in the resolution of fibrosis[54,55].

The high resolution and accuracy of mass spectrometry allows to study the acetylation sites of many proteins. As well as for phosphopeptides, it is important to enrich for the modified proteins, which, in the case of acetylated proteins, is done using antibodies directed against this modification.

Traditional methods used to characterize histone PTMs are mainly based on immunoassays, which include immunoprecipitation, immunofluorescence and western blotting. These methods are highly sensitive, but antibody specificity is affected by cross-reactivity and adjacent histone modifications[56].

Aim

Pharmacodynamic experiments require sensitive and accurate assays to evaluate the effect of drugs tested.

The aim of this work was to develop a method to evaluate the target engagement of a commercial HDAC inhibitor, used as reference for pulmonary disease, and study its effect during specific time points after administration.

A specific method was set up for the quantitative evaluation of acetylated histones H3 and H4 at different time points after the administration of the histone deacetylases pan-inhibitor in an in vivo pharmacodynamic experiment.

Results

Study design

The experimental design consisted of five time points (T0, 1h, 6h, 24h, 48h, 72h) after administration of the drug, an inhibitor of HDAC proteins (Panobinostat), in naïve rats. For each time point we analysed three biological replicates.

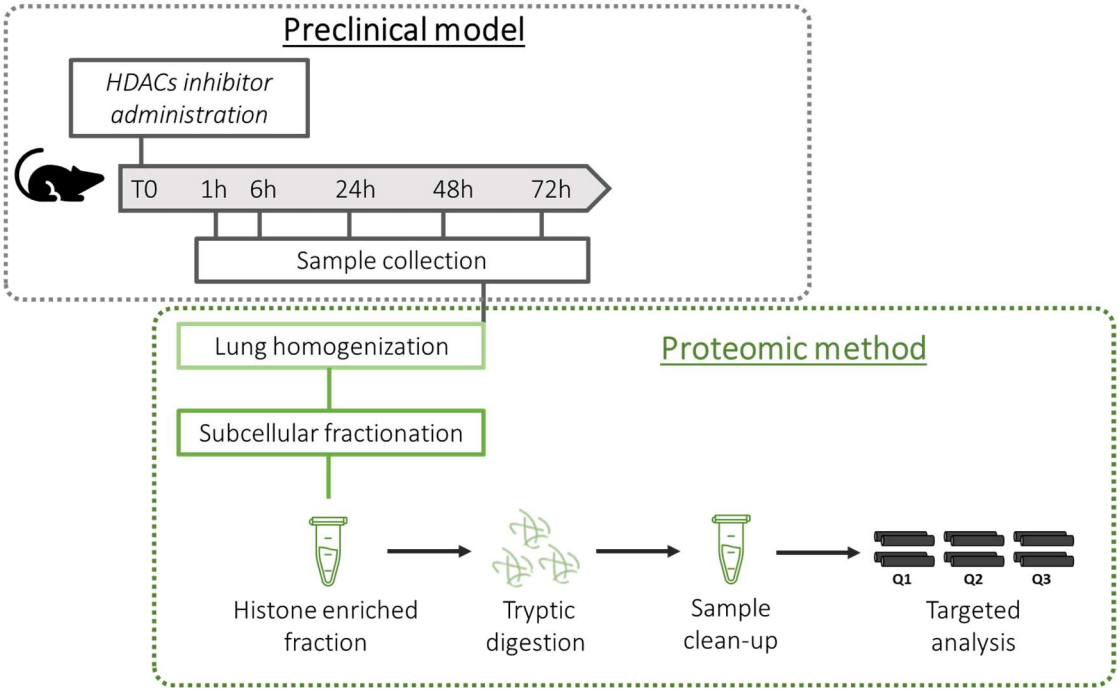


Figure 16:Experimental design and workflow for quantitation of acetylated histones in a PK/PD model testing a HDAC pan-inhibitor

Subcellular fractionation to enrich chromatin-bound proteins

To improve the detection of histones and, in particular, of acetylated histones, we added to sample preparation a subcellular fractionation step to enrich and isolate the chromatin-bound proteins.

To perform this additional step, we tested samples processed with the *EpiQuik™ Total Histone Extraction Kit (Abcam)*, a complete set of optimized buffers and reagents for extracting total core histone proteins (H2A, H2B, H3, and H4) from mammalian cells or tissues.

In the SDS-PAGE (Figure 17), representative samples enriched with extraction kit for each time point are showed. The protein profile is clearly enriched in histones (kDa range 12-16, red box in figure).

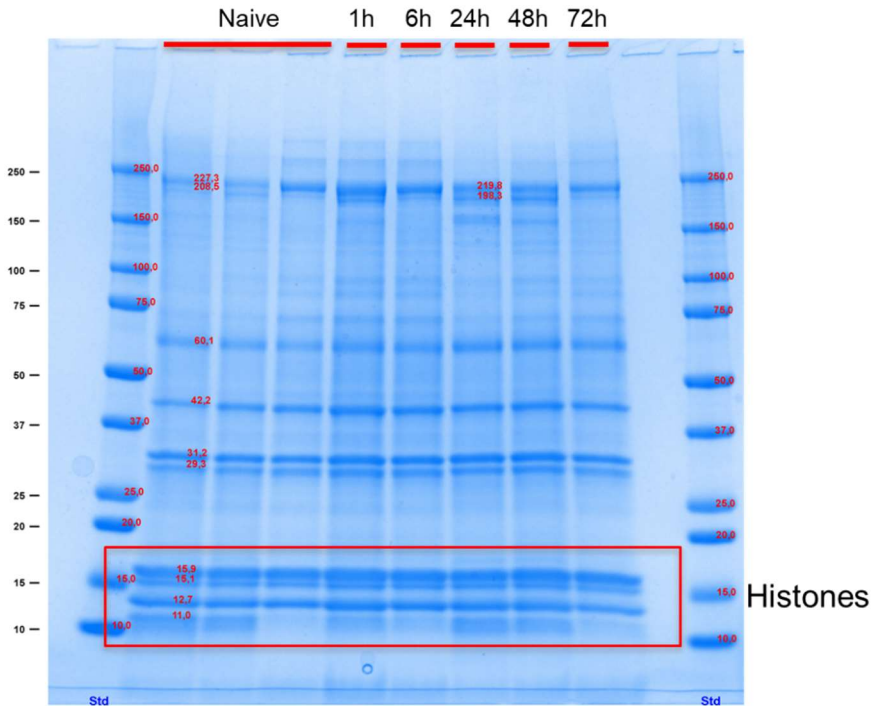


Figure 17: SDS-PAGE of lung homogenate enriched with Epiquick total histone extraction kit. Lane 1 and 11: Molecular weight marker Precision Plus Protein™ Dual Color Standards. The red box indicates histone bands.

For this study, lung samples were homogenized in PBS, the samples were then enriched with *the EpiQuik™ Total Histone Extraction Kit (assay performed by the Pharmacology Department).*

The sample, containing histone proteins, was denatured, reduced, alkylated and digested, as described in experimental section.

Targeted Quantitation of histone acetylation in PD model

The peptides of histone H3 and H4 with acetylated lysine (H3K14, H4K5, H4K8 and H4K12) were selected through the analysis of a pilot study performed to test TMT labelling. In this case, analysis with Proteome Discoverer was performed searching for the acetylation of lysine as dynamic modification.

Table 4: Proteotypic peptide selected for Histone H3 and Histone H4

Histone	*Peptide Sequence*	*Modification*	*Charge*	*Precursor (M/Z)*	*Transition (M/Z)*	
H4	DAVTYTEHAK		2	*567.774*	*849.4*	748.36
H4	DNIQGITKPAIR		3	*442.589*	*456.29*	685.43
H4	ISGLIYEETR		2	*590.814*	*810.39*	980.5
H4	GLGKGGAK	K12(Acetyl)	2	365.21	502.29	*559.31*
H4	GKGGKGLGK	K5(Acetyl);K8(Acetyl)	2	*443.26*	*658.38*	601.36
H3	STGGKAPR	K14(Acetyl)	2	408.22	728.4	*627.35*
H3	YRPGTVALR		3	344.86	559.35	*458.3*
H3	STELLIR		2	416.25	744.46	*643.41*
H3	EIAQDFK		2	425.71	721.38	*608.3*

To set-up the method, we selected also peptides of histones without acetylation, in order to normalize the abundance of acetylated peptides on the total protein abundance (Figure 18).

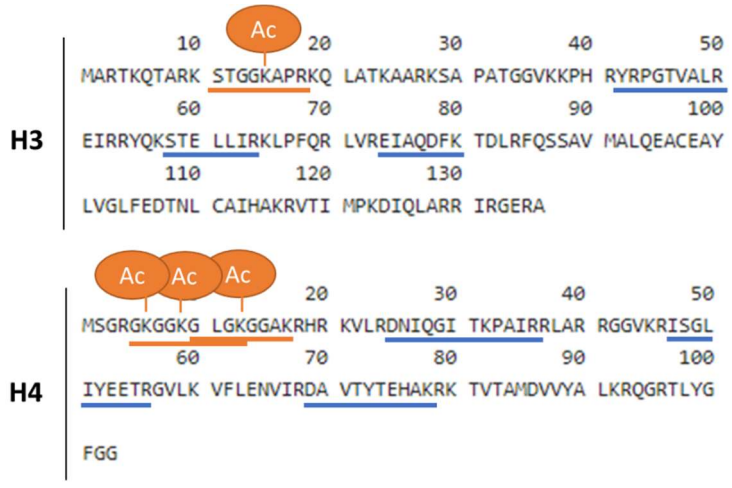


Figure 18: Amino acid sequence of Histone H3 and Histone H4: the peptides used in the targeted method are underlined. Acetylated peptides are underlined in orange, with the details of acetylated lysine detected, and the other peptides are underlined in blue

The targeted method was developed for the Q-trap 4000, coupled with HPLC.

The chromatographic conditions were set as described in the experimental section.

The abundance of acetylated peptides was normalized on the total abundance of the respective histones.

The results showed an increase in acetylation over time with a maximum at 24 hours after administration of the HDAC pan-inhibitor. At 72 hours after treatment a slow decrease of acetylation is detected, and high variability between animals (black dots) can be observed (Figure 19).

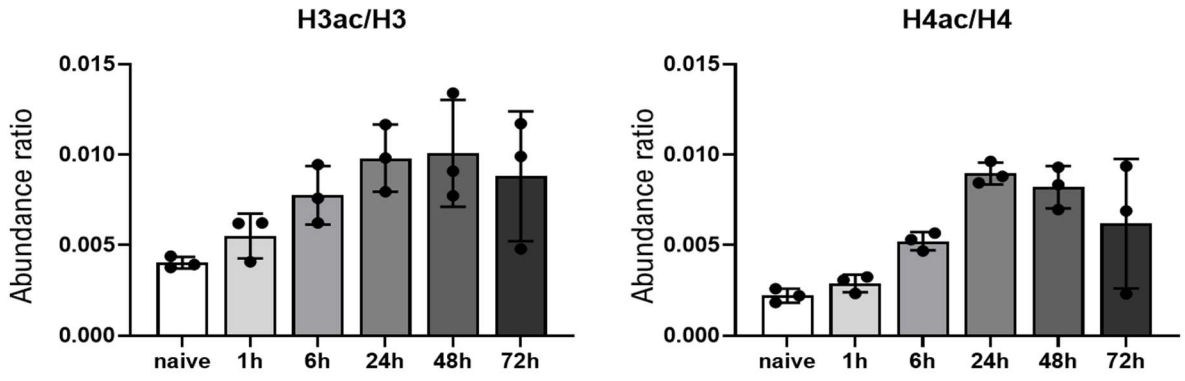


Figure 19: Acetylated peptide upon total histone ratio during time points after HDACs pan-inhibitor administration

Discussion

Histone acetylation quantitation with targeted proteomics

Early drug development's main goal is to choose promising compounds and possibly establish safe and effective dosages and dosing regimens.

Effective and successful pharmacokinetics/pharmacodynamics (PK/PD) studies during drug discovery are driven by inputs from complementary scientific disciplines involved in preclinical research.

Proteomic approaches can be used to identify relevant biomarkers that closely associate with the examined drug[57]. With this purpose, we developed a targeted method based on MRM experiment to evaluate the target engagement of a histone deacetylase pan-inhibitor tested in a PK/PD study in rat.

HDACs remove the acetyl groups from acetylated histones, increase chromatin condensation, and suppress gene transcription. Inhibition of HDACs induces general histone acetylation.

Histone modification is a reversible process and is commonly described as epigenetic modification. HDACs, as well as HATs, are involved in many cellular processes. Dysfunction of the HDACs is associated with many diseases (cancer, diabetes, tissue fibrosis and cardiac hypertrophy)[55].

Different MS-based methodologies have been employed to study histone and histone PTMs. Liquid chromatography (LC) MS is a sensitive, accurate and fast method that can be used for intact protein analysis, instead peptide mass finger-printing is an efficient method to map histone PTMs using mass correlation based on different modification groups. To obtain sequence information or site position of PTMs, the best experiments to perform are tandem MS (MS/MS).

From untargeted data, obtained with a Orbitrap spectrometer, we selected the peptides of interest for H3 and H4 with specific lysine acetylation on the N-terminal tail. We developed a targeted MRM method to simplify the experiment and data analysis, and to focus on interesting proteotypic peptides.

In order to focus on histones, as did for collagen, a fractionation method was used to extract the chromatin-bound proteins. The goal of fractionation is to simplify the biological samples, especially for lung homogenates, which are composed by more than 40 cellular types and thousands of different proteins.

Many studies have been performed to evaluate the effect of the HDAC inhibitor here considered, Panobinostat, commonly with western blot assays, and the results have shown a progressive increase in histone H4 acetylation correlating with duration of treatment with the drug[58], while flow cytometry results

on patient samples confirmed that HDAC pan-inhibitor significantly induces acetylation of histone H3 and histone H4 at 6 hours post HDAC pan-inhibitor dose, with a slow decrement after 24 hours[59].

The results obtained with the targeted MRM method developed in this thesis work confirmed the effect of the HDAC pan-inhibitor tested on naïve rats. Acetylation of histone H3 and histone H4 increased until 24 hours post treatment, and the decreasing of acetylation was detectable after 48 and 72 hours. These results can help to evaluate the dosing regimen for this HDAC inhibitor. Moreover, this method was used to study other pharmacologic models testing HDACs inhibitors.

The *plus* that this method brought to the evaluation of HDACs target engagement is the exact position of the acetylated lysine monitored, unlike the antibody-based technique tested in the pharmacology department.

Lung development Proteome profiling

Background

Quantitative proteomics: TMT labeling

Isobaric stable isotope labeling with tandem mass tags (TMTs) is being increasingly adopted in large-scale proteomic investigations. The reproducible measurement of single peptides across samples is extremely useful in experiments focusing on protein profiling during drug time course or development.

The tags and the analysis methods make it easy and accurate to identify peptides from various samples and measure their relative abundance. Using a collision induced dissociation (CID)-based analytical approach, the TMTs enable simultaneous determination of both the identity and relative abundances of peptide pairs. TMT labeling enables investigation with peptide-based protein characterization techniques[60].

Lung development

A detailed knowledge of the respiratory system development and of the biological process involved could help in research on preterm-related diseases such as Bronchopulmonary dysplasia (BPD), chronic obstructive pulmonary disease (COPD) and respiratory distress syndrome (RDS)[61]. The respiratory system is composed by trachea and airways, which ramify extensively within the lungs. Lungs are characterized by a branched network of epithelial tubes forming bronchi, bronchioles and alveoli, where blood gas exchange happens. These complex structures are made of different specialized epithelial cell types, which cooperate for air intake and gas exchange, protection from pathogens and maintenance of fluid and electrolyte homeostasis.[62]

The complex branched structure and the associated vasculature of the respiratory system develop through the spatiotemporal coordination of more than 40 different cell types and their molecular interactions.[63]

In order to expand information about the developmental stages of mammalian respiratory system, different animal models, from rodents to baboons, are studied.

Lung developmental stages

Mammalian lung development is traditionally described as a process of five stages (embryonic, pseudoglandular, canalicular, saccular and alveolar) characterized by specific histological features, cell type organization and morphology (Figure 20).

The lung development initiates with the embryonic stage: lung buds arise from the ventral foregut endoderm and, subsequently, septation takes place from the esophagus.

The bronchial tree expands during the pseudoglandular phase. This stage is characterized by branching morphogenesis of the bronchi and by the initial formation of alveolate saccules [63]. Branching morphogenesis continues the with the growth of the lungs and the elongation and subdivision of the terminal bronchioles, generating the bronchial tree.

The canalicular stage completes the formation of the terminal bronchioles of the five lung lobes. The pulmonary parenchyma is formed by the expansion of canaliculi with the associated capillary network.

Saccular stage starts with the increase in vascularization and consequent establishment of saccules. In this stage the formation of primitive respiratory sites for gas exchange take place, along with the expansion of lungs and the related lymphatic network.

The last stage is the alveolar stage and occurs post-birth. In this stage the surface for gas exchange increases and the alveolar ducts became mature thanks to the maturation of the surfactant system.[63,64]

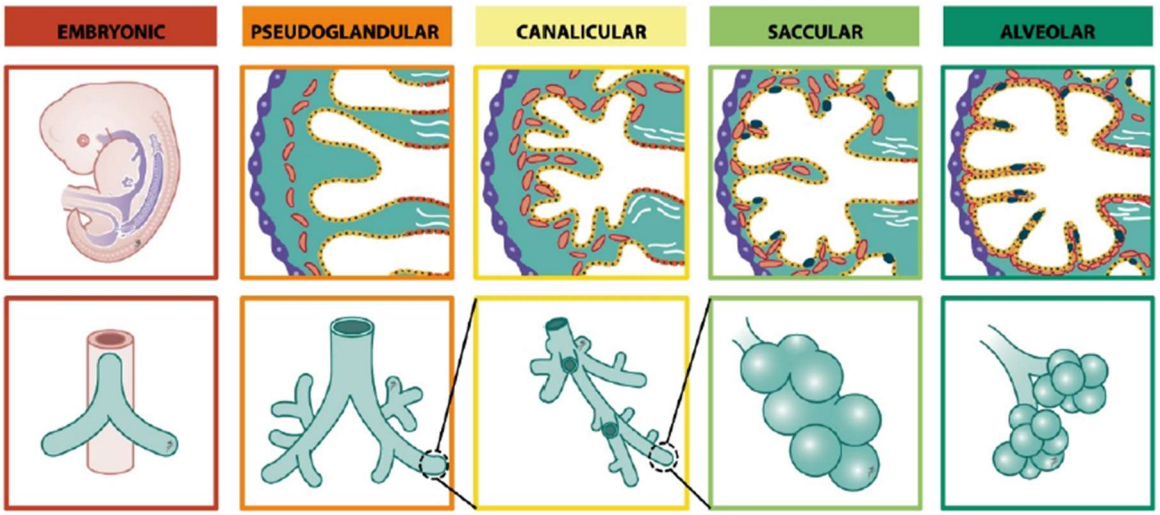


Figure 20: Mammalian lung development, from Salaets et al., 2017 [64].

Rabbit models to study lung development

Rabbits are medium-sized animals and their main advantage as a model to study lung development is that, besides having a short gestation time (31 days), they strictly reflect what happens in humans. Differently from smaller rodents (mice, rats) and similarly to humans, in rabbits alveolarization starts in-utero and term rabbit pups are born in the alveolar phase of lung development, thus with lungs presenting both structural and functional features of mature lungs[64]. As a result, the major strength of the rabbit over the smaller rodent

models is the possibility to induce prematurity in order to investigate the pathophysiology of preterm birth diseases.

Preterm rabbit pups born at day 28 are in the saccular stage and show signs of respiratory distress at birth, nevertheless they can survive without surfactant administration or mechanical ventilation[61]. Preterm pups are unable to upregulate antioxidant enzymes in defence of a hyperoxia insult. In rabbits, the foetal increase in antioxidant levels coincides with a steep maturation of the surfactant system and a swift development of the foetal lung from late canalicular, over saccular, to the alveolar stage.

Limitations of rabbit models regards the poor availability of antibodies to perform Western Blot analysis or ELISA assay. Additionally, rabbit proteome is nowadays described in Proteome Database, but the reviewed proteins are only 894, over the 41459 described in the UniprotKB (data from https://www.uniprot.org/proteomes/UP000001811).

Aim

The specific aim of the lung development study in rabbit model was to characterize the proteome profile of developmental stages and identify key molecular players and their modulation within time. The study of proteome in the developing lung is an expanding field, above all in murine model, but less is known about the rabbit model despite the similarity with the human lung development. Proteome profiling studied within a time course experiment of a developing organ, in this case lung, helps to understand all the different biological processes involved. Moreover, proteome profiling helps in the investigation of interesting proteins highlighting the role in growth, metabolic changes and development.

The general aim of this part of the work was to implement the use of quantitative proteomics, an emerging field of science, to respond to the numerous preclinical questions related to different animal models.

The following study is used as a case study to describe the development of this new tool to study diseases related to premature birth.

Before performing proteome profiling of developing rabbit lung, a pilot study was conducted on the IPF bleomycin-induced rat model previously described, in order to evaluate the applicability of TMT proteomics to a preclinical model, considering a limited number of samples.

