UNIVERSITÀ DEGLI STUDI DI SASSARI

Scuola di Dottorato di ricerca in Scienze Biomolecolari e Biotecnologiche

Indirizzo Biotecnologie molecolari e cellulari

XXII CICLO

Direttore: Chiar.mo Prof. Bruno Masala

Novel methods for proteomics analysis of formalin-fixed, paraffin-embedded tissues (FFPE), and their application for biomarker discovery

Tutor:
Prof. Sergio Uzzau

Tesi di dottorato di:
Dott. Alessandro Tanca

ABSTRACT

A wealth of information on proteins involved in many aspects of disease is encased within formalin-fixed paraffin-embedded (FFPE) tissue repositories stored in hospitals worldwide. However, FFPE full-length protein extracts, as described to date, often exhibit a low pattern complexity, together with a poor suitability for downstream gel-based proteomic techniques. Thus, an optimised method for extraction of full-length proteins from FFPE tissues was developed. The results obtained analysing FFPE muscle, liver, and thyroid extracts by GeLC-MS/MS, western immunoblotting, protein arrays, ELISA, are presented and discussed. An evaluation of the extent of modifications introduced on proteins by formalin fixation and crosslink reversal, and their impact on quality of MS results, is also reported.

Moreover, 2D-PAGE, and 2D-DIGE, protein maps of formalin-fixed samples, which show satisfactory proteomic information and comparability to fresh tissues, have never been illustrated so far. In the present study, 2D-PAGE and 2D-DIGE separation and mass spectrometry identification of full-length proteins extracted from fixed skeletal muscle and liver tissues are reported. Furthermore, interesting features are detailed, such as the composition of formalin-induced protein artificial complexes, the constant acidic shift of FFPE proteins p*I*, as well as the inverse correlation between percentage of lysine residues of a protein and preservation of signal intensity of its 2D-spot.

Finally, the application of 2D-DIGE-MS and GeLC-MS/MS for differential proteomic investigation of FFPE diseased samples was pursued. First, a comparison between the proteomes of sheep liver samples, affected by heavy metal intoxication at two different levels, is presented. Several differential stress biomarkers were detected, and data achieved by the two gel-based proteomic methods exhibit a high consistency. Subsequently, a biomarker discovery study was conducted, analysing two human lung neuroendocrine tumour forms (typical carcinoid, TC, and small cell lung carcinoma, SCLC). Duplicate GeLC-MS/MS analysis of 3 TC and 3 SCLC cases led to the identification of over 400 unique proteins per tumour class, which were classified according to molecular function, biological process, and cellular localisation. A panel of

over 30 differentially expressed putative biomarkers was identified. In addition, a 2D-DIGE experiment is described, whose results clearly support previous biomarkers identification.

TABLE OF CONTENTS

1. AIM OF THE STUDY

I. Preliminary remarks

The research field denominated *biomarker discovery* can be considered as a hub of biomedical translational research. A plethora of degenerative and infectious diseases can be subject of a biomarker discovery study. Indeed, the crucial challenge is to make molecular information achieved for a specific illness state useful for clinical applications.

Biomarkers can be classified according to different points of view. Under a clinical approach, diagnostic, prognostic, and predictive biomarkers can be distinguished. Diagnostic biomarkers facilitate to discriminate a pathological from a healthy state; prognostic ones allow identifying severity degree of a disease and predicting its future course; last, but not least, a predictive marker enables to select the more appropriate drug for a particular subtype of a disease, in relation to resistance forms or peculiar physiopathological features.

Moreover, several biological molecules can be targeted as disease markers, namely genomic sequences, transcripts (mRNA, miRNA), proteins, and metabolites. Protein biomarkers reflect more truly the functional properties of a tissue and can be measured by cheap and widespread methods, such as immunoassays. Recent technological development makes available a wide range of molecular techniques, which represent an optimal tool for mining the proteome in search for robust and suitable markers. Differential proteomics projects, carried out by electrophoresis- and/or mass spectrometry-based approaches, are the best way for identifying molecule panels which can be associated with a typical pathological state.

However, when the aim is to characterise the proteome of a diseased tissue, poor availability of clinical samples often makes difficult to perform thorough and statistically supported researches. Therefore, it is very important to find new methods to elucidate molecular signature of important diseases with a low, or lowest, population prevalence.

II. Primary and secondary objectives

The objective of this study was to develop and apply proteomics methodologies useful for biomarker research in rare disease samples. Proteomic techniques allow the identification of differential protein profiles between pathological and healthy tissues, or between different degrees or subtypes of a particular disease.

Yet, the only feasible way to investigate rare disease samples relies on the valuable repositories of fixed and embedded bioptic tissues, stored in hospitals worldwide. Unfortunately, the fixation procedure, most frequently performed with formalin, causes an "entrapment" of macromolecules within the tissue matrix, due to intra- and intermolecular crosslinking, which dramatically impairs protein extraction efficiency and antigen immunoreactivity. The methods developed so far to extract proteins from fixed tissues samples exhibit a poor suitability for downstream proteomics techniques. Thus, the primary objective of this work was to develop an efficient and versatile procedure for protein extraction from formalin-fixed, paraffin-embedded tissues (FFPE), that would allow differential proteomic analysis by widespread laboratory methods.

In keeping with this effort, the present choice for a diagnostic model fell on lung neuroendocrine tumours (lung NETs), because of their heterogeneity, clinic impact and diagnostic complexity. Indeed, since some of them (especially typical carcinoid, TC) show a very low frequency, it is not realistic to envisage that a whole biomarker discovery study could be carried out analysing only fresh biopsies. Hence, to perform an in-depth proteomic characterisation of model FFPE tissues, a secondary objective of this study was to apply the above mentioned method to the investigation of diagnostic and prognostic markers of lung NETs.

2. EXPERIMENTAL METHODOLOGY

I. State of the art in biomarker discovery methodologies

1. Biomarkers: essential concepts

The term *biomarker* (biological marker) was introduced in 1989 as "measurable and quantifiable biological parameter which serves as index for health- and physiology-related assessments". In 2001, an NIH working group standardised its definition as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention". Though historically associated with a physical trait or a physiological measure, the term is now typically synonymous to *molecular biological marker* [1,2].

Biomarkers can be indicators of disease trait (risk factor), disease state (preclinical or clinical), or disease rate (progression). Accordingly, biomarkers can be classified as *antecedent* biomarkers (identifying the risk of developing an illness), *screening* biomarkers (screening for subclinical disease), *diagnostic* biomarkers (recognizing overt disease), *staging* biomarkers (categorizing disease severity), *prognostic* biomarkers (predicting future disease course), *predictive* biomarkers (forestalling response to therapy), or *monitoring* biomarkers (monitoring efficacy of therapy). In the biopharmaceutical industry, biomarkers define molecular taxonomies of patients' diseases and serve as *surrogate endpoints* in early-phase drug trials [1-4].

Biomarker discovery is a systematic process in which a series of experimental aims need to be considered. Currently, the most commonly applied pipeline for biomarker development involves a global discovery phase on small numbers of samples followed by validation of the potential biomarker with large numbers of patient samples, before it is eventually adopted as a clinical tool. A plan including five distinct phases for the development and testing of disease biomarkers, established by the US National Cancer Institute, is shown in Table 2.1 [5-6].

Table 2.1. The five phases involved in the development and testing of disease biomarkers (Issaq, 2007)

Phase	Description
1	Discovery phase: Exploratory studies to identify potentially useful biomarkers.
2	Validation phase: Biomarkers are studied to determine their capacity to distinguish between diseased and healthy people.
3	Studies to assess the capacity of a biomarker to detect preclinical disease by testing the marker against tissues collected longitudinally from research cohorts.
4	Prospective screening studies.
5	Definitive large-scale population studies to determine the overall impact of screening on health outcomes in the target populations.

An ideal biomarker should be found in a sample whose acquisition is as noninvasive as possible (blood, urine, saliva, *etc.*), and sample preparation is minimal. The required technology for biomarker identification should be simply to operate, easily accessible, and low-cost. The results should be rapidly attainable and easily interpreted, and should give a high degree of accuracy with no false-positives or false-negatives (*i.e.*, 100% specificity and sensitivity) [6].

2. "Omics" revolution and biomarker discovery

The sequencing of the human genome, together with the introduction of new technologies allowing rapid, wide range, quantitative analysis of gene and protein expression in cells and tissues, were perceived as a major revolution in biomedical research. Besides genomics, several holistic approaches aimed at studying entities in groups or aggregates ("omics") have been introduced and developed more or less successfully, including epigenomics, transcriptomics, microRNomics, proteomics, metabolomics, gylcomics, lipidomics, interactomics, and so on.

The concept beyond omics approaches is that a given biological or molecular system can be best determined and understood by considering it in its globality, rather than studying its components individually. This concept is also the founding principle of *systems biology*, in which a given system is monitored in its entirety ("holistically") in response to changes, so that it can be better modeled and thoroughly understood compared to a single gene, or single pathway, approach. Omics studies generate high-dimensional data sets, which often require the implication of mathematical, statistical and computational efforts (i.e. bioinformatics) and the development of novel analytical tools [7-10].

All omics sciences, in view of aforenamed features, can represent a useful tool for the identification of new biomarkers. Comparing molecular profiles of healthy and pathological samples, or those of different subtypes of a disease, according to a genomic, transcriptomic, proteomic, and/or metabolomic approach, can highlight differential patterns, which may have a significative clinical application.

3. Genomics, transcriptomics, and metabolomics approaches

Recent advances in microarrays and deep sequencing technologies allow to thoroughly investigate DNA and RNA peculiarities of a tissue. Genomics, epigenomics, transcriptomics, and miRNomics, can exploit these two high-throughput approaches in order to characterise and quantify cellular nucleic acids and their modifications.

Genomics focuses on determination of DNA sequence variations among a population, which can be strongly correlated to disease risk. Genome-wide association studies investigate millions of common DNA sequence variants in the human genome, whose most common type are single nucleotide polymorphisms (SNPs), which can informative antecedent biomarkers. Also sequencing of tumour cell genomic DNA has been applied to identify “driver mutations”, which might clearly have an important impact on cancer molecular diagnosis and prognosis assessment [12-14].

Epigenomics studies essentially DNA methylation and post-translational modifications of histones, which can both be related to various illness states. Recent technology development has enabled the analysis of DNA methylation in a genome-wide scale, by array-based or non-array-based approaches, allowing the detection of epigenetic cancer biomarkers [12, 15-16].

Transcriptomics has been used to describe the global mRNA expression of a particular tissue, yielding information about the transcriptional differences between two or more healthy or disease states. Genome-wide expression studies empowered by microarray analysis enable the systematic analysis of complex biological systems. Differentially expressed genes demonstrated the potential to serve as prognostic biomarkers and therapeutic targets [12, 17].

Also microRNAs (miRNA) have been implicated in many human diseases, including tumor initiation and progression. Differential expression of miRNAs has revealed

diagnostic, prognostic and therapeutic implications. There are several approaches for genome scale miRNA expression profiling, such as miRNA oligo arrays, multiplexed q-RT–PCR assays and bead-based methods [18-24].

Metabolomics is a discipline that aims to identify and quantify the global composition of 'metabolites' of a biological fluid, tissue, or organism. A 'metabolite' may be defined as a native small molecule that participates in general metabolic reactions and is required for the maintenance, growth, and normal function of cells. The primary aim of global profiling of the endogenous and exogenous metabolites in biological samples is to discover single or associated biomarkers that aid in early diagnosis [25-26].

The experimental approach for global metabolic analysis has evolved over time. Nuclear magnetic resonance spectroscopy (NMR) has been used long since for the study of the chemical composition of biological fluids; recently, other techniques has been introduced, like gas chromatography (GC) (for nonpolar compounds) and liquid chromatography (LC) (for polar compounds), followed by MS detection, as well as Fourier transform infra-red spectroscopy, Raman spectroscopy, and direct infusion ESI-MS. There is no current method capable to detect and quantitate the entire metabolome of a tissue; indeed, subdivisions of metabolomics are now emerging, such as peptidomics, glycomics, and lipidomics [27-30].

4. Proteomics techniques for biomarker discovery: gel-based *vs* gel-free

The term proteome was first coined to describe the set of proteins encoded by a genome in a specific cell type or condition. The study of the proteome, called proteomics, now evokes not only all the proteins in any given cell, but also the set of all protein isoforms and modifications, the interactions between them, the structural description of proteins and their higher-order complexes. Since proteins represent the preponderance of the biologically active molecules responsible for most cellular functions, it is believed that the direct measurement of protein expression can more accurately indicate cellular dysfunction underlying the development of disease; thus, proteomics holds special promise for biomarker discovery. Proteomics techniques useful for biomarker research can be divided into two principal groups: gel-based proteomics methods and gel-free proteomics methods [4, 31].

A gel-based proteomic experiment comprises a first separation of a protein mixture through gel electrophoresis, followed by MS identification. The two most common forms of gel electrophoresis are 2D-PAGE and 2D-DIGE. In both methods, proteins are resolved in two sequential steps, according to their charge in the first dimension and to their mass in the second dimension. Good resolution of proteins is obtainable; however, the technique is not amenable to automation or direct coupling to the mass spectrometer. In 2D-PAGE, each sample is analysed in a distinct gel. After separation, staining is carried out to reveal protein expression; protein signals are compared, spots of interest are excised, proteins within the spot are enzymatically digested, and resultant peptides are extracted and analysed using MS [32-25] .

Ünlü *et al.* introduced 2D-DIGE to improve the reproducibility and reliability issues encountered with 2D-PAGE. In a simple DIGE experiment, proteins from two samples are labelled, each with a different cyanine fluorescent dye that has a different excitation and emission wavelengths. The two samples are mixed and loaded in the same gel. The same protein labelled with any of the dyes will migrate to the same position on the gel, because the dyes do not affect the size or the isoelectric point of the protein. Proteins are visualized by subjecting the gel to the excitation wavelengths of each of the dyes. Spots of interest are excised and treated as mentioned earlier. (Issaq, 2009) Several benefits of 2D-DIGE for proteomic analysis of biological samples in search of a biomarker can be highlighted. First, the use of the internal standard, which allows reliable comparison between different gels; second, separation of the two protein extracts of interest employing the same gel improves reproducibility because 50% fewer gels are required; third, differences in protein expression between two samples of proteins are easier to compare and can be more accurately imaged; fourth, DIGE requires less time to detect the protein spots because the labelling reaction in DIGE is faster than visualization using other staining methods. Hence, 2D-DIGE can be considered as the method of choice when absolute protein variation between two biological samples is the primary target, as in biomarker discovery projects. (32, 33, 36-39)

An alternative simple and easily performed approach is one-dimensional (1D)-SDS-PAGE protein separation, combined with nano-LC-MS/MS of in-gel-generated peptides

for protein identification. This method, named GeLC-MS/MS, constitutes an interesting compromise between the advantages and shortcomings of the aforementioned techniques; actually, it gains access to categories of proteins typically not accessible via 2D-PAGE (notably, large proteins and those with extreme pIs), and exploits excellent sensitivity of LC-MS/MS identification with samples of low complexity, as prefractionated by an electrophoretic separation [40-43].

Among gel-free methods, "shotgun proteomics", also named multidimensional protein identification technology (MudPIT), has achieved a great success and a widespread application in the last few years. This technique consists of an early enzymatic digestion of the entire protein mixture, and a subsequent multidimensional separation of peptides, followed by on-line mass spectrometric peptide detection and sequencing. Many approaches, using two or more orthogonal HPLC procedures (for example, IEX/RP, Affinity/IEX/RP), or capillary electrophoresis separation (namely, CZE, CEC, CIEF), have been tried to separate complex peptide mixtures such as cell lysate digest [44-48]. Relative quantitation of proteins in a bottom-up LC/MS-based proteomics experiment is carried out by a variety of stable isotope labelling techniques. Typically, control and compared samples are differentially stable isotope tagged using *in vivo* or *in vitro* labelling. An alternative approach has been developed, that does not require differential stable isotope labelling, and is thus referred to as label-free method. This approach is attractive because of its simplicity, cheapness and easiness to be carried out. However, it requires high reproducibility of HPLC separation and high sensitivity to detect low abundance peptides [33, 49-52]

Furthermore, proteomic analyses have been recently extended to whole tissue sections by using MALDI imaging. This approach allows detection of proteins *in situ* in the tissue sections, and analysis of their spatial distribution. Each MALDI imaging data set containing large number of mass spectra may be evaluated using hierarchical clustering, that is helpful in classification and interpretation of complex tissue proteomic patterns [53-54].

Another MS-based technology, useful for proteomic patterns comparison for diagnostic purposes, is SELDI TOF-MS. In this technique, sample is captured by the chromatographic surface of a protein chip arrays and then analysed by MS. The result

is a mass spectrum of all the bound proteins/peptides, which serves as a "fingerprint" of tissue proteome [55-56].

Trying to summarise principal benefits and drawbacks of gel-based and gel-free methods, it can be noted that gel-based strategies offer the advantage of visualisation of proteins and, to some extent, their modifications, and therefore preserves the protein context. On the other hand, gel electrophoresis has the intrinsic limitation of reducing sample complexity to more limited MW and pI ranges, if compared to shotgun approach. By contrast, early digestion strategies generate peptides upstream in the analysis workflow; peptides are more easily amenable to separation and analysis and behave more uniformly than proteins. Another difference between the two methods is that the relative abundances of proteins in the samples are compared before MS analysis (2D-PAGE), while in LC/MS the comparison is carried out after the MS data have been acquired [57-58].

II. Methods for extracting and analysing proteins from FFPE tissues

1. Formalin fixation, antigen retrieval and protein extraction methods

Formaldehyde fixation and paraffin embedding of tissue biopsies are routinely performed in hospitals worldwide. Formaldehyde fixation allows long-term stability and maintenance of tissue architecture, and paraffin embedding facilitates handling and microtome slicing of samples [59]. The combined FFPE treatment is aimed to histological examination of tissues, and results in highly stable tissue blocks which can be easily stored for years at room temperature in form of "biopsy banks". In the last decades, large repositories of healthy and pathological FFPE tissues have been generated worldwide, with associated information concerning diagnosis, treatment, and outcome of the patient. These FFPE tissue archives might therefore represent an important resource for retrieving information on proteomic repertoires associated to many disease processes.

Unfortunately, however, the FFPE procedure leads to what is regarded as an irreversible "entrapment" of macromolecules within the tissue matrix, consequent to

the establishment of extensive intra-and intermolecular crosslinks. Formaldehyde fixation acts on basic and amidic side chains through a serial process involving generation of methylol (hydroxymethylol) adducts, formation of Schiff bases, and finally introduction of apparently irreversible, stable methylene bridges. These bonds impair protein extraction efficiency, reduce immunoreactivity, and may result in alterations leading to misidentification of proteins with MS techniques [60-63].

For several years, immunohistochemistry (IHC) has been the only technique able to gather information from FFPE samples. In 1991, Shi and coworkers demonstrated that a brief incubation at high temperatures in suitable buffers allowed a considerable improvement in the IHC signal [64]. This simple and efficient empiric method, named *antigen retrieval*, is still regarded as the only procedure able to induce a strong increase of the immunohistochemical signal. The mechanism underlying the antigen retrieval process seems to rely on restoration of antigenic properties through a series of protein conformational changes, probably facilitated by the hydrolysis of formalin-induced crosslinks [65-70].

Almost all FFPE full-length protein extraction methods developed so far (also patented ones) are based on antigen retrieval concept; they necessarily provide a heating step, in order to release an adequate amount of proteins from the crosslinked matrix [71-79].

2. FFPE proteomics: shotgun *vs* full-length analysis

Many efforts are now underway by several research groups, aimed to release the proteomic information entrapped within this tissue matrix. To date, two main strategies have been applied, with varying degrees of success. One is based on direct proteolytic digestion followed by MS characterization of the complex peptide mixture by means of sophisticated mass spectrometry instrumentation, according to a shotgun proteomics experimental workflow. The second approach is aimed to recover full-length proteins by reversing the formaldehyde-induced crosslinks, through various modifications of the seminal antigen retrieval technique above mentioned. Reaching this goal would open the way to observe differential protein expression by combining 2D-PAGE/MS analysis, immunostaining of 2D-PAGE separated proteins, staining for

phosphorylated or post-translationally-modified proteins, and, last but not least, differential in gel electrophoresis (DIGE), on proteins extracted from FFPE tissues.

In a pioneering work published in 1998, Ikeda and coworkers [71] used the heat-induced antigen retrieval process to obtain protein extracts from FFPE tissues, but failed to obtain results comparable to fresh tissues. However, their method allowed the identification of eight diagnostically useful proteins at the expected molecular weight by western immunoblotting. Since then, several studies have been undertaken with various modifications to obtain proteins from FFPE tissues. Quality and viability of FFPE protein extracts have been evaluated by classical protein analysis methods such as SDS-PAGE, western immunoblotting, and protein arrays, with variable results [73, 79]. Recently, Becker and coworkers [80-83] developed a patented heat-extraction method and demonstrated its applicability to expression studies based on reverse phase protein microarrays. In these studies, despite the intensive efforts to settle the issue, quality of FFPE extracts is overall far from enabling differential proteomic analysis and application of mass spectrometry procedures on SDS-PAGE separated proteins.

Furthermore, only few studies to date reported the tentative application of 2D-PAGE to protein extracts from FFPE tissues [84-85]. Noteworthy, the aim of one of these two reports was to demonstrate the purported unsuitability of FFPE tissues to 2D proteomic studies. Therefore, application of 2D-PAGE separation followed by in-situ protein digestion and MS or MS/MS analysis, the fundamentals of gel-based proteomics, still appears to be challenging on FFPE tissues.

To overcome the limitations due to protein quality, several authors recently provided a rational development of methods to extract peptides from FFPE tissues followed by direct mass spectrometry identification. Taken together, these studies provide several methods enabling peptide MS analysis on FFPE extracts, which have been fruitfully applied for biomarker discovery investigations on a vast set of tissues and diseases [74-75, 77-78, 86-97]. However, information about the entire protein is lost, protein extract is not suitable also for subsequent immunological assays, and, most of all, advanced instrumentation is required for direct MS analysis of the extracted peptides, such as Orbitrap mass spectrometers or 2D chromatography-MS/MS facilities. At the

time of this writing, the application of these methods by use of more widespread analytical platforms, such as LC-ESI-QTOF instrumentations, has restricted identification to a very limited number of proteins [76]. Recently, also MALDI imaging has been successfully employing for the proteomic characterisation of FFPE tissues [98-103].

III. Gel-based proteomics on FFPE tissues: what, how and why

In view of the above mentioned state of the art in FFPE proteomics field, gel-based proteomics techniques were chosen to be applied for the characterisation and biomarker discovery-oriented investigation of formalin-fixed tissues. As described in the previous section, this is an almost unexplored field in FFPE research, and the potential of the wide gamut of methods, ranging from GeLC-MS/MS to 2D-DIGE, have still to be thoroughly and effectively exploited. Moreover, working with full-length proteins bears several advantage, such as the opportunity to use the same protein extract for electrophoretic, MS, and immunological analysis, to search for post-translational modifications, and to evaluate the overall "entire proteins" profile of a tissue. The pre-fractionation step carried out by 1D- or 2D-electrophoresis leads also to the possibility to achieve protein identification without the need of highly sophisticated MS instrumentation.

According to this reasons, the development of an optimised full-length protein extraction method from FFPE tissues was pursued. Such an extract is suitable for direct PAGE analysis, and allows the application of a vast and complete assortment of downstream proteomic techniques and immunoassays. Hence, several animal and human tissues (healthy skeletal muscle, liver, and thyroid, as well as pathological liver, lung carcinoid, and small cell lung carcinoma) were subjected to a proteomic investigation by means of various gel-based approaches, as GeLC-MS/MS, 2D-PAGE-MS, and 2D-DIGE-MS, with the final purpose of identifying useful putative disease biomarkers.

3. EXPERIMENTAL ACTIVITIES

I. Experimental design

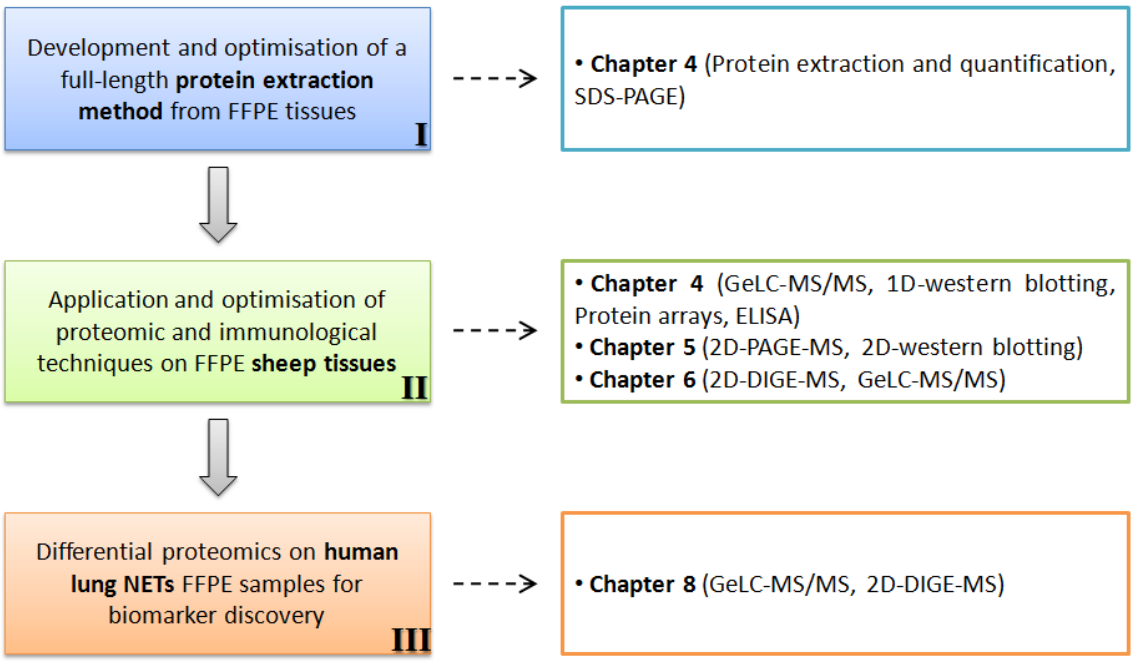

Figure 3.1. Flowchart representing overall experimental design.

II. Samples

1. Sheep samples

Sheep tissues used for optimisation of extraction and analysis procedures were obtained as follows. Immediately after slaughtering, tissue cubes smaller than 1 cm^3 were cut from femoral muscle and liver, and multiple replicate samples were immediately frozen at -80°C, or fixed in 10% buffered formalin for 48 h, dehydrated with a graded series of alcohol, and embedded in paraffin, respectively.

2. Human samples

Samples of human hyperplastic thyroid gland were obtained from routine hospital surgical procedures and immediately frozen at -80°C. Bioptic human hyperplastic

thyroid gland, small cell lung carcinoma, and typical carcinoid sections were cut from 1 to 60 months old FFPE tissue blocks stored in hospital archives.