Results

Pulmonary fibrosis bleomycin-induced rat model proteome profiling

Study design

Three rat samples of IPF bleomycin-induced model (Vehicle (V), Bleomycin (B) and Nintedanib (N), sacrified at day 28) were treated with TMT10plex method, after tryptic digestion, and the analysis was performed with nanoLC-MS analysis with Dionex Ultimate 3000 nano RSLC/Lumos at IRB in Barcelona. The stablished IPF bleomycin-induced fibrosis was evaluated with histological analysis (not shown) and with collagen quantitation as described in the chapter "Quantitation of Fibrillar collagens in a pulmonary fibrosis rat model".

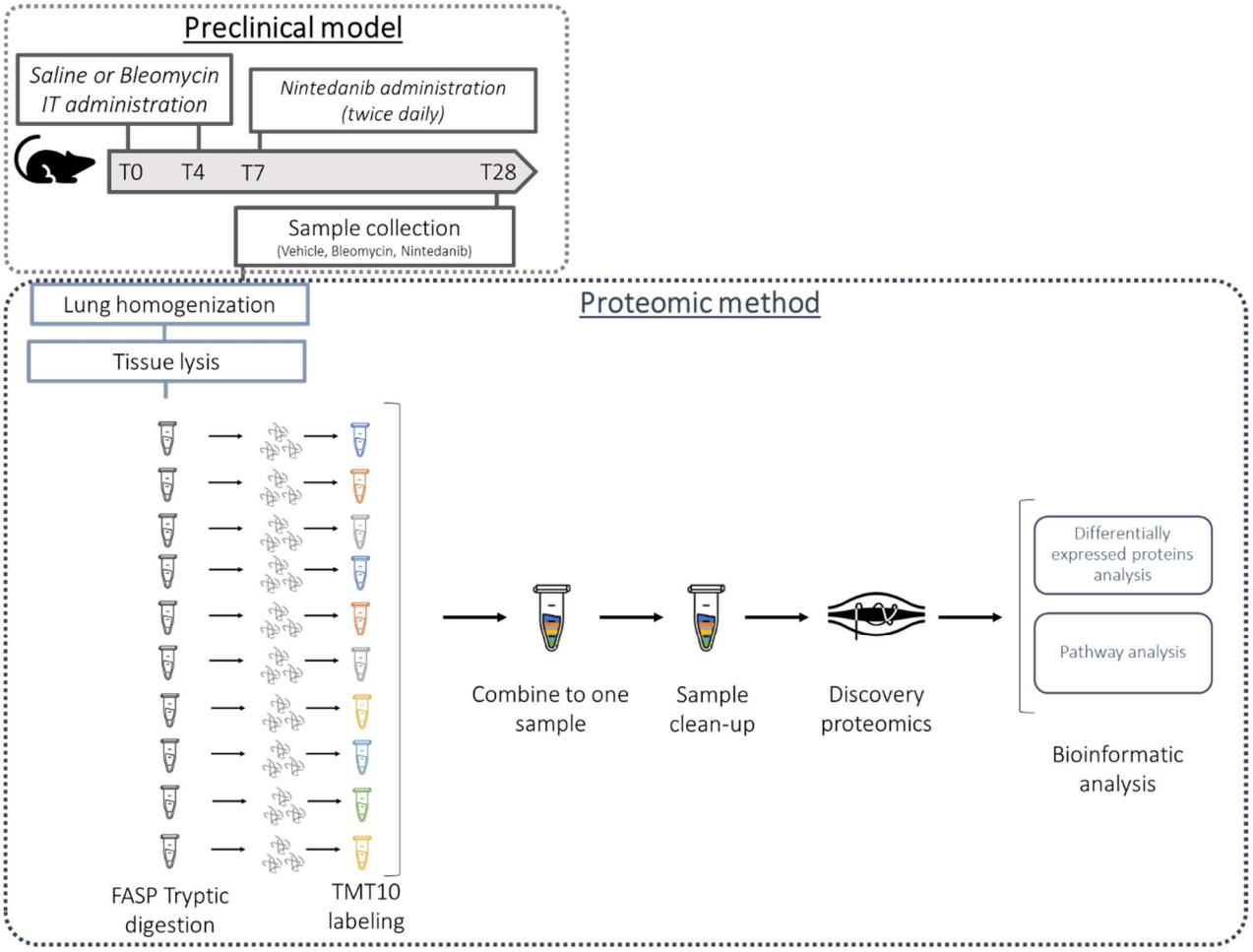


Figure 21: Experimental design and workflow for TMT pilot study on IPF bleomycin-induced rat model

The three samples were subdivided into three analytical replicates, then digested with the FASP protocol and, after sample clean-up, they were labelled with the TMT10 plex kit. The tenth sample was a pool of the three labelled samples.

TMT	126	127N	127C	128N	128C	129N	129C	130N	130C
Sample	Pool	N3	B1	B2	V3	B3	N2	V2	N1

Peptides were analyzed using an Orbitrap Fusion Lumos™ Tribrid mass spectrometer (Thermo Scientific) equipped with a Thermo Scientific Dionex Ultimate 3000 ultrahigh pressure chromatographic system (Thermo Scientific) and an Advion TriVersa NanoMate (Advion Inc. Biosciences) as the nanospray interface.

Database searches were performed with Proteome Discoverer v2.1.0.81 software (Thermo Scientific) using Sequest HT search engine and UniProt *Rattus norvegicus* (2018_09) and contaminants. Search parameters included trypsin, allowing for two missed cleavage sites, carbamidomethylation of cysteine and TMT 10plex modification of peptide N-terminus as static modifications and TMT 10plex at lysines, methionine oxidation and acetylation at protein N-terminus as dynamic modifications.

Differential expression analysis and pathways analysis

From the seventy thousand spectra detected in this experiment, Proteome Discoverer identified and quantified 3234 proteins.

Protein quantitative measurements were $\log_2$ transformed. Analysis of variance (ANOVA) was performed at protein level using a linear model. Weighting function was used to allow data variability to depend on data value. Comparisons of Bleomycin and Nintedanib samples upon Vehicle samples were performed. Finally, p-values were adjusted for multiple testing using the Benjamini & Hochberg FDR correction.

Differential expressed proteins were determined using an adjusted p-value cutoff of 0.05 and a fold change value lower than 0.8 (down-regulated) or higher than 1.25 (up-regulated).

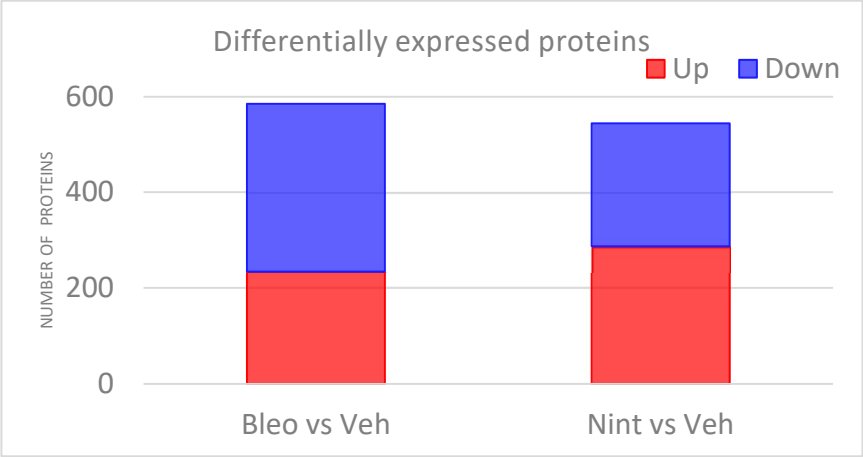


Figure 22: Differentially expressed proteins in IPF bleomycin-induced model

The lists of differentially expressed proteins were submitted to Metascape software to identify pathways modulated in the IPF bleomycin-induced rat model and following treatment with Nintedanib. Several processes appeared to be resolved after Nintedanib treatment, such as the collagen formation. Others, like extracellular matrix organization, were still significant in the N vs V comparison but to a much lesser extent.

Table 5: Pathways analysis results

ID [1]	Description	Log p-val B vs V	Log p-val N vs V
Up-regulated proteins			
GO:0002283	neutrophil activation involved in immune response	-19	-11
GO:0043312	neutrophil degranulation	-19	-11
R-HSA-6798695	Neutrophil degranulation	-19	-11
GO:0002446	neutrophil mediated immunity	-18	-10
GO:0036230	granulocyte activation	-18	-10
M5930	HALLMARK EPITHELIAL MESENCHYMAL TRANSITION	-19	-12
R-HSA-1474244	Extracellular matrix organization	-19	-11
GO:0030198	extracellular matrix organization	-18	-8.5
GO:0043062	extracellular structure organization	-18	-8.2
M5884	NABA CORE MATRISOME	-18	-13
M5921	HALLMARK COMPLEMENT	-8.2	-10
R-HSA-3000178	ECM proteoglycans	-10	-4.5
M18	PID INTEGRIN1 PATHWAY	-8	
R-HSA-1474228	Degradation of the extracellular matrix	-11	-4.4
R-HSA-1474290	Collagen formation	-6.7	
hsa04142	Lysosome	-9.9	-6.8
GO:0006457	protein folding	-6.6	
M5924	HALLMARK MTORC1 SIGNALING	-6.3	
R-HSA-2129379	Molecules associated with elastic fibres	-5.9	-6.8
R-HSA-1566948	Elastic fibre formation	-5.4	-6.2
hsa04610	Complement and coagulation cascades		-7.8
GO:0070613	regulation of protein processing		-6.1
GO:1903317	regulation of protein maturation		-6.1
Downregulated proteins			
CORUM:3055	Nop56p-associated pre-rRNA complex	-20	
R-HSA-2408557	Selenocysteine synthesis	-18	
CORUM:306	Ribosome, cytoplasmic	-17	
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	-16	
GO:0006613	cotranslational protein targeting to membrane	-15	
R-HSA-422475	Axon guidance	-15	
R-HSA-109582	Hemostasis	-9.6	-18
R-HSA-76002	Platelet activation, signaling and aggregation	-4.6	-8.8
GO:0042060	wound healing	-6.2	-10
GO:0009611	response to wounding	-5.8	-9.9
GO:0007596	blood coagulation	-7.4	-6.3
GO:0007599	hemostasis	-7.3	-6.2
GO:0050817	coagulation	-7.2	-6.2
GO:0030029	actin filament-based process	-9	-4.1
GO:0030036	actin cytoskeleton organization	-7.9	
GO:0006936	muscle contraction	-7.7	-6.8
GO:0003012	muscle system process	-7	-5.8
hsa04540	Gap junction	-6.3	-6.5
GO:0031032	actomyosin structure organization	-6.2	
GO:0008015	blood circulation		-9.7
GO:0003013	circulatory system process		-9.6
GO:0010817	regulation of hormone levels		-8.5

[1]Blue pathway IDs highlighted in Table 5 indicate pathways that resolve following Nintedanib treatment. Yellow pathway IDs indicate pathways specific for the Nintedanib vs vehicle comparison.

TMT quantitative analysis with nanoLC coupled with Lumos, a high-resolution mass spectrometer, was compared with the analysis performed on the Vanquish/Q-Exactive system, described in the paragraph *Preliminary untargeted analysis to identify specific targeted peptides*, in order to find out the advantages of each technique.

The parameters obtained from results highlight the superiority of Dionex Ultimate 3000 nano RSLC/Lumos range, based on the number of MS/MS spectra collected, that reflects the number of proteins identified and the quality of identification, on the quantity of protein needed to perform the analysis, and based on the time of sample analysis (Table 6).

For these reasons, we decided to use this instrument to perform the experiments of discovery proteomics.

Table 6: Comparison between two high resolution instruments for discovery proteomics

	Vanquish/Q-Exactive	**Dionex Ultimate 3000 nano RSLC/Lumos**
N° of MS/MS spectra collected	7000/8000	~70,000
µg of proteins	5-10	0.6
Time for analysis of 10 sample	1700 min (~28 hours)	300 min (5 hours)
N° of identified proteins	1000/1200	~3000
Full quantitative analysis	Not applicable	TMT
Pathway analysis (IPA)	Not applicable	Metascape

Lung development proteome profiling

Study design and proteomic workflow

To perform a quantitative untargeted proteomics analysis, the TMT method was applied to rabbit lung samples, derived from a study of lung development, in collaboration with Mass Spectrometry and Proteomics core facility at IRB, Barcelona.

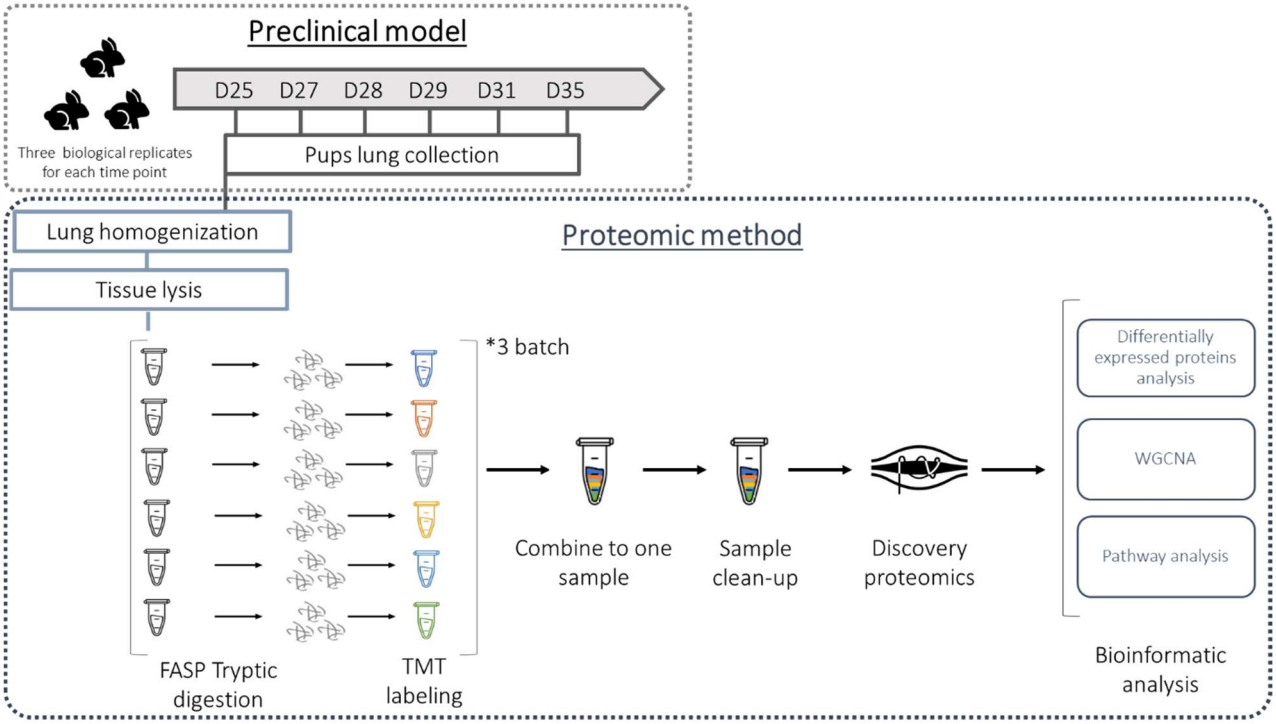


Figure 23: Experimental design and workflow for TMT quantitative proteomics of rabbit lung development

Samples were homogenized in phosphate buffered saline. SDS and DTT were added and samples were sonicated, heat-denatured and centrifuged to extract proteins. After protein quantitation, sample were digested with Trypsin following FASP protocol.

An aliquot of each sample was dried, reconstituted and labelled with TMT6plex (Thermo Scientific), which allows 6 different labels. The 18 samples were labeled using the TMTs described in Table 7. Three batches were created, being careful to insert one biological replicate for each time point per batch and to change the correspondence with the label.

Table 7: TMT groups created for mass spectrometry analysis

TMT	126	127N	128C	129N	130C	131
Group 1	D25.6	D29.6	D28.5	D31.5	D35.4	D27.5
Group 2	D29.9	D25.15	D27.9	D28.8	D31.12	D35.8
Group 3	D28.10	D31.16	D29.13	D35.12	D27.13	D25.19

Peptides were analyzed using an Orbitrap Fusion Lumos™ Tribrid mass spectrometer (Thermo Scientific) equipped with a Thermo Scientific Dionex Ultimate 3000 ultrahigh pressure chromatographic system (Thermo Scientific) and an Advion TriVersa NanoMate (Advion Inc. Biosciences) as the nanospray interface.

The mass spectrometer was operated in a data-dependent acquisition (DDA) mode.

For the MS3 analyses for TMT quantification, multiple fragment ions from the previous MS2 scan (SPS ions) were co-selected and fragmented by HCD using a 65 % collision energy and a precursor isolation window of 2 Da.

Protein quantitation and differentially expression analysis

Dataset consisted of six time points (D25, D27, D28, D29, D31, D35) and three biological replicates per day.

Database searches were performed with Proteome Discoverer v2.4.1.15 software (Thermo Scientific, new version released in 2020) using Sequest HT search engine against UniProt *Oryctolagus cuniculus* reference proteome (2020_04) and common contaminants. Search parameters included trypsin cleavage (allowing for two missed cleavage sites), carbamidomethylation of cysteine and N-terminal TMT 6plex peptide as static modification; TMT 6plex at lysine, hydroxylation of proline (to better identify the collagens), methionine oxidation and acetylation at protein N-terminus as dynamic modifications. Peptide mass tolerance was 10 ppm for MS1, 0.6 for the MS2 and 20 ppm for the MS3. SPS (Synchronous Precursor Selection) percentage was set at 55%. Peptides with a q-value lower than 0.1 and FDR < 1% were considered as positive identifications with a high confidence level. TMT reporter ion intensities were used for protein quantification. Data were normalized in the three TMT-6plex experiments using quantile normalization and statistical analysis was performed using the Perseus software.

TMT reporter ion intensities were $\log_2$-transformed and filtered to retain only proteins with valid quantification values in at least 4 over 6 samples per development stage (25-27: canalicular; 28-29: saccular; 31-35: alveolar). Missing values were imputed by replacing missing values from normal distribution with a width of 0.3 and a down-shift of 1.8 standard deviations. A t-test was performed on the following comparisons: D25-D35, D27-D28, D28-D31, canalicular-saccular, saccular-alveolar and canalicular-alveolar.

P-values were corrected with a permutation-based FDR of 0.05. Differentially expressed proteins (DEPs) were determined using an adjusted p-value cutoff of 0.05 and a fold change > 1.5.

The database searching for the new proteome identified a thousand more proteins:

- 4367 proteins were identified with the new database searching.
- 64035 peptide spectrum matches, corresponding to 28427 peptide groups, were found.
- 2796 proteins were quantified in at least 4 out of 6 samples per developmental stage.

Protein groups quantified in at least 4 out of 6 samples per developmental stage were used for the quantitative and statistical analysis. Data visualized using principal component analysis are shown in Figure 24.

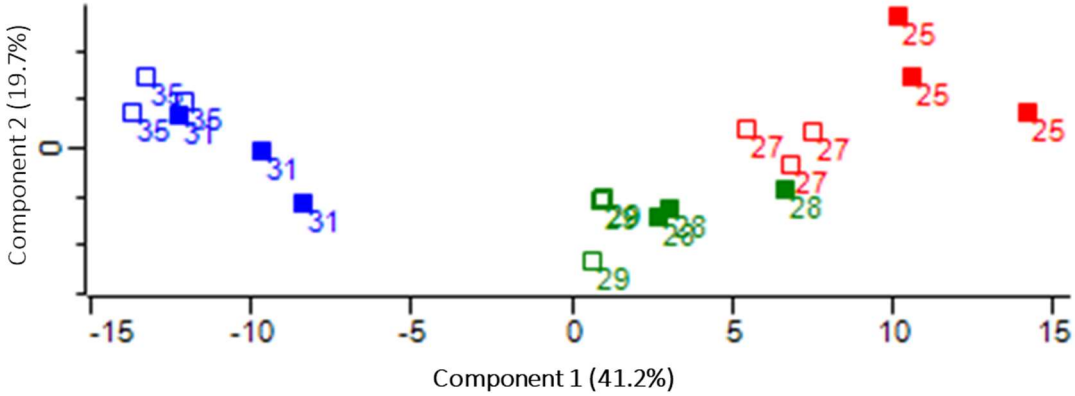


Figure 24: PCA of quantitative results. From the right side: D25 red filled square, D27 red square, D28 green filled square, D29 green square, D31 blue filled square, D35 blue square.

The highest number of DEPs could be found between the more distant time points (25-35) and the more distant developmental stages analyzed (Table 8).

Table 8: Number of differentially expressed proteins found with new proteome annotation

	25-35	28-31	27-28	CANALICULAR-SACCULAR	CANALICULAR-ALVEOLAR	SACCULAR-ALVEOLAR
UP	99	43	7	47	88	52
DOWN	146	118	14	44	143	134

Protein profiles analysis along lung development

The WGCNA analysis was performed on abundance-normalized values derived from TMT data, after a further normalization with IRS-TMM method, using different R packages according to an existing pipeline (github.com/pwilmart/IRS_normalization).

Only protein quantified in all the 18 samples were used for module construction. Modules were identified using the blockwise-modules function. The identification of pathways and processes enrichment in modules was performed using Metascape software, as described previously.