III. Methodologies

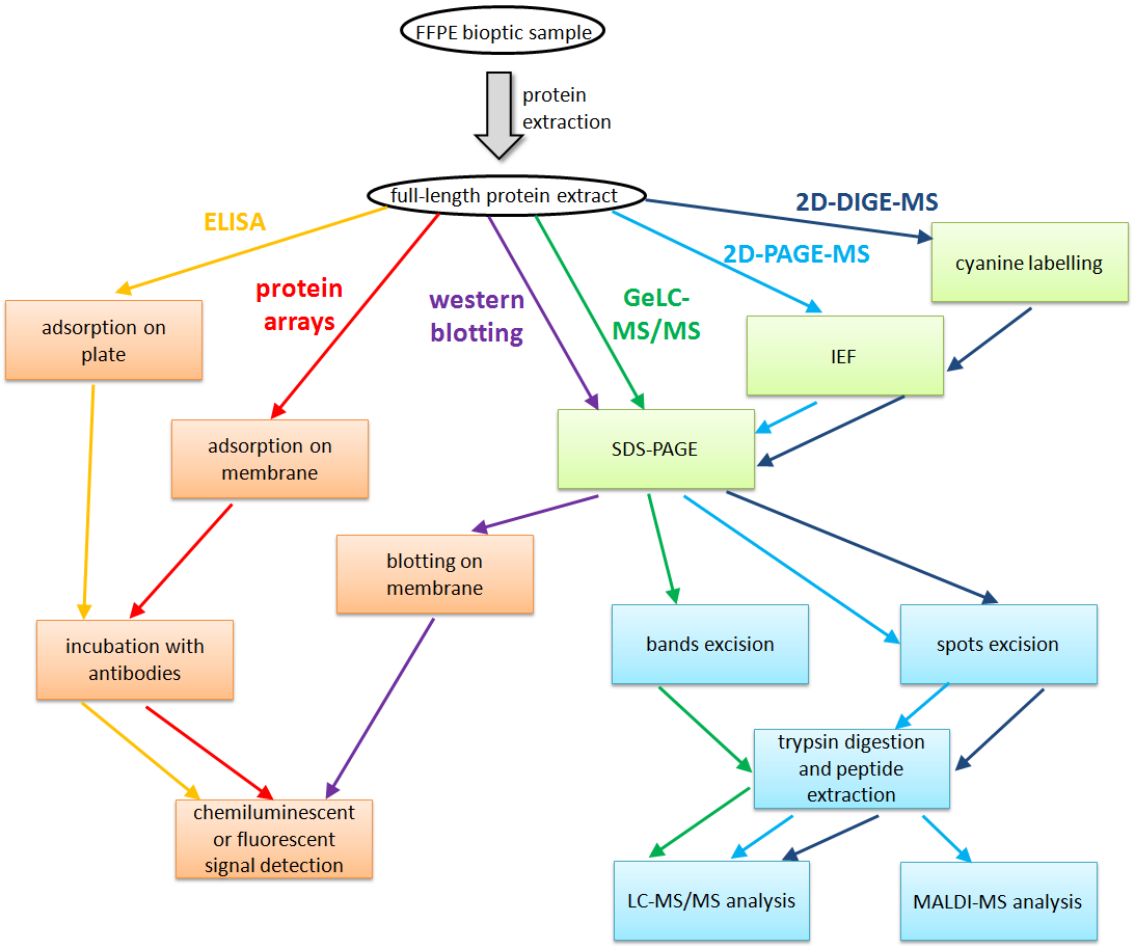

Figure 3.2. Methodologies flowchart

1. Protein extraction and quantitation

Replicates of 10 FFPE tissue microtome sections, 10 µm thick, 80 mm^2 wide, were placed in Eppendorf safe-lock tubes (Eppendorf, Hamburg, Germany) and deparaffinised by incubation at RT in xylene for 10 min. After each incubation, the tissue was pelleted at 12,000 x g for 3 min, and incubation/centrifugation steps were repeated two more times. The deparaffinised tissue pellets were then rehydrated with a graded series of ethanol. Replicates of fresh-frozen tissues were minced with a sterile

scalpel and placed in Eppendorf safe-lock tubes. All samples were then weighed, and matched amounts of tissues were each immersed at a 20% w/v ratio in a 20 mM Tris HCl pH 8.8, 2% SDS, 200 mM DTT, extraction buffer (EB). All samples were subjected to high-temperature extraction at 100°C for 20 min, and then at 80°C for 2 h with shaking.

For extraction efficacy and protein recovery evaluation, replicates of ten 80 mm^2 wide, 10 µm thick, FFPE skeletal muscle and liver tissue sections, and replicates of ~35 mg of fresh-frozen matched tissues were treated as follows. FFPE tissue replicates were deparaffinised, rehydrated, and weighed as above. Three replicates of deparaffinised, rehydrated FFPE muscle and liver tissue pellets, and three aliquots of fresh-frozen muscle and liver tissue were centrifuged to dryness using a Centrivac system (Heraeus, Hanau, Germany) and reweighed in order to assess average dry weight compared to wet weight. Three replicates of each deparaffinised, rehydrated tissue and three replicates of each fresh-frozen tissue were then subjected to comparative extraction as above or with the method described by Shi et al. [75]. Tissue pellets were alternatively immersed in of EB or retrieval solution (RS, 20 mM Tris HCl pH 7, 2% SDS), with a ratio of 5 µl for 1mg of wet weight. Extraction was then carried out at 100°C for 20 min, and then at 80°C with agitation (EB samples), or at 60°C without agitation (RS samples), for 2 h.

For preparation of protein extracts of fresh-frozen tissues with mechanical lysis, replicates of fresh-frozen tissues were fragmented into small pieces with a sterile scalpel, placed in 2 mL Eppendorf safe-lock tubes (Eppendorf, Hamburg, Germany), immersed at 5% w/v in EB, and subjected to three cycles of 3 min at 30 cycles/s in a TissueLyser mechanical homogenizer (Qiagen, Hilden, Germany).

All extracts were clarified for 15 min at 12,000 x g at 4°C, quantified by the Bradford method [104] and with the EZQ protein quantitation kit (Molecular Probes, Eugene, OR), and stored at -80°C until needed.

2. Electrophoretic techniques: SDS-PAGE, 2D-PAGE, 2D-DIGE

Protein extracts obtained from fresh-frozen and FFPE tissues were subjected to SDS-PAGE according to Laemmli [105]. Proteins were stained with Brilliant blue G 250

according to Westermeier [106] and gel images were digitalised with an ImageScanner (GE Healthcare, Little Chalfont, UK).

For two-dimensional electrophoresis, protein extracts obtained from fresh-frozen and fixed tissues were precipitated with TCA-Acetone [20]. To 200 µl of FFPE protein extract in EB, 40 µl of 60% TCA were added and the mixture was incubated overnight in ice. The mixture was centrifuged at 10,000 x g at 4°C for 30 min, the supernatant was removed, and 100 µl of ice-cold acetone were added to the pellet. The sample was incubated on ice for 15 min and centrifuged as above. The acetone-containing supernatant was removed and the pellet was air-dried. Protein pellets were resuspended at the desired protein concentration in 9 M urea, 4% CHAPS, 1% 3-10 NL IPG buffer, and 50 mM DTT.

Resuspended proteins were then absorbed overnight into 7 cm or 18 cm 3-10 NL IPG strips. Focusing was carried out for a total of 25,000 and 55,000 Vh for 7 cm and 18 cm strips, respectively. After focusing, strips were equilibrated in 50 mM Tris HCl, pH 6.8, 2% SDS, 7 M urea, 10% glycerol, supplemented with 2% DTT for 10 min, and then with 2.5% iodoacetamide for 10 min. Proteins were subjected to SDS-PAGE on 12% polyacrylamide gels with the mini and midi formats (10 cm and 20 cm, respectively) on the Mini Protean Tetra Cell and on the Protean Xi Cell (Bio-Rad Laboratories, Hercules, CA), respectively, following the manufacturer instructions.

Proteins were stained with colloidal Coomassie staining [106], mass-compatible silver staining [107], or with Sypro Ruby protein gel stain (Molecular Probes, Eugene, OR). Images were digitalized with an Image Scanner (GE Healthcare, Little Chalfont, UK). Image analysis was carried out with ImageMaster Platinum 6.0 (GE Healthcare), and with PDQuest Advanced 8.0.1 (Bio-Rad). For ImageMaster Platinum, values of smooth, min. area, and saliency were of 7, 10, and 250, respectively, while for PDQuest Advanced, values of sensitivity, size scale, and min. peak were of 4, 11, and 1500, respectively.

For 2D-DIGE analysis, protein extracts obtained from fresh-frozen and fixed tissues were precipitated with TCA-Acetone as described above, and resuspended at a concentration of 1-5 mg/ml in a buffer composed by urea (7M), Thiourea (2M), and CHAPS (4%).

Then, samples were labelled with CyDye DIGE Fluors (GE Healthcare), according to the minimal labelling protocol provided by the manufacturer. Briefly, after CyDye reconstitution with dimethylformamide (DMF) and preparation of a working solution (200 pmol/μl), 1 μl of diluted CyDye was added to a volume of protein sample equivalent to 50 μg. Samples were left on ice for 30 minutes in the dark, and then 1 μl of 10 mM lysine was added to stop the reaction. After 10 minutes on ice in the dark, labelled samples were directly mixed and analysed by DIGE or stored at -80°C until needed.

After mixing the labelled protein extracts to be loaded on the same gel, IPG buffer (corresponding to the desired pH range, at a 1% final concentration) and DeStreak Rehydration Solution (GE Healthcare) were added to a final volume of 450 μl. 24 cm IPG strips (GE Healthcare) were passively rehydrated for at least 6 hours. IEF was carried out on a Ettan IPGphor II (Amersham) for a total of 90-100.000 Vh. After focusing, strips were equilibrated as stated above. Proteins were then subjected to SDS-PAGE in 10-18% gradient polyacrylamide gels on the Ettan DALTsix system (GE Healthcare), following the manufacturer instructions. Gel image acquisition was performed on a Typhoon 9400 laser scanner, and image analysis was carried out using DeCyder 2D software (version 7.0, GE Healthcare).

3. In-gel trypsin digestion

For 1D-PAGE bands, visible protein bands were excised, or all lane was fractionated into gel slices, independently from band position and intensity. Bands were destained by repetitive washings with 50 mM NH_4HCO_3, pH 8.0, and acetonitrile. Samples were reduced and carbamidomethylated in 50 mM NH_4HCO_3 buffer, pH 8.0, first with 10 mM DTT at 56°C, and then with 55 mM iodoacetamide at RT in the dark. Tryptic digestion of the alkylated samples was performed at 37°C overnight using 60 to 100 ng of trypsin per band or gel slice, according or not to band intensity, respectively.

For 2D-PAGE spots, protein spots were excised from the gel and destained by repetitive washings with 50 mM NH_4HCO_3, pH 8.0, and acetonitrile. Tryptic digestion was performed at 37°C overnight using 50 to 100 ng of trypsin, according to spot intensity.

4. Mass spectrometry techniques: LC-MS/MS and MALDI-MS

LC-MS/MS analyses of tryptic digests were performed on a Q-TOF hybrid mass spectrometer equipped with a nano lock Z-spray source, and coupled on-line with a capillary chromatography system CapLC (Waters, Manchester, UK). After loading, the peptide mixture (6 µL) was first concentrated and washed at 20 µL/min onto a reverse-phase pre-column (Symmetry 300, C_{18}, 5 µm, NanoEase, Waters) using 0.2% formic acid as eluent. The sample was then fractionated onto a C_{18} reverse-phase capillary column (Symmetry, 75 µm x 15 mm, Waters) at a flow rate of 250 nL/min, using a linear gradient of eluent B (0.2% formic acid in 95% acetonitrile) in A (0.2% formic acid in 5% acetonitrile) from 7 to 50% in 40 min. Mass spectrometer was set up in a data-dependent MS/MS mode where a full scan spectrum (*m*/*z* acquisition range from 400 to 1600 Da/e) was followed by a tandem mass spectrum (*m*/*z* acquisition range from 100 to 2000 Da/e). Peptide ions were selected as the three most intense peaks of the previous scan. A suitable collision energy was applied depending on the mass and charge of the precursor ion. Argon was used as the collision gas.

ProteinLynx software, provided by the manufacturers, was used to analyze raw MS and MS/MS spectra and to generate a peak list which was introduced in the in-house Mascot MS/MS ion search software (Version 2.2, Matrix Science, Boston, MA) for protein identification. Search parameters were as follows: peptide tolerance 50 ppm, MS/MS tolerance 0,4 Da, charge state +2 and +3, enzyme trypsin, allowing up to 1 or 2 missed cleavages.

In regard to MALDI-MS analyses, mass spectra were recorded on a MALDI micro (Waters, Manchester, UK) equipped with a reflectron analyser and used in delayed extraction mode. Peptide samples were mixed with an equal volume of α-cyano-4-hydroxycynnamic acid as matrix (10 mg/mL in acetonitrile/0.2% TFA) (70:30, v/v), applied to the metallic sample plate, and air dried. Mass calibration was performed by using the standard mixture provided by manufacturer. Raw data, reported as monoisotopic masses, were then introduced into in house MASCOT Peptide Mass Fingerprint search software (Matrix Science, Boston, MA), and used for protein identification.

5. Data analysis

Total peptide hits were used as a parameter for estimating and comparing protein abundance. Total peptide hits value for each protein was calculated by summing "queries matched" number (as indicated by Mascot) of all gel bands corresponding to a single sample. Unique peptides and sequence coverage values reported in all tables have to be intended as the best value obtained in a single LC-MS/MS run (always among all bands analysed for a sample). Only proteins which reported at least one peptide ranked by Mascot with a value equal to 1 were included. Queries matched with more than a protein hit (among protein hits obtained from a single LC-MS/MS run) were counted for each of the proteins. Skin keratins were excluded from the final protein list.

Gene Ontology (GO) assignments were carried out using GoMiner [108] and DAVID [109] softwares, and graphical representations (histograms, pie charts, plots) were generated with Microsoft Excel.

Assessment of aminoacidic residues percentage were carried out by means of ProtParam tool, ExPASy [110].

Statistical analysis of data illustrated in Chapter 6 was performed by exploiting a Wilcoxon's Rank-Sum test [111]. Statistical analysis of data illustrated in Chapter 8 was carried out using a Kruskal-Wallis H test, which is a nonparametric version of one-way analysis of variance [111]. Considering k samples of sizes N1, N2, .., Nk, with N = sumNi, ranking all the data, regardless to the k samples, and indicating with R1, R2, ..., Rk the sums of the ranks for the k samples, the following statistic is defined:

H = {12/N(N+1)}*sum {j=1 to k} (Rj^2/Nj) - 3(N+1).

6. Immunological techniques: Western blot, Protein arrays, ELISA

For western blot, electrophoresed proteins were transferred to nitrocellulose using the Bio-Rad MiniTransBlot apparatus according to standard procedures. After transfer, membranes were blocked in Phosphate Buffered Saline, 0,05% Tween 20 (PBS-T), and 5% skim milk or 3% BSA for 1 hour to overnight, and then processed for immunological detection. Antibody reactivity was detected using HRP-conjugated secondary goat anti-

rabbit IgG, and visualized with a chemiluminescent peroxidase substrate (Sigma-Aldrich). Blot images were processed with QuantityOne (Bio-Rad, Hercules, CA).

In order to perform protein arrays experiments, fresh-frozen and FFPE tissue extracts were spotted on nitrocellulose sheets at 1 µg of total protein extract per spot, and then in 2 serial dilutions (1:2, 1:4) in extraction buffer, followed by extraction buffer without protein as a negative control. Nitrocellulose sheets were allowed to dry, and then blocked and processed for immunological detection as explained for Western blot.

ELISAs were carried out as follows. Total proteins from fresh-frozen and FFPE tissue extracts were precipitated with trichloroacetic acid/acetone following standard procedures, quantified, and resuspended in PBS at a concentration of 100 µg/mL. One hundred µL were dispensed in each well of a Maxisorp ELISA plate (NUNC, Thermo Fisher Scientific, Rochester, NY), and incubated at room temperature until dry. Wells were then washed and blocked with PBS-T plus 5% BSA for 1 hour and then processed for immunological detection. Controls for reactivity after incubation with water only, PBS only, and secondary antibody only, were always run in parallel with each experiment. A chromogenic substrate was then added to all wells (Sigma Peroxidase ELISA Substrate) and absorbance was read in wells at 650 nm every 15 min for 1 h. Assays were repeated in triplicate, and mean and SD were calculated and plotted with Microsoft Excel (Microsoft, Redmond, WA).

Antibodies and dilutions used were as follows: rabbit polyclonal antibodies (Sigma) against actin (1:50,000), GAPDH (1:50,000), tropomyosin (1:5,000), vinculin (1:5,000), and mouse monoclonal antibodies against E-cadherin (Zymed, 1:1000), and myosin (Millipore, 1:50,000).

4. GENERATION OF HIGH-QUALITY PROTEIN EXTRACTS FROM FORMALIN-FIXED, PARAFFIN-EMBEDDED TISSUES

I. Introduction

Formalin-fixed, paraffin-embedded (FFPE) tissue banks generated for histopathological analysis and stored in hospitals worldwide can be considered as a "hidden treasure" which could open the way to retrospective studies for biomarker discovery and validation. Unfortunately, the fixation procedure causes an "entrapment" of macromolecules within the tissue matrix, due to intra- and intermolecular crosslinking, which dramatically impairs protein extraction efficiency and antigen immunoreactivity [60]. Antigen retrieval (AR) is still considered as the only procedure able to induce a considerable improvement in antigenic reactivity of FFPE tissues, through a high-temperature treatment which probably induces cleavage of protein-protein bonds [64-66]. Current methodologies that aim to extract full-length proteins from FFPE tissues, useful for gel-based proteomics or immunological studies, are based on the antigen retrieval concept [71-79]. However, protein extracts produced by the methods described to date exhibit a low pattern complexity, together with a poor suitability for downstream gel-based proteomic techniques. Thus, the aim of this research was to develop an extraction method with a high degree of efficiency and versatility, which allows the application of a wide gamut of analytical approaches, such as GeLC-MS/MS, western blotting, protein arrays, and ELISA.

II. Results

1. High-quality proteins can be extracted with high yield from FFPE tissues

After slaughtering, muscle and liver sheep tissues were cut into blocks, and immediately fixed to avoid autolysis of proteins. Parallel samples were also frozen in order to obtain valid tissue and subject-matched controls for viable, unfixed proteins.

Starting from one month after fixation and embedding, microtome sections were cut from the FFPE blocks and used for optimisation of the extraction protocol. The best extraction performance was obtained upon deparaffinisation with xylene, rehydration with a graded series of alcohols, and serial incubation at 100°C and 80°C with a buffered detergent solution in a reducing environment. Extraction efficacy was assessed comparatively and in triplicate for FFPE and fresh-frozen muscle and liver tissues, by relating the amount of extracted proteins to the wet and dry weights of starting tissues. Results are reported in Table 4.1.

Table 4.1. Comparative yield of protein extraction from fresh-frozen and FFPE sheep tissues relative to initial wet and dry tissue weights.

Tissue		Protein concentration (µg/µl)	Extracted protein/ tissue wet weight (µg/mg)	Extracted protein/ tissue dry weight (µg/mg)
Skeletal muscle	*Fresh-frozen*	10.2	48.6	145.7
	FFPE	2.1	8.2	40.9
Liver	*Fresh-frozen*	10.8	50.9	152.6
	FFPE	7.0	28.8	143.8

This extraction procedure yielded an efficient and reproducible extraction of proteins, especially from liver; the extraction efficacy from FFPE samples reached levels comparable to the fresh-frozen samples, especially when dry weights were compared. Extraction efficacy was also reproducible but slightly lower for muscle. For comparison, protein extraction was also performed in triplicate and in parallel by using the method described by Shi et al [75]. Our method produced average protein yields of 16.3 µg and 86,8 µg per 80 mm^2 FFPE muscle and liver tissue slice, respectively, whilst extraction with the method described in literature produced average protein yields of 4,6 µg and 2,4 µg per 80 mm^2 FFPE muscle and liver tissue slice, respectively. However, with this method, recovery was difficult since the tissue pellet generated was less compact, and did not separate from the supernatant as well as the tissue pellet generated with our method. For comparison, the yields stated for the main commercial method suited for the extraction of full-length proteins from FFPE tissues, the Qproteome FFPE Tissue Kit

(Qiagen, Hilden, Germany), are of 15 µg per 10 µm thick, 80 mm^2 wide, liver tissue slice (comparison in Figure 4.1).

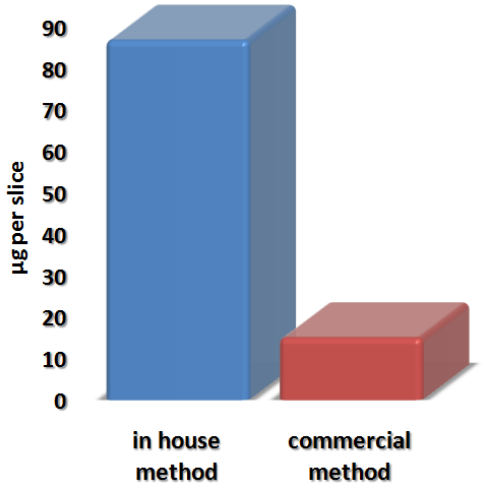

Figure 4.1. Quantitative comparison between values of extraction efficiency (expressed in terms of micrograms of proteins extracted per tissue slice) achieved using the method developed by our laboratory and those stated by a commercial method.

In order to assess integrity, preservation of relative abundance, and quality of electrophoretic separation, proteins extracted from FFPE tissues were subjected to SDS-PAGE. Figure 4.2 shows the SDS-PAGE profiles obtained upon separation of proteins extracted from fresh-frozen and FFPE sheep skeletal muscle and liver tissues. Although with few differences, and presence of a light background, the protein patterns corresponding to FFPE tissue extracts appear comparable to those corresponding to fresh-frozen tissue extracts, both in terms of position and relative abundance of protein bands.

In order to assess consistency and reproducibility of the extraction method when applied to FFPE human tissue bank samples, FFPE tissue sections from fresh-frozen and FFPE human hyperplastic thyroid gland were subjected to the same extraction protocol, producing consistent protein extract concentrations and acceptable SDS-PAGE banding patterns, although with a lower resolution compared to fresh tissues.

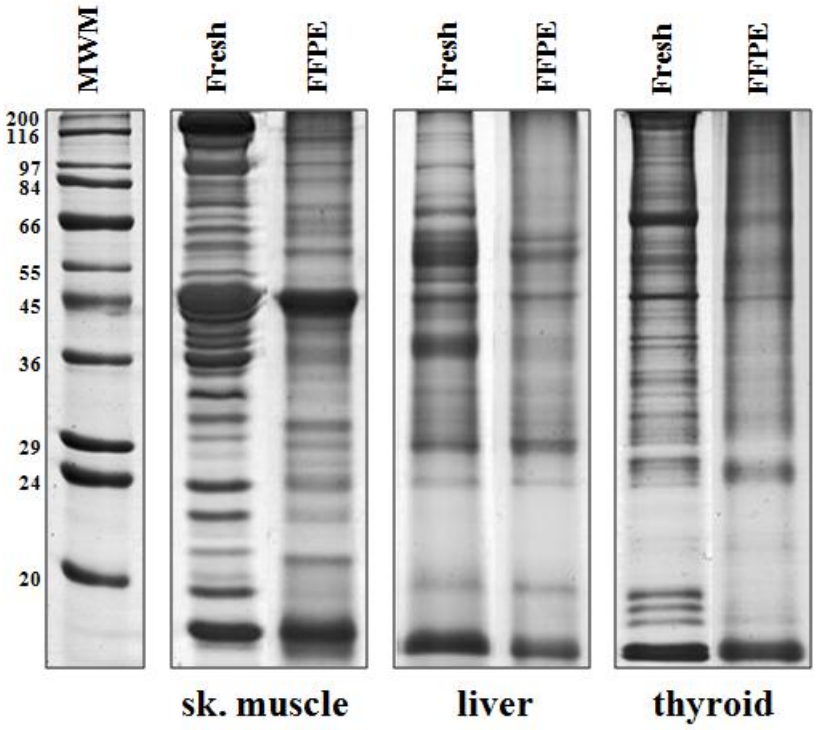

Figure 4.2. Representative SDS-PAGE patterns of fresh-frozen and FFPE tissue extracts. Sheep skeletal muscle, sheep liver, and human thyroid tissue extracts, obtained from fresh-frozen and FFPE tissues, are shown.

2. Proteins extracted from FFPE tissues can be reliably identified by 1D-GeLC-MS/MS

As stated above, the SDS-PAGE protein separation patterns of fresh-frozen and FFPE tissue extracts were of acceptable quality, and resolved protein bands were present. Therefore, the SDS-PAGE separation was used as a "pre-fractionation" step for an *in situ* hydrolysis approach, allowing MS/MS analysis of FFPE tissues with a nanoHPLC-nanoESI-QTOF tandem mass spectrometer. A comparative analysis of fresh-frozen and FFPE tissue extracts was then performed by staining the electrophoresed proteins with Coomassie, cutting corresponding areas from both lanes (particuarly, 21 matched bands), performing *in situ* digestion, and then subjecting tryptic peptides to MS/MS analysis.

Overall, 85 and 66 proteins were identified from fresh-frozen and FFPE tissue, respectively. Fifty-four (64%) of proteins identified in the fresh-frozen tissue were also identified in the FFPE tissue; on the other hand, 18% of proteins identified in the FFPE tissue were not identified in the fresh-frozen tissue (Figure 4.3, left). The identification of some proteins only in FFPE tissues was observed mostly for proteins with one or

two peptide hits. In fact, when considering the top 50 proteins based on the number of queries matched, 80% were identified from both tissue extracts (Figure 4.3, right). Total peptide hits from FFPE and fresh-frozen tissue sections were 1706 and 1824, respectively. It is important to note that these proteins and peptides represent only a part of the whole muscle proteome, since only visible gel bands were cut and analyzed; nevertheless, when compared with the recent data concerning MS analysis of FFPE extracts, which were carried out with more sophisticated instrumentation (such as the Orbitrap mass spectrometer), the number of peptide hits obtained from the FFPE extract is a satisfactory result.

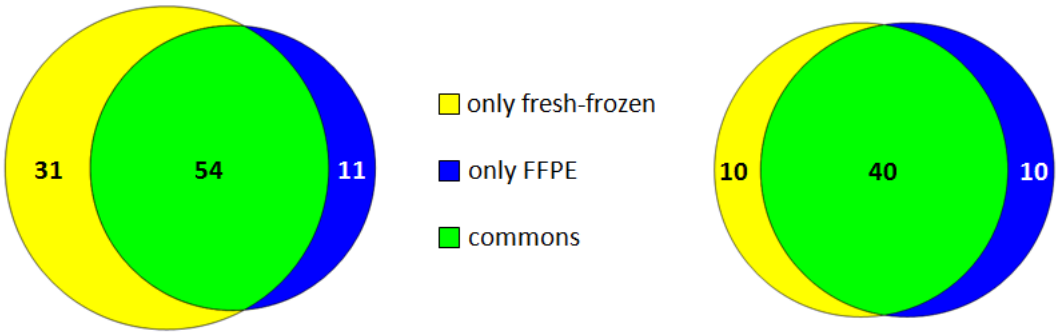

Figure 4.3. Venn diagrams illustrating distribution of proteins identified from fresh-frozen and FFPE tissues. Left: 64% of proteins identified from fresh-frozen tissues were also identified from FFPE tissues, and 85% of proteins identified from FFPE tissue are also identified from fresh-frozen tissues. Right: Among the top 50 proteins, 80% were identified from both fresh-frozen and FFPE tissues.

Initially, we expected the extraction of basic proteins to be a difficult challenge, since formalin fixation acts mainly on basic aminoacid residues and therefore basic proteins should bear more crosslinks compared to acidic proteins. Yet, even proteins with high isoelectric points could be easily identified from FFPE tissue extracts with high levels of confidence (for a complete list of GeLC-MS/MS data see Apeendices 1 and 2).

The mass spectrometry data from 10 selected gel-separated proteins, covering molecular weights from 100 to 15 kDa, and isoelectric points from 4 to 9, were also used to evaluate the extent of crosslinking reversal and the degree of chemical alteration introduced by fixation. Accordingly, the K-to-R ratio, the extent of methionine oxidation, and the number of missed cleavages, were evaluated. Results

Table 4.2. Evaluation of the extent of modification introduced by formalin fixation and crosslink reversal and of its impact on quality of the mass spectrometry results.

Protein			**Unique peptides**		**% common unique peptides**		**K:R ratio**		**% peptides with oxidized methionines**		**% peptides with internal missed cleavage(s)**	
Name	*MW*	*pI*	*FrFr*	*FFPE*	*FrFr*	*FFPE*	*FrFr*	*FFPE*	*FrFr*	*FFPE*	*FrFr*	*FFPE*
Alpha actinin-2	104	5.31	27	12	44.4	100	10:17	6:6	3.7	0	3.7	0
Heat shock 70 kDa protein	70	5.54	10	7	40	71.4	4:6	3:4	0	0	10	14.2
ATP synthase subunit alpha, mitochondrial	60	9.16	8	9	75	66.6	3:5	3:6	0	0	12.5	0
Beta-enolase	47	7.60	11	13	81.8	69.2	9:2	10:3	0	15.3	27.3	7.7
Actin, alpha skeletal muscle	42	5.23	15	14	93.3	100	7:8	6:8	26.6	28.6	26.6	21.4
Fructose-bisphosphate aldolase A	40	8.31	14	10	71.4	100	7:7	5:5	0	0	35.7	20
Triosephosphate isomerase	27	6.45	9	8	88.9	100	8:1	7:1	11.1	0	11.1	12.5
Myosin light chain 1, skeletal muscle isoform	21	4.97	6	6	83.3	83.3	2:4	3:3	0	0	16.7	16.7
Alpha-crystallin B chain	20	6.76	6	8	100	75	5:1	7:1	0	0	16.7	42.8
Troponin C, skeletal muscle	18	4.06	8	3	37.5	100	5:3	2:1	25	66.7	0	0
Average					72	87	60:54 (1.11)	52:38 (1.36)	7	11	16	13

are summarised in Table 4.2. Overall, only minor differences were observed for these parameters among fresh-frozen and FFPE tissue proteins.