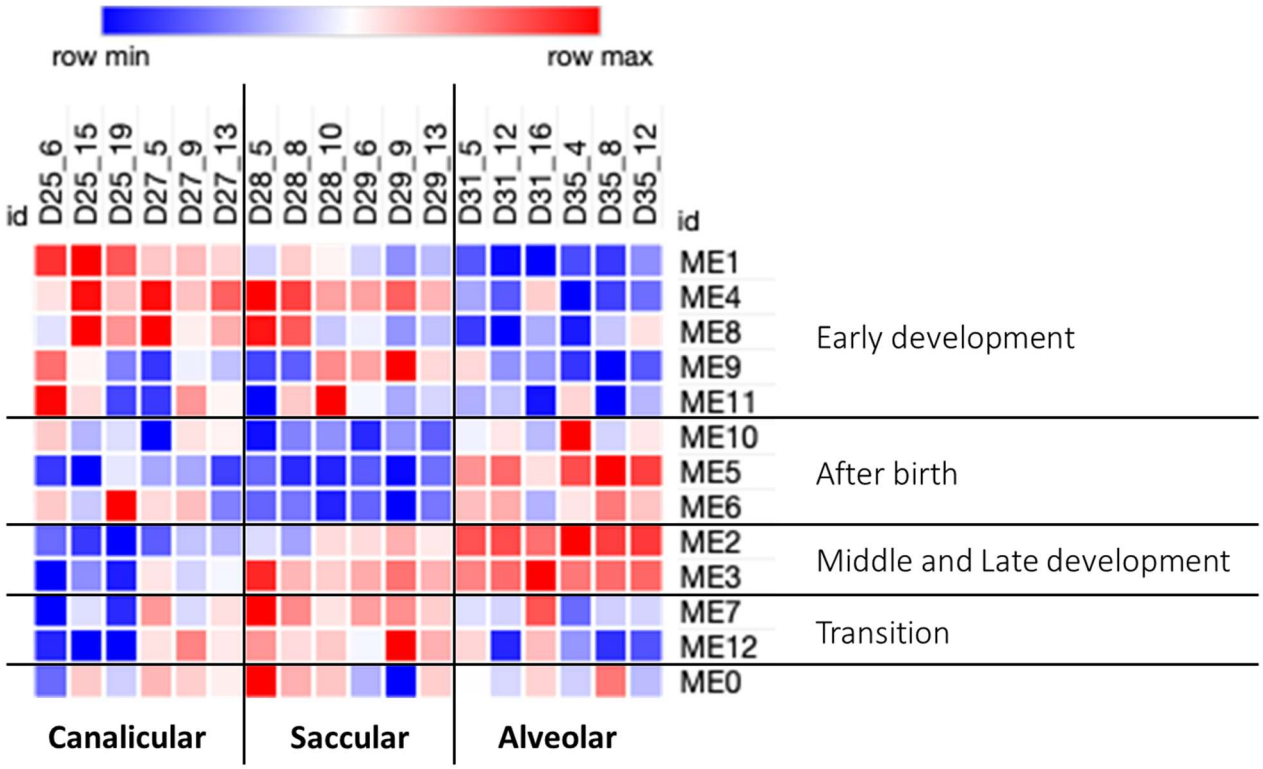


Figure 25: Module profiles. Module Eigengene heat map with new proteome database searching

The eigengene heat map identifies four different module subgroups from the twelve modules (Figure 25).

Protein modules 1, 4, 8, 9 and 11 show a decrease of abundance during lung development, with high abundance in canalicular stage (days 25-27), intermediate in saccular stage and a lower abundance in alveolar stage. These modules could be considered enriched in proteins specific for early development of rabbit lung (Table 9). The most significative pathways enriched in these modules are related to metabolism of RNA and cell cycle, post-translational protein modification, the citric acid (TCA) cycle and respiratory electron transport (especially for module 9) and chromatin assembly (for module 11).

Proteins of modules 10, 5 and 6 increase in abundance after birth, in alveolar phase (days 31-35). The most significative pathways enriched concern vesicle mediated transport, cellular detoxification, focal adhesion, supramolecular fiber organization and extracellular matrix organization (Table 10).

Table 9: Pathways significantly enriched in the early development modules cluster

	Early development				
Description	**ME1**	**ME4**	**ME8**	**ME9**	**ME11**
Metabolism of RNA	-84	-14	-2.8		-6.3
Translational initiation	-55	-4.2	-2.3		
RNA splicing	-51	-3.7			-2.9
Cell Cycle	-19	-12			-10
RNA transport	-16	-10	-2.4		
Nuclear transport	-14	-14	-2.8		
DNA conformation change	-13	-1.7			-11
Protein processing in endoplasmic reticulum	-9.4	-6.4		-1.6	
Chromatin assembly	-3.5				-11
Cellular amino acid metabolic process	-3.5	-23	-5.3		
Golgi-to-ER retrograde transport	-2.9		-10		
Post-translational protein modification	-2.6	-20	-2.1		
The citric acid (TCA) cycle and respiratory electron transport		-2.4		-33	
Cytosolic trna aminoacylation		-17	-5.3		

Table 10: Pathways significantly enriched in modules characterizing processes after birth

	After birth		
Description	**ME5**	**ME6**	**ME10**
Vesicle-mediated transport	-9.9	-2.2	-2
Detoxification of Reactive Oxygen Species	-8.3		
apoptotic signaling pathway	-7.8	-2.5	-2.8
supramolecular fiber organization	-7.2	-2.8	-6.3
Extracellular matrix organization			-33
Focal adhesion			-19
Assembly of collagen fibrils and other multimeric structures			-14

The proteins involved in middle and late development are described in Modules 2 and 3, and their quantity increases from saccular to alveolar stages. The enriched pathways for these modules are regulated

exocytosis, leukocyte mediated immunity, cell junction organization, regulation of actin in cytoskeleton and wound healing (Table 11). Surfactant proteins belong to module 2.

Modules 7 and 12 represent a transition between stages. Indeed, proteins belonging to these modules reach the highest abundance during saccular stage. Regulation of Insulin-like Growth Factor (IGF) transport and uptake, vesicle-mediated transport and endocytosis are the pathways enriched for these two intermediate modules (Table 12).

Table 11: Pathways significantly enriched in middle and late development modules cluster

	Middle and late development	
Description	**ME2**	**ME3**
Regulated exocytosis	-25	-22
Myeloid leukocyte mediated immunity	-24	-20
Regulation of cytoskeleton organization	-9.9	-20
Supramolecular fiber organization	-18	-14
Response to wounding	-14	-8.4
Hemostasis	-16	-6.2

Table 12: Pathways significantly enriched in transition modules cluster

	Transition	
Description	**ME7**	**ME12**
Vesicle-mediated transport	-9	-3.8
Generation of precursor metabolites and energy	-6	
Starch and sucrose metabolism	-5.5	
Endocytosis	-5.2	
Regulated exocytosis	-2.5	-7.2
Regulation of peptidase activity	-2.2	-4.6
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)		-6.6

Analysis of surfactant proteins

Within the thousands of proteins identified, the pulmonary surfactant proteins B and C deserve a specific focus because of their role in lung development. Both are processed by cathepsin H into their mature forms that have a key role in lung development and function. Indeed, surfactant proteins B (SP-B) and C (SP-C) are small hydrophobic proteins that are deeply embedded into the surfactant phospholipids; they enhance

interfacial adsorption of surface-active molecules into the air-liquid interface and contribute to mechanical stability of the interfacial films[65].

The quantitative proteomic analysis allows to investigate the proteome, but also the peptide abundance for each identified protein. This focused analysis is very useful for discriminating the mature form of interesting proteins, such as surfactant proteins.

Surfactant protein B

In ATII cells, SP-B is synthesized as a 40–42 kDa precursor that is cleaved by Napsin A. The resulting intermediate of 23–26 kDa reaches multivesicular bodies (MVB) through the Golgi. Two enzymes, Cathepsin H and Pepsinogen C, seem to be involved in the final processing of SP-B which takes place in intermediate composite bodies to lamellar bodies[65]. The mature protein is an 8 kDa peptide, residues 185 - 263 in the rabbit protein. Cathepsin H performs cleavage at residues 172, 263, 296, 333 and 362, instead Napsin A performs cleavage at residue 162[66] (Figure 26).

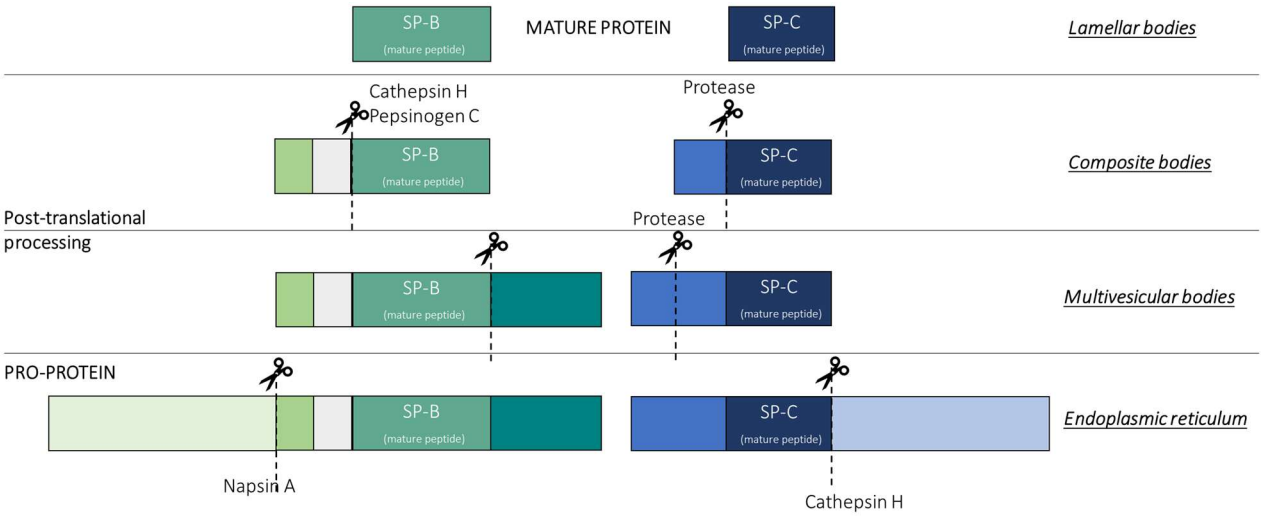


Figure 26: Post-translational processing of Surfactant protein B and Surfactant protein C

The abundance values of the SP-B protein showed an increase over time during lung development, with a maximum in the alveolar stage, at day 31 and 35 (Figure 27).

The peptide abundance analysis revealed the differences between the mature peptide and the pro-peptides, cleaved by different proteases. Propeptide 86 - 107 showed an increase from day 29, instead propeptide 136 - 159 and propeptide 290 - 299 did not show variations over time (Figure 28). The mature peptide showed a sharp increase from day 31, at birth (the beginning of alveolar stage) (Figure 29).

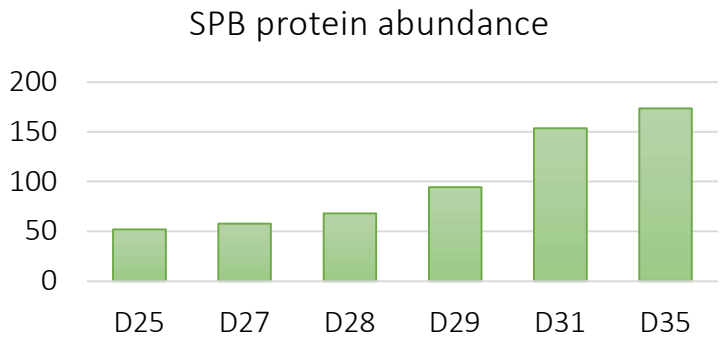


Figure 27: SPB protein abundance during lung development

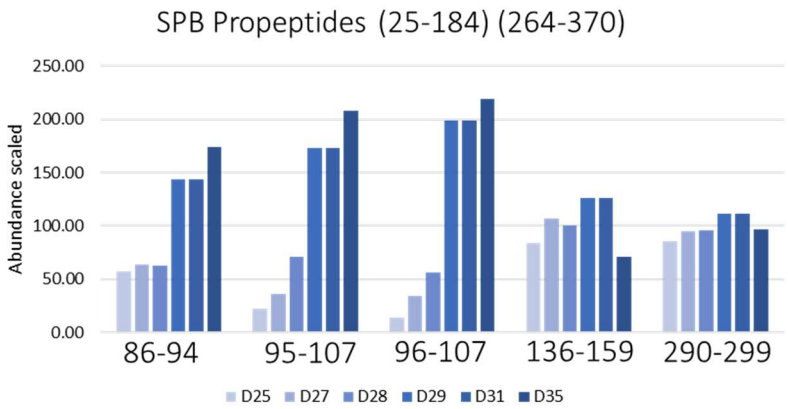


Figure 28: Abundance of SPB pro-peptides during development

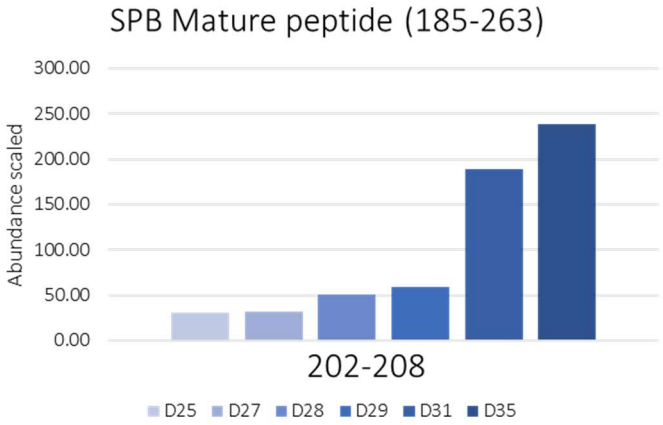


Figure 29: Abundance of SPB mature peptide during lung development

Surfactant protein C

SP-C is also synthesized in the endoplasmic reticulum as a 21 kDa precursor that is palmitoylated in the Golgi apparatus. Unknown enzymes cleave the C-terminal domain and later Cathepsin H probably cleaves the N-terminal domain resulting in the mature protein at the lamellar bodies (Figure 26). The absence of SP-B alters

processing and sorting of SP-C, reflecting the connection between processing and trafficking of both proteins in ATII cells[65].

The abundance values of the SP-C protein showed an increase in the course of lung development, with the maximum at day 29, and reached a plateau in alveolar stage (Figure 30).

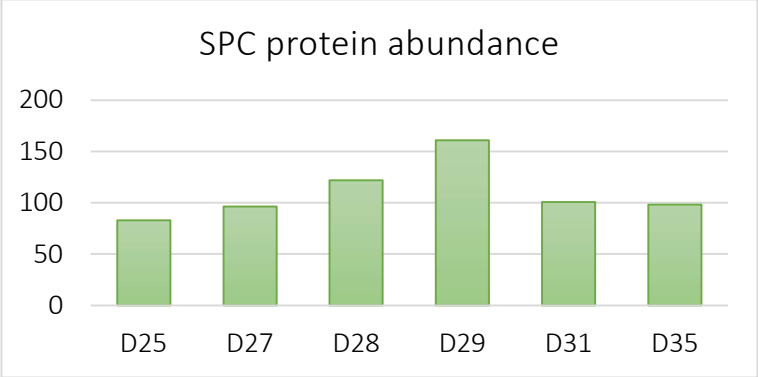


Figure 30: Surfactant protein C abundance during lung development

The peptide abundance analysis revealed differences between the mature peptide and the pro-peptides, cleaved by different proteases. Pro-peptides showed an increase from day 28 to day 29, and a decrease during alveolar stage (Figure 31). The mature peptide showed a sharp increase from day 29, with a maximum at day 31, at birth (Figure 32). The mature peptide was identified without the double palmitoylation on cysteine 28 and 29, a specific post-translational modification of SP-C.

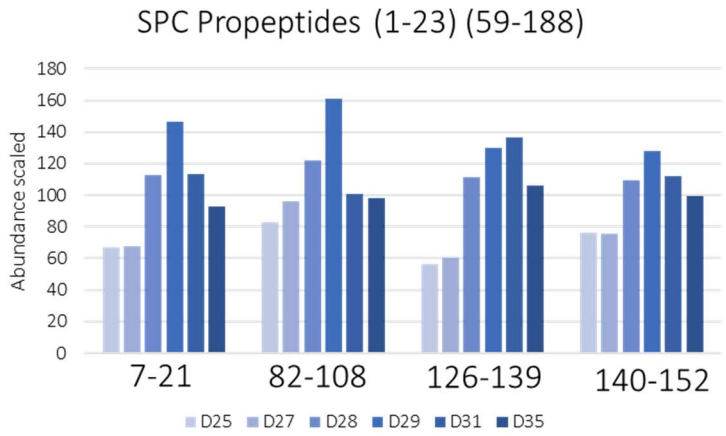


Figure 31: Abundance of SPC pro-peptides during development

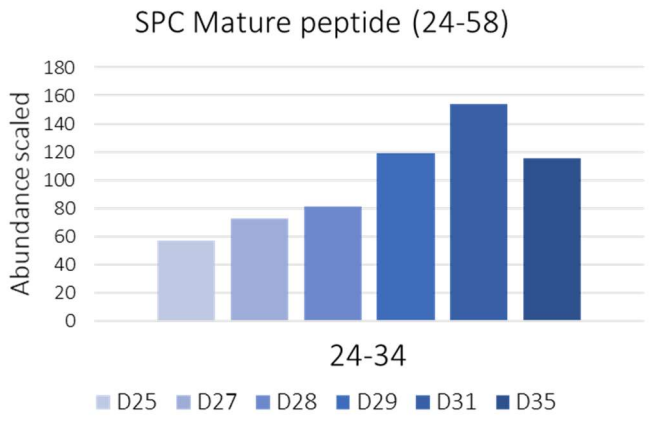


Figure 32: Abundance of SPC mature peptide during lung development

Discussion

Proteomic profile of Bleomycin-induced pulmonary fibrosis in rat model

In this study, we identified a cohort of pathways that are deregulated in the Bleomycin rat model of pulmonary fibrosis, using a combination of discovery proteomics and bioinformatic analysis.

More importantly, using Nintedanib as a potential therapeutic agent, we identified a set of important pathways that can serve as possible therapeutic targets against this chronic and progressive disease.

Some modulated pathways were found to be involved in the progression and attenuation of fibrosis induced by bleomycin.

It has been proposed that IPF pathogenesis involves complex interactions of several pathways, including oxidative stress, endoplasmic reticulum stress, unfolded protein response, coagulation system, inflammation, abnormal wounding, fibroblast proliferation, fibrogenesis, and extracellular matrix deposition, based on numerous proteomics studies in human and animal models[67].

With the LC-MS instrument and the TMT labeling, more than three thousand proteins were identified within the samples. Comparing this result to another quantitative proteomic analysis of a Bleomycin-induced pulmonary fibrosis mouse model[68] we identified and quantified one thousand more proteins, but a lower number of differentially expressed proteins.

Another proteomic analysis on Bleomycin-induced Pulmonary Fibrosis in rat based on tag labeling reported 146 differentially expressed proteins, 88 of which displayed increased abundance and 58 were downregulated in the bleomycin group. Most of these proteins were associated with extracellular matrix, inflammation, damage response, vitamin A synthesis and metabolism[69].

Besides the different animal model investigated in this pilot study, to evaluate the applicability of TMT proteomics to the preclinical model, we studied only three animals (one vehicle, one bleomycin and one Nintedanib), using technical replicates for statistical analysis. Therefore, differences in number of differentially expressed proteins and in enriched pathways could be explained with the lack of biological replicates.

The pathway analysis performed on the differentially expressed proteins showed that the key pathways emerging from the up-regulated proteins in bleomycin sample over vehicle sample included immune response neutrophil mediated, complement system, extracellular matrix organization, and collagen production. Most of these biological processes were also described in the study on the rat bleomycin model by Yang et al [69].

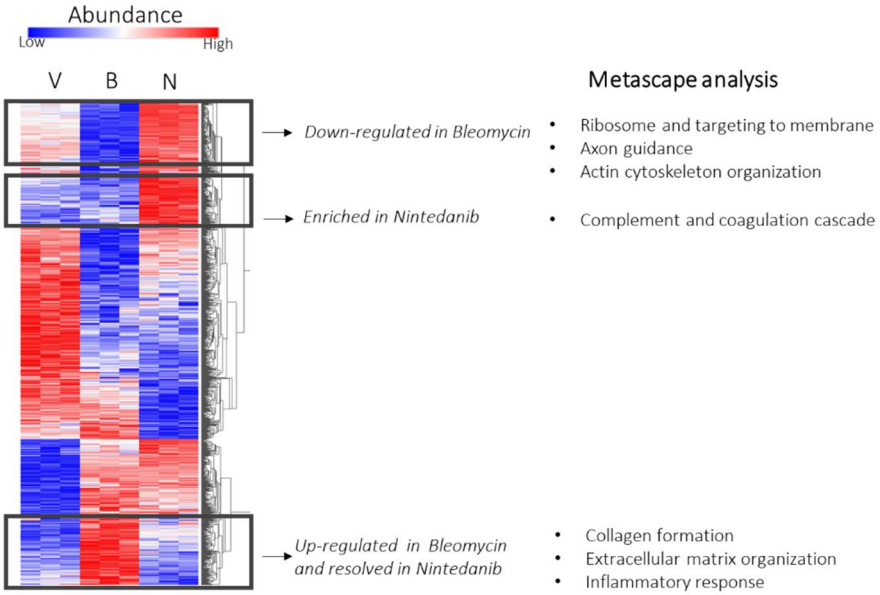


Figure 33: Main pathways involved in bleomycin-induced pulmonary fibrosis rat model evaluated with TMT analysis

Several processes, such as collagen production, appeared to be resolved after Nintedanib treatment. Others, such as extracellular matrix organization, were still significantly deregulated, although to a lower level, in the Nintedanib vs Vehicle comparison. Down-regulated proteins were enriched in pathways including axon guidance, actin cytoskeleton organization, response to wounding and platelet activation (Figure 33).