Overall, quality of MS identifications was comparable for both extracts types. Figure 4.4 presents a comparison of protein distribution according to sequence coverage, showing that identification data of FFPE tissue proteins have a confidence level comparable to proteins extracted from fresh-frozen tissues.

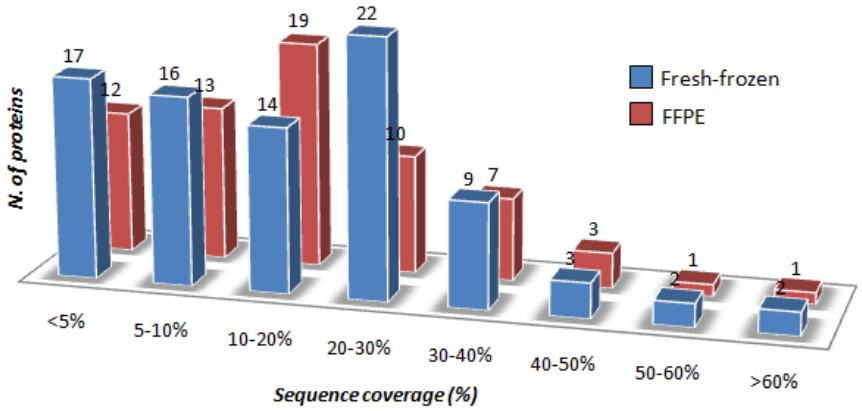

Figure 4.4. Histogram representing distribution of proteins identified from fresh-frozen tissues and FFPE tissues according to their sequence coverage (expressed in percentage).

Peptides produced after trypsin digestion of proteins extracted from FFPE tissues generated nanoLC-nanoESI-Q-TOF fragmentation spectra comparable to those generated from fresh-frozen tissue extracts (Figure 4.5). Formalin fixation seemed to have no detrimental effects on the quality of the MS/MS spectra obtained from FFPE and fresh-frozen tissues. These results suggest that formalin fixation either does not involve all proteins, and therefore the ones which are not damaged produce a number of peptides sufficient for identification, or that crosslink reversal does not impair the identification of peptides after extraction with the conditions described in this paper.

For comparison of extraction quality and reproducibility, all protein identifications obtained from fresh-frozen and FFPE tissue extracts were analysed for molecular function and cellular localisation. Figure 4.6 reports the GO classification obtained for both extracts. Although few differences can be observed, the distribution of classes and their relative abundances are mostly preserved; therefore FFPE fixation performed

before extraction of proteins from tissues does not seem to introduce a significant bias over particular protein functions or localisations.

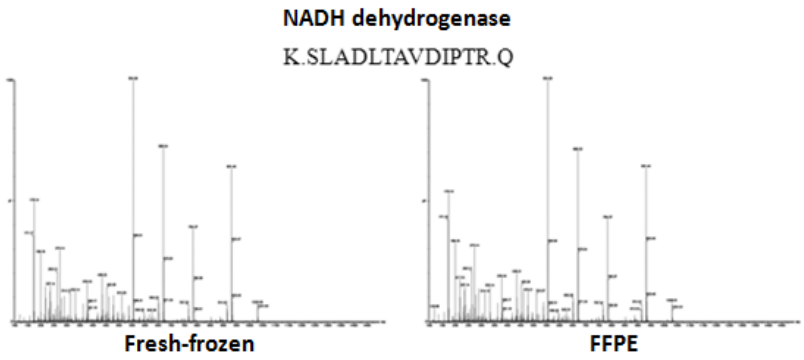

Figure 4.5. Comparison of representative MS/MS spectra of a peptide identified in extracts obtained from FFPE and fresh-frozen sheep skeletal muscle tissue.

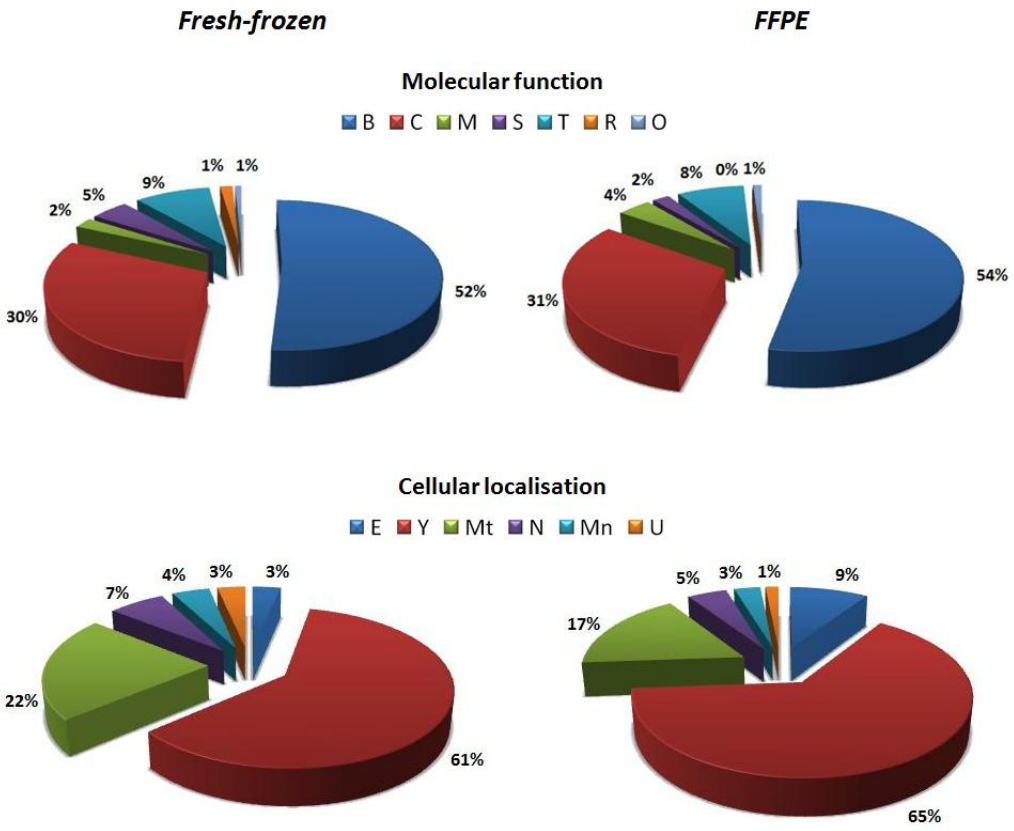

Figure 4.6. Comparison of gene ontology (GO) ratios for proteins identified from fresh-frozen (left) and FFPE (right) protein extracts. The upper diagrams indicate functional classification as follows: B, binding; C, catalytic; M, motor; S, structural; T, transport, R: transcription/regulation, O: other. The lower diagrams indicate subcellular localisation as follows: E, extracellular; Y, cytoplasmic; Mt, mitochondrial; N, nuclear; Mn, membrane; U, other/undefined.

3. Standard immunoassays can be applied to proteins extracted from FFPE tissues

The protein extraction method evaluated in this study stems from the antigen retrieval procedure used in IHC. Other authors applied western blotting protocols to FFPE protein extracts obtained with the antigen retrieval strategy, although cross-reactivity problems and a limited molecular range for protein identification were often reported. In view of these results, and to assess the reactivity of extraction products, we initially probed all tissue extracts with anti-actin and anti-GAPDH antibodies. As shown in Figure 4.7, clear, defined, and marked bands corresponding to actin and GAPDH were observed at the expected positions in all protein extracts, confirming that proteins retained their molecular weight and that immunoreactivity was preserved.

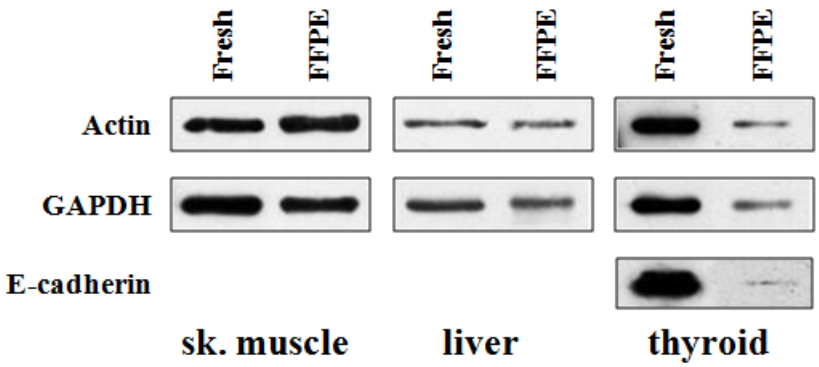

Figure 4.7. Reactivity of actin, GAPDH, and E-cadherin with the tissue extracts reported in the corresponding positions of Figure .

Reproducibility and consistency of western immunoblotting results was assessed in triplicate ad evaluated by densitometric quantitation. Results are reported in Table 4.3, in terms of band volume for actin and GAPDH in all tissues. Mean, SD, and CV are also reported. A high reproducibility was observed for the animal model tissues, whilst the human tissue showed a higher degree of variability among replicates. However, this could be explained by the higher tissue heterogeneity of thyroid compared to liver or muscle.

Many of the heat-induced antigen-retrieval techniques for intact proteins have the disadvantage of reduced or incomplete membrane protein extraction. In order to evaluate the extraction performance for this difficult subset of proteins, the human thyroid tissue extracts were also probed with antibodies to E-cadherin. As shown in

Figure 4.7, E-cadherin could be detected by western immunoblotting in an FFPE extract generated from an archival human biopsy, although with a significantly lower intensity compared to fresh-frozen human thyroid tissue extract.

Table 4.3. Evaluation of western blotting reproducibility measured by densitometric quantitative comparison with QuantityOne (BioRad). Values correspond to band volume, defined as sum of the intensities of the pixels inside the volume boundary x area of a single pixel (in mm^2). Average (AV), standard deviation (SD), and coefficient of variation (CV), are reported for each tissue.

Antigen	*Tissue*	*Exp. 1*	*Exp. 2*	*Exp. 3*	*AV*	*SD*	*CV*
Actin	*Skeletal muscle*	54.6	42.4	41.9	46.3	7.2	0.2
	Liver	8.2	8.1	10.3	8.9	1.2	0.1
	Thyroid	10.4	7.0	3.8	7.0	3.3	0.5
GAPDH	*Skeletal muscle*	11.1	13.7	19.1	14.7	4.1	0.3
	Liver	23.2	23.2	26.8	24.4	2.1	0.1
	Thyroid	2.7	5.4	8.0	5.3	2.7	0.5

In view of these results, FFPE skeletal muscle tissue extracts were then used for comparative evaluation of reactivity by western immunoblotting, ELISA, and protein arrays with antibodies against GAPDH, tropomyosin, vinculin, and myosin.

As regards to western blotting results (as shown in Fig. 4.8), GAPDH reactivity was clear and consistent with the protein molecular weight (about 36 kDa). Tropomyosin reactivity was also consistently observed at the expected position (about 33 kDa), although with a lesser defined reactivity compared to fresh tissues, and with the presence of a cross-reactive smear at higher molecular weight (as described in [79]). Vinculin, a 120 kDa protein, showed an unexpected and interesting behaviour. The western immunoblot reactivity of the anti-vinculin antibodies in sheep skeletal muscle extracts was always comparable for FFPE and fresh tissues; however, instead of a 120 kDa band, two bands of about 45 and 55 kDa, often accompanied by a lighter band of about 28 kDa, were consistently observed in all experiment replicates performed during this study. This could be explained by presence of splice variants or truncated forms in this individual, by inability of this particular antibody to react with the full-length sheep protein, or by other reasons. In any case, however, this only adds to the demonstration that FFPE tissue extracts generate results comparable to fresh tissue extracts, even when unexpected results or unusual behaviours are observed. Myosin, a high molecular weight protein, could not be detected in FFPE tissue extracts with the

antibody used in this study. Poor immunological detection of high molecular weight proteins can either be related to an alteration of antigenicity in FFPE tissue extracts, blotting efficiency at high molecular weights, or simply to the limited SDS-PAGE resolution of highly crosslinked protein structures.

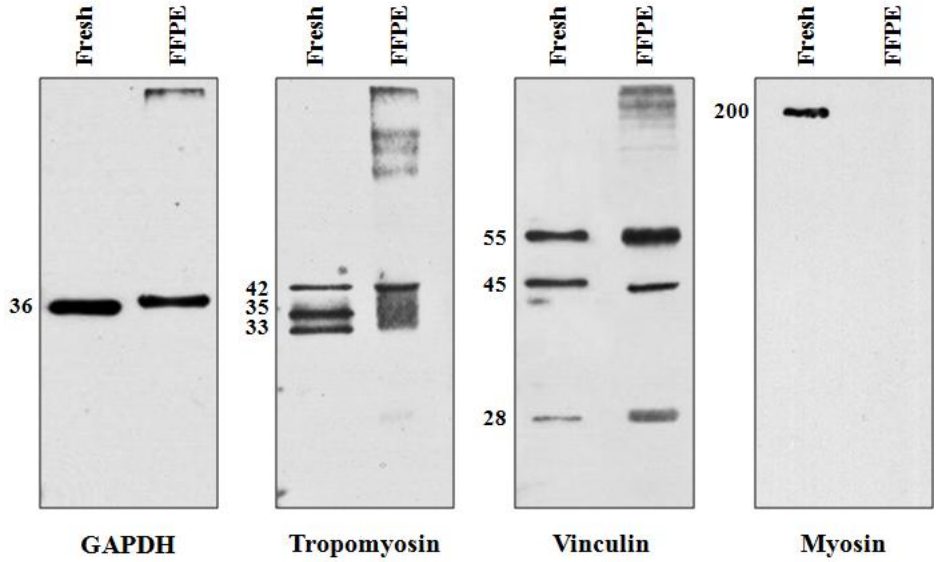

Figure 4.8. Summary of western blot reactivities obtained with proteins from fresh-frozen and FFPE tissues. Molecular weights of the reactive protein bands are shown.

Recently, reverse phase protein arrays have proven to be a very efficient and versatile technology for studying protein abundance and function, and have been applied to screening and quantitative profiling of disease-associated proteins. Applicability of FFPE protein extracts to protein arrays was introduced by Chu [73] and it was recently confirmed by Becker and co-workers [80]. Therefore, we evaluated reactivity of the FFPE extracts obtained in this study with the array technique (Figure 4.9).

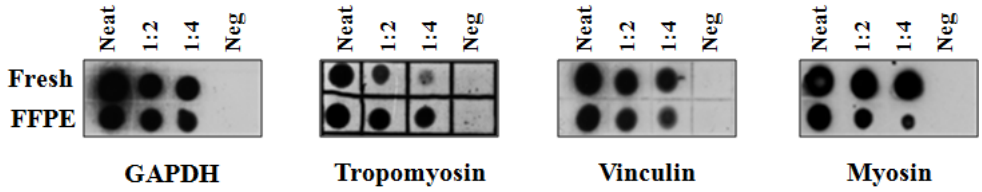

Figure 4.9. Summary of protein array reactivities obtained with serial dilutions of proteins from fresh-frozen and FFPE tissues.

All antibodies tested produced a strong, reproducible, and quantitative signal in FFPE extracts. Even myosin, although problematic by western blotting, was clearly detected by the protein array method.

ELISA is a widespread immunoassay technique presenting several important advantages, such as amenability to automation, diffusion of compatible instrumentation and assay platforms, and availability of antibodies optimised for this methodology. Therefore, we evaluated its applicability to FFPE protein extracts. In this case, proteins were precipitated and resuspended in an adequate buffer, since the detergent content of the extract would have interfered with the coating process. Encouraging results were obtained with almost all antibodies tested, as shown in Figure 4.10. These results show that the application of ELISA might be successfully pursued.

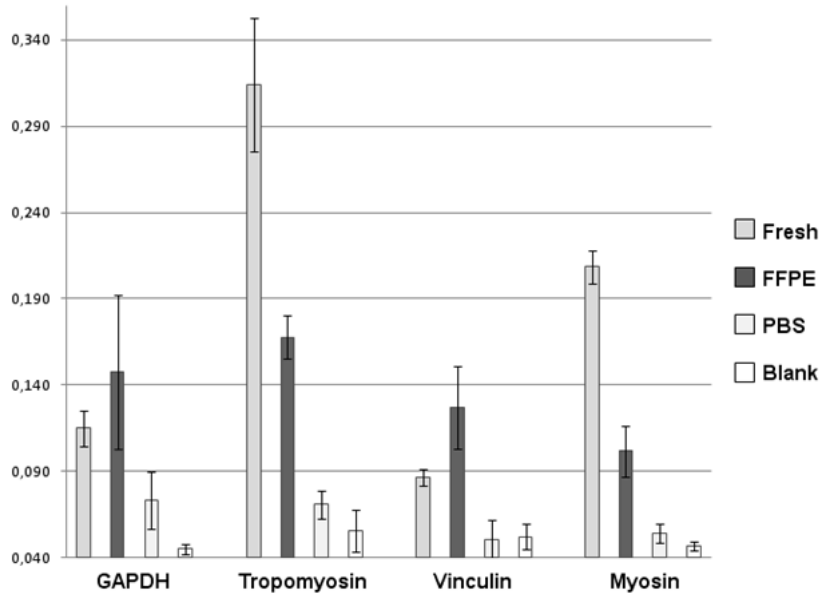

Figure 4.10. Summary of ELISA reactivities for proteins from fresh-frozen and FFPE tissues. The mean and SD values of OD650 for three independent experimental replicates are reported.

III. Discussion and concluding remarks

The primary purpose of this work was the development and optimisation of an efficient, reproducible, and versatile procedure for extracting proteins from FFPE

tissues. Therefore, in order to eliminate experimental variability due to storage duration of FFPE tissue blocks, differences in fixing times, autolysis due to delayed fixing, individual differences in samples, anatomical or regional heterogeneity, and other possible sources of variability, an animal model was established. Sheep was chosen as a tissue source, since the large animal size allowed the collection of significant amounts of tissue with a high uniformity of organ cuts, and it was possible to access tissues from a healthy animal monitored by a veterinarian during slaughtering for meat production. Skeletal muscle and liver of the same individual were used for all the studies concerning evaluation of extraction ability, reproducibility, and flexibility, because of their different densities, cell compositions, and protein abundance distributions. Subsequently, skeletal muscle tissue extracts were used for application of GeLC-MS/MS, protein arrays, and ELISA, due to the lower proteome complexity and variability in physiological conditions, as compared to other tissues.

After evaluation of several parameters, such as reagents for deparaffinisation, composition and pH of extraction buffer, temperature values of heat incubation and their sequence and duration, an optimised procedure was developed. The combination of an *antigen retrieval*-based high-temperature treatment and a Laemmli-based extraction buffer allowed the achievement of a noteworthy extraction efficiency, particularly from liver, significantly higher than that attained by commercial methods. The full-length protein extracts generated by this novel protocol exhibit a high pattern reproducibility, demonstrated by SDS-PAGE analysis, besides a remarkable versatility, since the same extract can be subjected to a wide gamut of proteomic and immunological analytical approaches.

Previous investigators showed SDS-PAGE images of proteins extracted from FFPE material; however, protein patterns were globally of low quality and definition, and mostly related to medium to low molecular weight proteins. [75, 80] As showed in Figure 4.2, 1-D profiles present a good definition, and a substantial preservation of band intensities relative ratio, although combined with an evident background, which seems to decrease proportionally to molecular weight. Even if this fact has modest

analytical consequences, it indirectly represents an index of protein extract overall quality.

1D-GeLC-MS/MS approach represents an interesting hybrid between gel-based and shotgun proteomics. It allows performing robust and in-depth proteomic profiling, thanks to a combination of two common proteomic techniques, such as SDS-PAGE and LC-MS/MS, and without the need of sophisticated instrumentation, for instance Orbitrap mass spectrometer. In the case of this study, only visible gel bands were cut and analysed, therefore only a quite partial "snapshot" of the skeletal muscle proteome was done. Nonetheless, data are undoubtedly promising, above all for the effective comparability between fresh and FFPE results. Proteins covering a wide range of pIs, molecular weights, functions, relative abundances were identified and characterised also in FFPE extracts. Total number of peptide hits, sequence coverage, MS/MS spectra quality, percentage of chemically modified residues, and K/R ratio of fresh-frozen and FFPE proteins were compared, and no significant differences were found. In our experience, protein identification is achievable also analysing gel regions which did not show a visible Coomassie stain (data not shown); thus, subjecting to a GeLC-MS/MS investigation all the gel lane should considerably improve numerousness and completeness of proteomic data. Although it may appear lengthy and labour-intensive, the ability to GeLC-MS/MS might open new and interesting perspectives for biomarker discovery and proteomic profiling in a wealth of diseases. Particularly, still notable differences between fresh-frozen and FFPE tissue profiles may be overcome performing a comparative analysis between two or more FFPE extract, which share the same drawbacks and bias.

In the FFPE extract, some of the most abundant muscle proteins (such as actin, tropomyosin, myoglobin, GAPDH, creatine kinase, and so on) were detected along the whole gel area, beyond their expected molecular weight. In fact, peptides with a sequence corresponding to these proteins were found in tryptic mixtures together with peptides corresponding to proteins at the "correct" MW. It appears that small peptides belonging to the most represented proteins might form a "background" of fragments which remain linked to other proteins, probably as a result of an incomplete reversal of crosslinks, and are released only after trypsin hydrolysis [88]. However,

these represent a sort of “background noise” which can be processed and subtracted, and does not interfere with correct identification of co-migrating proteins. This background can be seen in the western blotting patterns, in the “trail” of reactivity at higher molecular weights (Figure 4.8).

In order to assess the actual restoration of immunoreactivity of fixed proteins, fresh-frozen and FFPE extracts were subjected to the best-known immunoassays, i.e. western blotting, protein arrays and ELISA. These kinds of techniques represent the preferential choice, together with immunohistochemistry, for assessing and validating differential expression of proteins. Mostly protein arrays and ELISA, because of their elevated reproducibility, high-throughput feature and suitability to automation, can be considered as the optimum approaches for validation of protein putative biomarkers. Antigens to be probed, namely myosin, vinculin, GAPDH, and tropomyosin, were selected with the intention to enclose proteins with different molecular weights, isoelectric points and levels of expression. Apart from vinculin, which falls in a gel area that was not cut, the other proteins were all previously identified by GeLC-MS/MS. Both MS and blotting data underline that is difficult to obtain satisfactory results with very high-molecular proteins, such as myosin heavy chain, whilst GAPDH and tropomyosin show similar blot signal and total peptide hits number between fresh and FFPE. This phenomenon can be explained by the fact that western blotting and GeLC-MS/MS share the first electrophoretic step, and bulky proteins, maybe extensively crosslinked, can difficultly enter the gel matrix.

Among all immunological techniques evaluated, protein arrays bear several advantages. Compared to ELISA, protein extracts obtained with this method can be used directly without further processing. The extraction buffer described here does not interfere with protein binding, so making precipitation unnecessary. Extracted proteins can be spotted directly on membranes, and therefore there is no selective loss, thus allowing maintenance of relative ratios. This is especially remarkable when performing quantitative studies. However, crossreactivity issues need to be taken into account and carefully evaluated, since they might interfere with quantitation. In this respect, more reliable results can be obtained by western blotting, but high molecular weight proteins can be problematic due to transfer and migration issues, as already

discussed. Moreover, western blotting it is a low-throughput technique requiring a consistent amount of work and user experience compared to protein arrays.

The peculiarities seen in antibody reactivity of FFPE protein extracts have been addressed by Sompuram and co-workers [112]. These authors demonstrated that antibodies reacting with proteins subjected to antigen-retrieval techniques recognize linear epitopes; therefore, only these antibodies should be expected to display reactivity with FFPE tissue extracts. The different protein folding occurring in ELISA, caused by removal of SDS and resuspension in buffers such as PBS, might “hide” or refold the linear epitopes recognised by these antibodies. The current Human Protein Atlas initiative, in this respect, may represent an invaluable resource for identification of the most suitable antibodies to be used for investigation of FFPE protein extracts with different immunoassays [113].

As reported by a few researchers, various factors influence extraction efficacy and protein extract properties, such as duration of formalin fixation and bioptic sample “aging” (67,96). This study has been conducted deliberately reducing as more as possible all source of variability, as asserted above. Actually, when collecting samples for a proteomic study, i.e. for biomarker discovery aims, it is not ever possible to keep a significant degree of homogeneity as far as these features are concerned. A thyroid sample was subjected to the same procedure carried out on sheep tissues, in order to demonstrate the applicability of the method also to a “true” human bioptic specimen. Satisfactory SDS-PAGE and western blotting results were obtained, even though with a significantly lower reproducibility and 1D-PAGE sharpness. Anyhow, it is realistic to foresee that the quality of the protein extracts and the efficacy of the extraction procedure might vary in a considerable measure in relation to the peculiarities of selected samples. Further studies, conducted getting out proteomic information from routinely processed biopsies (also from “aged” ones), may support and definitively ascertain the actual effectiveness and widespread applicability of this method.

5. 2D-PAGE AND MS ANALYSIS OF PROTEINS FROM FORMALIN-FIXED, PARAFFIN-EMBEDDED TISSUES

I. Introduction

FFPE tissue banks represent a potentially invaluable resource for retrospective studies aimed to biomarker discovery and validation, both in cancer and in other important diseases. Application of standard proteomic profiling techniques to FFPE protein extracts would unveil a wealth of statistically supported information concerning functions involved in many aspects of disease, enabling identification of meaningful biomarkers related to specific disease diagnosis, prognosis, monitoring, and treatment.

A method enabling extraction of full-length proteins from FFPE tissues based on the antigen retrieval strategy was optimized in our laboratory, as described in Chapter 4. The protein extracts generated with this method produced a well-resolved protein pattern by SDS-PAGE and were successfully subjected to MS/MS characterization following in gel trypsin digestion.

To date, only few studies reported the tentative application of 2D-PAGE to protein extracts from FFPE tissues [84-85]. Noteworthy, the aim of one of these two reports was to demonstrate the purported unsuitability of FFPE tissues to 2D proteomic studies. Therefore, application of 2D-PAGE separation followed by in-situ protein digestion and MS or MS/MS analysis, the fundamentals of gel-based proteomics, still appears to be challenging on FFPE tissues.

In view of these facts, 2D-PAGE separation of FFPE tissue proteins extracted with the previously described method was pursued. Specifically, 2D-PAGE separation, 2D-western immunoblotting, MALDI-MS, and nano-HPLC-nano-ESI-Q-TOF mass spectrometry of in-situ digested spots were successfully applied to FFPE skeletal muscle tissue protein extracts.

II. Results

1. Proteins can be extracted from FFPE tissues, and separated by 2D-PAGE

Formalin-fixed, paraffin embedded tissue samples are largely available as "tissue banks" in hospitals worldwide. If accessible to proteomic investigations, these samples would represent an invaluable source of information on disease biomarkers. In order to investigate the possibility of performing 2D-PAGE and MALDI/MS analysis on FFPE extracts, fresh-frozen and FFPE sheep skeletal muscle tissue samples were used. For optimisation of the various protocol steps, an animal model was established. Sheep was chosen as a tissue source, since the large animal size allowed to collect a significant amount of tissue with a high uniformity of organ cuts, and facilitated optimisation of the extraction and separation steps starting from plenty of reproducible samples. Skeletal muscle was chosen for its availability in large amounts, its low proteome variability in physiological conditions, and its relatively lower proteome complexity compared to other organs.