In the study on the mouse model, the pathways resulting from proteins downregulated in BLM-treated samples compared to controls are quite similar to the pathways detected in our pilot study.

In conclusion, we evaluated the use of TMT labeling for discovery proteomics in a preclinical model and found out the pro and cons of this technique. The number of proteins detected, and the depth of proteome

coverage come first, followed by the possibility of multiplex analysis, the efficient use of instrument time, and further technical variation controls.

Biological replicates are a key point to perform a robust bioinformatic analysis and to discover biological insights, moreover the resolution of mass spectrometer used determines whether peptides can be identified and quantified with accuracy.

Proteomic analysis of extracellular matrix proteins in presence of bleomycin-induced pulmonary fibrosis will have to be performed to deeply investigate the main biological process involved in fibrotic process.

Profiling of rabbit lung proteome during developmental stages

Based on the consideration made for TMT pilot study, we performed a TMT labeling experiment to characterize the proteome profile of developing rabbit lung, during three stages.

The ventral anterior foregut gives rise to the mammalian respiratory system (trachea and lungs), which undergoes a series of complex morphological changes during prenatal development and even after birth to become a highly specialized organ. Several molecular mechanisms involving epithelial, mesenchymal, and neuronal cells influence lung development and function. In all mammals, the process of lung formation is divided into five morphologically distinct stages: embryonic, pseudoglandular, canalicular, saccular, and alveolar.

We used LC-MS-based proteomics to characterize the molecular constituents of the rabbit lung during the last phases of normal development. Our proteome analysis of lungs from healthy rabbit at canalicular (day 25, 27), saccular (day 28, 29) and alveolar stage (day 31, 35) (n= 3 individual rabbit per time point) identified four thousand proteins.

A similar study, based on TMT labeling technique, was conducted on mouse lungs by Moghieb et al. in 2018 [70] and represents one of the proteomic analyses with deeper coverage reported for normal lung development, ranging from prenatal to postnatal time points.

The lung proteome showed considerable alterations during development, with the three developmental stages divided into three well-defined clusters, according to principal component analysis of the proteomics data.

To further explore the temporal dynamics of the developing lung proteome, WGCNA clustering was applied. Twelve modules of proteins with similar expression profile across developmental stages were identified. These protein modules were classified into four groups: early development, i.e. expression modules that exhibit a general decreasing abundance profile through time; after birth, i.e. expression modules that exhibit

a general increasing abundance profile through time; middle and late development, i.e. expression modules exhibiting induction from saccular stage to alveolar stage; and transition, i.e. expression modules with peak abundance during saccular stage.

A very similar protein profiling emerges from the analyses conducted on mouse data for normal lung development characterization[70].

To identify dynamic changes in biological processes occurring in the transition from the prenatal to postnatal period, we performed a functional pathway analysis using Metascape on the module protein lists. Modules in which expression increased through time concerned vesicle mediated transport, cellular detoxification, focal adhesion, supramolecular fiber organization and extracellular matrix organization, while modules involved in middle and late development are enriched for regulated exocytosis, leukocyte mediated immunity, cell junction organization, regulation of actin in cytoskeleton and wound healing.

Surfactant proteins, which are abundantly expressed in type II epithelial cells and crucial for many aspects of surfactant structure, function, and metabolism[65], were among the proteins found in the middle and late development modules.

These findings are consistent with what is known about the saccular and alveolar stages of human lung development. Alveolar type II cells become extremely proliferative during the saccular and alveolar stages and produce specialized organelles (vesicles) dedicated to surfactant production, called lamellar bodies. Meanwhile, new cell types emerge in the alveolar mesenchyme that secrete extracellular matrix (ECM) and remodel it, assisting in the formation of saccules and laying the groundwork for future septation during alveologenesis[62].

Clusters of proteins that decreased through time, therefore much abundant in canalicular stage, were functionally enriched in annotation terms related to metabolism of RNA and cell cycle, post-translational protein modification and chromatin assembly.

In the mouse study, the same processes were reported for the canalicular stage cluster. Early lung development is characterized by an accelerated rate of cell proliferation, which explains these processes[62,70].

Annotation enrichment analysis of the transition cluster, characterized by a peak abundance profile in saccular stage, showed as main terms the regulation of Insulin-like Growth Factor (IGF) transport and uptake, vesicle-mediated transport and endocytosis.

Targeted mutation of IGF genes in mice has revealed that IGF signaling is important for lung tissue development, homeostasis, and repair. According to the literature, the failure of progression from canalicular

to saccular structures and increased proliferation in perinatal fetuses are the principal consequences of total IGF-1R inactivation during mouse lung development[71].

Overall, the results obtained with pathway analysis confirmed the information about normal lung development deriving from proteome profile of other species and from human samples (Figure 34).

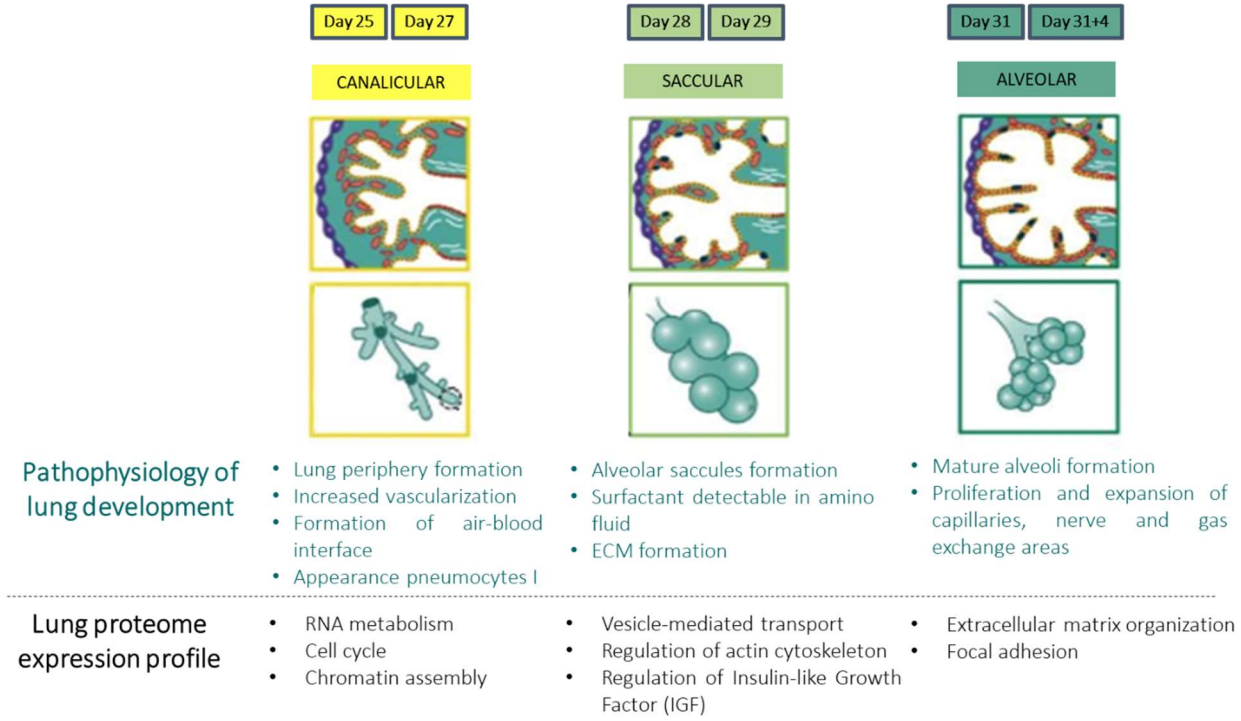


Figure 34: Lung proteome expression profile compared to pathophysiology knowledge of normal lung development. Adapted from Salaets et al.[41]

The study of surfactant proteins demonstrated how proteomics can be used to assess structural changes in proteins of interest.

As previously stated, alveolar type II epithelial cells generate and store surfactant in intracellular inclusion organelles termed lamellar bodies, which have typical lamellated features. Surfactant is made up for 90% of lipids and from 5% to 10% of proteins. Surfactant protein (SP)-A, SP-B, SP-C and SP-D have been found to have a variety of roles in surfactant's physicochemical properties, function, and metabolism, as well as in the management of host defenses and inflammation in the lungs. SP-B and SP-C are small hydrophobic proteins that are made from higher molecular weight precursor proteins via proteolytic processing[65,72].

Processing pathways of SP-B and SP-C in alveolar type II cells were extensively described in papers by Perez-Gil and Ueno[65,66]. SP-B is made from a 40 kDa precursor, which is cleaved by Napsin A. Through the Golgi, the

resultant intermediate of 23 kDa reaches multivesicular bodies. In the final processing of SP-B, which occurs in intermediate composite bodies to LB, two enzymes, Cathepsin H and Pepsinogen C, appear to be involved. SP-C, on the other hand, is synthesized as a 21 kDa precursor in the endoplasmic reticulum and palmitoylated in the Golgi apparatus. The C-terminal domain is cleaved by unknown enzymes, while the N-terminal region is cleaved by Cathepsin H, resulting in the mature protein at the lamellar bodies.

The absence of SP-B affects SP-C processing and sorting, indicating that both proteins' processing and trafficking are linked in alveolar type II cells.

Therefore, based on this information of surfactant protein maturation, we evaluated the maturation of SP-B and SP-C using peptide abundance analysis and discriminating between pro-peptides and mature peptides. The diverse forms of SP displayed different profiles, with mature peptides abundance much higher at birth and after birth.

On the whole, maturation process starts in the saccular stage to reach a maximum in the alveolar phase, especially at birth (day 31). This finding aligns with the relationship between surfactant activity and normal lung development.

Investigation of Nintedanib mode of action through phosphoproteomics

Background

Phosphoproteomics

Phosphoproteomics is a branch of proteomics that identifies, catalogs, and characterizes proteins containing a phosphate group as a post-translational modification.

Phosphorylation is the most studied post-translational modification because of its central role in living cells: phosphorylated proteins regulate cell growth, cell signalling, apoptosis, and differentiation. Many studies investigated cancer processes predominantly regulated by kinases and phosphatases, the key enzymes for phosphorylation.

Mass spectrometry-based proteomics has revolutionized the global analysis of phosphoproteins leading to identify more than ten thousand phosphorylation sites in proteomes [73].

As mentioned before, the analysis of PTMs necessitates of an enrichment step for phosphopeptides. Phosphopeptides are low-abundant, have a high complexity and have a highly dynamic nature; for these reasons is important to enrich peptides with phosphorylation on serine (S), threonine (T) or tyrosine (Y). Different methods for enriching phosphopeptides or phosphoproteins from a complex background of unphosphorylated peptides have been developed. Fractionation at protein level is based on immunoprecipitation or isoelectric focusing. Instead, enrichment at peptide level could be performed with different types of affinity capture:

- Immobilized metal ion affinity chromatography (IMAC);
- Metal oxide affinity chromatography (MOAC);
- Phospho-serine/tyrosine/threonine antibody.

The most commonly used MOAC phosphopeptide enrichment strategy employed titanium dioxide (TiO_2), in fact TiO_2 affinity to organic phosphates has been known since 1990 and nowadays TiO_2-MOAC is the most promising phosphopeptide-enrichment protocol. Moreover, MOAC have a better tolerance to EDTA as well as to other detergents, and in general it have been found to have a good selectivity and sensitivity[74].

In 2018 Humphrey et al. issued a Nature Protocols article describing the EasyPhos platform. The EasyPhos workflow enables the analysis of hundreds of phosphorylation sites and is based on the enrichment through the titanium dioxide beads[73].

Nintedanib, a multikinase inhibitor

Nintedanib, one of the clinical approved drug for pulmonary fibrosis, is an indolinone derivative, derived from a chemical lead optimization program for receptor tyrosine kinase inhibitors (Patent application WO2001027081, example 473). Nintedanib was originally designed as an anti-angiogenic drug targeting the receptor tyrosine kinases VEGFR, FGFR and PDGFR for cancer treatment.

Nintedanib blocks kinase activity by occupying the intracellular ATP-binding pocket of specific tyrosine kinases (FGFRs, PDGFRs and VEGFRs). This binding results in blockage of the autophosphorylation of these receptors and the downstream signaling cascades. Inhibition by Nintedanib ultimately results in reduced proliferation, migration, survival of fibroblasts and, potentially, also to attenuated angiogenesis in the lung (Figure 35).

The *in vivo* efficacy of Nintedanib has been explored in three models of pulmonary fibrosis: bleomycin-induced lung fibrosis in mice and rats and silica-induced lung fibrosis in mice.[20,75]

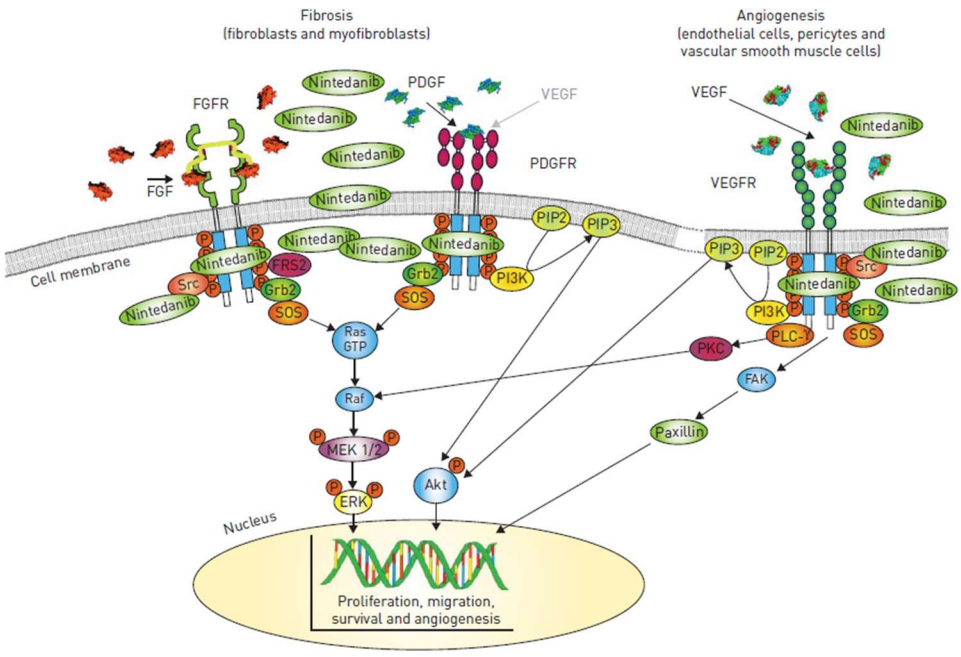


Figure 35: Mode of action of Nintedanib and the downstream signalling pathways. Image from Wollin et al., 2015 [75].

Aim

The main goal of this experiment was to study the mode of action of a multi-kinase inhibitor used in clinical on patients suffering for idiopathic pulmonary fibrosis, highlighting its multi-pharmacology action on many kinases and signaling cascades. This aim was achieved by developing a protocol workflow starting from the

sample preparation, with the phosphopeptides enrichment, to the mass spectrometry analysis till the data elaboration.

Furthermore, the methodology presented here was designed to implement in the analytic pipeline a method suitable for analyzing phosphorylation cascade for various preclinical models, beginning with sample preparation and ending with extensive data analysis of phosphorylated sites.

In this thesis work I used a modified protocol for phosphoproteomics sample preparation derived from the Nature protocol[73]. The changes were implemented in the Chiesi Farmaceutici analytic department to align the sample preparation to the pharmacologic questions.

Results

Study design

A quantitative label-free phosphoproteomic analysis was performed on lung samples from non-treated rats (naïve, T0) and animals treated with Nintedanib for 1, 3 or 24 hours. All conditions were assayed in biological triplicate. Nintedanib was chosen for its mechanism of action, being an inhibitor of different tyrosine kinases.

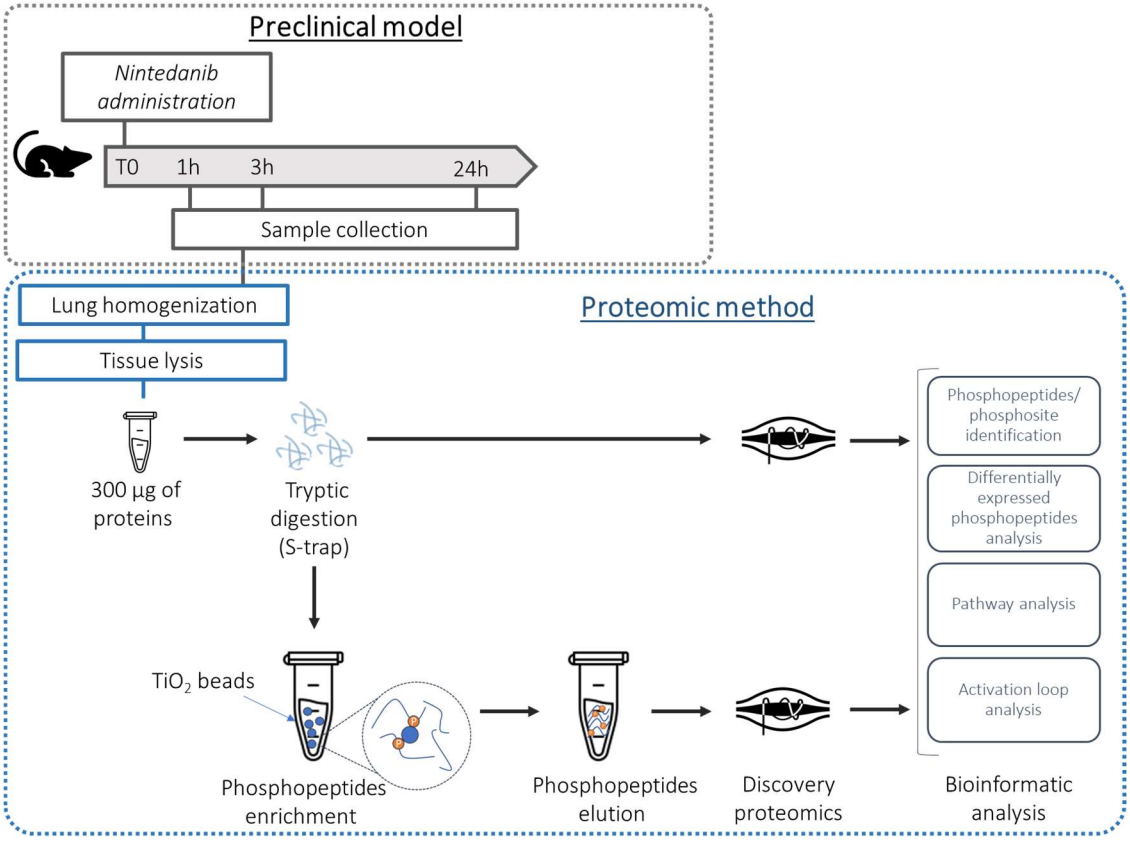


Figure 36: Experimental design and workflow for phosphoproteomics of a PD in vivo experiment for Nintedanib treatment

Sample preparation and phosphopeptide enrichment

Lung homogenization was performed in RIPA buffer with the addition of protease and phosphatase inhibitors. Protein content was assayed with Bicinconic acid assay (BCA). Samples were diluted 1:20 in Ammonium bicarbonate for the assay, because of the RIPA compatibility. Tryptic digestion was performed following the s-trap mini protocol, as described in the experimental section.

After digestion, a small aliquot of peptide solution was stored and used for the MS analysis, to evaluate the whole lung proteome.

For the phosphoproteomic analysis, the protocol was optimized from the paper by Humphrey et al [73].

The peptide solution was enriched in phosphopeptides through specific TiO_2 beads, following the protocol described in the experimental section.

The enriched peptides and the aliquot of digested whole lung were analyzed with nanoLC-MS. The same MS method was used to analyse the phospho-enriched fraction and the whole proteome for each sample.

Identification of proteins, peptides and phosphopeptides

Database searches were performed with SEQUEST HT (as a node in PD 2.4) with few modifications: trypsin digestion, 2 maximum missed cleavages, precursor mass tolerance of 10 ppm, fragment mass tolerance of 0.8 Da, a fixed modification of +57.021 Da (carbamidomethylation) on cysteine, and variable modifications of +15.995 Da (oxidation) on methionine and +79.966 Da (phosphorylation) on serine, threonine, and tyrosine. A reverse database search was performed using SEQUEST HT to determine the spectral false discovery rate (FDR), and subsequent results were filtered by FDR ≤ 1%. The FASTA database used was a *Rattus norvegicus* proteome downloaded from Uniprot (2020) with addition of 73 common contaminants such as proteases and keratin. The MS1 intensity quantitative tool Minora Feature Detector was utilized for label-free quantitation for area under the curve calculations for up to the top three unique plus razor peptides detected per unique protein. Normalization was performed using the total peptide amount to account for random errors.

The analysis was performed on all the samples, 12 whole lungs digested and 12 phospho-enriched samples.

The analysis identified 6645 proteins as master proteins with high confidence identification.

PCA reveals the differences between whole samples and phospho-enriched samples (Figure 37).

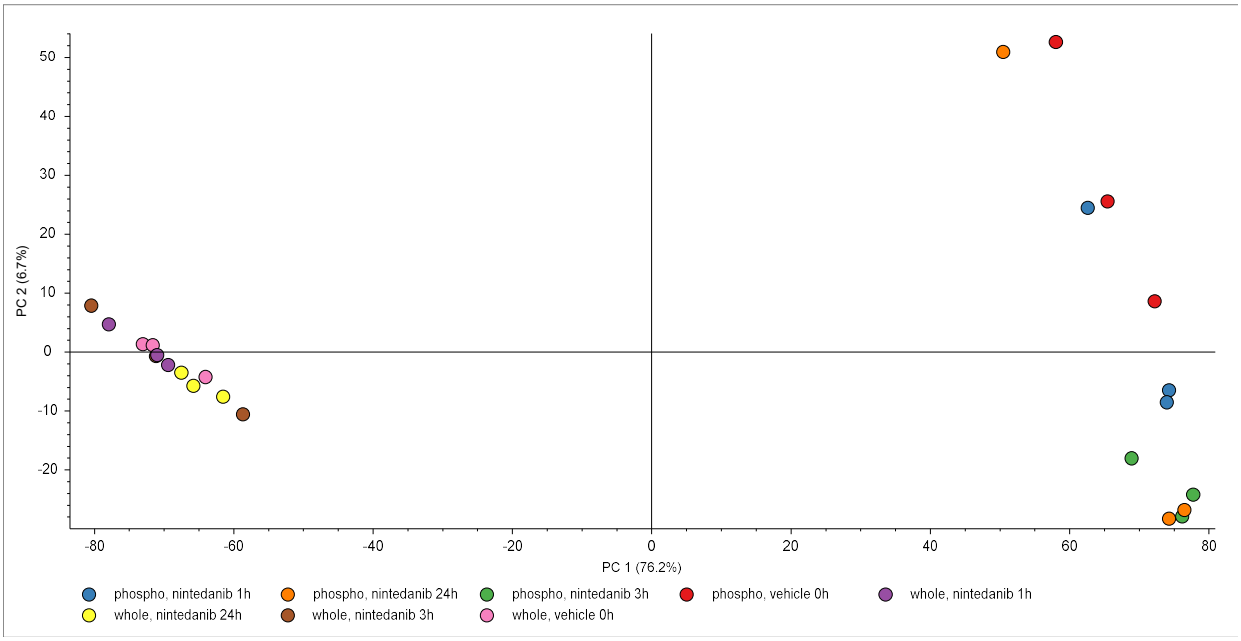


Figure 37: PCA of all samples analysed

A differential expression analysis was performed with t-test on proteins quantified in the whole samples (Figure 38) and the differentially expressed proteins were analyzed with Metascape to make a biological inference. No significative pathways were detected for the whole samples. The cut off used for DEPs are FC 1.5, adj. p-value 0.05.

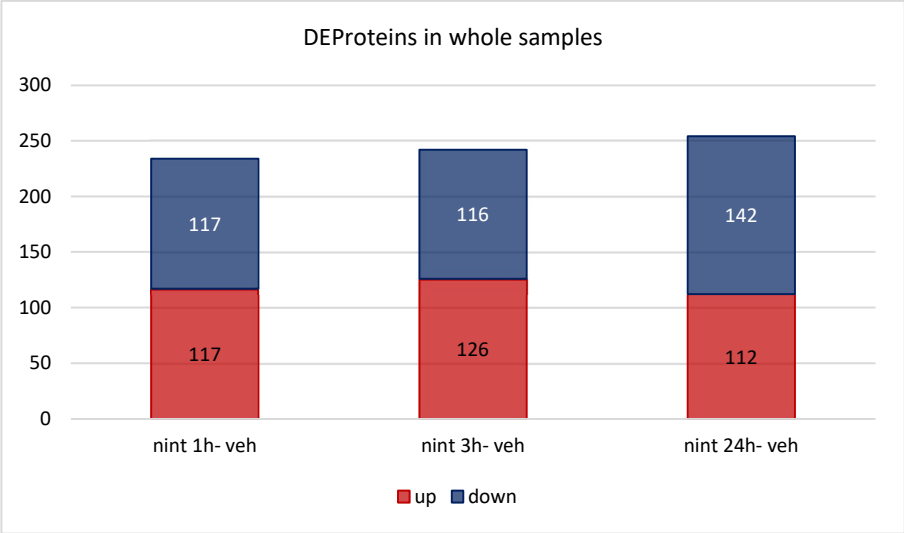


Figure 38: Differentially expressed proteins in whole samples

Phosphopeptide analysis

Focusing on the Proteome Discoverer analysis for phosphopeptides, we obtained these main results:

- 79449 peptides identified and quantified
- 10508 phosphopeptides (with at least one phosphorylation)
- 9213 phosphopeptides with defined position of PTM
- 5591 peptides with 1 phosphorylation
- 2975 peptides with 2 phosphorylation
- 647 peptides with 3 phosphorylation

The PCA of peptides detected in phospho-enriched samples (Figure 39) revealed a weak clusterization of vehicle (naïve, t0) and a wide dispersion of samples for Nintedanib treatment, mostly at 24 hours.

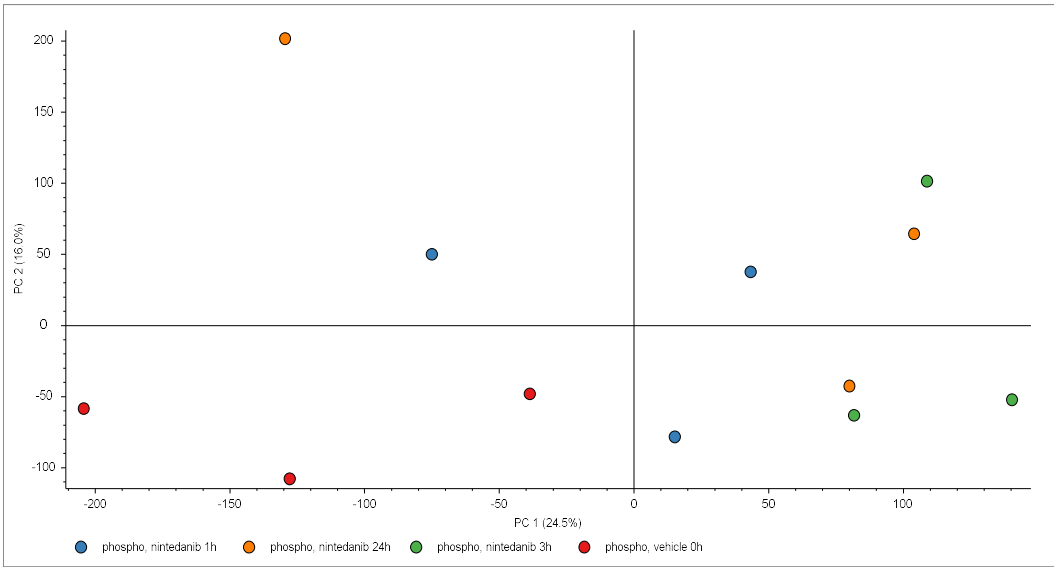


Figure 39: PCA of peptides detected in phospho-enriched samples

A differential expression analysis was performed also for phosphopeptides through Proteome Discoverer (Figure 40). According to literature, the cut off selected to consider peptides significant was FC 2 and adj. p-value 0.05. The phosphopeptides taken in account have the PTM at a defined position.

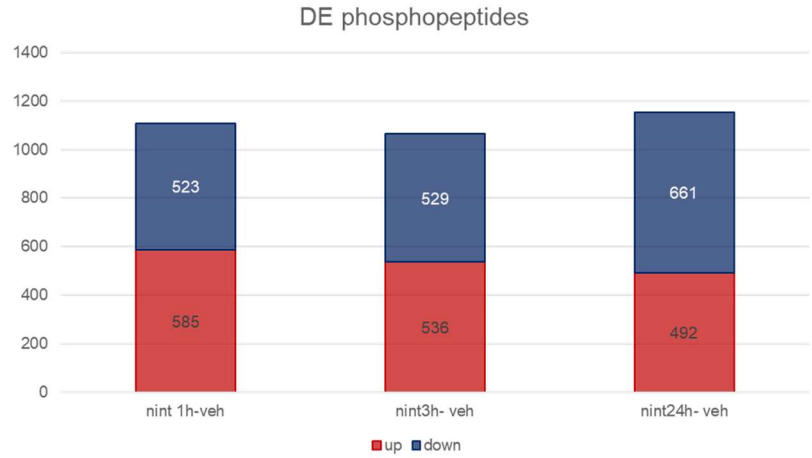


Figure 40: Differentially expressed phosphopeptides identified in phospho-enriched samples

Bioinformatic analysis for phosphoproteomics

Besides the phosphopeptides, it was interesting to evaluate the phosphosites detected in the peptides, because of the 3500 phosphopeptides with 2 or 3 phosphorylations.

Phosphosites were extrapolated from the description of phosphopeptides. For each phosphosite the software reports information about the aminoacidic position and the site probability, expressed in percentage. All the phosphosites with site probability ≥ 95% were considered. The analysis of phosphosites was performed using different bioinformatic tools.

I. *Phospho enrichment tool* (http://phomics.jensenlab.org/phospho_enrichment)

This tool analyses enrichment of biological terms from phosphoproteomics datasets taking into consideration the number of phosphorylation sites. The data elaboration takes three steps:

- Determination of a foreground and a background phosphoproteomics datasets of interest: the differentially expressed phosphosites for each time point were used as foreground database, the phosphosites not differentially regulated were used as background database;
- Use of DAVID "Functional Annotation" tool to get a Functional Annotation table for each dataset.
- Upload the foreground and background phosphoproteomics datasets and of DAVID output files in the phospho-enrichment tool page.

The result is a list of pathways with associated p-value, FDR and Bonferroni correction. The phosphorylation enrichment for down regulated phosphosites describes very few nonspecific biological processes. Moreover, the main drawback of this analysis is the lack of information about which phosphosites are involved in the processes described.

II. *Activation loop analysis (*http://phomics.jensenlab.org/activation_loop_sites)

Phosphorylation of the activation loop is a fundamental step in the activation of most protein kinases. Phosphorylation in the activation loop is known to rigidify the structure and contribute to the switch from the inactive to a fully active form. This software allows to identify phosphosites belonging to the activation loop of kinases. We performed the analysis on the differentially expressed phosphosites, the result is showed in Table 13. The down regulated activation loop sites belong to the MAPK signaling pathway, and these kinases are at the end of the signaling cascade starting with the binding of growth factors to their receptors. Among these receptors there are the main targets of Nintedanib, such as FGFR, VEGFR and PDGFR.

Table 13: Phosphosites belonging to activation loop differentially expressed over time

	nint 1h	nint 3h	nint 24h
up regulated phosphosites	3	3	5
down regulated phosphosites	1	1	

III. *Metascape analysis of proteins with differentially expressed phosphosites*

A pathway enrichment analysis was performed using Metascape on the proteins with differentially expressed phosphosites. The results obtained indicate the down regulation of many pathways known as part of Nintedanib mechanism of action, such as Signaling by receptor tyrosine kinases. Pathways related to VEGF, PDGF, FGF, and cellular response to growth factor stimulus emerge only in the lists of down-regulated proteins. Many known kinases involved in Nintedanib mechanism emerge as actors for these pathways related to growth factors, such as non-receptor tyrosine kinases (Src and Lyn), MAPK kinases and Akt.

Discussion

Phosphoproteome profiling of a multikinase inhibitor

The last application of mass spectrometry-based proteomics covered in this thesis work is phosphoproteomics.

The case study presented for the phosphoproteomics methodology regards the characterization of a PD rat model testing a multikinase inhibitor, Nintedanib, widely used for pulmonary fibrosis.

Phosphorylation of proteins is one of the most common and widespread change that affects every element of a cell. As a result, it is critical to figure out exactly what role this alteration plays in signaling systems and to learn more about it. The identification of phosphorylation events in a cell has become standard thanks to advances in technology and enrichment techniques.

In 2005, an offline setup for TiO_2 chromatography was published, which included the use of strong buffer conditions and a high concentration of trifluoroacetic acid (TFA) in the loading buffer, which significantly reduced unspecific binding from non-phosphorylated peptides. Furthermore, ammonia solution at pH > 11.3 was observed to elute phosphorylated peptides from the TiO_2 column more efficiently than pH 9, resulting in better phosphopeptide recovery[76].

The TiO_2 purification procedure is easy to follow. Even from highly complex biological materials, it is quick and effective at enriching phosphopeptides[77].

The protocol for phosphopeptide enrichment tested in this work was modified from EasyPhos platform designed by Humphrey et al[73].

We applied the protocol to digested lung homogenate, starting from 300 µg of peptides.

The enrichment protocol proved effective, and we identified nine thousand phosphopeptides in the phospho-enriched fraction of the samples.

The goal of the study was to confirm the Nintedanib mode of action on kinases and related phosphorylated proteins, at three different time points following drug administration.

PCA results revealed a poor clustering of the time points studied, with a substantial factor of biological variability. Above all, a sample at 24 hours after treatment seemed to be an outlier in the phospho-enriched samples.

Despite the estimated variability, we analyzed phosphoproteomics datasets for phosphosites that reside inside a kinase activation loop.

Phosphoproteomics data provide a variety of levels of information that can be retrieved or inferred to evaluate the functional relevance of individual kinases in a given biological context. In general, kinase roles can be investigated by examining specific phosphosites or finding which kinases are more globally active in the conditions studied. Based on the amino acids in the surrounding sequence, a responsible kinase for each phosphorylation event can be predicted[78].

Furthermore, kinases' enzymatic activity is frequently regulated by their phosphorylation. The activation of most protein kinases begins with the phosphorylation of the activation loop. The activation loop is known to be rigidified by phosphorylation, which aids in the transition from an inactive to a fully active state[79].

Besides the ones used in this work, there are numerous other bioinformatic tools available to analyse phosphoproteomics data that organize and offer information about known phosphosites, as well as about sequence-based kinase-substrate relationships. The goal of such software is to figure out which enzymes are accountable for the modulated phosphosites so that the signaling cascade may be reconstructed. Unfortunately, these bioinformatic tools, such as NetworKIN or KSEA or Netphorest, are available for few model species, commonly human, yeast and mouse.

The results of activation loop analysis revealed the down regulation of phosphosites after drug administration in extracellular signal-regulated kinases (ERK, also named Mapk).

The MAPK pathway is a three-layer signaling cascade in which phosphorylation of tyrosine and threonine within a conserved Thr-Xxx-Tyr pattern in the kinase domain activates the MAPK components. Dual-specificity kinases, MAPK kinases (MAPKK, MEK), catalyze this phosphorylation, which is regulated by serine/threonine phosphorylation inside a common pattern in the kinase activation loop. The extracellular signal-related kinases (Mapk1/Mapk3, also known as Erk1/Erk2), Jun amino-terminal kinases (Jnk1/2/3), p38-MAPK, and Erk5 are four different MAPK signaling cascades. Although growth factor signaling regulates all these MAPK cascades, growth factors are thought to be the most important regulators of the Mapk1/Mapk3 cascade[80].

Nintedanib's exact mechanism of action is unknown, however it is known to inhibit platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), and vascular endothelial growth factor receptor tyrosine kinases (VEGF). The inhibition of one or more of these receptor tyrosine kinases is thought to underlie Nintedanib's antifibrotic actions observed in preclinical research[75,81,82]. Pathways analysis of phosphoproteomics data confirm the inhibition of signalling by receptor tyrosine kinases. From this signalling cascade inhibited by Nintedanib emerge some known kinases[82]: the potency of non-receptor tyrosine kinases of the Src family were determined in enzymatic assays confirming their inhibition with Nintedanib treatment[83].

Furthermore, Nintedanib was shown to inhibit PDGFR phosphorylation and Akt and Erk phosphorylation in lung tissue from mouse[84] leading to a regulation of MAPK pathway[81]. The activation loop results obtained in this theses work confirm these findings described by Wollin and Rangarajan. Nintedanib inhibits growth factor receptors resulting in the inactivation of extracellular signal-related kinases (Erk). The extracellular

cascade of signal-related kinases is involved in a number of cellular functions, including proliferation, differentiation, development, and, in some cases, apoptosis and survival, all of which can be influenced by Nintedanib therapy[82].

Conclusions

Targeted proteomics

Candidate biomarkers are frequently discovered using mass spectrometry-based proteomics methods. In a complex biological matrix, targeted proteomics has emerged as a powerful tool for quantifying several interesting proteins.

Targeted MS methods enable high sensitivity and precise quantification of proteins of interest across samples. These strategies use data-dependent collection to provide highly reproducible and quantitatively accurate data of a predetermined set of proteins.

Although quantitative protein measurement is one of the most basic analytical tasks in preclinical research, widely utilized immunochemical methods frequently have low specificity and large measurement variation.

The ability to target specific peptide sequences, including variations and post-translational modification, and the capacity for multiplexing, which allows the study of dozens to hundreds of peptides, are two key features of MRM assays. Different quantitative standardization procedures offer options that strike a compromise between precision, sensitivity, and the cost of the assay[9].

Mass spectrometry-based targeted proteomics is nowadays an alternative technique to detect interesting proteins, to quantify isoforms, to evaluate post-translational modification in complex biological samples.

The methods described in this work were used for different preclinical applications and proved to be robust, reproducible and in line with other types of assays, such as histologic readout for collagen content and western blot assay for histone acetylation.

Proteomics has the advantage of allowing researchers to create customized methods for each protein of interest or for specific PTMs, eliminating the problem of limited antibodies availability and specificity or the need for (aspecific) colorimetric assays.

The methods developed in this part of the thesis are now consistently implemented to investigate preclinical experiments performed by the pharmacology department of Chiesi Farmaceutici:

- About twenty studies have been conducted in the last two years using the collagen detection method. Furthermore, the same workflow was adapted to quantify collagen in a mouse model.
- The method to identify the actin isoform was adapted to monitor the proteins in other tissues, such as heart, to respond to new preclinical questions.
- The acetylation of histones was also analysed in different study of pulmonary fibrosis bleomycin-induced

Discovery proteomics

Mass spectrometry-based proteomics is the method of choice to detect proteome alterations that occur during disease progression, developmental phases, or therapeutic treatments. Untargeted profiling of hundreds of proteins from complex biological samples acquired from tissues of various animal models is possible using these discovery methodologies.

In preclinical research, protein differential expression and changes can provide reliable diagnostic indicators.

Proteomics analysis can reveal not only the peptide sequence, but also protein quantity, post-translational modifications (PTMs) and protein interactions, depending on the focus[85].

In this thesis work two different methodologies of discovery proteomics were described and were tested in different preclinical models.

Mass spectrometry-based proteomics has proven able to adapt to different aims and contexts, from protein expression profiling during developmental stages to identification of specific post-translational modifications for target engagement study. For this reason, proteomics was used as tool for this thesis work to examine preclinical animal models, spanning from illness characterization to proteome profiling, to investigation of drug action signaling cascades, thanks to the flexibility of the methodology.

The mentioned preclinical proteomic investigations led to a better understanding of disease processes, developmental phases, and signaling cascades.

Proteomics allows to obtain integrated biology information by knowing the complete cell protein network rather than individual proteins, with the ultimate goal of understanding protein function and of determining gene expression.

Bioinformatic tools are the key to understand the data matrix obtained with the high-throughput proteomic methods.

The advantage of using proteomics data for pathways analysis is that, based on systemic proteome measurements rather than transcriptome measurements, the investigation of other features such as differential translation rate and PTMs, occurring after the transcription process, can be included in such analysis.

Similarly to the targeted method, also the pipeline described both for TMT and phosphoproteomics have been consistently implemented in preclinical research in Chiesi Farmaceutici laboratories:

- TMT analysis is used for many experiment sessions and allowed a comparison of different studies conducted over the course of a year, confirming the robustness of the technique.
- Phosphoproteomics pipeline is used routinely to investigate pharmacodynamic preclinical model of kinase inhibitors.