Replicates of FFPE muscle tissue samples were treated for full-length protein extraction in presence of detergents and reducing agents (see Chapter 4). In order to eliminate interferences by these compounds during the focusing steps, protein extracts were subjected to precipitation with TCA/acetone, and resuspended in a suitable buffer. Figure 5.1 shows the 18 cm 2D-PAGE map of proteins extracted from fresh-frozen (top) skeletal muscle tissue using mechanical lysis, and FFPE (bottom) skeletal muscle tissue using the high-temperature extraction protocol. Up to 250 different spots could clearly be detected after 2D-PAGE separation and mass-compatible silver staining of 300 µg of proteins extracted from FFPE skeletal muscle tissue, corresponding approximately to 18 FFPE microtome slices, 10 µm thick, 80 mm^2 wide (according to statements reported in Chapter 4). Most importantly, the molecular mass and the isoelectric point of proteins extracted from fresh-frozen tissue sections (Fig. 5.1, left), appeared comparable to those subjected to the fixation and extraction procedure (Fig. 5.1, right). For comparison, up to 397 spots were detected in matched fresh-frozen tissue 2D-PAGE maps. Although fewer spots were present in the FFPE tissue map compared to the fresh-frozen tissue map, the FFPE muscle tissue protein

pattern could be matched to the fresh-frozen protein pattern. To our knowledge, this is the first report showing comparable 2D-PAGE protein profiles for formalin-fixed and fresh-frozen replicate sample tissues.

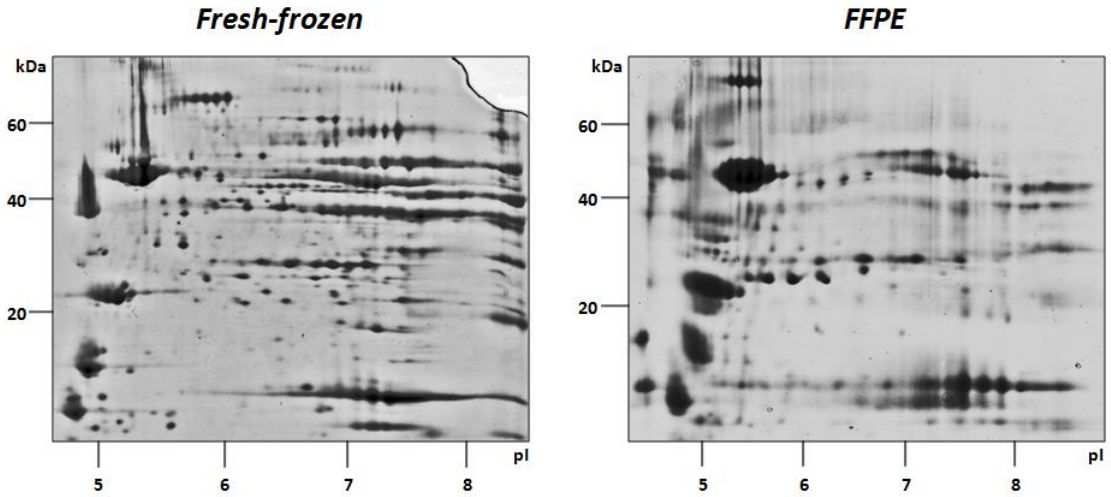

Figure 5.1. 2-DE analysis of proteins from fresh-frozen (left) and FFPE (right) sheep skeletal muscle tissue on 18 cm gels. Proteins were separated in 3-10NL pH ranges, and then in 12% polyacrylamide gels. Gels were stained with mass-compatible silver staining. Up to 250 spots were detected in the FFPE 2D-PAGE gels.

2. Proteins can be revealed by western immunoblotting at comparable molecular weights and isoelectric points

Proteins extracted from FFPE fixed tissues have been shown to recover their antigenicity, and can be detected with immunoassay techniques [71, 80]. Hence, fresh-frozen and FFPE skeletal muscle tissue proteins were subjected to 2D western immunoblotting, in order to assess whether specific proteins could be detected from both extracts, at comparable molecular weights and isoelectric points. Polyclonal antibodies were used to probe actin (*pI* 5.3) and GAPDH (*pI* 8.5) in fresh-frozen and FFPE muscle tissue extracts. As shown in Figure 5.2, anti-actin antibodies detected a protein spot at the expected MW and *pI* ranges, demonstrating that actin extracted from FFPE tissue retains its focusing characteristics. When the blot was probed with anti-GAPDH antibodies, a clear and defined positivity was detected at the expected MW and *pI* values. GAPDH produced a pattern formed by a “spot train” of presumptive isoforms, although a slight crossreactivity was present at higher MW (Figure 5.2,

bottom, right), probably due to residual crosslinked peptides bound to higher molecular weight proteins (as stated in Chapter 4). A comparable pattern of reactivity was observed for the fresh-frozen tissue extracts (Figure 5.2, bottom, left). Taken together, these data suggest that proteins extracted from FFPE tissues preserve substantial antigenic features, and that protein isoforms retain their focusing properties.

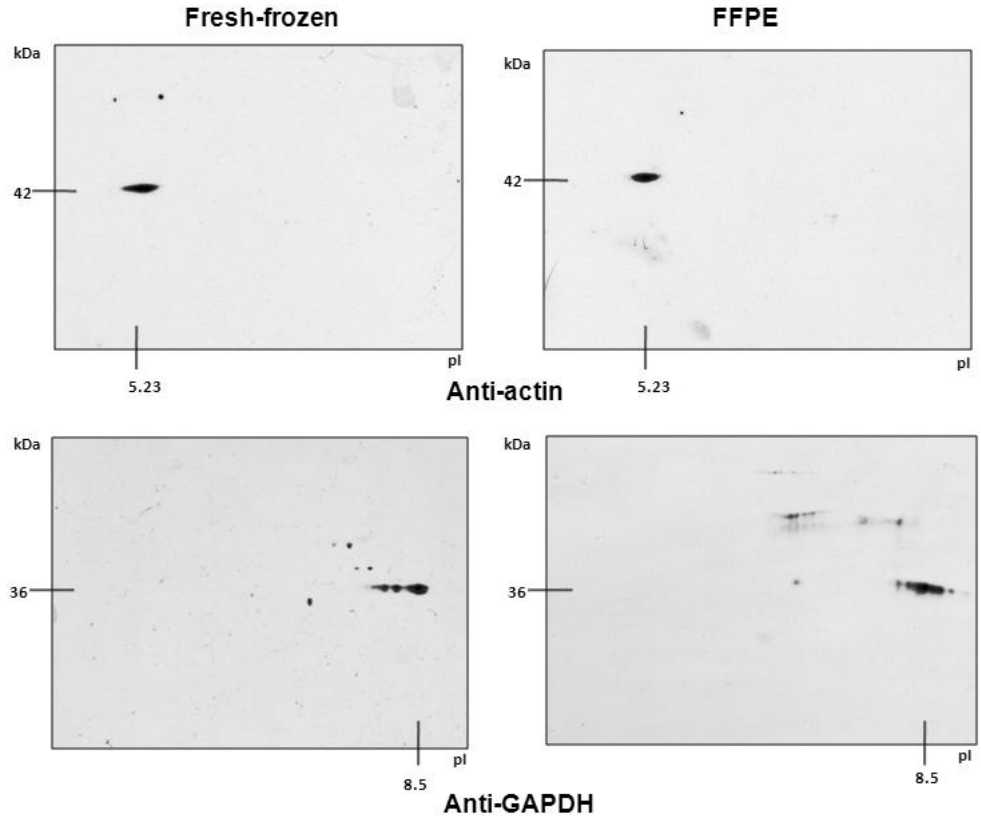

Figure 2. Western immunoblots of proteins from fresh-frozen (left) and FFPE (right) skeletal muscle tissue, separated in 7 cm 2D-gels, transferred to nitrocellulose, and developed with antibodies for actin (top) and GAPDH (bottom).

3. 2D-PAGE separated protein spots can be reliably identified by MALDI-MS and by MS/MS analysis

In order to evaluate whether FFPE protein extracts resolved by 2D-PAGE were suitable to mass spectrometry identification, matched protein spots generated from fresh-frozen and FFPE sheep skeletal muscle were subjected to comparative MS analysis. Replicate 7 cm 2D-PAGE gels were run with corresponding fresh-frozen and FFPE protein loads, namely, about 80 µg of proteins, obtained from 5 FFPE tissue slices, and

then stained with the colloidal Coomassie staining procedure (Figure 5.3). After staining, gels were processed with the ImageMaster Platinum 6.0 image analysis software for spot detection and matching.

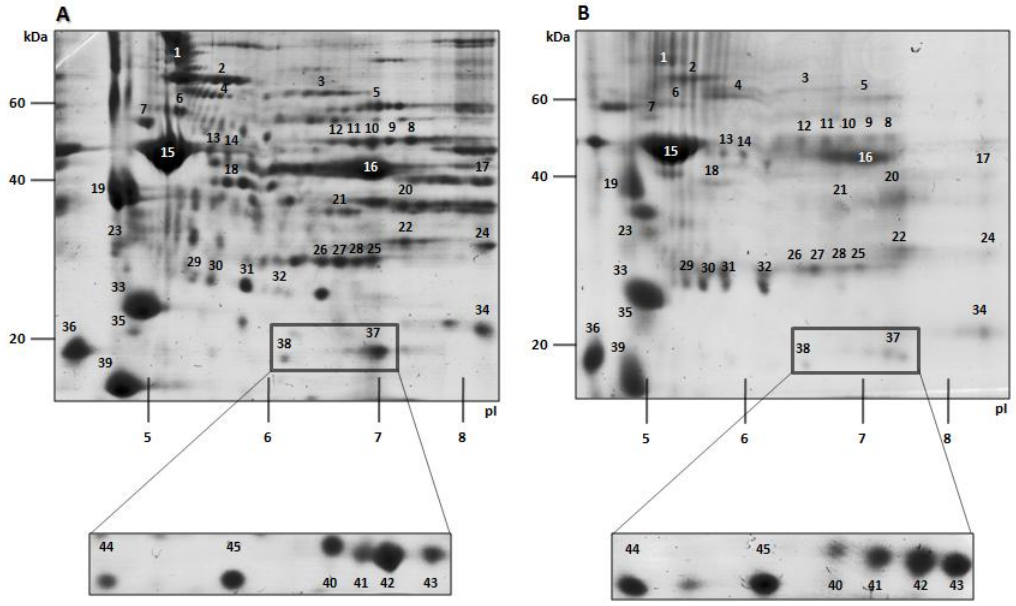

Figure 5.3. 2-DE analysis of proteins from fresh-frozen (A) and FFPE (B) sheep skeletal muscle tissue performed on 7 cm gels. Proteins were separated in 3-10NL pH ranges, and then in 12% polyacrylamide gels. Gels were stained with colloidal Coomassie. Numbers indicate picked and identified spots. Rectangles illustrate an enlargement of the selected zone obtained in replicate gels with higher protein loads. Spots 40-43 (corresponding to spot 37 in the larger map) were identified as alpha-crystallin isoforms; spots 44-45 (corresponding to spot 38 in the larger map) were identified as heat-shock protein beta-6 isoforms.

Among matched spots detected in fresh-frozen and FFPE gels, 39 were excised and subjected to MALDI/MS identification (Figure 5.3). Identities of fainter spots or uncertain identifications were confirmed by nanoLC-nanoESI-Q-TOF mass spectrometry.

As shown in Appendix 3, proteins were identified with high levels of confidence from 2D-PAGE spots corresponding to FFPE extracts, and estimated MW and *pI* on gel were always comparable to those corresponding to proteins from fresh-frozen tissue extracts. Moreover, many spot "trains", probably corresponding to protein isoforms, showed a very good match between fresh-frozen and FFPE gels. This is representatively illustrated in Figure 5.3 (detailed frames) where FFPE-extracted isoforms of the same proteins, identified in this case as alpha-cristallin (spot N.38,

isoforms are spots 40-43) and heat-shock protein 6 (spot N. 39, isoforms are spots 44-45), form distinct spot “trains” equivalent to those observed in gels from fresh-frozen tissues.

Quality of mass spectra generated from FFPE and fresh-frozen tissue extracts was comparable. Figure 5.4 reports four representative MALDI-MS spectra. Spectra at the top correspond to creatine kinase M-type, a highly abundant protein, and spectra at the bottom correspond to carbonic anhydrase 3, a protein with low-medium abundance in skeletal muscle. Proteins were identified in two matched spots from fresh-frozen and FFPE extracts. In both cases, quality of the protein spectra was extremely high and similar to that of fresh-frozen extracts.

Following in-situ hydrolysis and MS/MS analysis, peptides belonging to other proteins of lower molecular weight were detected together with proteins of the expected molecular weight (Table 5.1, shaded). The number of these artificially complexed peptides was significantly more represented in the acidic-high molecular weight region, whilst it was less pronounced in other areas of the 2D map. This phenomenon had already been observed for FFPE protein extracts after SDS-PAGE separation (see Chapter 4).

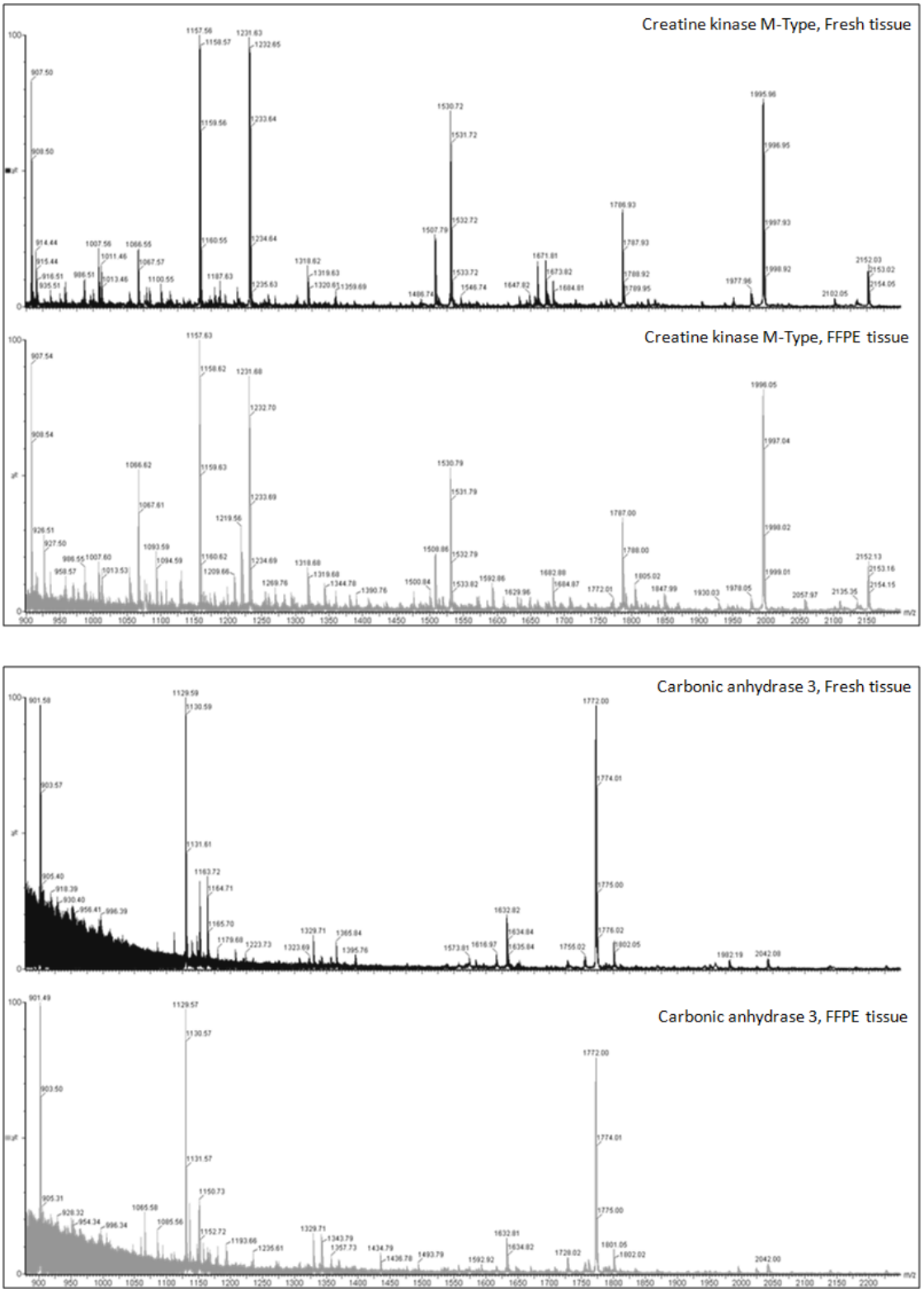

Figure 5.4. Representative MALDI-MS spectra of proteins from fresh-frozen and FFPE muscle tissue extracts. Top, creatine kinase M-Type, bottom, carbonic anhydrase 3. Dark grey, fresh-frozen; light grey, FFPE.

Table 5.1. Proteins identified in three representative high molecular weight spots. "Unexpected" proteins are shaded.

Spot number	Estimated MW on gel	Estimated p*I* on gel	*Protein*	Queries matched	Sequence coverage (%)	Mowse score	Theoretical MW	Theoretical *pI*
1	83781	5.42	Actin, alpha skeletal muscle	62	35	2132	42366	5.23
			Myosin light chain 1, skeletal muscle isoform	11	27	389	209930	4.97
			Creatine kinase M-type	6	14	279	43190	6.63
			Myoglobin	2	19	111	17044	6.87
			Tropomyosin-1 alpha chain	2	14	89	32746	4.69
			Serum albumin precursor	2	2	72	71139	5.80
			Triosephosphate isomerase	1	4	50	26901	6.45
			Hornerin	2	1	95	283140	10.05
			Fructose-bisphosphate aldolase A	1	2	48	39851	8.30
			Creatine kinase, sarcomeric mitochondrial precursor	1	2	42	47714	8.45
3	63677	6.54	Actin, alpha skeletal muscle	15	19	440	42366	5.23
			Phosphoglucomutase-1	11	10	417	61805	6.58
			Creatine kinase M-type	6	14	243	43190	6.63
			Fructose-bisphosphate aldolase A	7	14	219	39851	8.30
			Triosephosphate isomerase	4	19	196	26901	6.45
			Carbonic anhydrase 3	5	17	194	29637	7.71
			Pyruvate kinase isozyme M1	5	8	175	58522	7.24
			Myosin light chain 1, skeletal muscle isoform	5	16	149	20993	4.97
			Myoglobin	3	26	119	17044	6.87
			Heat-shock protein beta-1	2	4	86	22436	5.98

			Fructose-bisphosphate aldolase C	1	6	75	39830	6.41
			L-lactate dehydrogenase A chain	1	3	75	36916	8.12
			Beta-enolase	6	10	64	47409	7.60
			ATP synthase subunit alpha, mitochondrial precursor	2	3	58	59828	9.16
			Myosin light polypeptide 4	1	11	53	21665	4.97
			Glyceraldehyde-3-phosphate dehydrogenase	1	4	51	36073	8.50
5	58522	7.24	Pyruvate kinase isozyme M1	31	22	886	58522	7.24
			Creatine kinase M-type	7	17	346	43190	6.63
			Glyceraldehyde-3-phosphate dehydrogenase	5	13	294	36073	8.50
			Fructose-bisphosphate aldolase A	7	27	238	39851	8.30
			Carbonic anhydrase 3	4	14	203	29637	7.71
			Fructose-bisphosphate aldolase C	2	6	185	39830	6.41
			Triosephosphate isomerase	2	10	108	26901	6.45
			Myoglobin	2	9	94	17044	6.87
			Hemoglobin subunit alpha-1/2	3	16	86	15212	8.72
			Actin, alpha skeletal muscle	2	5	62	42366	5.23
			Creatine kinase, sarcomeric mitochondrial precursor	1	2	55	47714	8.45
			Beta-enolase	3	7	53	47409	7.60

III. Discussion and concluding remarks

Full-length proteins were extracted from fresh-frozen and FFPE muscle tissue, separated by 2D-PAGE, and comparatively identified by means of MALDI and nanoHPLC-nanoESI-Q-TOF mass spectrometry, producing matched 2D-PAGE maps and matched protein identifications. This goal was reached through adaptation of a recently developed method for generation of well-defined bands after SDS-PAGE separation of FFPE tissue protein extracts (described in Chapter 4). The application of a precipitation step allowed to remove interfering substances, namely SDS and salts, and to achieve the successful focusing of proteins in an immobilised pH gradient, allowing 2D-PAGE separation. To our knowledge, this is the first report presenting the differential comparison of 2D-PAGE protein extracts obtained from a fresh-frozen and a replicate FFPE tissue. As shown in Fig. 5.1, Fig. 5.3, and Appendix 3, it is possible to separate by 2D-PAGE, and to identify by MS, full-length proteins from FFPE tissues, obtaining datasets that are comprehensive and comparable to those produced with fresh-frozen tissues. Nevertheless, evident differences in spot intensity or focusing quality can still be seen between fresh-frozen and FFPE tissue 2D maps. During this study, however, we observed that the 2D maps generated from FFPE tissues were highly reproducible. Therefore, although a perfect match may not be obtained between fresh and FFPE 2D maps, differential proteomics studies could be performed among FFPE 2D maps, allowing comparison of healthy and pathological tissues stored in bioptic tissue banks.

The successful 2D-PAGE separation of FFPE extracts suggests that a sufficient amount of proteins can be recovered with unaltered or restored physical properties, even from limited amounts of tissue, allowing electrophoretic migration under both conditions, IEF and SDS-PAGE. The migration of proteins at the expected MW and pI ranges was also suggested by the western immunoblotting studies. In fact, as shown in Figure 5.2, the pattern of reactivity observed with anti-actin and anti-GAPDH antibodies when probing FFPE skeletal muscle extracts is comparable with the pattern of reactivity obtained when probing fresh-frozen tissues, notably, also in terms of the isoelectric series of spots. The result obtained with GAPDH (*pI* 8.5) is of particular relevance, since

basic proteins are able to engage a high number of formaldehyde-induced crosslinks. GAPDH, therefore, is expected to be extracted and separated less efficiently compared to actin or other acidic proteins. The results reported here demonstrate that the FFPE protein extraction method used efficiently reverts formaldehyde-induced crosslinks, and that such reversal does not impact significantly neither focusing properties nor immunoreactivity of the extracted proteins. This opens the way to many applications, such as immunostaining for post-translational modifications, for example. However, potential crossreactivity problems caused by residual crosslinked peptides need to be taken into account in this type of assays.

Following 2D-PAGE separation, focused protein spots can be successfully excised from the gel and identified by MALDI-MS analysis after in-situ hydrolysis. The MS spectra generated are of high quality, and are comparable to those obtained with proteins from fresh-frozen tissue extracts. Alterations of the peptide mass fingerprint, if present, are minimal, and do not impair identification of the protein. Mass spectrometry results were always comparable in matched protein spots scattered throughout acidic, basic, high MW, and low MW areas of the map.

Notably, MS/MS analysis of many spots from FFPE maps led to detect a significant number of peptides belonging to the most abundant skeletal muscle proteins (such as actin, myoglobin, creatine kinase, and others) besides those coming from the expected proteins. This phenomenon was especially present in the high molecular weight and acidic region (Table 5.1). This might suggest the presence of short, residual fragments that remain linked to medium-high MW proteins, as a consequence of an incomplete reversal of crosslinks and protein breakdown. These peptides could be released owing to trypsin digestion and codetected by MS/MS, without impairing MALDI-MS identification of the “correct” protein species present within the spot.

To date, only two reports described the application of 2D-PAGE to proteins extracted from FFPE tissues [84-85]. The 2D-PAGE map published by Ahram and his colleagues was virtually devoid of focused spots, and was indeed specifically aimed to demonstrate the purported unsuitability of FFPE protein extracts to 2D-PAGE-MS studies [84]. This study, in fact, suggested the use of ethanol for tissue fixation if the biopsies were to be subjected to proteomic studies. On the other hand, a recently

published study described application of 2D-PAGE for the discovery of biomarkers in cervical carcinoma [85]. Although this is a valuable work, and interesting results were obtained by these investigators, the 2D-PAGE maps generated were of poor quality in terms of number and resolution of protein spots. Moreover, these authors used cyanine dyes for staining, while the 2D maps reported here allowed to detect a higher number of spots by means of far less sensitive staining chemistries. Nevertheless, the results obtained by Ono et al. are of great interest, since the possibility of staining 2D-PAGE FFPE maps with cyanine dyes strongly encourages the application of the recently developed DIGE protocols.

2D-PAGE maps and MS identifications presented here demonstrate that it is possible to carry out gel-based differential proteomics experiments on FFPE samples. Standardisation and optimisation of this protocol may open the way to biomarker discovery studies, as well as systematic proteomic profiling of difficultly available tissues, such as rare tumour samples. Enlarging and improving the range of techniques useful for the molecular characterisation of FFPE tissues can be considered a crucial point of biomedical translational research for the next few years.

6. DIGE ANALYSIS AND CHARACTERISATION OF PROTEINS FROM FORMALIN-FIXED, PARAFFIN-EMBEDDED TISSUES

I. Introduction

FFPE tissues represent an invaluable source of information in the field of disease molecular characterisation, both for diagnostic and therapeutic purposes. A critical target of proteomics in the few past and next years might be to improve the ability to carry out biomarker discovery projects on this complicated kind of biological samples. 2D-DIGE is nowadays considered as the most refined technique for performing gel-based differential proteomics studies [36, 38]. As described in Chapter 5, recent data demonstrated the skill of separating and identifying proteins from FFPE tissues, by means of 2D-PAGE-MS. Thus, the following step became the implementation and validation of this result by the application of DIGE technique to FFPE samples. As well, when a limited amount of sample is available, as for fixed tissues, 2D-DIGE allows to obtain significant results because of its property of minimizing experimental variability. Thus, here we present a 2D-DIGE characterisation of healthy sheep skeletal muscle and liver tissues, and a comparison of two pathological FFPE liver samples, still by a DIGE approach, in order to investigate usefulness and reliability of this technique for biomaker discovery studies on fixed samples.

II. Results

1. FFPE extracts can be successfully separated and analysed by 2D-DIGE

In order to evaluate the suitability of 2D-DIGE for FFPE protein extracts, an animal model was used. Indeed, performing optimisation experiments on human bioptic samples would have been unreasonable, owing to their scarceness and preciousness. Thus, skeletal muscle and liver tissues samples were collected from a single sheep individual, with the aim of reducing biological variability. Every tissue block was divided

into two parts: the first was subjected to formalin fixation and subsequent paraffin embedding, following standard laboratory procedures; the second was immediately frozen at -80°C. FFPE blocks were microtomised into 10 µm sections and kept at room temperature. For protein extraction, FFPE sections were subjected to a high-temperature protein extraction procedure, as described in previous chapters. Fresh-frozen tissue samples were minced with a sterile scalpel, immersed in the same buffer used for FFPE samples, and then subjected to an extraction procedure based on mechanical lysis. Protein extracts were treated with 2-D Clean-Up Kit (GE Healthcare), in order to eliminate SDS and other substances which could interfere with IEF separation; final protein pellets were resuspended in a buffer suited for DIGE analysis, and quantified by EZQ fluorescent method (Molecular Probes). For evaluating FFPE extraction technical variability and its impact on 2D-DIGE, three replicates, each consisting of ten adjacent FFPE skeletal muscle or liver sections, were subjected to distinct extraction procedures. As internal standard, a pool comprising 25 µg of each FFPE replicate and 75 µg of fresh-frozen extract was prepared. Then, sample pool, FFPE replicates and fresh-frozen extract of each tissue were labelled with CyDyes, and mixed for performing three DIGE experiments per tissue, as summarised in Table 6.1. Proteins were isoelectrofocused along two different pH ranges: 3-10 for skeletal muscle, and 4-7 for liver. SDS-PAGEs were carried out on gradient (8-18%) polyacrilamide 24 cm gels.