Experimental section

Sample preparation

Lung homogenates

Applicable when:

- Lung homogenates prepared with GentleMax are thawed from -80°C (from rat, mice, rabbit)
- Lung homogenates are stored in eppendorf (2 ml or 1.5 ml)
- Minimum volume required: 60 µL

Materials required:

- Tungsten Carbide Beads, 3 mm, Catalogue Number 69997, QIAGEN
- TissueLyser II, QIAGEN (lab C1)

Procedure:

1) Insert a Bead in each sample to be homogenized
2) Homogenization with TissueLyser, 2 minutes at 30 Hz
3) Keep on ice until use

Lung tissues

Applicable to frozen lungs

Materials required:

- PBS sterile solution, pH 7.4
- Dry ice
- Scalpel
- Tungsten Carbide Beads, 3 mm, Catalogue Number 69997, QIAGEN
- TissueLyser II, QIAGEN

Procedure:

1) Do not thaw the sample to maintain the properties and to facilitate cutting
2) Cut a piece of tissue, from 20 to 60 mg, and put in an eppendorf tube (2 ml or 1.5 ml)
3) Keep the sample in dry ice
4) Insert a Bead in each sample to be homogenized
5) Add 100 µl of PBS per 10 mg of tissue weighed
6) Homogenization with TissueLyser, 5 minutes at 30 Hz
7) Keep on ice until use

Protein quantification

Bradford Assay

Materials required:

- Coomassie (Bradford) Protein Assay Kit (Product Number 23200, Thermo)
- Microplate (e.g. Pierce TM 96-Well Plates, Product Number 15041, Thermo)
- MultiskanGO plate reader

Procedure:

1) Prepare dilutions of BSA standard as described in Table 14
   NOTE: preferably use the sample buffer to make BSA dilutions.

Table 14. Serial dilutions of BSA standard for Bradford Assay.

Dilution Scheme for Standard Test Tube and Microplate Protocols (Working Range = 125–1500µg/mL)			
Vial	**Volume of Diluent**	**Volume and Source of BSA**	**Final BSA Concentration**
A	0	300µL of Stock	2000µg/mL
B	125µL	375µL of Stock	1500µg/mL
C	325µL	325µL of Stock	1000µg/mL
D	175µL	175µL of vial B dilution	750µg/mL
E	325µL	325µL of vial C dilution	500µg/mL
F	325µL	325µL of vial E dilution	250µg/mL
G	325µL	325µL of vial F dilution	125µg/mL
H	400µL	100µL of vial G dilution	25µg/mL
I	400µL	0	0µg/mL = Blank

2) Add 250 µL of the Coomassie Reagent to each well and mix by pipetting.
3) Pipette 5 µL of each standard or unknown sample into the appropriate microplate wells.
   NOTE: if bubbles appear on the surface of the well, blow them up using a tip or a needle.
4) Place the plate in the MultiskanGO reader and incubate for 10 minutes at room temperature with interval mixing.
5) Measure the absorbance at 595 nm.
6) Elaborate the results.
   a. Subtract the average 595 nm measurement for the Blank replicates from the 595 nm measurements of all other individual standard and unknown sample replicates.
   b. Prepare a standard curve by plotting the average Blank-corrected 595 nm measurement for each BSA standard versus its concentration in µg/mL.
   c. Use the standard curve to determine the protein concentration of each unknown sample. To interpolate the curve, use a four-parameter curve.

BCA Assay

Materials required:

- BCA Protein Assay Kit (Product Number 23225, Pierce)
- Microplate (e.g. Pierce TM 96-Well Plates, Product Number 15041, Thermo)
- MultiskanGO plate reader

Procedure:

1) Prepare dilutions of BSA standard as described in Table 15.
   NOTE: preferably use the sample buffer to make BSA dilutions.

Table 15. Serial dilutions of BSA standard.

Dilution Scheme for Standard Test Tube Protocol and Microplate Procedure (Working Range = 20-2,000µg/mL)			
Vial	**Volume of Diluent (µL)**	**Volume and Source of BSA (µL)**	**Final BSA Concentration (µg/mL)**
A	0	300 of Stock	2000
B	125	375 of Stock	1500
C	325	325 of Stock	1000
D	175	175 of vial B dilution	750
E	325	325 of vial C dilution	500
F	325	325 of vial E dilution	250
G	325	325 of vial F dilution	125
H	400	100 of vial G dilution	25
I	400	0	0 = Blank

2) Prepare the required volume of BCA Working Reagent (WR): Prepare WR by mixing Reagent A with Reagent B in a 50:1 ratio
3) Prepare the Microplate to be assayed:
   c. Pipette 25 µL of each standard or unknown sample replicate into a microplate well.
   d. Add 200 µL of the WR to each well (sample to WR ratio = 1:8) and mix
4) Cover the Microplate with aluminium foil and incubate at 37 °C for 30 minutes, then cool the plate at room temperature
5) Put the plate it in the MultiskanGO reader and measure the absorbance at 562 nm.
6) Elaborate the results:
   e. Calculate the average absorbance measurement of the Blank standard replicates
   f. Subtract the average absorbance of the Blank from the measurements of all other individual standard and unknown sample replicates
   g. Calculate the average Blank-corrected measurement for each BSA standard and unknown sample
   h. Prepare a standard curve by plotting the average Blank-corrected measurement for each BSA standard vs. its concentration in µg/mL
      NOTE: a four-parameter (quadratic) or best-fit curve will provide more accurate results than a purely linear fit.
   i. Use the standard curve to determine the protein concentration of each unknown sample

SDS-PAGE

Materials required:

- Electrophoresis chamber
- Power supply
- Criterion TGX Precast Protein Gel, 18 wells:
  - Any kD (Product Number 5671124, Bio-Rad)
  - 4-20% (Product Number 671094, Bio-Rad)
  - 12% (Product Number 5671044, Bio-Rad)
  - 7.5% (Product Number 5671024, Bio-Rad)
- 10x Tris/Glycine/SDS Running Buffer (Product Number 1610732, Bio-Rad)
- 2x Laemmli Sample Buffer (Product Number 1610737, Bio-Rad)
- 2-Mercaptoethanol (Product Number 1610710, Bio-Rad)
- Molecular weight standards:
  - Precision Plus Unstained Protein Standards (Product Number 1610363, Bio-Rad)
  - Precision Plus Dual Color Standards (Product Number 1610374, Bio-Rad) – optional
- Coomassie Brilliant Blue R-250 (Product Number 1610436, Bio-Rad)
- Thermomixer
- ChemiDoc for gel image acquisition
- ImageLab software for image analysis.

Advice on the selection of SDS-PAGE gel: Gel selection based on the expected molecular weight distribution (Figure 41): for complex samples it is recommended to use Any kD or 4-20%;

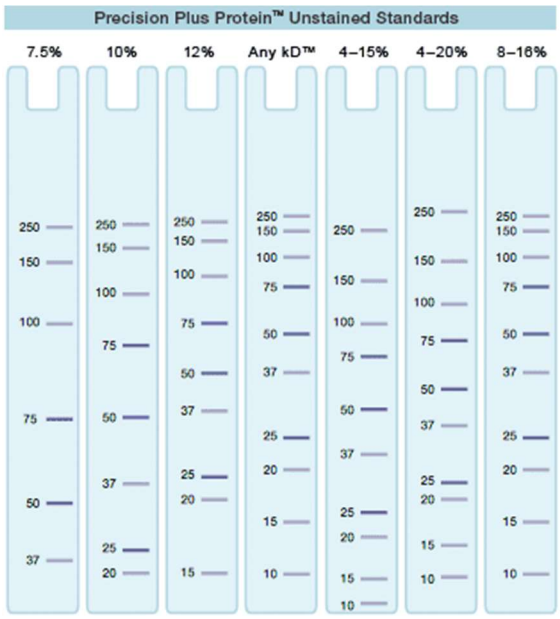


Figure 41. Distribution of molecular weight of proteins in different type of Polyacrylamide gels.

Procedure:

1) Heat up the thermomixer at 95 °C.
2) Prepare the reducing sample buffer by adding 5% β-mercaptoethanol to the 2X Laemmli Sample Buffer
3) Mix each sample 1:1 with the 2X Sample Buffer for a final volume of 20 µL.
   NOTE: load 15 µg of total protein per biological sample.
4) Denature the samples at 95 °C for 5 minutes.
5) Prepare the electrophoresis cell:
   a. Prepare 1 L of the running buffer by diluting the 10X the commercial solution with water.
   b. Put the gel in the electrophoretic cell.
   c. Gently pour a small volume of the running buffer in the upper chamber and on the outer side of the cell
   d. Remove the comb and add running buffer to fill the upper chamber completely.
6) Load the samples and Molecular Weight Markers.
7) Gently fill the outer chamber with the remaining running buffer.
8) Close the electrophoretic cell with the cap.
9) Connect the cell to the power supply: red cable with red socket, black cable with black socket.
10) Connect the power supply to the socket and switch the device on.
11) Set constant voltage (200 V) and start the run.
    NOTE: the electrophoretic run will last about 30-45 minutes, depending on the gel type.
12) At the end of the run stop the device, switch it off and disconnect all the cables.
13) Open the cell and put the gel in a bowl with MilliQ water to rinse for 15-30 minutes.
14) Remove the water and add Coomassie Brilliant Blue stain, enough volume to cover the gel, then wait at least 1 hour.
15) Remove the staining solution and rinse the gel in water until the background colour faints and protein bands appear.

Sample preparation for Targeted proteomics

Extracellular matrix digestion

Materials required:

- hMMP1 SRP6269-10µg (SigmaAldrich)
- Thermomixer
- PBS

Procedure:

1) Resuspend hMMP1 powder in the vial with 200 µl of PBS (concentration 50µg/ml)
2) Add enzyme in an E/S ratio in a range 1:150 (w/w).
3) Incubate for 16 hours at 37°C.

Thermal denaturation

Incubate the sample at 95°C for 5 minutes.

Reduction with DTT

Materials required:

- DTT, powder; MW = 154.25 g/mol

Procedure:

1) Weight a small amount of DTT powder in a tube
2) Solubilize the DTT as stock solution at 500 mM, using H_2O
3) Add DTT to the sample, making a 100X dilution of the stock solution (final concentration 5 mM)
4) Incubate the sample at 37 °C for 45 minutes.

Alkylation with Iodoacetamide

Materials required:

- IAA, powder; MW = 184.96 g/mol

Procedure:

1) Weight a small amount of IAA powder in a tube
2) Solubilize the IAA as stock solution at 600 mM, using H_2O.
   NOTE: keep the solution in the dark because IAA is light-sensitive.
3) Add IAA to the sample, making a 40X dilution of the stock solution (final concentration 15 mM)
4) Incubate the sample at room temperature for 30 minutes in the dark (for example in the thermomixer with the cover).
5) Quench the reaction with 5 mM DTT, 15 minutes at room temperature.

Digestion with Trypsin

Materials required:

- Trypsin from porcine pancreas, proteomics grade (Product Number T6567-5X20UG, Sigma-Aldrich)
- HCl. Prepare a 1 mM solution by diluting HCl in H_2O
- Formic Acid (FA)
- Thermomixer

Procedure:

1. Resuspend trypsin powder in the vial with 20 µl of HCl 1mM (concentration 1 mg/ml)

2. Add enzyme in an E/S ratio in a range 1:100 - 1:20 (w/w).
3. Incubate for 16 hours at 37°C.
4. Quench the reaction with 0.1% FA.

Sample preparation for TMT experiments

Sample lysis

1. Add SDS and DTT at a final concentration of 2% and 0.1M, respectively.
2. Sonicate samples 30s (cycle 0.5, amplitude 50%) and store on ice 30s (x5)
3. Incubate 5 min at 95ºC
4. Sonicate 30s (cycle 0.5, amplitude 50%) and store on ice 30s (x3)
5. Centrifuge 20 min at 16000 g and 13ºC.

Digestion

Samples were digested with Trypsin following FASP protocol[86].

Materials required

a. NaOH: 0.1 M NaOH (Mw 39.997 g/mol) in water.
b. UA: 8 M urea (Mw 60.06 g/mol) in 0.1 M TEAB pH 8.5: 14.4144 g urea + 30 ml 0.1 M TEAB.
c. IAA solution: 50 mM iodoacetamide (Mw 184.96 g/mol) in UA: 175 µl 200 mM IAA in UA + 525 µl UA.
d. Trypsin: 2%: 5 µg: 5 µl trypsin 1 µg/µl per sample.
e. TEAB: 100 mM triethyl ammonium bicarbonate (TEAB) in water, pH 8.0-8.5: 3 ml TEAB 1M + 27 ml water.
f. Amicon Ultra-0.5mL 30k (Millipore, Ref. number UFC503024)
g. Refrigerated Bench-top centrifuge, temperature 15ºC

Procedure (**200 µg** of each sample are going to be digested).

Filter washing

1. Add 400µL of 0.1M NaOH to the filter and centrifuge for 10 min at 14000 x g.
2. Add 400µL of MilliQ water to the filter and centrifuge for 10 min at 14000 x g.
3. Add 400µL of UA to the filter and centrifuge for 10 min at 14000 x g

Sample processing

1. Incubate samples in the Thermomixer for 5 min at 95ºC.
2. Dilute the samples with UA buffer to a final concentration of 0.1% SDS, in this case it is a dilution 1/40.

3. Add diluted samples to the filter unit and centrifuge at 14,000 x g for 10 min. In each step, a maximum of 450µL of sample can be loaded. Repeat this step until all the sample is loaded on the filter.
4. Add 200 µl of UA to the filter unit and centrifuge at 14,000 x g for 15 min.
5. Add 100 µl IAA solution, mix for 1 min and incubate without mixing for 30 min.
6. Add 200 µl of UA to the filter unit and centrifuge at 14,000 x g for 15 min.
7. Add 200 µl of TEAB to the filter unit and centrifuge at 14,000 x g for 15 min (x3). Keep the third wash in the tube for the incubation.
8. Add 100 µl TEAB with trypsin (enzyme to protein ration: 1 µg trypsin for 50 µg sample) and mix.
9. Incubate the units at 37°C, 300 rpm and overnight.
10. Transfer the filter units to new collection tubes and centrifuge at 14,000 x g for 15 min.
11. Add 100 µl of TEAB and centrifuge the filter units at 14,000 x g for 15 min (x2).
12. Transfer elutes to LB eppendorfs 1.5 ml.
13. Add TFA to 1% final concentration.

TMT labeling

70 µg of each sample were dried, reconstituted in 100µL of TEAB 100mM (triethyl ammonium bicarbonate) and labelled with TMT6plex (Thermo Scientific).

Materials required

- 100 mM TEAB, pH 8.0-8.5: 3 ml TEAB 1M + 27 ml H_2O
- TMT 6plex Reagent
- Acetonitrile (LC-MS)
- 50% hydroxylamine
- 5% hydroxylamine: 20 µl 50% hydroxylamine + 180µl H_2O

Procedure

1. Wash the syringes with 2 ml acetonitrile (ACN).
2. Prepare sample in 100 µl of 100 mM TEAB.
3. Equilibrate the TMT Label Reagents to RT for 5 min before opening and spin.
4. Add 41 µL of anhydrous acetonitrile to each TMT tube with the corresponding syringe. Allow the reagent to dissolve for 5 minutes with occasional vortex.
5. Spin the solution.
6. Carefully add the 41 µL of the TMT Label Reagent with 100 mL glass syringe to each 100 mL of sample.
7. Vortex, spin and incubate the reaction for 1 hour at room temperature.
8. Add 8 µL of 5% hydroxylamine to the sample and incubate for 15 minutes to quench the reaction at RT. (V_f = 149 mL).
9. Combine samples (20 µg for each sample).
10. Dry the sample in SpeedVac and reconstitute it in 100 µl of 1% TFA.

Sample preparation for phosphoproteomics

Digestion protocol using Protifi S-Trap

S-Trap is a sample preparation technology for proteomic analysis marketed by Protifi. Such approach is compatible with SDS (up to 5%), low pH and organic solvents (methanol and chloroform in case) in order to maximize protein denaturation/solubilization and contaminants removal. Protein samples is trapped into S-Trap resin and then subjected to proteases for digestion.

S-trap is very versatile platform and can be used to prepare samples for proteomic analysis obtained from different sources, such as cellular growth media, cell lysates and tissue homogenates. All samples can be treated as dry or in solution samples.

Depending on the amount of protein required for your analysis, 3 S-Trap sizes are available: MICRO (loading capacity 1-100ug), MINI (loading capacity 100-300 ug) and MIDI (loading capacity > 300 ug).

Material required:

- 1.7 mL sample tubes
- Mini S-trap
- Benchtop centrifuge
- Vortex mixer
- Heat block
- Thermoblock
- SpeedVac or lyophilizer
- Sonicator.

Sample preparation:

1) Add equal volume of 2X lysis buffer to liquid sample to yield a final volume containing from 300 µg of protein.
2) Protein denaturation, reduction and alkylation:

Materials required:

- DTT, powder; MW = 154.25 g/mol
- IAA, powder; MW = 184.96 g/mol

Procedure:

- Solubilize the DTT as stock solution at 800 mM, using H_2O
- Add DTT to the sample, making a 40X dilution of the stock solution (final concentration 20 mM)
- Incubate the sample at 95 °C for 10 minutes.
- Solubilize the IAA as stock solution at 800 mM, using either H_2O
- Add IAA to the sample, making a 20X dilution of the stock solution (final concentration 40 mM)

- Incubate the sample at room temperature for 30 minutes in the dark.

3) Add phosphoric acid to a final concentration ~1.2%). Vortex.

This step is essential to completely denature proteins and trap them efficiently. The pH will be ≤ 1.

Trap proteins

4) Add binding/wash buffer to the sample and mix.
5) Apply sample to S-Trap column. Transfer all sample including anything insoluble into the S-Trap.
6) Centrifuge the S-Trap column at 4,000 g for 30 sec to trap proteins.

Clean protein

7) Add 400 µL binding/wash buffer; centrifuge at 4,000 g for 30 sec. Repeat 3 times and discard flow through as necessary.
8) Transfer S-Trap micro column to a clean 1.7 mL sample tube for the digestion. Incubate and digest protein
9) Add 125 µL of digestion buffer containing trypsin at a 1:15 weight to weight ratio into the top of the S-Trap (20µg of trypsin for each sample).
10) Incubate overnight at 37 °C.

Elute peptides for analysis

** Three elution steps described below, must be collected in the same collector.*

11) Add 40 µL of 50 mM TEAB (or 50 mM ammonium bicarbonate) to the S-Trap then centrifuge (4,000 rcf, 1 min).
12) Add 40 µL of 0.2% formic acid to the S-Trap then centrifuge (4,000 rcf, 1 min).
13) Add 40 µL of 50% ACN – 0.2% formic acid to the S-Trap then centrifuge (4,000 rcf, 1 min). This elution assists in recovery of hydrophobic peptides. Other organics may also be used as needed.
14) Pool eluted peptides, dry down using speedvac.

Phosphopeptides enrichment

Adapted from Easyphos platform[73].

Materials required:

- 2-mL tubes
- Titansphere TiO_2, 5 µm, 500 mg, Product Number 5020-75000, GL Sciences
- C18 Spin Columns
- Acetonitrile (ACN)
- Isopropanol (ISO)

- Trifluoroacetic acid (TFA)
- KH_2PO_4
- Ammonium hydroxide (NH_4OH)
- Centrifuge
- Thermomixer
- SpeedVac

Procedure:

1) Prepare the required solutions:

SOLUTION	COMPOSITION
ISO	100% isopropanol
RESUSPENSION BUFFER	5% ACN and 0.1 % TFA
ENRICHMENT BUFFER	48% TFA and 8 mM KH_2PO_4
LOADING BUFFER	6% TFA and 80% ACN
WASH BUFFER	5% TFA and 60% ISO
TRANSFER BUFFER	0.1% TFA and 60% ISO
ELUTION BUFFER	Add 200 µL of NH_4OH to 800 µL of 40% ACN

2) Resuspend sample with 200 µL of Resuspension Buffer
3) Add 400 µL of ISO to each sample and mix thoroughly (1,500 rpm, 30 s)
4) Add 100 µL of EP enrichment buffer and mix thoroughly (1,500 rpm, 30 s)
5) Resuspend the TiO_2 beads in EP loading buffer at a concentration of 1 mg/µL and pipette an aliquot of suspended beads into each sample (recommended ratio: 9 mg for 300 µg of protein). Incubate the beads at 40 °C with shaking (2,000 rpm) for 5 minutes
6) Pellet the beads by centrifugation (2,000 g for 1 minute at RT) and discard the supernatant
7) Add 1.5 mL of EP Wash Buffer to the samples, incubate it at RT with shaking (2,000 rpm) for 30 s, pellet the beads by centrifugation (2,000 g for 1 minute at RT) and discard the supernatant. Repeat this step a total of five times.
8) After the final wash, resuspend the beads in 75 µL of EP Transfer Buffer and transfer to the top of a C18 Spin column. Repeat.
9) Place the spin column into centrifuge adapters and spin through to dryness (1,500 g for 8 minutes at RT)
10) Elute the phosphopeptides with 40 µL of EP Elution Buffer, centrifuge (1,500 g for 4 minutes at RT) to dryness and collect the eluates into clean tubes. Repeat the step and collect the eluates in a single tube.
11) Immediately place the tubes into an evaporative concentrator and concentrate under vacuum without drying the samples completely.

Cellular fractionation

Sample type: homogenized tissue or frozen tissue

ProteoExtract® Subcellular Proteome Extraction Kit

- For homogenized tissue in PBS: dry under nitrogen 500 µL of homogenate and resuspend as described in Procedure 1.
- For frozen tissue: cut and weigh 50 mg of tissue and homogenized with beads for 5 minutes in 1000 µL of Extraction buffer I.