Table 6.1. Summary of DIGE experiments.

exp.	tissue	pH range	Cy2	Cy3	Cy5
M1	Skeletal muscle	3-10	Pool standard	FFPE repl. 1	Fresh-frozen
M2	Skeletal muscle	3-10	Pool standard	Fresh-frozen	FFPE repl. 2
M3	Skeletal muscle	3-10	Pool standard	FFPE repl. 3	Fresh-frozen
F1	Liver	4-7	Pool standard	FFPE repl. 1	Fresh-frozen
F2	Liver	4-7	Pool standard	Fresh-frozen	FFPE repl. 2
F3	Liver	4-7	Pool standard	FFPE repl. 3	Fresh-frozen

Representative 2-D maps are shown in Figure 6.1. Skeletal muscle patterns are similar or even slightly better to those obtained by silver-staining, and previously published by our group. It was not an obvious result, because only 50 µg of FFPE proteins were loaded; moreover, cyanine minimal labelling involves lysine residues, that were

demonstrated to be one of the favourite targets of formalin in the context of fixation reactions [60-62]. Liver extract behaviour was really surprising. Unlike skeletal muscle, which is characterised by the presence of a few acidic highly abundant proteins, well conserved in FFPE pattern, liver proteome shows a narrower dynamic range, and a plenty of medium-low abundant proteins. In view of these features, IPG strips with a more limited pH range were chosen for liver DIGE experiments, in order to reduce pattern density. In spite of these expectations, liver FFPE 2D-profile exhibits a good complexity, a notable resolution, as well as an outstanding comparability with fresh-frozen pattern. Indeed, it is noteworthy the preservation of isoform "trains", also for medium-high molecular weight proteins. Up to 250 and 350 spots could be detected in skeletal muscle and liver FFPE maps, respectively, compared to a mean of about 450 spots for fresh-frozen samples.

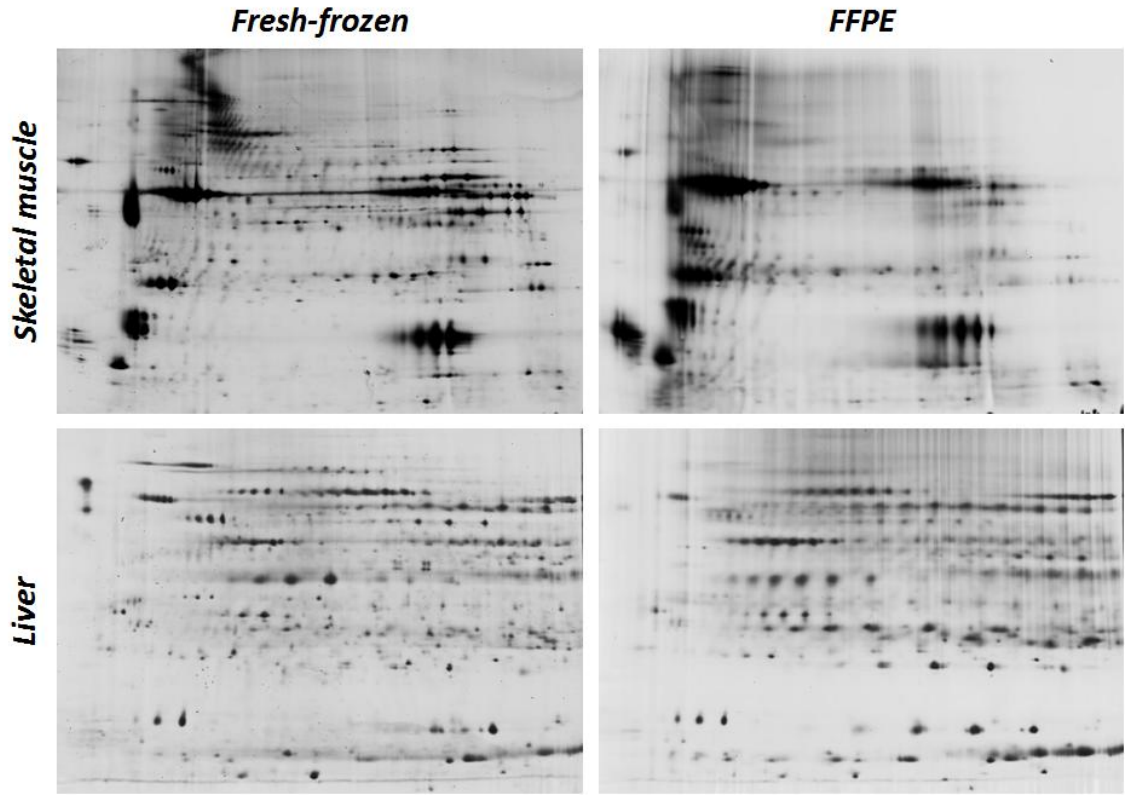

Figure 6.1. 2D-DIGE analysis of proteins from sheep skeletal muscle and liver tissues. FFPE and fresh-frozen extracts of each tissue were run in the same gel. Proteins were separated first along 3-10 NL (muscle) and 4-7 (liver) pH ranges, and then on 24 cm polyacrilamide gradient gels (8-18%).

In order to discover the identity of the main proteins expressed by liver and skeletal muscle, fresh-frozen and FFPE extracts of both tissues were separated by 24 (or 7) cm gel electrophoresis; globally, 41 spots were excised, subjected to *in situ* hydrolysis with trypsin, and tryptic digests were analysed by MALDI-MS or LC-MS/MS.

2. FFPE extract replicates show a high reproducible 2D-DIGE pattern

High experimental reproducibility is an essential condition for performing differential proteomics studies. Thus, one of the aims of this work was to assess the level of reproducibility between 2D-DIGE maps of FFPE technical replicates. For technical replicates, identical samples subjected to distinct protein extraction procedures are intended. Therefore, a triplicate DIGE analysis were conducted, as already displayed in Table 6.1. 2D-DIGE images of FFPE replicates are comparatively shown in Figure 6.2.

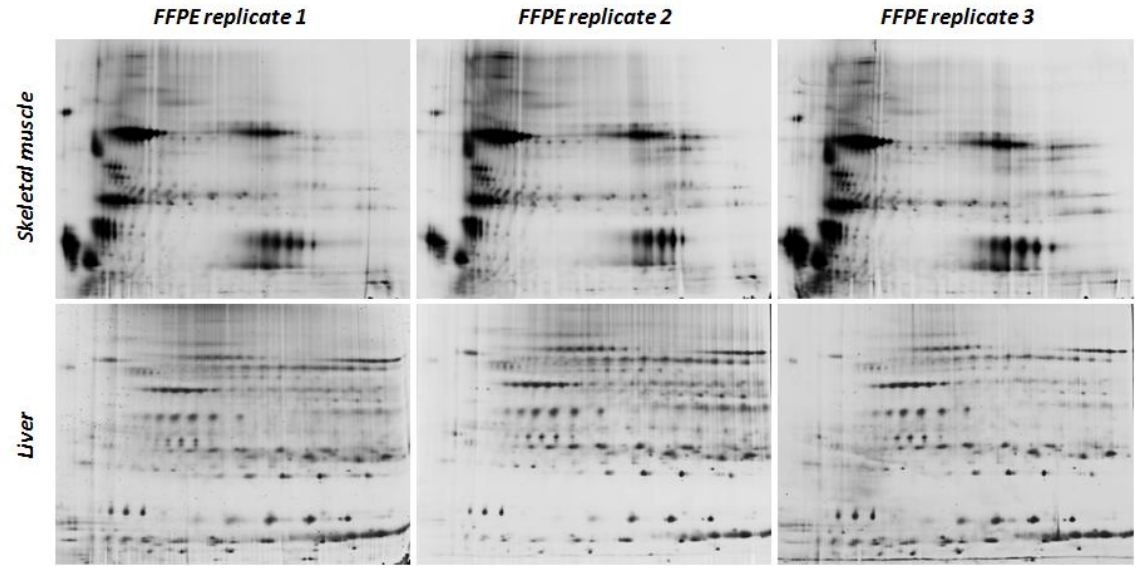

Figure 6.2. 2D-DIGE profiles of FFPE extracts technical replicates from skeletal muscle (upper line) and liver (lower line) tissues.

Reproducibility rate was measured by two different methods: comparing, between fresh-frozen and FFPE maps, standard deviation and coefficient of variation of overall spot number (see Table 6.2), or coefficient of variation of spot volume for ten selected and matched spots for each tissue (see Figure 6.3 and Appendix).

Comparing gel images, and data presented in Table 6.2, overall 2D-profile of FFPE muscle and liver replicates appears highly reproducible, in terms of quantity and

quality. Coefficient of variation value is evidently low and very similar with fresh-frozen.

Table 6.2. Overall spot number of triplicate 2D-DIGE maps for skeletal muscle (left) and liver (right) fresh-frozen and FFPE extracts. Mean value (MV), standard deviation (SD) and coefficient of variation (CV) of spot number are also reported for each tissue.

muscle	***FrFr***	***FFPE***
exp. 1	456	224
exp. 2	423	251
exp. 3	444	216
MV	441	230
SD	16,7	18,3
CV	0,04	0,08

liver	***FrFr***	***FFPE***
exp. 1	440	345
exp. 2	512	315
exp. 3	471	323
MV	474	328
SD	36,1	15,5
CV	0,08	0,05

Furthermore, ten widespread distributed spots per tissue were selected; their volumes among the three replicates were measured, and respective means, standard deviations and coefficients of variation were calculated. Considering the average of coefficients of variation, variability assumes a more significant value than that computed with the previous method. However, FFPE and fresh-frozen data are widely comparable (see Fig. 6.3 and Appendix 4).

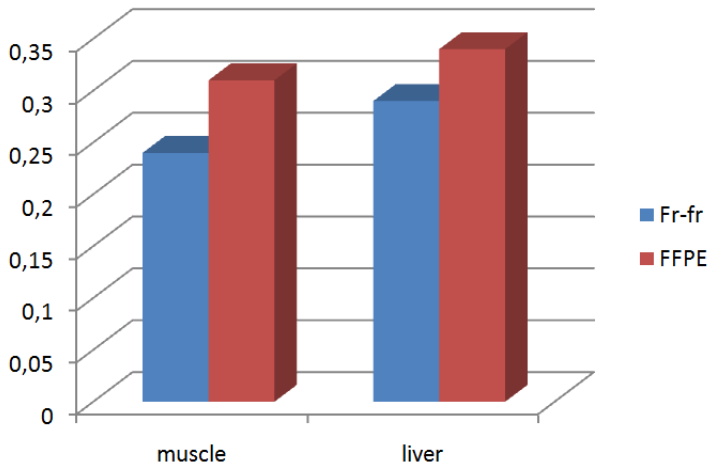

Figure 6.3. Histogram comparing average coefficient of variation (CV, in ordinate) of spot volumes (calculated for ten selected spots per tissue), as a measure of technical variability.

It is not trivial to observe that, whereas fresh-frozen replicates shared the same extract, each FFPE replicate derived from a different tissue portion and was extracted

in a distinct time. Therefore, the greater part of the variability amongst FFPE maps was due to electrophoresis, while the contribution of extraction procedure or sample heterogeneity was almost negligible.

3. FFPE proteins are acidic-shifted compared to corresponding fresh-frozen ones

The purpose of 2D-DIGE experiments was to compare FFPE and fresh-frozen protein extracts, not only for investigating the degree of comparability between them, but also for determining peculiar pattern differences, which would be taken into account for further studies. 2D-DIGE overlay patterns of skeletal muscle and liver tissues (presented in Fig. 6.4) show an evident, unambiguous acidic-shift that affects almost all FFPE spots and spot "trains". Considering for the moment only the conserved trains (specific preservation or depletion of FFPE spots will be treated in the next paragraph), a typical fashion can be noticed: in the basic side of the train the "fresh" protein seems to prevail, in the middle a substantial balance can be highlighted, whilst the acidic isoforms appear to be highly more abundant in FFPE extract.

In order to elucidate more in depth this interesting phenomenon, six selected spot trains, widely distributed for molecular weight and isoelectric point within skeletal muscle and liver maps, were analysed, as shown in Figure 6.5. The colour gradient is extremely evident, and pI values point out how ample is the shift. Considering as a measure of the shift the difference between the mean isoelectric point of fresh-frozen trains and that one of FFPE trains, the top value, among the six selected proteins, coincides with 0.28 pI units of shift, determined for liver superoxide dismutase [Cu-Zn] train (Fig. 6.5, panel E).

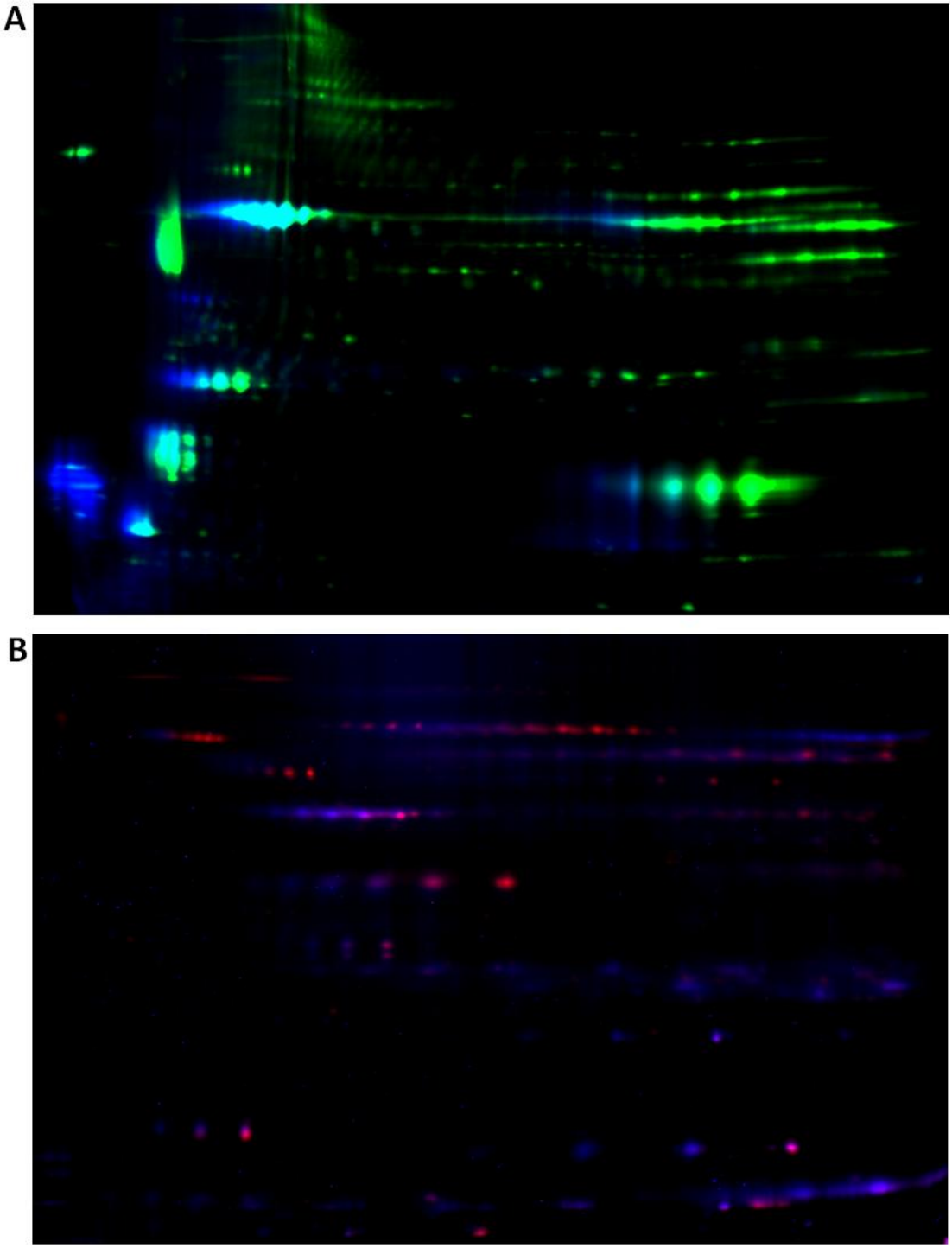

Figure 6.4. 2D-DIGE analysis of skeletal muscle (A) and liver (B) protein extracts. In A, fresh-frozen, FFPE and overlay patterns are marked in green, dark blue, and light blue, respectively. In B, fresh-frozen, FFPE and overlay patterns are marked in red, dark blue and pink, respectively.

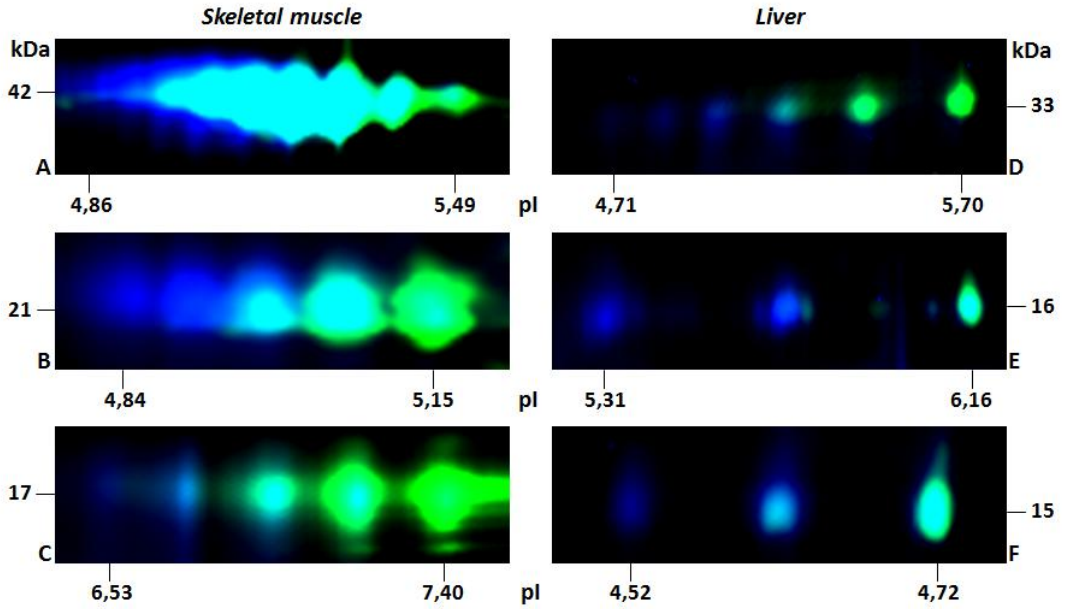

Figure 6.5. Selected spot "trains" from skeletal muscle (left) and liver (right) 2D-DIGE maps, corresponding to actin (A), myosin light chain 1 (B), myoglobin (C), regucalcin (D), Cu-Zn superoxide dismutase (E), and cytochrome b5 (F). Fresh-frozen, FFPE and overlapped patterns are depicted in green, dark blue, and light blue, respectively. Estimated molecular weights and isoelectric points are also indicated.

4. Formalin fixation detrimental impact on preservation of electrophoretic protein pattern seem to be correlated with lysine residues percentage

Another point to be investigated was the preservation (or the depletion) of specific proteins in FFPE patterns, and whether any protein physico-chemical feature might be correlated with this event. Therefore, image analysis of 2D-DIGE gels was carried out with DeCyder software, with the aim of identifying FFPE (or fresh) protein spots whose intensity was significantly changed, compared to fresh-frozen (or FFPE). The first ten "overexpressed" spots for each gel are highlighted with red (more abundant in fresh) or blue (more abundant in FFPE) circles in Figure 6.6. Hatched circles indicate spots pointed out in the corresponding sample (always fresh *vs* FFPE), for comparison purposes. Spots distribution according to molecular weight and pI shows interesting features. First, proteins more abundant in fresh-frozen extract (and, consequently, depleted in FFPE) are mostly confined in the medium-high molecular weight gel region, and only a few high MW proteins are properly conserved in FFPE gels; it may be due to the fact that bulky proteins are more difficult to be extracted, to enter the gel or to be correctly separated and visualised in electrophoresis. Second, in regard to skeletal

muscle map, acidic proteins appear more easily preserved than basic. The same consideration cannot be done about liver, because of the restricted pH range of the map.

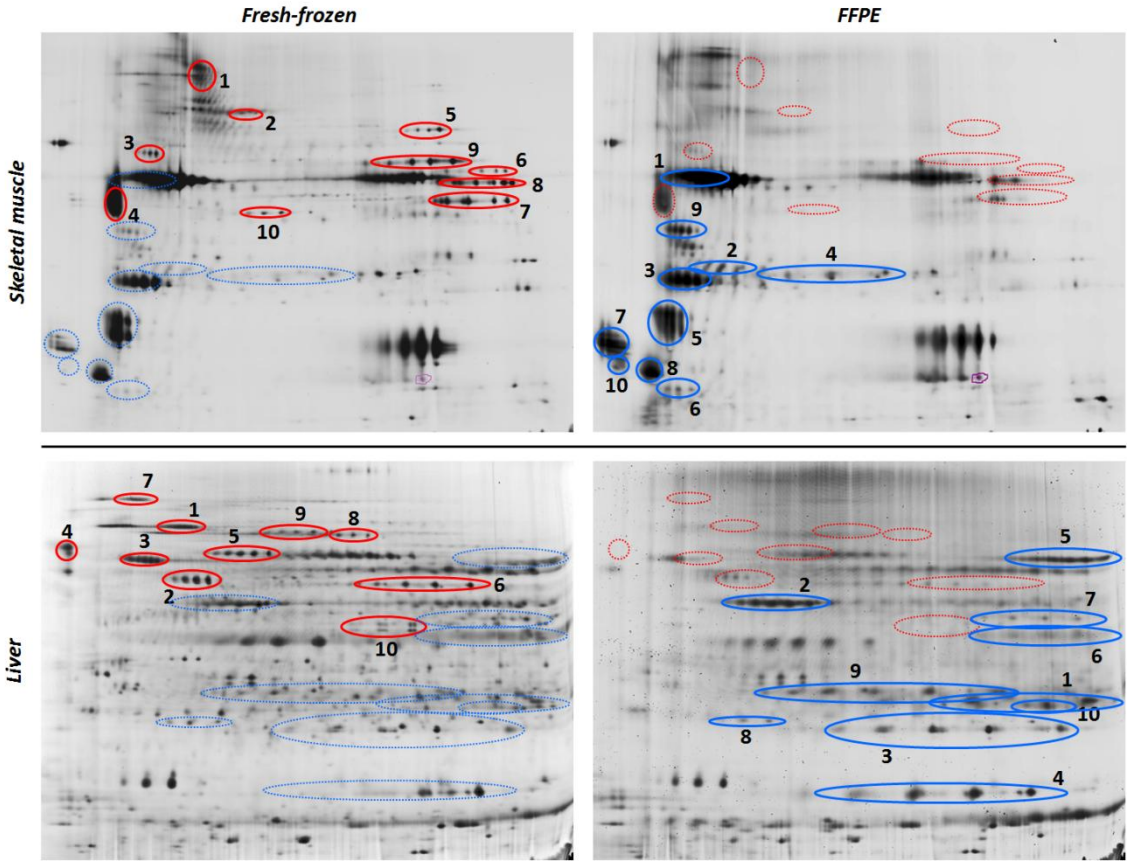

Figure 6.6. Skeletal muscle (top) and liver (bottom), fresh-frozen (left) and FFPE (right) 2D-DIGE maps. Circles and numbers indicate differentially "represented" proteins (ordered by decreasing spot volume ratio) in comparison with matched fresh-frozen or FFPE tissue. Spots which show higher volume in fresh-frozen maps are marked in red, while those more represented in FFPE extracts are marked in light blue. Hatched circles refer to spots that are overexpressed in the corresponding FFPE, or fresh-frozen, map.

In order to understand what structural factors influence preservation of fixed proteins, some physico-chemical features were considered for the forty differentially expressed proteins mentioned above. Particularly, lysine and arginine percentages, as well as their ratio, were calculated, with the aid of ProtParam tool [110], and the trend of these values for fresh and FFPE proteins were statistically analysed. Wilcoxon's Rank-Sum test results demonstrate a significant variation between the two groups for lysine percentage ($p = 0.005$) and lysine to arginine ratio ($p = 0.01$) of liver proteins, while only lysine percentage ($p = 0.035$) shows a significantly difference in skeletal muscle

proteins (Figure 6.7, and Appendices 5-6). Therefore, depletion of lysine-rich proteins observed in FFPE 2D-maps might be caused by a probable more extensive reactivity with formalin [60-62], which should affect protein extraction and/or separation.

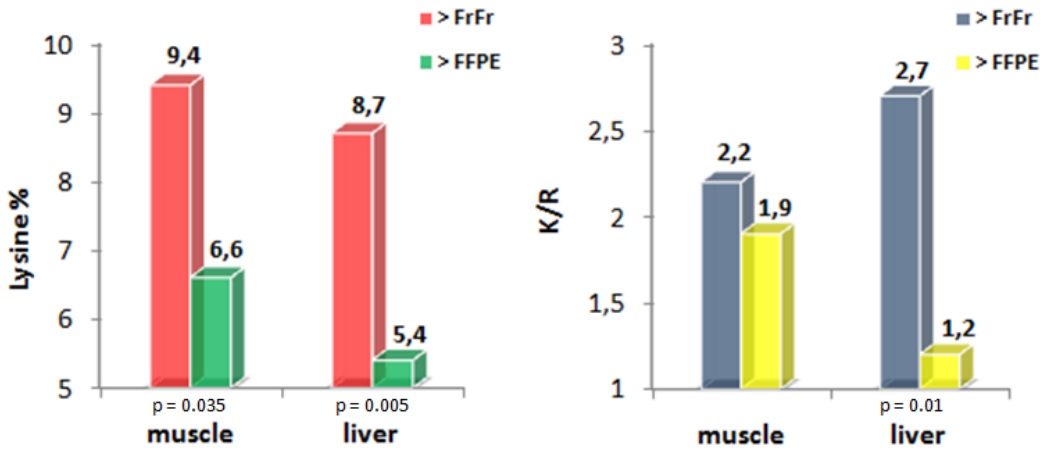

Figure 6.7. Comparison of mean lysine percentage and K to R ratio between a group of ten spots more represented in fresh-frozen map and a group of ten spots more represented in FFPE map.

5. 2D-DIGE and GeLC-MS/MS allow the identification of several heavy metal damage biomarkers in sheep liver FFPE tissues

The first purpose of this work was to demonstrate that proteomic differential studies could be performed through gel-based DIGE technology. Once proved adequate reproducibility and efficacy of the method, in comparison with fresh-frozen tissue, the following step was to apply extraction procedure and 2D-DIGE workflow to differentially analyse FFPE samples. In view of the satisfactory results achieved with sheep liver, two sheep samples affected by heavy metal liver intoxication at different stages, and routinely fixed and embedded, were chosen. Hence, the goal of this part of the survey was to recognise a few disease biomarkers overexpressed by the more damaged liver (hereafter called A), in comparison to the case with a less advanced stage of degeneration (hereafter called B).

In order to reach this outcome, proteins were extracted from A and B FFPE samples, and separated by SDS-PAGE; subsequently, the same extracts were labelled with Cy2 and Cy5, respectively, with the aim of performing a preliminary 1D-DIGE analysis. Comparing the two patterns, placed side by side in Figure 6.8, four differential bands were detected, excised and processed by a GeLC-MS/MS approach.

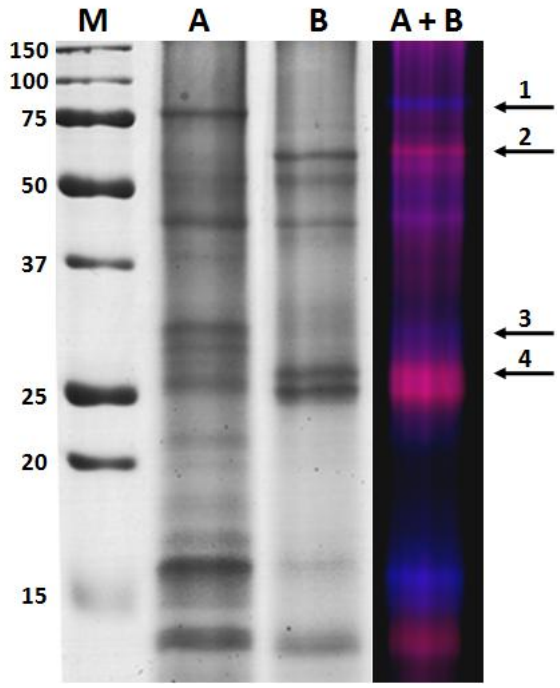

Figure 6.8. SDS-PAGE (left, grey-scale) and 1D-DIGE (right, in colour) of sheep pathological liver extracts. SDS-PAGE gel was stained with colloidal Coomassie blue. Lane M represents Precision Plus Protein Standard (Bio-Rad). For DIGE analysis, sample A was labelled with Cy2 (in blue), and sample B with Cy5 (in red); overlapped bands are in pink. Arrows indicate picked bands.