1) Mix 1 ml ice-cold Extraction Buffer I with 5 µl Protease Inhibitor Cocktail and immediately add the mixture to the fragmented tissue or frozen cell pellet. Gently resuspend the fragmented tissue by gently flicking the tube. Incubate for 10 min at 4°C under gentle agitation.
2) Pellet the insoluble material by centrifugation for 10 min at 500-1000 g, 4°C.
3) Carefully transfer the supernatant (fraction 1) to a clean tube. Store fraction 1 on ice.
4) Mix 1 ml ice-cold Extraction Buffer II with 5 µl Protease Inhibitor Cocktail and immediately add the mixture to the pellet. Resuspend the pellet by gently flicking the tube. Incubate for 30 min at 4°C under gentle agitation. The use of a rotary shaker is recommended to avoid formation of cell clumps.
5) Pellet the insoluble material by centrifugation for 10 min at 5000-6000 g, 4°C.
6) Carefully transfer the supernatant (fraction 2) to a clean tube. Store fraction 2 on ice.
7) Mix 500 µl Extraction Buffer III with 5 µl Inhibitor Cocktail and 1.5 µl (≥ 375 U) Benzonase® nuclease and immediately add the mixture to the pellet. Resuspend the pellet by gently flicking the tube. Incubate for 10 min at 4°C under gentle agitation. The use of a rotary shaker is recommended to avoid formation of cell clumps.
8) Pellet the insoluble material by centrifugation for 10 min at 7000 g, 4°C.
9) Carefully transfer the supernatant (fraction 3) to a clean tube. Store fraction 3 on ice.
10) Mix 500 µl room temperature Extraction Buffer IV with 5 µl Protease Inhibitor Cocktail and immediately add the mixture to the pellet. Resuspend the residual particles by pipetting up and down (fraction 4).

Subcellular Protein Fractionation Kit for Tissues

- For homogenized tissue in PBS: dry under nitrogen 500 µL of homogenate and resuspend in 500 µL of CEB.
- For frozen tissue: cut and weigh 50 mg of tissue and homogenized with beads for 5 minutes in 500 µL of CEB.
- Increase volume of buffer for different tissue amount follow the volume ratio described in the table

Tissue weight (mg)	CEB (µL)	MEB (µL)	NEB (µL)	NEB (µL) +$CaCl_2$, MNnase*	PEB (µL)
50	500	325	110	80	60
100	1000	650	225	170	125
200	2000	1300	450	340	250

*MNase = Micrococcal Nuclease

Materials required:

- Subcellular Protein Fractionation Kit for Tissues (Product Number 87790, Thermo Scientific)
- Thermomixer
- Centrifuge

Procedure:

1) Centrifuge the sample homogenized in CEB in the tube at 500 × *g* for 5 minutes. Immediately transfer the supernatant (cytoplasmic extract, F1) into a clean pre-chilled tube on ice.
2) Add ice-cold MEB containing protease inhibitors to the pellet. Vortex the tube for 5 seconds on the highest setting and incubate the tube at 4°C for 10 minutes with gentle mixing.
3) Centrifuge at 3000 × *g* for 5 minutes.
4) Transfer the supernatant (membrane extract, F2) into a clean pre-chilled tube on ice.
5) Add ice-cold NEB containing protease inhibitors to the pellet. Vortex the tube on the highest setting for 15 seconds and incubate the tube at 4°C for 30 minutes with gentle mixing.
6) Centrifuge the tube at 5000 × *g* for 5 minutes. Transfer the supernatant (soluble nuclear extract, F3) fraction into a clean pre- chilled tube on ice.
7) Prepare the chromatin-bound extraction buffer by adding 5µL of 100mM CaCl2 and 3µL of Micrococcal Nuclease (300 units) per 100µL of room-temperature NEB.
8) Add room-temperature NEB containing protease inhibitors, $CaCl_2$ and Micrococcal Nuclease to the pellet. Vortex on the highest setting for 15 seconds.
9) Incubate the tube at room temperature for 30 minutes.
10) Vortex the tube on the highest setting for 15 seconds and centrifuge at 16,000 × *g* (highest microcentrifuge setting) for 5 minutes.
11) Transfer the supernatant (chromatin-bound nuclear extract, F4) fraction into a clean pre-chilled tube on ice.
12) Add room-temperature ABC with 1% of SDS to resuspend the extra-cellular matrix fraction.

Sample purification at peptide level, prior to LC-MS

C18 Spin Columns

Applicable when:

- The sample is free of excess organic solvents such as acetonitrile (ACN) or methanol. If organic solvents are present, dry the sample in a vacuum evaporator. Carefully resuspend sample in 20 µL of 0.5% trifluoroacetic acid (TFA) in 5% ACN before processing with Pierce C18 Spin Columns.
- The peptide solution is clear, without cellular debris.
- The total peptide amount is ≤ 30 µg.

Materials required:

- Pierce C18 Spin Columns (Product Number 89870, Thermo)
- Acetonitrile
- Ultrapure H_20
- Trifluoroacetic acid (TFA)
- Methanol
- Centrifuge

Material Preparation

- Activation Solution: 50% Methanol; 400 µL per sample
- Equilibration Solution: 0.5% TFA in 5% ACN; 400 µL per sample
- Sample Buffer: 2% TFA in 20% ACN; 1 µL for every 3 µL of sample
- Wash Solution: 0.5% TFA in 5% ACN; 400-800 µL per sample; wash volume will be dependent upon amount and type of contaminants present in sample
- Elution Buffer: 70% ACN; 40 µL per sample

A. Sample Preparation

Each Pierce C18 Spin Column can process 10-150µL of sample. Mix 3 parts sample to 1 part of Sample Buffer. The final sample will contain 0.5% TFA in 5% ACN.

B. Column Preparation

1) Tap column to settle resin. Remove top and bottom cap. Place column into a receiver tube.
2) Add 200µL of Activation Solution to rinse walls of the spin column and to wet resin.
3) Centrifuge at 1500 × *g* for 1 minute. Discard flow-through.
4) Repeat steps B.2-B.3.
5) Add 200µL Equilibration Solution. Centrifuge at 1500 × *g* for 1 minute. Discard flow-through.
6) Repeat step B.5.

C. Sample Binding

1) Load sample on top of resin bed.

2) Place column into a receiver tube. Centrifuge at 1500 × *g* for 1 minute.
3) To ensure complete binding, recover flow-through and repeat steps C.1-C.2.

Note: Flow-through may be retained to confirm sample binding.

D. Wash

1) Place column into a receiver tube. Add 200µL Wash Solution to column and centrifuge at 1500 × *g* for 1 minute. Discard flow-through.
2) Repeat step D.1.

Note: If sample contains high levels of contaminants (i.e., 2M urea or ≥100mM ammonium bicarbonate), repeat the wash step one to two additional times.

E. Elution

1) Place column in a new receiver tube. Add 20µL of Elution Buffer to top of the resin bed. Centrifuge at 1500 × *g* for 1 minute.
2) Repeat step E.1 with same receiver tube.
3) Gently dry sample in a vacuum evaporator. For MALDI-MS analysis, carefully suspend sample in 1-2µL of matrix solution prepared just before use. For LC-ESI applications, suspend sample in 0.1% formic acid or the appropriate buffer.

Quantification at peptide level

Materials required:

- 96-well microplate
- MultiskanGO
- Pierce Quantitative Colorimetric Peptide Assay (Product Number 23275, Thermo)

Procedure:

1) Preparation of Standards

Use the procedure in Table 1 to prepare a dilution series of the Peptide Digest Assay Standard to generate a standard curve. Dilute the Peptide Digest Assay Standard into clean vials, preferably using the same diluent as the sample(s). This will provide sufficient volume to run an 8-point standard curve (from 0-1000µg/mL) in triplicate.

2) Preparation of Working Reagent

→ Use the following formula to determine the total volume of Working Reagent (WR) required: [(# of standards) + (# of unknowns)] × (# of replicates) × (0.18mL of WR per sample) = total volume [WR required Example: For an 8-point standard curve and 5 unknowns in triplicate:

(8 standards + 5 unknowns) × (3 replicates) × (0.18mL) = 7.02mL of WR required
Note: For the above example, make at least 7.5mL of WR to ensure sufficient reagent for use.]

Prepare WR by mixing:

a. 50 parts of Colorimetric Peptide Assay Reagent A
b. 48 parts of Colorimetric Peptide Assay Reagent B
c. 2 parts of Colorimetric Peptide Assay Reagent C
d. For the above example, combine 3.75mL of Peptide Assay Reagent A, 3.6mL of Peptide Assay Reagent B and 0.15mL of Peptide Assay Reagent C.
e. **Note:** After preparation, the WR can be kept at room temperature for ≤ 30 minutes.

3) Procedure

a. Prepare samples diluted to a concentration between 500 and 60 µg/ml, based on estimated concentration from total protein digested.
b. Pipette 20µL of each standard or unknown sample replicate into a well of a 96-well microplate. Triplicate for each standard or unknown sample are recommended.
c. Add 180µL of the WR to each well and mix plate thoroughly on a plate shaker for 30 seconds to 1 minute.
d. Cover plate and incubate at 22°C (room temperature) for 30 minutes into the MultiskanGO.
e. Read the plate at 480 nm with MultiskanGO.
f. Prepare a standard curve by plotting the average blank-corrected 480nm measurement for each standard versus its concentration in µg/mL. Use the standard curve to determine the peptide concentration of each unknown sample based on the average blank-corrected absorbance value of the samples.

Mass spectrometry analysis

Preliminary untargeted analysis

For untargeted analysis has been used Q-Exactive, a high-resolution mass spectrometer composed by a quadrupole and a Orbitrap mass analyser. The mass analyser is coupled with a Vanquish.

The HPLC method used is described in the following paragraph (HPLC conditions) related to targeted analysis.

The parameters used for data-dependent analysis are the following:

Full MS Resolution: 70000

Scan range: 300 to 1500 m/z

Maximum IT: 100 ms

MS2 Resolution: 17500

Maximum IT: 100 ms

Loop count: 5

Scan range: 200 to 2000 m/z

Isolation window 2.0 m/z

MRM analysis

For targeted analysis has been used 4000 Q-TRAP (AB Sciex) in a triple quadrupole mode defined Multiple Reaction Monitoring (MRM).

The software connected to this mass spectrometer is Analyst.

HPLC conditions

Chromatograph: Agilent 1100

- Analytical column: XSelect® Peptide CSH™ C18, XP column, 4.6 mm X 150 mm, 2.5 µm (Waters)

Eluents:

A. H_2O 0.1 % formic acid

B. CH_3CN 0.1 % formic acid

Gradient:

MINUTES	A %	B %
0	97	3
5	97	3
85	55	45
90	10	90
100	10	90
100.5	97	3
110	97	3

Flow rate: 200 µL/min

Oven temp: 50 ºC

Injection volume: 5 µL

Autosampler temp: 10 ºC

MS conditions

Mass spectrometer: Q-TRAP4000

Spray voltage: 5000

Temperature.: 550ºC

Data acquisition method: MRM

Time: 100 ms

Declustering potential: 70

Collision energy: 30

Actin

The proteotypic peptides selected and analysed for the alpha smooth muscle actin are reported in Table 16.

Table 16: Actin proteotypic peptides

PEPTIDE SEQUENCE	RESIDUE	M/Z [DA]	CHARGE	TRANSITION	
GYSFVTTAER	199-208	565,77	2	Y5: 577,29	Y6: 676,36
VAPEEHPTLLTEAPLNPK	98-115	652,68	3	Y5: 568,36	Y8: 869,47

Collagen

The proteotypic peptides selected and analysed for collagen type I and collagen type III are reported in Table 17.

The proline in lowercase are hydroxyproline. For some sequences, more status of hydroxyprolination were monitored.

Table 17: Proteotypic peptides of collagen type I

CHAIN	PEPTIDE SEQUENCE	START	END	M/Z [DA]	CHARGE	TRANSITION	
Col1a1	GVVGLpGQR	948	956	449,7591	2	y6:643.35	y7: 742.41
Col1a1	GQAGVMGFpGpK	564	575	589,2875	2	y7: 765.35	y10:992.48
Col1a1	GQAGVMGFpGPK	564	575	581,2925	2	y7: 749.36	y9: 905.45
Col1a1	GFpGADGVAGPk	483	494	552,7699	2	y10 2+: 450.72	y9:787.39
Col1a1	GFpGADGVAGPK	483	494	544,7724	2	y10 2+: 442.72	y9:771.39
Col1a1	GLTGSpGSpGPDGK	528	541	629,7991	2	y9: 843.377	y6:586.28
Col1a1	GFSGLDGAK	258	265	426,2165	2	y7: 647,33	y6: 560,30
Col1a1	GFSGLDGAkGDTGPAGPK	258	274	549,937	3	y9: 799,39	y10: 943,48

Col1a2	GApGAIGApGPAGASGDR	682	698	755,8638	2	y10: 900,41	b6: 483,25
Col1a2	SGHpGPVGPAGVR	1074	1086	401,8795	3	b5: 452,18	y6: 556,32
Col1a2	GATGLpGVAGApGLpGPR	316	333	796,9222	2	y13: 1193.62	y10: 924.48
Col1a2	GLpGEFGLpGPAGPR	580	594	727,3752	2	y7: 667.34	y13: 1283.63
Col1a2	GLpGEFGLpGpAGPR	580	594	735,3732	2	y7: 683.34	y13: 1299,62
Col1a2	GLVGEPGpAGSK	349	360	542,785	2	y7:629,32	y5: 459,25
Col1a2	GEAGNIGFpGPK	493	504	580,2907	2	y6: 618.32	y4:414.23
Col1a2	GEAGNIGFpGPk	493	504	588,2875	2	y6: 634,31	y4: 430,22
Col3a1	DGSSGHPGpIGpPGPR	1150	1165	506,2433	3	y10: 976,52	y5: 539,29
Col3a1	GEAGSpGIpGPK	449	460	549,7741	2	y7: 697,38	y4: 414,23
Col3a1	GEMGpAGIpGApGLLGAR	395	412	834,9228	2	y10: 940,52	b8: 729,32
Col3a1	GEMGPAGIpGApGLLGAR	395	412	826,9232	2	y7: 699,41	b11: 954,44
Col3a1	GGpGGpGLpGPAGK	596	609	583,7939	2	y6: 542,29	b5: 342,14
Col3a1	GLAGPpGMpGPR	956	967	569,7879	2	y8: 840,40	y7: 743,34
Col3a1	GLAGppGMpGPR	956	967	577,7845	2	y8: 856,39	y7: 743,34
Col3a1	GPpGTAGTpGLR	683	694	556,7884	2	y10: 958,49	y6: 616,34
Col3a1	GPVGPHGPpGK	1139	1149	339,1825	3	y5: 471,25	b6: 545,28
Col3a1	GRpGLpGAAGAR	308	319	371,203	3	b5: 497,28	y7: 615,32
Col3a1	GEAGSpGIpGPK	449	460	549,7741	2	y7: 697,38	y4: 414,23

Histones

To investigate acetylation status of histones, some proteotypic peptides were selected to monitor Histone H3 and Histone H4 (Table 18). The lysine in lowercase represent the acetylated lysine.

Table 18: Proteotypic peptides of Histone H3 and Histone H4

HISTONE	PEPTIDE SEQUENCE	MODIFICATION	CHARGE	PRECURSOR (M/Z)	TRANSITION (M/Z)	
h4	DAVTYTEHAK		2	*567.774*	*849.4*	748.36
h4	DNIQGITKPAIR		3	*442.589*	*456.29*	685.43
h4	ISGLIYEETR		2	*590.814*	*810.39*	980.5
h4	GLGkGGAK	K12(Acetyl)	2	365.21	502.29	*559.31*
h4	GkGGkGLGK	K5(Acetyl); K8(Acetyl)	2	*443.26*	*658.38*	601.36
h3	STGGkAPR	K14(Acetyl)	2	408.22	728.4	*627.35*
h3	YRPGTVALR		3	344.86	559.35	*458.3*
h3	STELLIR		2	416.25	744.46	*643.41*
h3	EIAQDFK		2	425.71	721.38	*608.3*

TMT analysis

nanoLC conditions

Chromatograph: Thermo Scientific Dionex Ultimate 3000 nano RSLC

Column: Trap loading, separation on column

- Trap column: µ-precolumn 300 µm i.d. × 5 mm PepMap100, 5 µm, 100 Å, C18 (Thermo Scientific)
- Analytical column: NanoEase MZ HSS T3 column (75 µm × 250 mm, 1.8 µm, 100Å) (Waters)

Eluents:

A. H2O 0.1 % formic acid

B. CH3CN 0.1 % formic acid

Trapping: 3 min, 15µl/min, Solvent A

Gradient:

MINUTES	A %	B%
3	99	1
273	65	35
278	50	50
280	15	85
285	15	85
286	99	1
300	99	1

Flow rate: 250 nL/min

Oven temp: 40 ºC

Injection volume: 4 µL, mode ulPickup, 800 ng on column/sample

Autosampler temp: 10 ºC

LC-MS coupling

LC-MS coupling was performed with the Advion Triversa Nanomate (Advion BioSciences, Ithaca, NY, USA) as the nanoESI source performing nanoelectrospray through chip technology. The Nanomate was attached to an Orbitrap Fusion Lumos™ Tribrid mass spectrometer and operated at a spray voltage of 1.6 kV and a delivery pressure of 0.5 psi in positive mode.

MS conditions

Mass spectrometer: Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific)

Spray voltage: 1.6 kV

Ion transfer tube temp.: 275ºC

RF Lens: 30%

Data acquisition method: OT MS> CID IT MS2> SPS HCD OT MS3

MS1: full Orbitrap MS scan (Orbitrap)

Parent ion isolation in Q1

MS2 scan: CID in the Ion Trap

SPS HCD FTMS3 scan

Synchronous Precursor Selection in the Ion Trap

HCD MS3, FTMS detection of MS3 fragments

Phosphoproteomics analysis

nanoLC conditions

Chromatograph: Thermo Scientific Dionex Ultimate 3000 nano RSLC

Column: Trap loading, separation on column

- Trap column: µ-precolumn 300 µm i.d. × 5 mm PepMap100, 5 µm, 100 Å, C18 (Thermo Scientific)
- Analytical column: EasySpray (75µm x 500 mm, 2 µm, 100Å) (Thermo Scientific)

Eluents:

A. H_2O 0.1 % formic acid

B. CH_3CN 80%, 0.1 % formic acid

Trapping: 3 min, 15µl/min, Solvent A

Gradient:

MINUTES	A %	B%
0	95	5
3	95	5
203	71	29
247	50	50
248	1	99
264	1	99
265	99	1
300	99	1

Flow rate: 300 nL/min

Oven temp: 40 ºC

Injection volume: 1/15 µl

Autosampler temp: 4 ºC

MS conditions

Mass spectrometer: Orbitrap Fusion Lumos™ Tribrid (Thermo Scientific)

Spray voltage: 1.9 kV

Ion transfer tube temp.: 280ºC

RF Lens: 30%

Polarity: positive

Data acquisition method: OT MS> ddMS^2 OT HCD

MS OT resolution: 120000

Scan Range (m/z): 300-1300

Charge state: 2-6

ddMS^2 OT HCD

Orbitrap resolution: 30000

Activation Type: HCD

Detector type: Orbitrap

Maximum injection time: 130 ms

Bioinformatic analysis

Proteome discoverer

TMT

Software: Proteome Discoverer v2.1.0.81 software (Thermo Scientific) with SEQUEST HT algorithm

Database: UniProt Oryctolagus cuniculus (2019_10) and contaminants

Enzyme: Trypsin with 2 missed cleavages

Static modifications: Carbamidomethyl in cysteine and TMT 10plex peptide N-terminus

Dynamic modifications: TMT 10plex in K, Methionine oxidation and Acetylation in protein N-terminus

Precursor mass tol.: 10 ppm

MS/MS tolerance: 0.6

MS3 tolerance: 20 ppm

Phosphoproteomics

Two processing methods were performed: one for the whole samples, without S/T/Y phosphorylation, and one for phospho-enriched samples with the phosphorylation modification. Consensus method was used to quantify the proteins belonging from both the processing analysis.

Software: Proteome Discoverer v2.4 software (Thermo Scientific) with SEQUEST HT algorithm

Database: UniProt Rattus Norvegicus (2020_10) and contaminants

Enzyme: Trypsin with 2 missed cleavages

Static modifications: Carbamidomethyl in cysteine

Dynamic modifications: Methionine oxidation and Acetylation in protein N-terminus, Phosphorylation in S/T/Y

Precursor mass tol.: 10 ppm

MS/MS tolerance: 0.6 Da

Generic visualization tools

The Principal Component Analysis (PCA) is a multivariate statistical tool for reducing the number of variables of a big set of data, with the aim to increase the interpretability and minimize the loss of information. The plotting of the first two/three PC in a cartesian graph shows a graphical representation of the variability of the experiment. The PCA showed in this work were performed with Perseus and Proteome Discoverer on the normalized abundance value.

All the heatmap presented in this work were realized using Morpheus, a matrix visualization and analysis software (https://software.broadinstitute.org/morpheus).

The Venn Diagrams were performed using Venny, a web tool for comparing lists (Oliveros, J.C. (2007-2015) https://bioinfogp.cnb.csic.es/tools/venny/index.html).

Perseus

Statistical analysis was performed using the Perseus software vs1.6.10.43[14].