Table 6.3. List of differentially expressed proteins identified by GeLC-MS/MS, ordered decreasingly according to molecular weight. Total number of peptide hits for each sample (in columns A and B, respectively) and band(s) corresponding to protein identification are also reported.

Overexpressed proteins	MW	A	B	Band
78 kDa glucose-regulated protein	72	11	0	1
Protein disulfide-isomerase A2	59	11	0	1,2
Protein disulfide-isomerase	58	23	0	1,2
D-3-phosphoglycerate dehydrogenase	57	8	0	2
Protein disulfide-isomerase A3	57	6	0	2
Peroxiredoxin-4	31	10	0	4
14-3-3 protein epsilon	29	3	0	2,3
Chymotrypsin-like elastase family member 1	29	8	0	4
Proproteinase E	28	5	0	4
Chymotrypsinogen A	26	6	0	4
Endoplasmic reticulum resident protein ERp27	23	5	0	2,3,4
Peroxiredoxin-1	22	6	0	4

Underexpressed proteins	MW	A	B	Band
Glutamate dehydrogenase 1, mitochondrial	62	0	27	1,2,3,4
Catalase	60	0	54	1,2,3,4
Retinal dehydrogenase 1	55	0	5	2
Regucalcin	34	0	5	1,3
Enoyl-CoA hydratase, mitochondrial	32	0	8	2,4
3-hydroxyacyl-CoA dehydrogenase type-2	27	8	62	1,2,3,4

As previously described in Chapter 4, GeLC-MS/MS analysis of FFPE extracts proved to be a useful and robust method for the characterisation of proteomic profile. Here, a mean of 30 protein identifications per band was attained, and over 90 and 60 unique proteins were detected in total for A and B samples, respectively. Complete MS identifications are listed in Appendix 7. Table 6.3 reports main proteins that exhibited a differential abundance, estimated by total peptide hits number, between A and B. Interestingly, several of the proteins overexpressed in the most damaged tissue are clearly related to liver stress and intoxication states, as will be more widely discussed below.

Then, liver pathological samples were subjected to 2D-DIGE analysis. Protein extracts were firstly isoelectrofocused in 3-11 IPG strips, and secondly separated according to their molecular weight in (8-16%) gradient polyacrilamide gels. 2D-maps in Figure 6.9 reveal a quite acceptable complexity and resolution, at least for the zone corresponding to a molecular weight lower than 40 kDa. Image analysis was carried out by means of DeCyder software, and principal differential expressed spots are marked with blue circles (those overexpressed in A) or red rectangles (those underexpressed in A).

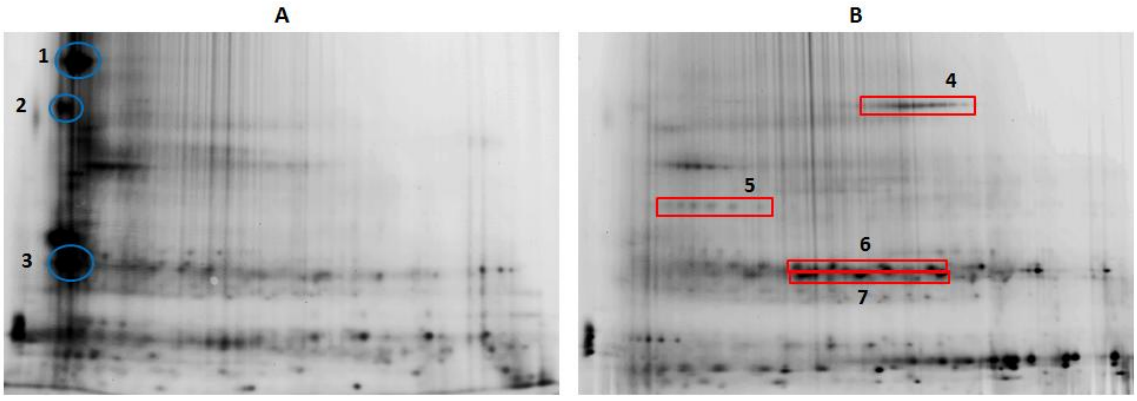

Figure 6.9. 2D-DIGE analysis of FFPE sheep liver samples affected by heavy metal intoxication at two different stages. Differentially expressed proteins are numbered and marked in light blue (spots overexpressed in A) and red (spots underexpressed in A).

Differential spots were cut from DIGE gels, digested with trypsin, and analysed by MALDI-MS and LC-MS/MS. Protein identities confirmed identifications achieved by GeLC-MS/MS, as illustrated in Table 6.4. All discovered putative biomarkers, such as 78

kDa glucose-regulated protein (also known as GRP78), protein disulfide-isomerases, 14-3-3 proteins, and ERp27, share functional properties and molecular pathways [114-116], and have been previously indicated as heavy metal intoxication markers [117-119], or liver stress signallers [120-122].

Table 6.4. Proteins identified by MALDI-MS and LC-MS/MS from 2D-DIGE differentially expressed spots.

Spot	AN	Overexpressed proteins	MW	pI	UP	SC
1	Q0VCX2	78 kDa glucose-regulated protein	72	5,07	8	15
2	P05307	Protein disulfide-isomerase	58	4,80	11	26
	Q13087	Protein disulfide-isomerase A2	59	4,89	6	8
3	P62261	14-3-3 protein epsilon	29	4,63	2	7
	Q32L47	ERp27	23	4,81	1	5
4	P00432	Catalase	60	6,78	6	14
5	Q9TTJ5	Regucalcin	34	5,54	4	16
6	Q58DM8	Enoyl-CoA hydratase, mitochondrial	32	8,82	2	7
7	O02691	3-hydroxyacyl-CoA dehydrogenase type-2	27	8,45	7	46

UP, unique peptides; SC, sequence coverage (in percentage).

III. Discussion and concluding remarks

In the past few years, 2D-DIGE has emerged as one the more suited and sophisticated techniques for gel-based differential proteomics studies. Utmost sensitivity of cyanine staining, noteworthy minimisation of experimental variability, opportunity to run up to three different samples in a single gel, make this system as a perfect way for the analysis of FFPE tissues. In fact, limited size of greater part of stored FFPE biopsies does not allow the access to an elevate amount of proteins; hence, the ability to perform a proteomic investigation without the need of a large number of experiments, or of a huge availability of protein extract, may represent an invaluable resource for rare disease research. In view of these reasons, the successful 2D-DIGE analysis of FFPE extracts described here can be considered as a significant advance in clinical proteomics.

FFPE 2D-maps, as described above, show a notable degree of comparability with fresh-frozen, a high reproducibility between replicates, and an unexpected pattern complexity, particularly for liver. Compared to the only attempt of applying DIGE to FFPE samples published so far [85], overall map quality has been considerably

improved. Furthermore, when applied to routinely collected biopsies, the combination of the FFPE protein extraction method developed in our laboratory and 2D-DIGE led to significant results. The experiment consisted in a comparison between proteomes expressed by two FFPE sheep liver samples, affected by different levels of heavy metal intoxication, in order to identify damage biomarkers. All main differentially expressed spots correspond to proteins which have been involved in liver stress pathways. GRP78 and protein disulfide-isomerase (PDI) are long since well-known liver damage biomarkers, and were found as overexpressed in a wide range of disease states [117-119, 122]. 14-3-3 isoforms have been related to liver cancer [121], while little is known about ERp27, except its interaction with PDI [116]. In addition, PDI, GRP78 and ERp27 are all localized at level of endoplasmic reticulum and share folding properties [114-115]; 14-3-3 has been associated to various signalling pathways, especially related to cell cycle and apoptosis regulation [120]. GeLC-MS/MS analysis results show an excellent consistency with 2D-DIGE-MS identifications, confirming robustness and trustworthiness of both methods, in spite of the historical “inaccessibility” of FFPE samples.

Undoubtedly, several remarkable limitations remain still unresolved. For instance, in high molecular weight gel regions, a moderately high background is present, combined with a lower spot intensity, if compared with the signal of smaller proteins; this may cause a few troubles in image analysis performances. It is very important to note that comparison must be done only between FFPE samples; it should be senseless, at present, to perform a “hybrid” differential proteomic investigation, with both FFPE and fresh samples. Instead, when comparing FFPE extracts, drawbacks, even if existing, are shared, and so can be somehow equally “subtracted”. Nonetheless, there is a certain need to further improve methodologies for the analysis of FFPE samples, here included extraction procedures and analytical methods.

A secondary aim of this study was to infer from 2D-DIGE data information about chemical modifications of fixed proteins. Two crucial evidences were argued: a diffuse acidic-shift, and a possible correlation between lysine percentage and spot intensity reduction, always as compared to fresh-frozen proteins. The first phenomenon provoked by fixation seems to affect spots trains all along the 2D map, and presents a

clear visual impact, highlighted by DIGE technology. Rait *et al.* [67] took notice of an isoelectric point decrease in bovine RNAse A, when treated with formaldehyde; however, a widespread p*I* shifting were never demonstrated before. A possible explanation of this event might be found in the impact on protein charge of chemical modifications, caused by formaldehyde, on basic residues and N-termini: bond formation probably modifies basicity of nitrogens, bearing a p*I* variation.

Also the involvement of lysine in fixation reactions has been previously described [60-62]; several works count lysine as one of the most reactive residues in fixation procedures. Therefore, it seems reasonable to state that lysine-rich proteins form a wider number of crosslinks, and this may make them more difficult to be adequately extracted and/or impair their correct migration on gel. Some of the FFPE more abundant spots, listed in Appendix 6, are actually truncated forms of larger proteins. According to sequence coverage information obtained from LC-MS/MS data, it was possible to estimate what part of the protein corresponded to the selected spot (data not shown). Intriguingly, all these protein "by-products" own a lower lysine residues number than the original molecules; perhaps not by chance, they result more abundant in FFPE, even if perfectly detectable also in fresh-frozen pattern.

As described in results section, FFPE protein preservation appears to depend on molecular weight and p*I*, with large size and basic proteins somehow "handicapped". Moreover, also extraction buffer pH has been indicated as a factor that can influence selectivity in protein extraction [123]. The extraction procedure used in this work is based on a basic buffer; according to preceding papers, it might favour the release of acid proteins, in despite of basic ones.

Another interesting point is represented by the fact that lysine is also the target of cyanines in minimal labelling procedure. It could be thought that FFPE modified lysines, reacting only partially with CyDyes, give a less intense staining; however, a rapid comparison with skeletal muscle silver-stained maps, previously published by our group, or with Coomassie-stained gels generated for spot picking and MS identification purposes (images not shown), clearly demonstrates how preservation/depletion trend of FFPE proteins is highly conserved, also in differently stained gels.

In conclusion, data reported here can be considered as the first stable step in 2D-DIGE characterisation of FFPE samples. A widespread application of this technology on rare disease samples, together with a refinement of FFPE protein extraction and analysis methods, might represent a precious weapon for biomarker discovery studies in the years to come.

7. ESTIMATION OF RELATIVE PROTEIN ABUNDANCE BY LABEL-FREE GELC-MS/MS PROTEOMICS

I. Introduction

Quantifying changes in protein abundance between samples is a key requirement for differential proteomic studies. In the last years, two principle methods for estimating protein quantitative variation were proposed: labelling approaches, which consist in isotope or mass tag labelling of peptides, and label-free approaches, which base themselves on MS/MS peak area or spectrum number measurement. Even if perhaps less sensitive, label-free methods count the advantage of low costs and lack of supplementary chemical reactions or data analyses, that are required by isotope-labeling.

GeLC-MS/MS represents an interesting "hybrid" between gel-based proteomics and shotgun proteomics; it allows to achieve a quite complex proteomic profiling, without the need of extremely sophisticated MS instrumentation and starting from a full-length protein characterisation step.

Thus, the aim of this work was to determine whether, following a typical GeLC-MS/MS workflow, an estimation of protein abundance could be successfully obtained, and to establish which MS parameter, among total peptide hits, unique peptides, and percentage sequence coverage, exhibits the best correlation with protein abundance.

II. Results

In order to investigate GeLC-MS/MS suitability to semi-quantitative studies, three protein standards were chosen (bovine serum albumin, yeast enolase, and equine myoglobin, ranging from about 15 to 70 kDa), and a total *E. coli* protein extract was prepared, as a complex protein mixture. A bacterial extract was preferred, since belonging to a species not phylogenetically closed to those where standards come

from, to limit MS identification overlap. Then, a SDS-PAGE was carried out as follows: 250, 100, or 40 nanograms of each protein marker were added to the *E. coli* extract, or loaded in the adjacent lane as "pure" proteins (Figure 7.1). Hence, the intention was to investigate the behaviour of each protein both alone and within a composite mixture, and its influence on MS identification, and then on protein abundance estimation.
Following a standard GeLC-MS/MS procedure, after Coomassie-staining, 9 gel slices, comprising all marker visible bands, and the corresponding gel regions along the *E. Coli* 1-D pattern, were excised; trypsin *in-gel* digestion was carried out, and tryptic peptides were extracted and analysed in LC-MS/MS.

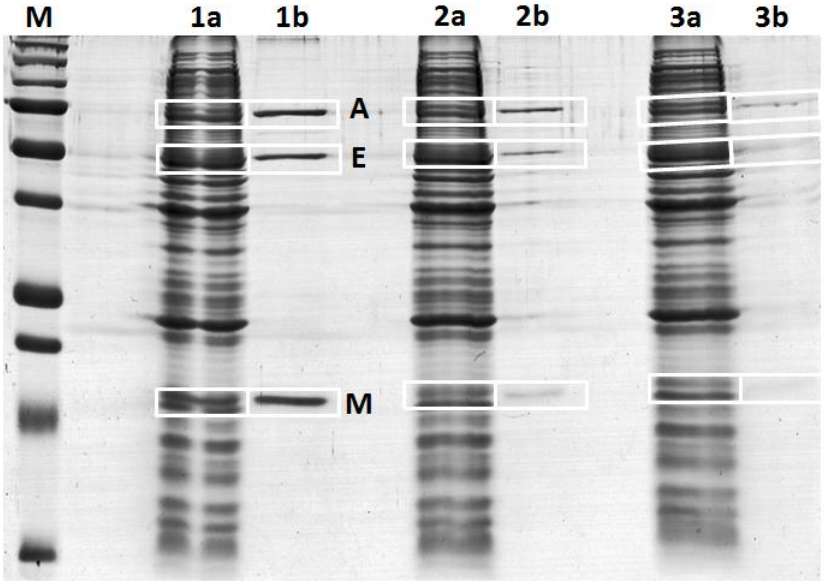

Figure 7.1. SDS-PAGE of standard protein dilutions. 250, 100, and 40 nanograms of each protein were loaded in lane 1, 2, and 3, respectively. In lanes marked with "a", protein standards were mixed with 20 micrograms of total *E. coli* protein extract.

MS identifications data are presented in Table 7.1, indicating for each experiment total peptide hits, unique peptides, and percentage sequence coverage. At a first sight, a substantial correlation between these parameters and protein quantity comes out, according to the hypotheses.

Table 7.1. Summary of GeLC-MS/MS identifications data, separated per protein standard.

ALB	Pept. hits	Unique pept.	Seq. coverage	ENO	Pept. hits	Unique pept.	Seq. coverage	MYO	Pept. hits	Unique pept.	Seq. coverage
1a	33	20	40	**1a**	11	7	21	**1a**	46	6	51
2a	10	9	16	**2a**	0	0	0	**2a**	9	3	22
3a	8	8	12	**3a**	0	0	0	**3a**	0	0	0
1b	94	30	52	**1b**	34	12	30	**1b**	60	6	51
2b	40	23	34	**2b**	21	11	30	**2b**	11	5	40
3b	17	12	19	**3b**	12	9	25	**3b**	4	2	14

A more in-depth data analysis shows a few interesting features. First, as shown in Table 7.2 and Figure 7.2 (Panel A, B, and C), and summarised in Figure 7.3, values concerning "pure" proteins change significantly in comparison to the "mixture". Paradoxically, in presence of a plenty of other proteins, MS data seem more consistent with actual protein quantity, at least for sequence coverage and unique proteins. However, total peptide hits parameter keeps in both conditions and for all three proteins an optimal correlation trend, never below 0,95 (R-squared value) and with a top of 1,000 for pure albumin. Instead, unique proteins and sequence coverage criteria demonstrate a wider variability, and only in one-fifth of the cases reach a higher coefficient of determination than total peptide hits method.

Enolase peptides, when mixed with *E. coli* ones, coming from co-migrating bacterial proteins, could not be detected from 100 nanograms of protein loaded downwards: it is an important issue, which will be discussed in the next paragraph.

Table 7.2. Summary of R-squared values obtained for each experiment.

	Sample	Total peptide hits	Unique peptides	Seq. coverage
Mixture	*A*	0,957	0,958	0,978
	E	-	-	-
	M	0,991	0,942	0,973
	main	0,974	0,950	0,976
Pure	*A*	1,000	0,970	0,964
	E	0,981	0,824	0,519
	M	0,973	0,750	0,793
	main	0,985	0,848	0,759
Overall	*main*	0,979	0,899	0,867
	main excl. E	0,980	0,905	0,927

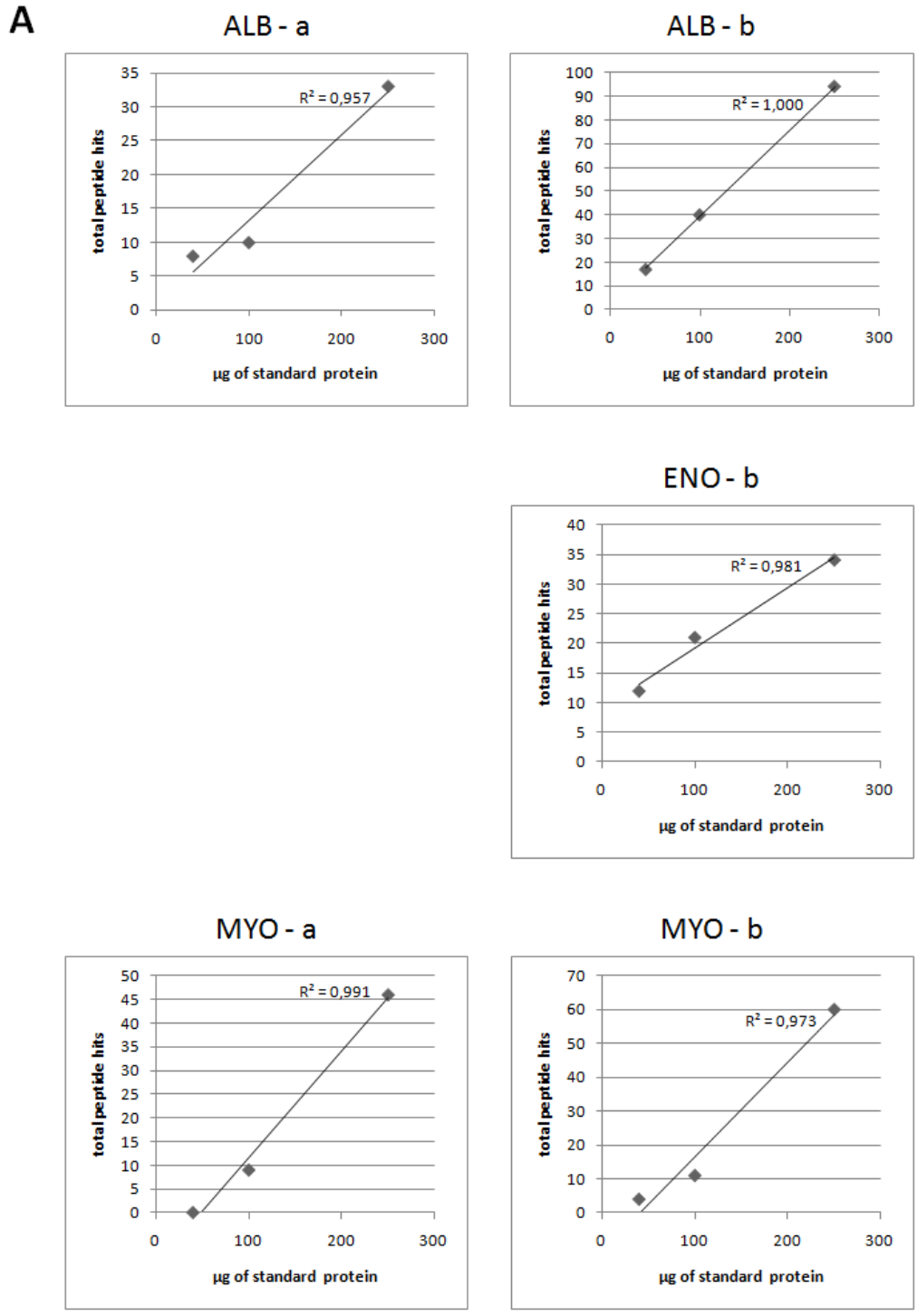

Figure 7.2. Panel A: Correlation between total peptide hits number and protein abundance for three protein standards, in a mixture with *E. coli* extract (a), or as pure proteins (b). R-squared values are reported for each plot.

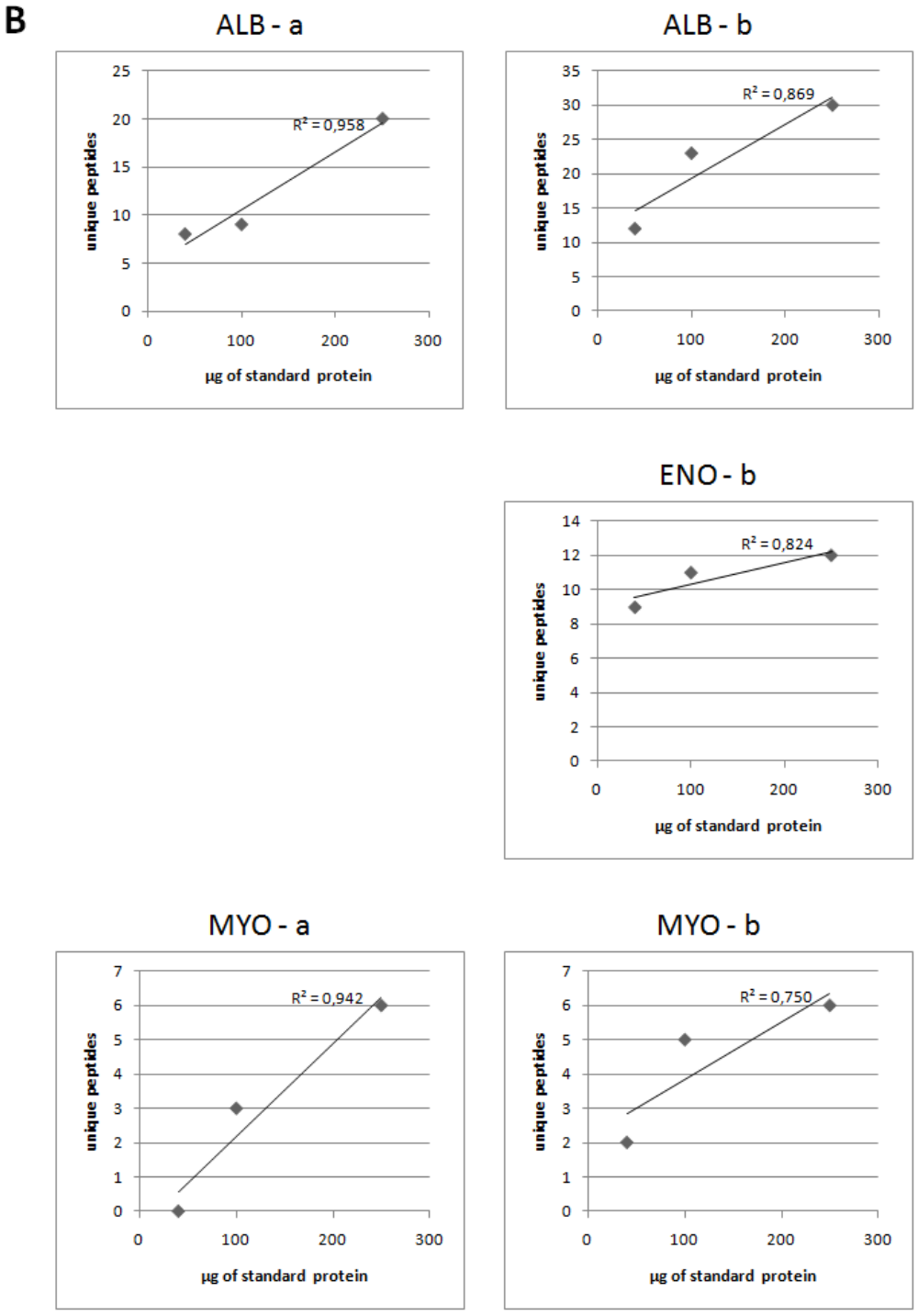

Figure 7.2. Panel B: Correlation between unique peptides number and protein abundance for three protein standards, in a mixture with *E. coli* extract (a), or as pure proteins (b). R-squared values are reported for each plot.

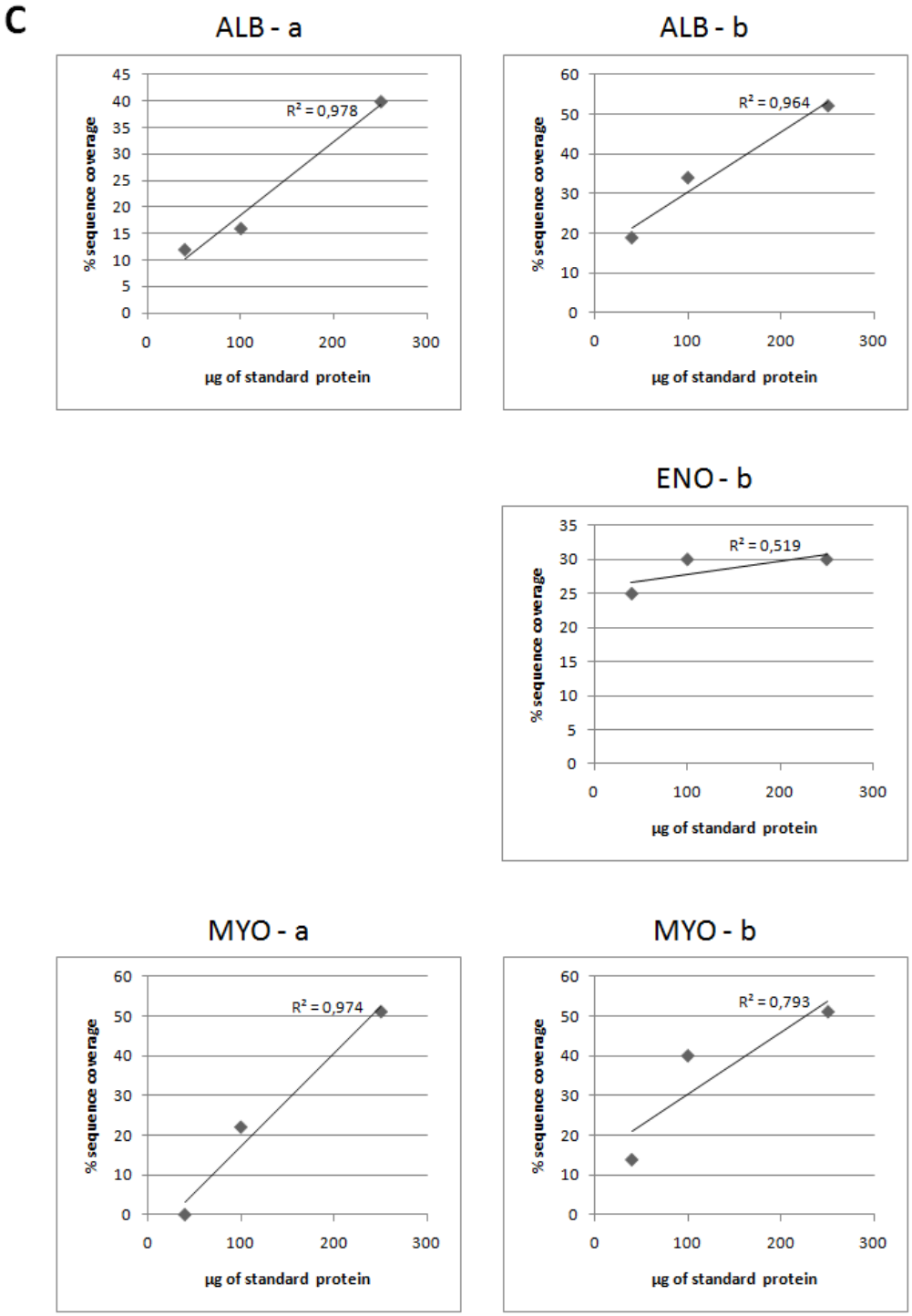

Figure 7.2. Panel C: Correlation between sequence coverage and protein abundance for three protein standards, in a mixture with *E. coli* extract (a), or as pure proteins (b). R-squared values are reported for each plot.