WeiGhted Correlation Network Analysis (WGCNA)

WGCNA is a tool for *omic* sciences able to define the correlation profile among genes or gene products across *omic* data. In this work, the correlation analysis was performed on protein abundance values. The WGCNA identifies clusters (modules) of highly correlated proteins using eigengene network methodology and it relates them one to another[87].

In the WGCNA, the network could be signed (modules correspond to positively correlated genes) or unsigned (modules correspond to clusters of genes with high absolute correlations). In this work, the modules were constructed on signed network. Only proteins that have abundances in all the 18 samples were kept for module construction.

In the study, modules of co-expressed proteins between the different time points were identified using the WGCNA package in R. Modules were identified using the blockwise-modules function (WGCNA package) set at default parameters, except for soft thresholding power (14), minimum module size (5) and mergeCutHeight (0.25).

Metascape

The pathway enrichment analysis was performed on module protein lists or on differentially expressed protein lists, using Metascape analysis resource available online[88]. Only pathways with a Log FDR < -4 were considered as significantly enriched.

Activation loop analysis

The phosphoproteomic data were investigated through the activation loop analysis at http://phomics.jensenlab.org, which provides a tool to analyze phosphoproteomics datasets for phosphosites that reside within a kinase activation loop.

Phosphorylation of activating phosphosites in the kinase domain's activation loop can be studied for kinase activity.

References

1. Wilkins M, Sanchez J-C, Gooley A, et al. Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It. *Biotechnol Genet Eng Rev*. 1996;13(1):19-50. doi:10.1080/02648725.1996.10647923

2. Aebersold Ruedi, Mann Matthias. Mass-spectrometric exploration of proteome structure and function. *Nature*. 2016;537(7620):347-355. doi:10.1038/nature19949

3. Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH. Proteomics: Technologies and Their Applications. *J Chromatogr Sci*. 2017;55(2):182-196. doi:10.1093/chromsci/bmw167

4. Rune Matthiesen. *Mass Spectrometry Data Analysis in Proteomics*. Humana; 2020.

5. Xian F, Hendrickson CL, Marshall AG. High Resolution Mass Spectrometry. *Anal Chem*. 2012;84(2):708-719. doi:10.1021/ac203191t

6. Matthiesen R, Carvalho AS. Methods and Algorithms for Quantitative Proteomics by Mass Spectrometry. *Methods Mol Biol*. 2020;2051:161-197. doi:10.1007/978-1-4939-9744-2_7

7. Zhang L, Elias JE. Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry. *Methods Mol Biol*. 2017;1550:185-198. doi:10.1007/978-1-4939-6747-6_14

8. Rozanova S, Barkovits K, Nikolov M, Schmidt C, Urlaub H, Marcus K. Quantitative Mass Spectrometry-Based Proteomics: An Overview. *Methods Mol Biol*. 2021;2228:85-116. doi:10.1007/978-1-0716-1024-4_8

9. Liebler DC, Zimmerman LJ. Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry*. 2013;52(22):3797-3806. doi:10.1021/bi400110b

10. Orsburn BC. Proteome Discoverer-A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes*. 2021;9(1). doi:10.3390/proteomes9010015

11. Tyanova Stefka, Temu Tikira, Cox Juergen. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols*. 2016;11(12):2301-2319. doi:10.1038/nprot.2016.136

12. Palomba A, Abbondio M, Fiorito G, Uzzau S, Pagnozzi D, Tanca A. Comparative Evaluation of MaxQuant and Proteome Discoverer MS1-Based Protein Quantification Tools. *J Proteome Res*. 2021;20(7):3497-3507. doi:10.1021/acs.jproteome.1c00143

13. Välikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform*. 2018;19(1):1-11. doi:10.1093/bib/bbw095

14. Tyanova Stefka, Temu Tikira, Sinitcyn Pavel, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*. 2016;13(9):731-740.

doi:10.1038/nmeth.3901

15. Toshiyuki Mikami, Mikio Aoki, Toru Kimura. The Application of Mass Spectrometry to Proteomics and Metabolomics in Biomarker Discovery and Drug Development. *Curr Mol Pharmacol*. 2012;5(2):301-316. doi:10.2174/1874467211205020301

16. Tuntland T, Ethell B, Kosaka T, et al. Implementation of pharmacokinetic and pharmacodynamic strategies in early research phases of drug discovery and development at Novartis Institute of Biomedical Research. *Front Pharmacol*. 2014;5:174. doi:10.3389/fphar.2014.00174

17. Bertoncello IPD, Rydell-Törmänen K, Johnson JR, Walker JM. The Applicability of Mouse Models to the Study of Human Disease. In: Bertoncello IPD, Rydell-Törmänen K, Johnson JR, Walker JM, eds. *Mouse Cell Culture : Methods and Protocols*. 1940th ed. ; 2019:3-22. doi:10.1007/978-1-4939-9086-3_1

18. Shukla SD, Swaroop Vanka K, Chavelier A, et al. Chronic respiratory diseases: An introduction and need for novel drug delivery approaches. *Targeting Chronic Inflammatory Lung Diseases Using Advanced Drug Delivery Systems*. January 2020:1-31. doi:10.1016/B978-0-12-820658-4.00001-7

19. Norman KC, Moore BB, Arnold KB, O'Dwyer DN. Proteomics: Clinical and research applications in respiratory diseases. *Respirology*. 2018;23(11):993-1003. doi:10.1111/resp.13383

20. Liu T, De Los Santos FG, Phan SH. The Bleomycin Model of Pulmonary Fibrosis. *Methods Mol Biol*. 2017;1627:27-42. doi:10.1007/978-1-4939-7113-8_2

21. Umezawa H, Ishizuka M, Maeda K, Takeuchi T. Studies on bleomycin. *Cancer*. 1967;20(5):891-895. doi:10.1002/1097-0142(1967)20:5<891::AID-CNCR2820200550>3.0.CO;2-V

22. Bauer Y, Tedrow J, de Bernard S, et al. A novel genomic signature with translational significance for human idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol*. 2015;52(2):217-231. doi:10.1165/rcmb.2013-0310OC

23. Moore BB, Hogaboam CM. Murine models of pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol*. 2008;294(2):L152-L160. doi:10.1152/ajplung.00313.2007

24. Bateman ED, Turner-Warwick M, Adelmann-Grill BC. Immunohistochemical study of collagen types in human foetal lung and fibrotic lung disease. *Thorax*. 1981;36(9):645-653. doi:10.1136/thx.36.9.645

25. McKleroy W, Lee T-H, Atabai K. Always cleave up your mess: targeting collagen degradation to treat tissue fibrosis. *Am J Physiol Lung Cell Mol Physiol*. 2013;304(11):L709-L721. doi:10.1152/ajplung.00418.2012

26. Yamauchi M, Sricholpech M. Lysine post-translational modifications of collagen. *Essays Biochem*. 2012;52:113-133. doi:10.1042/bse0520113

27. Scott I, Yamauchi M, Sricholpech M. Lysine post-translational modifications of collagen. *Essays Biochem*. 2012;52:113-133. doi:10.1042/bse0520113

28. Rydell-Törmänen Kristina, Andréasson Kristofer, Hesselstrand Roger, et al. Extracellular matrix alterations and acute inflammation; developing in parallel during early induction of pulmonary fibrosis. *Lab Invest*. 2012;92(6):917-925. doi:10.1038/labinvest.2012.57

29. Adams J, Rappu P, Salo AM, Myllyharju J, Heino J. Role of prolyl hydroxylation in the molecular interactions of collagens. *Essays Biochem*. 2019;63(3):325-335. doi:10.1042/EBC20180053

30. Henriksen K, Karsdal MA. Type I Collagen. In: Morten A. Karsdal, Diana J. Leeming, Kim Henriksen, Anne-Christine Bay-Jensen, eds. *Biochemistry of Collagens, Laminins and Elastin*. Academic Press; 2016:1-11. doi:10.1016/B978-0-12-809847-9.00001-5

31. Nielsen MJ, Karsdal MA. Type III Collagen. In: Morten A. Karsdal, Diana J. Leeming, Kim Henriksen, Anne-Christine Bay-Jensen, eds. *Biochemistry of Collagens, Laminins and Elastin*. Academic Press; 2016:21-30. doi:10.1016/B978-0-12-809847-9.00003-9

32. Dominguez R, Holmes KC. Actin Structure and Function. *Annu Rev Biophys*. 2011;40:169-186. doi:10.1146/annurev-biophys-042910-155359

33. Rockey DC, Weymouth N, Shi Z. Smooth muscle α actin (Acta2) and myofibroblast function during hepatic wound healing. *PLoS One*. 2013;8(10):E77166. doi:10.1371/journal.pone.0077166

34. Zeisberg M, Neilson EG. Biomarkers for epithelial-mesenchymal transitions. *Journal of clinical investigation, The*. 2009;119(6):1429-1437. doi:10.1172/JCI36183

35. Jenkins RG, Moore BB, Chambers RC, et al. An Official American Thoracic Society Workshop Report: Use of Animal Models for the Preclinical Assessment of Potential Therapies for Pulmonary Fibrosis. *Am J Respir Cell Mol Biol*. 2017;56(5):667-679. doi:10.1165/rcmb.2017-0096ST

36. Gorres KL, Raines RT. Prolyl 4-hydroxylase. *Critical Reviews in Biochemistry and Molecular Biology*. 2010;45(2):106-124. doi:10.3109/10409231003627991

37. Sun K-H, Chang Y, Reed NI, Sheppard D. α-Smooth muscle actin is an inconsistent marker of fibroblasts responsible for force-dependent TGFβ activation or collagen production across multiple models of organ fibrosis. *Am J Physiol Lung Cell Mol Physiol*. 2016;310(9):L824-L836. doi:10.1152/ajplung.00350.2015

38. Chung L, Dinakarpandian D, Yoshida N, et al. Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *EMBO Journal, The*. 2004;23(15):3020-3030. doi:10.1038/sj.emboj.7600318

39. Bailey AJ, Peach CM, Fowler LJ. Chemistry of the collagen cross-links. Isolation and characterization of two intermediate intermolecular cross-links in collagen. *Biochem J*. 1970;117(5):819-831. doi:10.1042/bj1170819

40. Huizen NA, Ijzermans JNM, Burgers PC, Luider TM. Collagen analysis with mass spectrometry. *Mass Spectrom Rev*. 2020;39(4):309-335. doi:10.1002/mas.21600

41. Nimptsch A, Schibur S, Ihling C, et al. Quantitative analysis of denatured collagen by collagenase

digestion and subsequent MALDI-TOF mass spectrometry. *Cell and Tissue Research*. 2011;343(3):605-617. doi:10.1007/s00441-010-1113-2

42. Naba A. The Matrisome: <IT>In Silico</IT> Definition and <IT>In Vivo</IT> Characterization by Proteomics of Normal and Tumor Extracellular Matrices. *Mol Cell Proteomics*. 2012;11. doi:10.1074/mcp.M111.014647

43. Liu W. Antiflammin-1 attenuates bleomycin-induced pulmonary fibrosis in mice. *Respir Res*. 2013;14:1-11. doi:10.1186/1465-9921-14-101

44. Decaris, Gatmaitan, FlorCruz, et al. Proteomic Analysis of Altered Extracellular Matrix Turnover in Bleomycin-induced Pulmonary Fibrosis▪. *Mol Cell Proteomics*. 2014;13(7):1741-1752. doi:10.1074/mcp.M113.037267

45. Yanagihara T, Chong SG, Vierhout M, Hirota JA, Ask K, Kolb M. Current models of pulmonary fibrosis for future drug discovery efforts. *Expert Opin Drug Discov*. 2020;15(8):931-941. doi:10.1080/17460441.2020.1755252

46. Murray LA, Knight DA, Laurent GJ. Fibroblasts. In: Peter J. Barnes, DM, DSc, et al., eds. *Asthma and COPD (Second Edition)*. Academic Press; 2009:193-200. doi:10.1016/B978-0-12-374001-4.00015-8

47. Perrin BJ, Ervasti JM. The actin gene family: Function follows isoform. *Cytoskeleton (Hoboken)*. 2010;67(10):630-634. doi:10.1002/cm.20475

48. Bergen HR, Ajtai K, Burghardt TP, Nepomuceno AI, Muddiman DC. Mass spectral determination of skeletal/cardiac actin isoform ratios in cardiac muscle. *Rapid Commun Mass Spectrom*. 2003;17(13):1467-1471. doi:10.1002/rcm.1075

49. Ravenscroft G, Colley SMJ, Walker KR, et al. Expression of cardiac @a-actin spares extraocular muscles in skeletal muscle @a-actin diseases - Quantification of striated @a-actins by MRM-mass spectrometry. *Neuromuscul Disord*. 2008;18(12):953-958. doi:10.1016/j.nmd.2008.09.010

50. Chicooree N, Unwin RD, Griffiths JR. The application of targeted mass spectrometry-based strategies to the detection and localization of post-translational modifications. *Mass Spectrom Rev*. 2015;34(6):595-626. doi:10.1002/mas.21421

51. Garcia BA, Shabanowitz J, Hunt DF. Characterization of histones and their post-translational modifications by mass spectrometry. *Curr Opin Chem Biol*. 2007;11(1):66-73. doi:10.1016/j.cbpa.2006.11.022

52. Choudhary C, Kumar C, Gnad F, et al. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science*. 2009;325(5942):834-840. doi:10.1126/science.1175371

53. Xu WS, Parmigiani RB, Marks PA. Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene*. 2007;26(37):5541-5552. doi:10.1038/sj.onc.1210620

54. Epigenetics in idiopathic pulmonary fibrosis. *Biochem Cell Biol*. 2015;93(2):159-170. doi:10.1139/bcb-2014-0126

55. Lyu X, Hu M, Peng J, Zhang X, Sanders YY. HDAC inhibitors as antifibrotic drugs in cardiac and pulmonary fibrosis. *Ther Adv Chronic Dis*. 2019;10. doi:10.1177/2040622319862697

56. Su X, Ren C, Freitas MA. Mass spectrometry-based strategies for characterization of histones and their post-translational modifications. *Expert Rev Proteomics*. 2007;4(2):211-225. doi:10.1586/14789450.4.2.211

57. Sako K-I, Haniu H, Hasegawa M. The Application of Proteomics to PK-PD Modeling and Simulation. *J Bioequiv Availab*. 2011;1(2). doi:10.4172/jbb.S2-002

58. Maiso P, Colado E, Ocio E M, et al. The synergy of panobinostat plus doxorubicin in acute myeloid leukemia suggests a role for HDAC inhibitors in the control of DNA repair. *Leukemia*. 2009;23(12):2265-2274. doi:10.1038/leu.2009.182

59. Wood P, Strong R, McArthur G, et al. A phase I study of panobinostat in pediatric patients with refractory solid tumors, including CNS tumors. *Cancer Chemotherapy and Pharmacology*. 2018;82(3):493-503. doi:10.1007/s00280-018-3634-4

60. Thompson, Schäfer J, Kuhn K, et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal Chem*. 2003;75(8):1895-1904. doi:10.1021/ac0262560

61. Salaets T, Gie A, Jimenez J, et al. Local pulmonary drug delivery in the preterm rabbit: feasibility and efficacy of daily intratracheal injections. *Am J Physiol Lung Cell Mol Physiol*. 2019;316(4):L589-L597. doi:10.1152/ajplung.00255.2018

62. Conway RF, Frum T, Conchola AS, Spence JR. Understanding Human Lung Development through In Vitro Model Systems. *BioEssays*. 2020;42(6). doi:10.1002/bies.202000006

63. Beauchemin KJ, Wells JM, Kho AT, et al. Temporal dynamics of the developing lung transcriptome in three common inbred strains of laboratory mice reveals multiple stages of postnatal alveolar development. *PeerJ*. 2016;4:E2318. doi:10.7717/peerj.2318

64. Thomas Salaets, Andre Gie, Bieke Tack, Jan Deprest, Jaan Toelen. Modelling Bronchopulmonary Dysplasia in Animals: Arguments for the Preterm Rabbit Model. *Curr Pharm Des*. 2017;23. doi:10.2174/1381612823666170926123550

65. Lopez-Rodriguez E, Perez-Gil J. Structure-function relationships in pulmonary surfactant membranes: From biophysics to therapy. *Biochim Biophys Acta Biomembr*. 2014;1838(6):1568-1585. doi:10.1016/j.bbamem.2014.01.028

66. Ueno, Linder, Na, Rice, Johansson, Weaver. Processing of Pulmonary Surfactant Protein B by Napsin and Cathepsin H*. *J Biol Chem*. 2004;279(16):16178-16184. doi:10.1074/jbc.M312029200

67. Khan T, Dasgupta S, Ghosh N, Chaudhury K. Proteomics in idiopathic pulmonary fibrosis: the

quest for biomarkers. *Mol Omics*. 2021;17(1):43-58. doi:10.1039/d0mo00108b

68. Kulkarni YM, Dutta S, Iyer AKV, et al. A proteomics approach to identifying key protein targets involved in VEGF inhibitor mediated attenuation of bleomycin-induced pulmonary fibrosis. *Proteomics*. 2016;16(1):33-46. doi:10.1002/pmic.201500171

69. Yang T, Jia Y, Ma Y, Cao L, Chen X, Qiao B. Comparative Proteomic Analysis of Bleomycin-induced Pulmonary Fibrosis Based on Isobaric Tag for Quantitation. *Am J Med Sci*. 2017;353(1):49-58. doi:10.1016/j.amjms.2016.11.021

70. Moghieb A, Clair G, Mitchell HD, et al. Time-resolved proteome profiling of normal lung development. *Am J Physiol Lung Cell Mol Physiol*. 2018;315(1):L11-L24. doi:10.1152/ajplung.00316.2017

71. Epaud R, Aubey F, Xu J, et al. Knockout of insulin-like growth factor-1 receptor impairs distal lung morphogenesis. *PLoS One*. 2012;7(11):E48071. doi:10.1371/journal.pone.0048071

72. Boggaram V. Regulation of lung sufactant protein gene expression. *Front Biosci*. 2003;8(4):D751-D767. doi:10.2741/1062

73. Humphrey Sean, Karayel Ozge, James David, Mann Matthias. High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform. *Nature protocols*. 2018;13(9). doi:10.1038/s41596-018-0014-9

74. Fíla J, Honys D. Enrichment techniques employed in phosphoproteomics. *Amino Acids*. 2012;43(3):1025-1047. doi:10.1007/s00726-011-1111-z

75. Wollin L, Wex E, Pautsch A, et al. Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. *Eur Respir J*. 2015;45(5):1434-1445. doi:10.1183/09031936.00174914

76. Larsen, Thingholm, Jensen, Roepstorff, Jørgensen. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns*. *Mol Cell Proteomics*. 2005;4(7):873-886. doi:10.1074/mcp.T500007-MCP200

77. Thingholm TE, Larsen MR. The Use of Titanium Dioxide for Selective Enrichment of Phosphorylated Peptides. *Methods Mol Biol*. 2016;1355:135-146. doi:10.1007/978-1-4939-3049-4_9

78. Munk S, Refsgaard JC, Olsen JV, Jensen LJ. From Phosphosites to Kinases. *Methods Mol Biol*. 2016;1355:307-321. doi:10.1007/978-1-4939-3049-4_21

79. Pucheta-Martínez E, Saladino G, Morando MA, et al. An Allosteric Cross-Talk Between the Activation Loop and the ATP Binding Site Regulates the Activation of Src Kinase. *Sci Rep*. 2016;6:24235. doi:10.1038/srep24235

80. Katz M, Amit I, Yarden Y. Regulation of MAPKs by growth factors and receptor tyrosine kinases. *Biochim Biophys Acta Mol Cell Res*. 2007;1773(8):1161-1176. doi:10.1016/j.bbamcr.2007.01.002

81. Rangarajan S, Kurundkar A, Kurundkar D, et al. Novel Mechanisms for the Antifibrotic Action of Nintedanib. *Am J Respir Cell Mol Biol*. 2016;54(1):51-59. doi:10.1165/rcmb.2014-0445OC

82. Wollin L, Wex E, Pautsch A, et al. Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. *Eur Respir J*. 2015;45(5):1434-1445. doi:10.1183/09031936.00174914

83. Hilberg F, Roth GJ, Krssak M, et al. BIBF 1120: Triple Angiokinase Inhibitor with Sustained Receptor Blockade and Good Antitumor Efficacy. *Cancer Res*. 2008;68(12):4774-4782. doi:10.1158/0008-5472.CAN-07-6307

84. Wollin L, Maillet I, Quesniaux V, Holweg A, Ryffel B. Antifibrotic and Anti-inflammatory Activity of the Tyrosine Kinase Inhibitor Nintedanib in Experimental Models of Lung Fibrosis. *J Pharmacol Exp Ther*. 2014;349(2):209-220. doi:10.1124/jpet.113.208223

85. Frantzi M, Latosinska A, Mischak H. Proteomics in Drug Development: The Dawn of a New Era? *Proteomics Clin Appl*. 2019;13(2). doi:10.1002/prca.201800087

86. Wisniewski Jacek R, Zougman Alexandre, Nagaraj Nagarjuna, Mann Matthias. Universal sample preparation method for proteome analysis. *Nature Methods*. 2009;6(5):359-362. doi:10.1038/nmeth.1322

87. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. December 2008:559. doi:10.1186/1471-2105-9-559

88. Zhou Yingyao, Zhou Bin, Pache Lars, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*. 2019;10(1):1523. doi:10.1038/s41467-019-09234-6