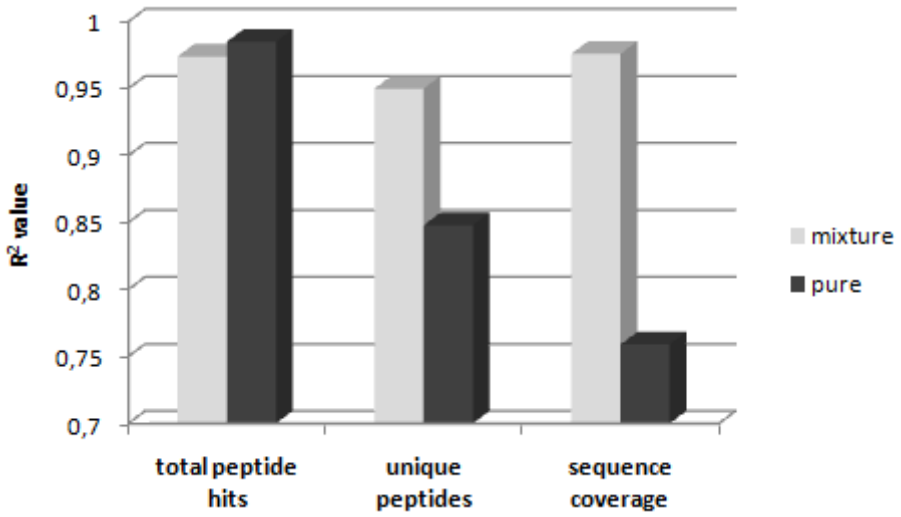

Figure 7.3. Histogram showing main R-squared values obtained for the three abundance index examined.

III. Discussion and concluding remarks

In view of the results obtained, even if referring to a not exhaustive survey, a clear, significant correlation between MS total peptide hits and protein abundance was revealed. Therefore, this parameter can be used as a valid (semi-)quantitative index for comparing differential expression of the same protein between two groups of samples analysed with a GeLC-MS/MS approach.

Enolase data demonstrate that analysis effectiveness largely depends on the context in which target protein is inserted. As can be noted in Figure 7.1, *E. coli* protein pattern shows a strongly intense band at the height of "pure" enolase band; probably, an ample difference in the order of magnitude between neighbouring bacterial highly expressed protein and enolase (present in a quantity not higher than 100 nanograms) hided enolase signal, preventing a correct identification, and quantification. As a consequence, it should not be reliable to compare with this method the abundance of different proteins. In fact, every gel area presents a proper protein profile, and so a dissimilar protein density, and this difference might impair a trustworthy comparison, even between two protein with an identical concentration. Nevertheless, an evaluation of total peptide hits of the same protein in two similar samples, being

surrounded by the same co-migrating proteins, should be surely more informative and allow a confident semi-quantitative analysis.

8. IDENTIFICATION OF LUNG NEUROENDOCRINE TUMOURS CANDIDATE BIOMARKERS BY GEL-BASED PROTEOMICS ANALYSIS OF FORMALIN-FIXED, PARAFFIN-EMBEDDED SAMPLES

I. Introduction

Lung neuroendocrine tumours (lung NETs) represent a spectrum of neoplastic tissues arising from neuroendocrine cells of lung epithelium, and comprise about 20% of all pulmonary cancers. Although sharing structural and morphological features, they are separated into four subgroups: typical carcinoid tumour (TC), atypical carcinoid tumour (AC), large-cell neuroendocrine carcinoma (LCNEC), and small-cell lung carcinoma (SCLC), which exhibit a considerably dissimilar biology [124]. The frequent mistakes in differential diagnosis between TC and SCLC have been described as one of the most important "pitfalls" in the management of lung cancers [125]. In combination with the histological appearance, a variety of markers, including Ki-67, chromogranin A (CgA), neuron-specific enolase (NSE), serotonin, synaptophysin, and adrenocorticotrophic hormone (ACTH), have some utility in establishing the differential diagnosis. However, a molecule with a satisfactory degree of sensitivity and specificity has still to be found. For instance, an increase of plasma CgA level is a relatively sensitive marker of lung carcinoids, but also it shows positivity also in about 60% of SCLC cases [124].

In the past few years, some gene expression studies comprising lung NETs were carried out, but no conclusive information about diagnostic or prognostic biomarkers was inferred [126-132]. As concerns proteomics, an in-depth systematic characterisation of these tissues has never been published so far; only a limited comparative 2D-PAGE survey has been conducted in 2005 by Cho *et al.* [133]. In fact, a thorough differential proteomic investigation, particularly comparing TC and SCLC protein patterns for biomarker discovery purposes, which might have a really significant diagnostic usefulness, has never been performed to date.

From the point of view of biomedical research, the largest hindrance in lung NETs is represented by their low frequency, especially as regards carcinoids [134]. Hence, the scarce availability of fresh biopsies makes extremely difficult to carry out a biomarker

discovery study, without the help of formalin-fixed, paraffin-embedded (FFPE) tissue banks, stored in hospitals worldwide. Our group recently optimised a novel FFPE protein extraction method, and described the application to these extracts of several proteomic techniques (as illustrated in previous chapters).

In view of this, a GeLC-MS/MS differential proteomic analysis was conducted on FFPE TC and SCLC samples, in order to molecularly characterise these tissues, and to identify a panel of putative protein biomarkers, to subsequently validate by means of immunological high-throughput techniques. Moreover, a 2D-DIGE-MS experiment was also performed, with the intention of confirming, or disproving, differential expression data achieved with the previous analytical approach.

II. Results

1. GeLC-MS/MS analysis leads to detect over 400 proteins for each kind of pathological tissue and to assess biological variability

In order to perform a differential analysis, six human FFPE bioptic samples, three diagnosed as typical carcinoids, and three as small cell lung carcinomas, were subjected to the protein extraction procedure previously developed in our laboratory. Two identical aliquots of each protein extract were separated by SDS-PAGE, and then the polyacrilamide gel were cut according one of the following methods: excision of the only 13 clearly visible bands (VB method), or fractionation of the whole lane into 38 homogeneous slices, independently from band intensities (AL method). Thus, a total of 306 gel portions were *in situ* digested with trypsin; tryptic peptides were extracted and analysed by LC-MS/MS; finally, MS identification data were processed, in order to obtain a complete proteomic profile for each probed extract. The complete experimental workflow of this study is illustrated in the flowchart of Figure 8.1.

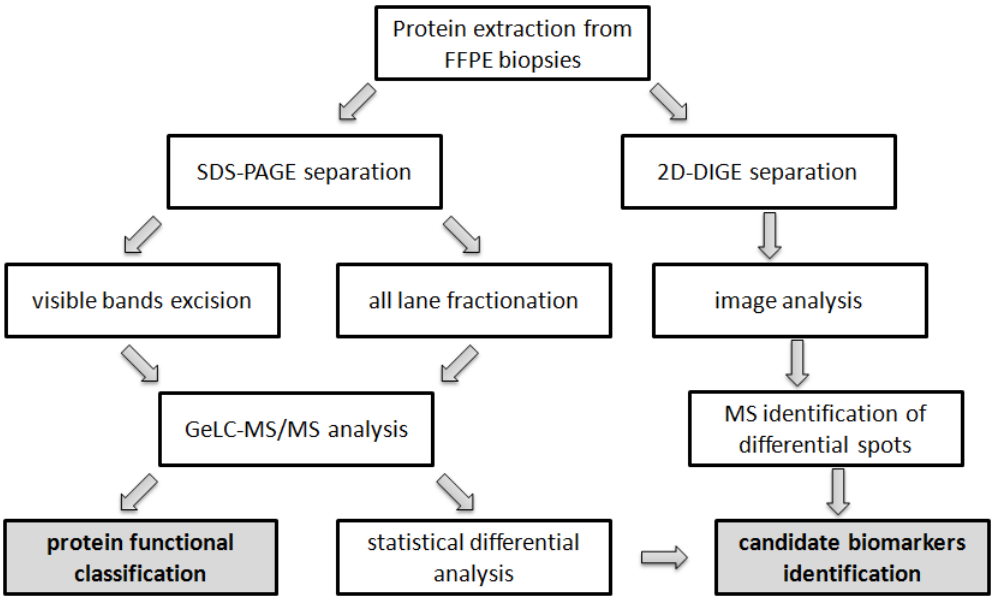

Figure 8.1. Experimental workflow.

Overall GeLC-MS/MS results were generated merging proteomic information of two replicates (obtained by VB and AL methods, respectively) for each sample, while total datum derives from combining profiles of single samples into classes. Complete lists of protein identifications, classified according to tumour subgroup, are presented in Appendices 8-9. As summarised in Table 8.1, a total of 420 and 442 unique proteins for TC and SCLC, respectively, and a mean of almost 240 proteins per sample, and 430 proteins per class, were detected, whereas total peptide hits value per sample was around 5000.

As concerns distribution of identified proteins among individual samples (see Table 8.2), no less than a half of total identified proteins belong only to a single sample. It might appear the consequence of a dramatic biological variability; actually, if considering mean peptide number per protein, it can be noted that they are often single-peptide proteins, whose “random” finding is largely caused by the coincidence between their low concentration and the detection limit of the instrument, rather than by a “real” varied abundance among samples. As a confirm of this, protein and peptide percentages in function of the number of sample in which they were detected highlight an opposite trend: common proteins are generally the most abundant, whilst those detected only in a little fraction of samples exhibit a low abundance. Thus, except for particular cases (such as hyaluronan and proteoglycan link protein 1, highly

expressed in a SCLC sample and absolutely undetected in all the others), variability seems to be much more technical than biological.

Table 8.1. Summary of GeLC-MS/MS data, according to sample classes

Sample	Unique proteins	Total peptide hits
TC1	224	5624
TC2	218	5146
TC3	261	4981
mean TC	*234*	*5250*
total TC	*420*	*15751*
SCLC1	262	5685
SCLC2	155	4709
SCLC3	295	3815
mean SCLC	*237*	*4736*
total SCLC	*442*	*14209*
mean NET	*236*	*4993*
total NET	*616*	*29960*

Table 8.2. Distribution of proteins and peptides, according to the number of samples in which they were identified.

No. of samples	Unique proteins	Total peptide hits	Average peptides per protein	% Proteins	% Peptides
1	305	884	3	50	3
2	97	578	3	16	2
3	78	766	3	13	3
4	49	1076	5	8	4
5	36	2204	12	6	7
6	51	24452	80	8	82

On the other hand, 51 proteins are shared by all samples (listed in Appendix 10); together with a few "housekeeping proteins", also forms belonging to distinctive tumoral pathways proteins can be identified, as will be illustrated in next paragraphs.

Figure 8.2 shows sharing and intersection of protein identifications among samples belonging to the same tumour subgroup. 102 and 79 proteins can be considered peculiar of TC and SCLC, respectively, since were detected in all cases of their proper disease type. Combining these data, identities of proteins always detected in a class and never in the other were found: 11 proteins were detected in all TC cases and in none of SCLCs, and only 3 proteins were detected in all SCLC cases and in none of TCs.

These 14 proteins, which can be more consistently judged as differentially expressed, are listed in Table 8.3.

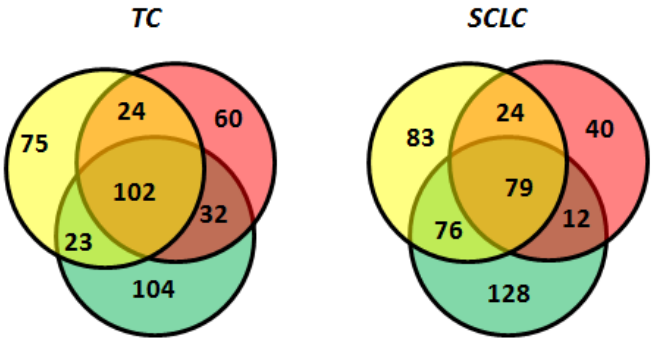

Figure 8.2. Venn diagrams illustrating distribution of identified proteins among the three samples of the same tumour subtype.

Table 8.3. List of differentially expressed protein identifications.

Accession no.	Protein	TC TPH	SCLC TPH
P10645	Chromogranin-A	66	0
Q8IWL2	Pulmonary surfactant-associated protein A1	15	0
P00441	Superoxide dismutase [Cu-Zn]	9	0
P08294	Extracellular superoxide dismutase [Cu-Zn]	8	0
P02794	Ferritin heavy chain	7	0
O60888	Protein CutA	7	0
P30044	Peroxiredoxin-5, mitochondrial	6	0
P02766	Transthyretin	6	0
Q96IU4	Abhydrolase domain-containing protein 14B	5	0
P07988	Pulmonary surfactant-associated protein B	5	0
P18669	Phosphoglycerate mutase 1	4	0
P16401	Histone H1.5	0	39
P09651	Heterogeneous nuclear ribonucleoprotein A1	0	27
Q13242	Splicing factor, arginine/serine-rich 9	0	9

2. "All lane" fractionation method assures a significant enlargement of proteomic data set, especially for low-abundant proteins

In a previous work, a GeLC-MS/MS investigation of FFPE sheep muscle extract was conducted, by cutting only band visibile after colloidal Coomassie staining. However, in our experience protein identification is achievable also analysing gel regions which does not show a visible signal; this brought to hypothesise that subjecting to a GeLC-MS/MS analysis gel slices excised from the whole lane might improve quantitatively

and qualitatively proteomic data. Thus, as secondary aim of this work, a comparison of visible band (VB) and all lane (AL) fractionation methods data was conducted.

Results, summarised in Table 8.4 and in Figure 8.3, demonstrate that also "transparent" gel regions can provide precious information, particularly regarding to low-abundant proteins; actually, over fifty percent of identified proteins were detected only by means of this fractionation procedure. It is a remarkable outcome, not forgetting that a "strange" matrix, as a FFPE tissue, was examined. Nevertheless, VB replicates allow to exclusively find 62 proteins; on account of this, both fractionation replicates were considered in GeLC-MS/MS results.

Table 8.4. Comparison of identification data between different fractionation methods.

Method	Data	Unique proteins	Total peptide hits	Average peptides per protein
VB	Total	286	6164	22
	Average per exp	108	1027	10
AL	Total	554	23796	43
	Average per exp	202	3966	20

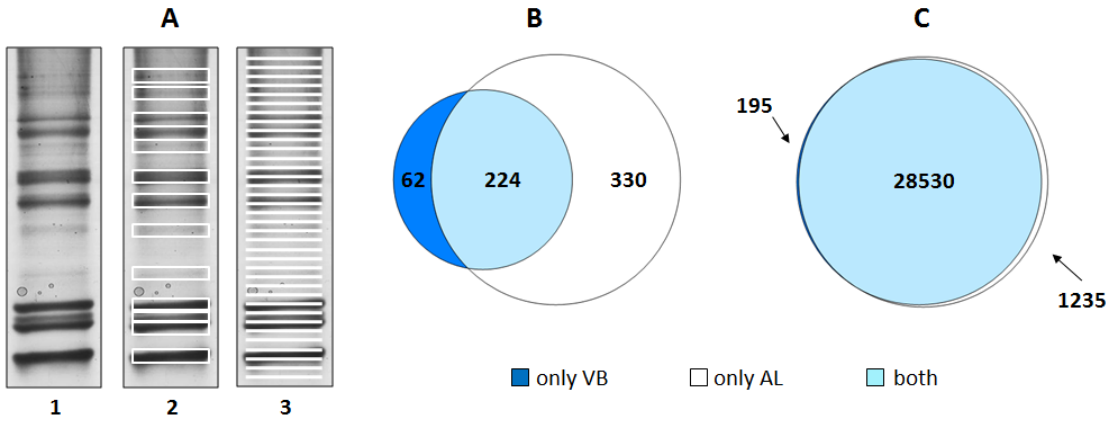

Figure 8.3. Panel A: Representative 1D-PAGE profile of a FFPE lung NET extract (lane 1). Illustration of fractionation methods by visible bands cutting (VB, lane 2), and all lane cutting (AL, lane 3). Panel B: Venn diagram showing distribution of proteins identified with both fractionation methods (light blue), only with VB (dark blue), or only with AL (white). Panel C: Venn diagram showing distribution of peptides corresponding to proteins identified with both fractionation methods (light blue), only with VB (dark blue), or only with AL (white).

3. Main tissues proteins form a sort of "background noise" in a high percentage of gel regions

As reported in our previous works, FFPE extracts seem to contain protein artificial complexes which co-migrate with restored "normal" proteins. These complexes appear to be composed by a combination of the most expressed proteins of the tissue. Also for human FFPE lung NETs extracts, GeLC-MS/MS analysis demonstrated this phenomenon: in fact, the overall protein abundance (according to total peptide hits, for all the bands examined) correlates quite well with the percentage of bands in which the protein was identified, as shown in Figure 8.4. For example, beta subunit of hemoglobin, which exhibits also the highest total peptide hits value, was found in nearly 95% of gel slices, also in the highest molecular weight ones (see Table 8.5). The most widely represented proteins are globins, histone isoforms and serum albumin. Anyhow, this "protein background" does not impair the detection of co-migrating low-abundant proteins.

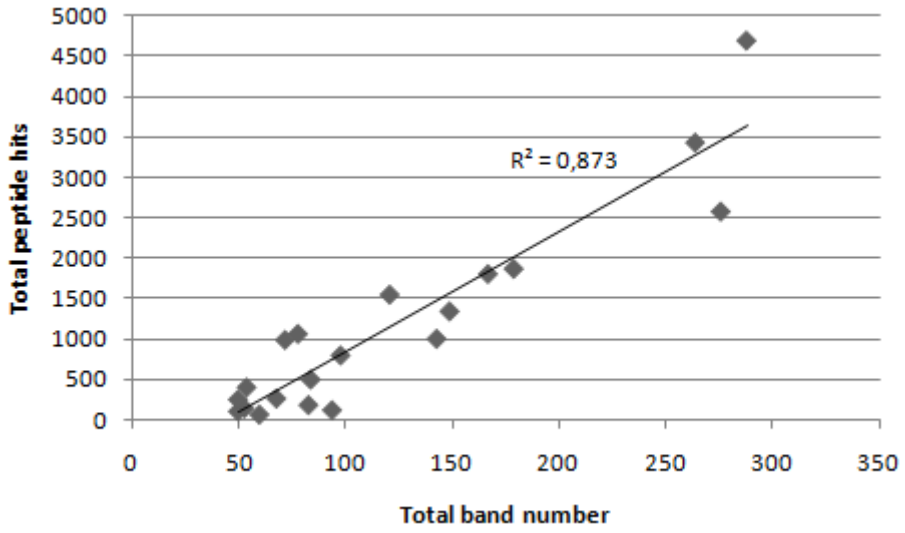

Figure 8.4. Plot indicating correlation between protein total peptide hits and number of bands in which protein was detected. The first twenty most represented proteins were considered.

Table 8.5. List of the five most represented proteins along gel bands. Total peptide hits in all samples analysed are also indicated.

Protein	No. of bands	% bands	Total peptide hits
Hemoglobin subunit beta	288	94,1	4693
Histone H4	276	90,2	2582
Hemoglobin subunit alpha	264	86,3	3434
Actin, cytoplasmic 1	179	58,5	1875
Serum albumin	167	54,6	1811

4. Gene ontology analysis allows to characterise functional features of TC and SCLC proteomes

Apart from a few gene expression studies, an evident lack of molecular information about lung neuroendocrine tumours has to be noticed. Small cell carcinoma has been seldom considered in the proteomic investigations described so far, whilst to our knowledge lung carcinoids were never subjected to a systematic proteomic survey. Thus, an in-depth classification of these tissues, following Gene Ontology parameters, has a reasonable degree of novelty. Pie charts in Figures 8.5 and 8.6 show percentage protein distribution for molecular function, biological process, and cellular localisation. Comparing functional data of tumour subgroups, as described in the following histogram (Figures 8.7), wider differences in protein distribution concern mRNA splicing factors, protein kinases (more abundant in SCLC), and oxidoreductases (present in a larger number in TC). Moreover, TCs show a larger amount of extracellular proteins, whereas SCLCs of intracellular proteins. Finally, almost 50% of proteins detected in both tissues are encoded by genes which are known to be expressed in brain tissue; this datum is consistent with neuroendocrine origin of this kind of cancers.

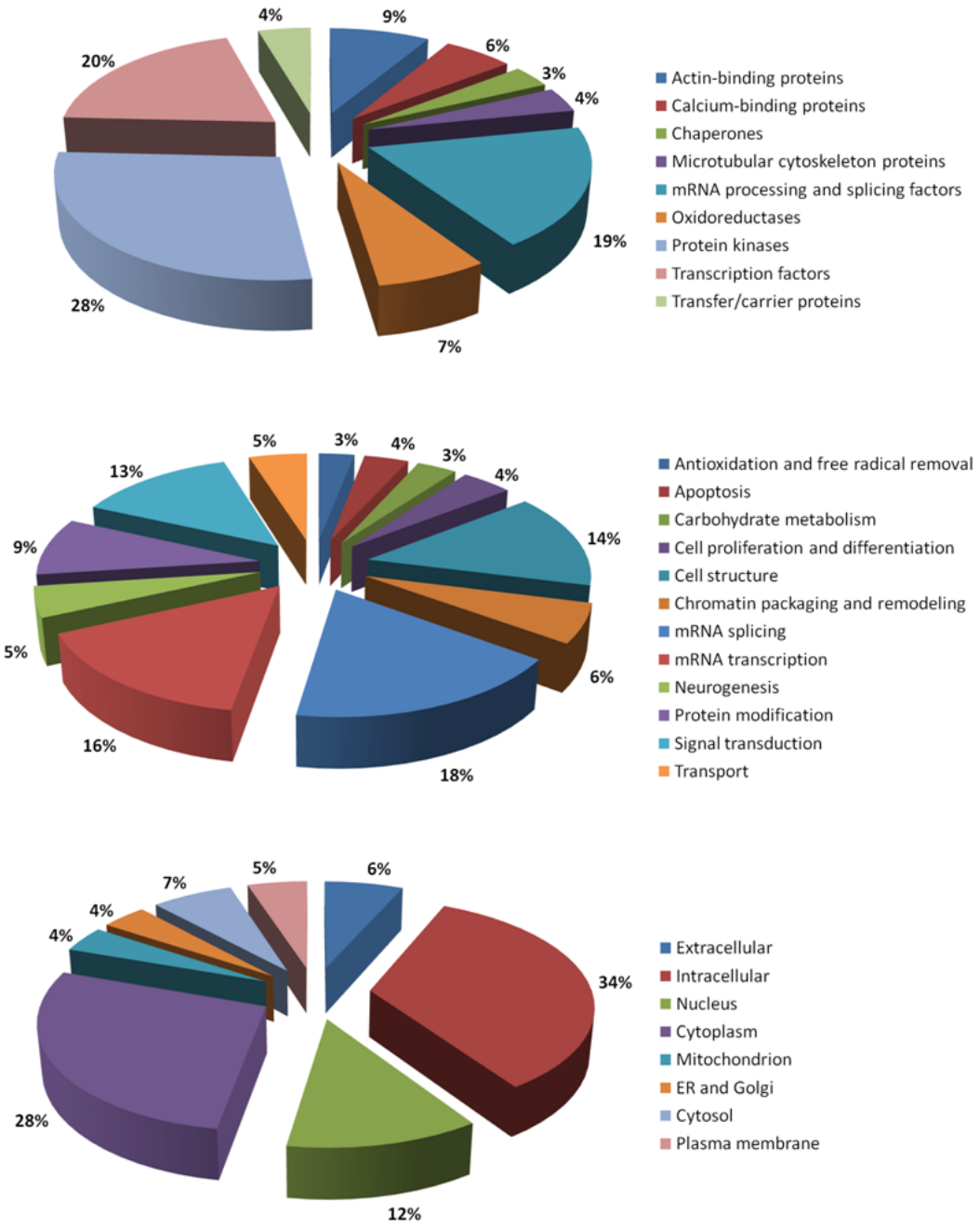

Figure 8.5. Pie charts representing distribution of TC proteins according to molecular function (top), biological process (middle), and cellular localisation (bottom).

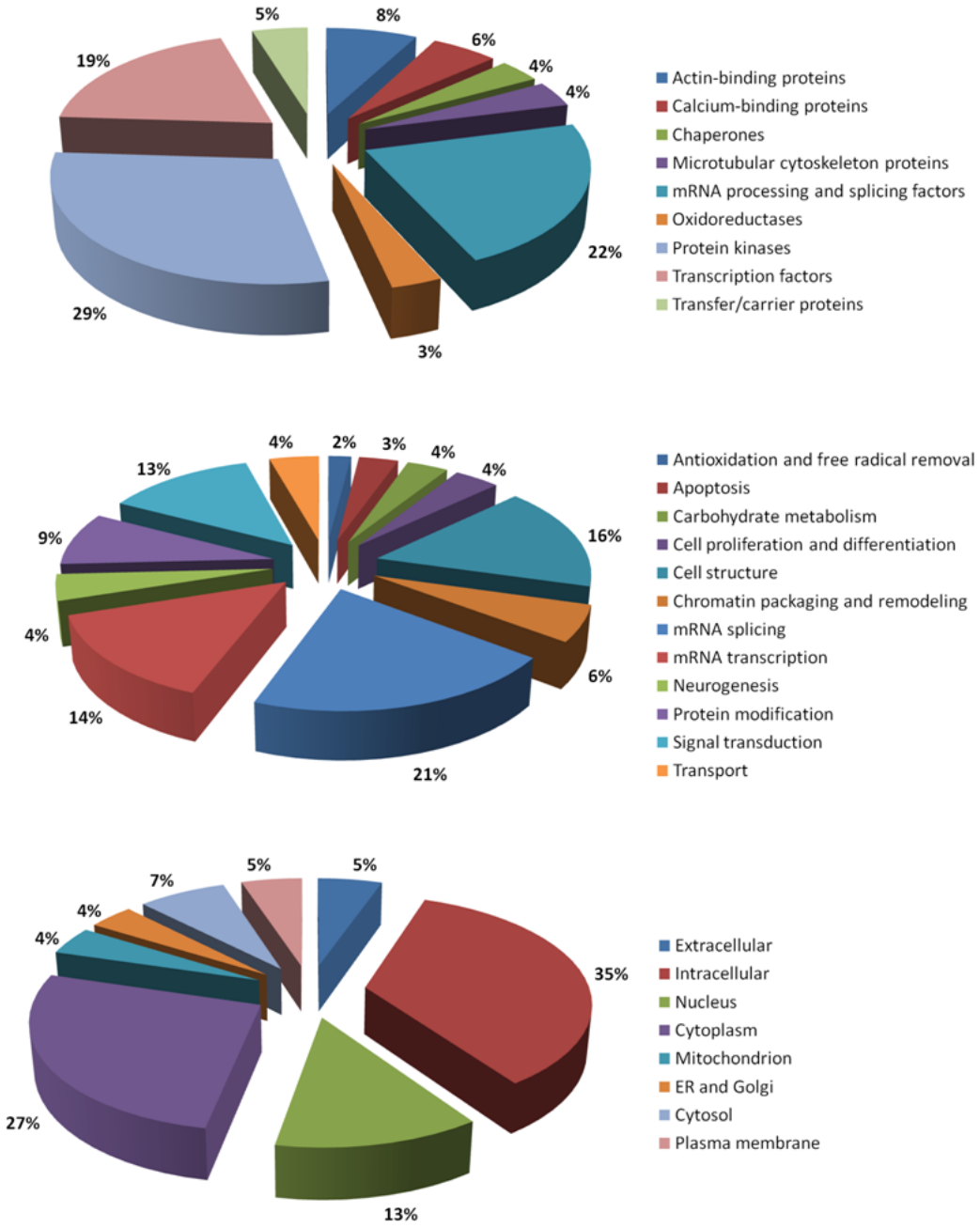

Figure 8.6. Pie charts representing distribution of SCLC proteins according to molecular function (top), biological process (middle), and cellular localisation (bottom).

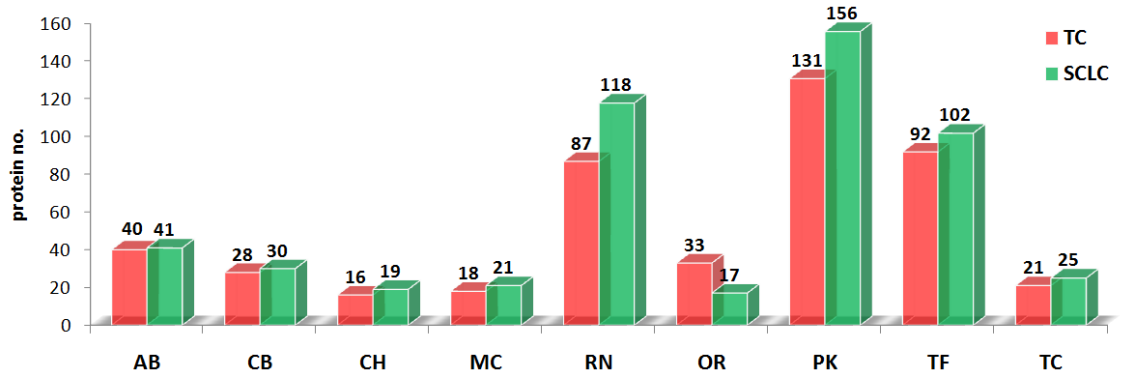

Figure 8.7. Histogram comparing number of TC and SCLC proteins belonging to the main molecular function classes. AB, actin-binding proteins; CB, calcium-binding proteins; CH, chaperones; MC, microtubular cytoskeleton proteins; RN, mRNA processing and splicing factors; OR, oxidoreductases; PK, protein kinases; TF, transcription factors; TC, transfer/carrier proteins.

5. Several protein families seem to exhibit a differential expression abundance between TC and SCLC

Total number of peptide hits was used in this work as a protein abundance index, according to spectral count label-free protein quantitation approach. A demonstration of linear correlation between protein quantity and peptide hits for GeLC-MS/MS analyses was described in chapter 7.

Considering variation in total peptides number among protein functional classes between TC and SCLC, RNA-binding proteins and microtubular cytoskeleton proteins exhibit a significant overexpression in SCLC, while calcium-binding proteins and iron homeostasis proteins are more abundant in TC (Fig. 8.8). More deeply, comparing peptide hits distribution in some specifically expressed protein families, it can be pointed out that peroxiredoxins, annexins, S100 proteins and transgelins seem more "typical" of TC, whilst plenty of hnRNPs and splicing factors appears as a peculiar signature of SCLC, as indicated in Figure 8.9. Furthermore, tubulins, histones, hnRNPs and splicing factors are present with a wider variability and abundance of isoforms in SCLC samples; all identified members belonging to these protein families are listed in Appendix 11.

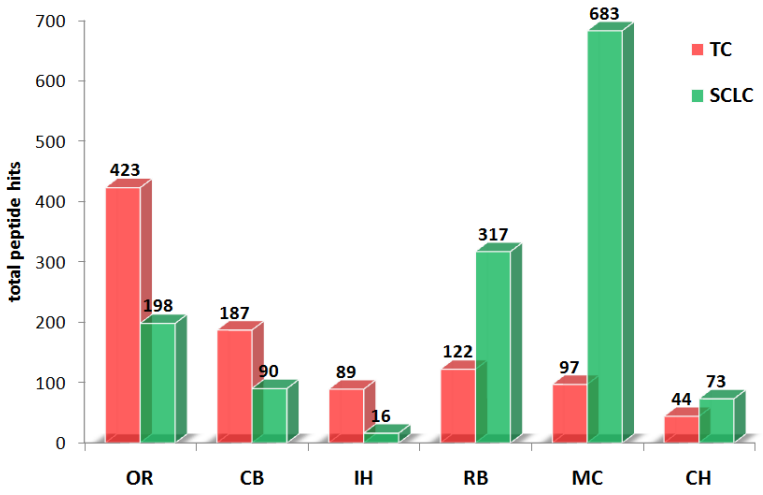

Figure 8.8. Histogram comparing total peptide hits of proteins belonging to some peculiar protein functional classes. OR, oxidoreductases; CB, calcium-binding proteins; IH, iron homeostasis proteins; RB, RNA-binding proteins; MC, mucrotubular cytoskeleton proteins; CH, chaperones.

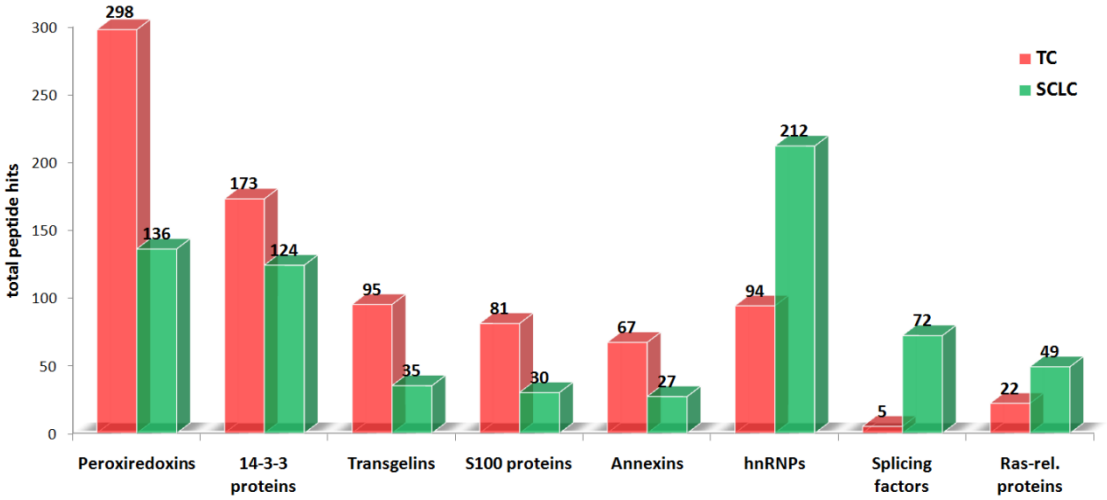

Figure 8.9. Histogram comparing total peptide hits of proteins belonging to main differentially distributed protein families.

6. Statistical analysis provides a list of 35 proteins significantly differentially expressed, which may represent potential disease biomarkers

The final purpose of this work was to identify a panel of putative biomarkers, which could help in lung NETs differential diagnosis. As semi-quantitative parameter, total number of peptide hits was used, as stated above. With the purpose of identifying a set of statistically-supported, robust, and authentically differentially expressed candidate biomarkers, a non-parametric Kruskal-Wallis test was performed. The task

of this statistical "filter" was to select proteins homogeneously abundant among the samples of a disease class, and absent (or significantly, evenly underexpressed) within the other subgroup. Proteins which fell into a 95% confidence interval are displayed in Table 8.6.

Table 8.6. List of proteins identified as differentially expressed by means of statistical analysis.

Entry name	Protein name	TC TPH	SCLC TPH	> expressed in
ABHEB_HUMAN	Abhydrolase domain-containing protein 14B	5	0	TC
ANXA5_HUMAN	Annexin A5	49	17	TC
CMGA_HUMAN	Chromogranin-A	66	0	TC
CUTA_HUMAN	Protein CutA	7	0	TC
EF1A1_HUMAN	Elongation factor 1-alpha 1	4	40	SCLC
FRIH_HUMAN	Ferritin heavy chain	7	0	TC
H12_HUMAN	Histone H1.2	18	135	SCLC
H15_HUMAN	Histone H1.5	0	39	SCLC
H2A1D_HUMAN	Histone H2A type 1-D	138	859	SCLC
H31T_HUMAN	Histone H3.1t	104	404	SCLC
H4_HUMAN	Histone H4	725	1857	SCLC
HBA_HUMAN	Hemoglobin subunit alpha	2580	854	TC
HBB_HUMAN	Hemoglobin subunit beta	3459	1234	TC
HBD_HUMAN	Hemoglobin subunit delta	929	142	TC
HNRPC_HUMAN	Heterogeneous nuclear ribonucleoproteins C1/C2	13	29	SCLC
HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	6	24	SCLC
PGAM1_HUMAN	Phosphoglycerate mutase 1	4	0	TC
PRDX5_HUMAN	Peroxiredoxin-5, mitochondrial	6	0	TC
PRDX6_HUMAN	Peroxiredoxin-6	19	3	TC
PSPB_HUMAN	Pulmonary surfactant-associated protein B	5	0	TC
RL18_HUMAN	60S ribosomal protein L18	2	12	SCLC
RL23A_HUMAN	60S ribosomal protein L23a	1	13	SCLC
ROA1_HUMAN	Heterogeneous nuclear ribonucleoprotein A1	0	27	SCLC
RS26L_HUMAN	Putative 40S ribosomal protein S26-like 1	1	9	SCLC
RS4X_HUMAN	40S ribosomal protein S4, X isoform	1	7	SCLC
S10A8_HUMAN	Protein S100-A8	15	1	TC
SFRS9_HUMAN	Splicing factor, arginine/serine-rich 9	0	9	SCLC
SFTA1_HUMAN	Pulmonary surfactant-associated protein A1	15	0	TC
SODC_HUMAN	Superoxide dismutase [Cu-Zn]	9	0	TC
SODE_HUMAN	Extracellular superoxide dismutase [Cu-Zn]	8	0	TC
STMN1_HUMAN	Stathmin	1	34	SCLC
TAGL2_HUMAN	Transgelin-2	46	22	TC
TBA1A_HUMAN	Tubulin alpha-1A chain	37	223	SCLC
TBB5_HUMAN	Tubulin beta chain	14	155	SCLC
TTHY_HUMAN	Transthyretin	6	0	TC

Besides not significative elements (globins, for instance), and well-known neuroendocrine markers (such as chromogranin A), an interesting list of proteins, which comprises superoxide dismutase, protein Cut A, peroxiredoxin-5, and protein

S100-A8, may represent a potential markers panel useful for specifically diagnosing TC. On the other hand, among SCLC candidate biomarkers, proteins like hnRNP A1, splicing factor arginine/serine-rich 9, and stathmin were discovered. Obviously, trustworthiness of differential expression has now to be validated with complementary approaches, such as immunohistochemistry, on a wider gamut of bioptic samples.

7. 2D-DIGE-MS/MS analysis confirms several candidate biomarkers previously identified by GeLC-MS/MS

As a final step, also a 2D-DIGE analysis was performed, with the aim of comparing TC and SCLC patterns. Two FFPE samples (different from those analsyed by GeLC-MS/MS), were subjected to protein extraction, protein precipitation (following a standard TCA/acetone method), CyDye labelling, and 2D-DIGE analysis, as described in *Experimental activities* section.

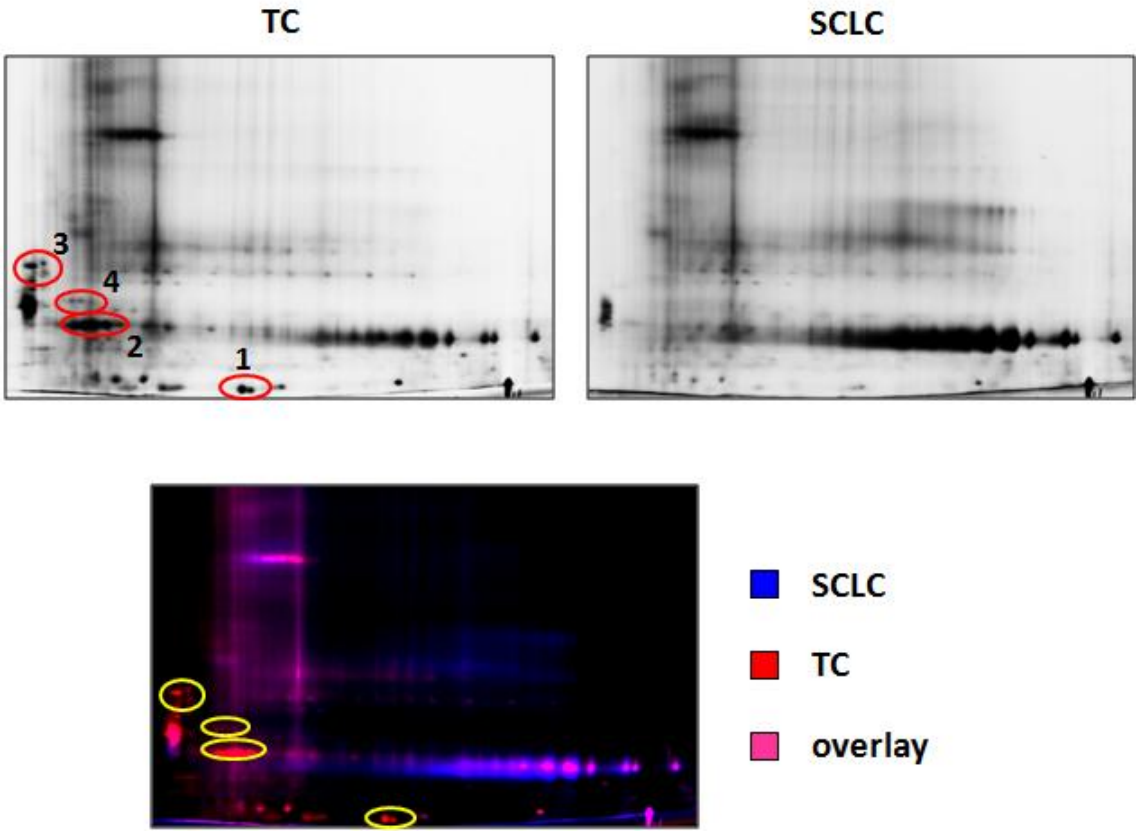

Figure 8.10. 2D-DIGE of TC and SCLC FFPE samples (upper, single channel images; lower, overlay image). Picked and MS-analysed spots are marked with red circles and indicated by numbers.

Image analysis with DeCyder software highlighted four spots, as shown in Figure 8.10, which exhibited a more intense signal in typical carcinoid 2D-map. Such spots were picked, digested with trypsin, and their derivate peptide mixtures were analysed by LC-MS/MS. MS identifications are in greater part consistent with those found by GeLC-MS/MS approach (as summarised in Table 8.7). In fact, some members previously inserted in Table 8.6 were detected, as chromogranin A (in truncated forms), protein CutA, and annexin A5. Moreover, a lot of proteins formerly identified in TC extracts, and presenting a higher peptides number in comparison with SCLC, were identified, such as S100 protein A11 and A6 isoforms, Peroxiredoxin-2, and ProSAAS.

Table 8.7. List of MS identifications from 2D-DIGE spots. Proteins stated as overexpressed in GeLC-MS/MS data are shaded; newly identified proteins are in bold type; probable truncated forms are asterisked.

Spot no.	Accession no.	Protein	MW	pI	Mowse score	Queries matched	Unique peptides	Sequence coverage (%)
1	Q9NP97	Dynein light chain roadblock-type 1	10915	6,58	74	1	1	12
	P14174	Macrophage migration inhibitory factor	12639	7,74	61	1	1	9
	P31949	Protein S100-A11	11847	6,56	61	1	1	8
	P06703	Protein S100-A6	10230	5,33	49	4	2	16
	P61604	10 kDa heat shock protein, mitochondrial	10925	8,89	44	1	1	9
	A6NHG4	D-dopachrome decarboxylase-like protein	14414	5,89	44	1	1	5
	P14854	Cytochrome c oxidase subunit 6B1	10414	6,54	41	1	1	9
2	P10645	Chromogranin-A*	50829	4,58	544	16	3	9
	P08758	Annexin A5*	35971	4,94	103	1	1	5
	O60888	Protein CutA	19218	5,42	85	1	1	7
	Q15121	**Astrocytic phosphoprotein PEA-15**	15088	4,93	68	1	1	7
	P12277	Creatine kinase B-type *	42902	5,34	59	1	1	2
	P52907	F-actin-capping protein subunit alpha-1*	33073	5,45	57	1	1	3
	P32119	Peroxiredoxin-2	22049	5,66	48	1	1	5
	Q15370	**Transcription elongation factor B polypeptide 2**	13239	4,73	46	2	2	18
3	P10645	Chromogranin-A*	50829	4,58	190	4	3	15
4	Q9UHG2	ProSAAS	19077	4,86	343	5	5	25
	Q9Y2B0	**Protein canopy homolog 2**	20981	4,81	100	1	1	8
	P10645	Chromogranin-A*	50829	4,58	81	3	2	4

III. Discussion and concluding remarks

After optimisation of FFPE extraction method, and its application on model sheep tissues (see Chapters 4-6), the following challenge was to perform a complete differential proteomics study on human bioptic fixed samples. Efficacy, suitability, and reproducibility of extraction procedure were widely confirmed by homogeneity, consistency, and overall quality of GeLC-MS/MS data. Taking into account the complexity of FFPE matrix, the amount of proteins and peptides detected, the sensitivity level of MS instrumentation used, and the high similarity between TC and SCLC, the achieved results are very satisfactory: a plethora of specific and comparative proteomic data were obtained, and a significative panel of candidate biomarkers to be validated were generated.

Label-free GeLC-MS/MS application for biomarker discovery exhibits several *pros* and *cons*. SDS-PAGE prefractionation allows to decrease sample complexity, and, unlike two-dimensional separation, information about proteins with extreme p*I*s is widely preserved; moreover, performing a LC-MS/MS analysis starting from full-length electrophoresed proteins facilitates correct peptide matching, increasing identification significance. On the other hand, compared to classical "shotgun" proteomics experiments, initial gel separation may deplete extracts of the highest- and/or lowest molecular weight proteins; in addition, an uneliminable variability in band picking and peptide extraction has to be considered; finally, label-free quantitation (at least for the simplified form utilised here) must be regarded as a (good) abundance estimation method, rather than an actual quantitation (see Chapter 7).

Proteomic profiling can undoubtedly aid the comprehension of lung neuroendocrine tumours molecular features; in particular, data about proteins (and protein families) differentially expressed between carcinoid and small cell carcinoma might open the way to a more in-depth characterisation of biological processes and molecular patways in which they are involved, such as cellular redox balance regulation, calcium homeostasis maintenance, mRNA splicing and processing, and so on.

As concerns identified putative biomarkers, several of them has been already described as correlated to neoplastic diseases. Overexpression of various isoforms of

heterogeneous nuclear ribonucleoproteins, included A1, was demonstrated for different kinds of tumour [135-138]; it is still debated whether this mRNA-binding proteins increase represents a driving force for, or a consequence of, cell cycle deregulation [139]. Some variants of arginine/serine rich splicing factors were identified as receptors for lung colonisation of cancer cells [140]. A few surveys indicate a correlation between stathmin expression and cancer development, in neuroendocrine types as well [141-142]. Annexin A5, unlike other members of its super-family, has been used as a probe to mesaure apoptosis *in vivo* and *in vitro*, also in tumoral tissues, but has not been still directly related to cancer development [143]. Among oxidoreductases, peroxiredoxins are expressed by several normal and tumoral tissues; especially, type 6 is typical of lung (both healthy and diseased) [144-147]. Superoxide dismutase isoforms are present at heterogeneous ratios in cancer tissues, mostly in the lung, as a consequence of its constant exposition to oxygen [148-149]. As regards calcium-binding proteins, S100 protein family members have been variously linked to tumours; S100 A8 were detected in colorectal, breast, and gastric cancer tissues [150-151]. Other proteins are more consistent with a neuroendocrine differentiation, such as transthyretin and protein CutA. Transthyretin plays a crucial role in senile systemic amyloidosis development [152]. Little is known about protein CutA, apart from its role in acetylcholinesterase secretion pathway [153]. Finally, transgelin was suggested as a tumour suppressor, but recently was demonstrated by proteomic investigations to be overexpressed in lung and gastric adenocarcinomas [154-156].

Even though being not an ample and robust investigation by itself, the 2D-DIGE experiment included in this work has the merit of supporting GeLC-MS/MS data, and demonstrates that also 2D-DIGE can be useful for biomarker discovery in FFPE tissues, at least as a complementary approach.

In conclusion, this study provides a global molecular characterisation of neuroendocrine tumours at a proteomic level, and makes available a list of assorted candidate biomarkers, useful for lung NETs differential diagnosis. The further objective in this research field will be to validate the most promising putative biomarkers, by means of immunological approaches, such as immunohistochemistry, on a wider

number of samples (including also intermediate subgroups, namely, atypical carcinoid and large cell carcinoma). On the other hand, carrying out a complementary transcriptomic investigation (by cDNA microarrays, in situ hybridisation, or qPCR) might represent a significant widening of the study, and a decisive corroboration of the results obtained in proteomics.

9. GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

I. General conclusions

A wealth of statistically supported information on disease is encased within FFPE tissue repositories. Recently, scientists have been trying to access this "locked proteome" with the development and application of various protein extraction strategies; however, so far, the spectrum and the reproducibility of downstream applications have been limited by yield and quality issues affecting proteins extracted from the formalin-fixed tissue matrix. The full-length protein extraction method described here is versatile, reliable, amenable to different analytical platforms, such as 1-D and 2-D PAGE, GeLC-MS/MS, 2D-PAGE-MS, 2D-DIGE-MS, 1-D and 2-D western blotting, reverse phase protein arrays, and ELISA.

Recent achievements in the field of proteomic characterization of FFPE tissues, mostly limited to gel-free shotgun approaches, the application of gel-based proteomics (i.e. GeLC-MS/MS, 2D-PAGE-MS, and 2D-DIGE-MS) to these fixed protein matrices had never been reported in a comparative fashion with matched fresh tissue samples. Rather, past efforts in this direction produced discouraging results. Nevertheless, the ability to apply the investigational power of gel-based proteomics to FFPE tissue repositories would enhance our knowledge on the proteomic repertoire of disease, and deserves further efforts. In this study, the applicability of 2D-PAGE-MS and 2D-DIGE-MS to FFPE tissues was demonstrated by using matched animal tissue model samples. This approach led to obtain satisfactory amounts of fresh tissues for comparison, and also to eliminate variability factors such as differences in fixing times, individual differences in samples, anatomical or regional heterogeneity, and other possible sources of variability, in FFPE samples. 1-D and 2-D FFPE protein profiles show a high reproducibility and a satisfactory quality, and MS identifications are largely comparable with fresh-frozen extracts. Moreover, DIGE and MS characterisation of FFPE samples allowed the discovery, or the confirmation, of a few peculiar chemical features of fixed proteins, such as pI acidic shift, artificial complexes existence and

their electrophoretical behaviour, and influence of lysine percentage in crosslinks formation and/or restoration.

Implications are noteworthy under several aspects. Surgical biopsies are extremely valuable for investigating disease processes, but require highly invasive procedures and cannot be collected for this purpose only. On the other hand, the production of high quality protein extracts from archival FFPE tissues allows not only to access a large number of diseased tissues, but also to recover samples collected along prolonged time intervals and for which the clinical course has been established. Furthermore, proteomic profiling of the immunohistochemistry diagnosis "leftovers" opens the way to statistically supported studies on low-incidence diseases, such as rare tumour forms. In addition, the extremely stable nature and biological safety of FFPE tissues greatly facilitates long distance exchange of samples for performing multicentric studies.

Once a robust and reliable FFPE proteomic analysis workflow was developed and validated, it was exploited for carrying out a biomarker discovery investigation in the field of lung neuroendocrine cancer. Differential diagnosis of these diseases represents often an arduous task for pathologists, because of the lack of specific and sensitive markers which can facilitate, and make unfailing, discrimination between several tumour subgroups. In view of such a need, the differential proteomics survey illustrated in this work, and the successful identification of a panel of putative diagnostic and/or prognostic biomarkers, may have crucial repercussions in terms of clinical significance and analytical usefulness. Moreover, given the want of proteomics study which elucidate molecular and functional profile of lung NETs, specifically for lung carcinoids, proteomic information provided here is a noteworthy contribution for the understanding of these diseases, and may open the way to in-depth investigations of signal transduction pathways and biological processes involved in neuroendocrine differentiation and cancer development.

II. Short-term perspectives

With regard to the "methodologic" issue of protein extraction and characterisation from FFPE tissues, a further challenge might be to evaluate correlation between fixation time, or bioptic section "age", and proteomic profile quality by 2D-PAGE and/or 2D-DIGE. Similar surveys were recently performed with shotgun proteomics approaches [96]; however, actual effects of fixation-related factors on full-length protein electrophoretic profile may be quite different, and their description would be a very interesting topic.

As extensively discussed in the previous sections of this writing, the following step of this biomarker discovery research will be candidate biomarkers validation. By means of immunohistochemistry, or other immunological high-throughput techniques, protein expression will be evaluated within a wide gamut of neuroendocrine tumour samples, in order to point out positive/negative correlation between antigen abundance (or cellular localisation) and cancer malignancy, therefore assigning to the marker a diagnostic or prognostic value. Also quantitative analysis of transcripts, with the help of *in situ* hybridisation, and/or real time RT-PCR, may contribute to strengthen proteomic evidences from a complementary point of view.

III. Long-term perspectives and future research directions

Another interesting issue to deal with, if owning a highly significative number of lung NETs cases, would be to carry out a transcriptomic biomarker discovery study by cDNA microarrays. It may provide a robust complementary data set, which would give an actual contribute to the research.

Furthermore, proteins belonging to the molecular pathways which appear to be involved in neuroendocrine differentiation might be subjected to functional studies, with both proteomics and genomics methods. For instance, their post-translational modifications may be studied, or relative ratios of protein isoforms may be determined.

Finally, once concluded validation phase, the most reliable candidate biomarkers might be searched for in patients' sera. This is the final purpose of every cancer biomarker discovery project; in fact, the possibility to make a lung tumour diagnosis simply taking a sample of blood would be an extraordinary and remarkable result.

ADDENDUM

Other experimental approaches which were applied or taken into account during the doctorate:

- genotyping by mass-spectrometry techniques (MALDI-MS detection of oligonucletotides and DNA restriction fragments)
- genotyping by DNA microarrays (by Illumina bead-arrays system)
- gene expression analysis by cDNA microarrays (by Illumina bead-arrays system)

ACKNOWLEDGMENTS

REFERENCES

1 Vasan RS. Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation* 2006; 113:2335-2362.

2 Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther*. 2001;69:89-95.

3 Fox N, Growdon JH. Biomarkers and surrogates. *Neuro Rx*. 2004;1:181.

4 Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology* 2006, 24(8):971–983.

5 Pepe MS, Etzioni R, Feng Z, Potter JD *et al.*. *J. Natl. Cancer Inst.* 2001; 93: 1054–1061.

6 Issaq HJ, Veenstra TD. The role of electrophoresis in disease biomarker discovery.
Electrophoresis 2007;28(12):1980-8.

7 Rüegg C, Tissot JD, Farmer P, Mariotti A. Omics meets hypothesis-driven research- partnership for innovative discoveries in vascular biology. *Thromb. Haemost.* 2008; 100:738-46.

8 Weinstein JN. 'Omic' and hypothesis-driven research in the molecular pharmacology of cancer. *Curr. Opin. Pharmacol.* 2002; 2: 361–365.

9 Evans GA. Designer science and the „omic“ revolution. *Nat. Biotechnol.* 2000; 18:127.

10 Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* 2001; 2:343–372.

11 Yu U, Lee SH, Kim YJ, *et al.*. Bioinformatics in the post-genome era. *J. Biochem. Mol. Biol*. 2004; 37:75–82

12 Chen X, Jorgenson E, Cheung ST. New tools for functional genomic analysis. *Drug Discovery Today* 2009; 14:754-760.

13 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273:1516–1517.

14 Parmigiani G, *et al.*. Design and analysis issues in genome-wide somatic mutation studies of cancer. *Genomics* 2009; 93:17–21.

15 Shen L, Waterland RA. Methods of DNA methylation analysis. *Curr. Opin. Clin. Nutr. Metab. Care* 2007; 10:576–581.

16 Zilberman D, Henikoff S. Genome-wide analysis of DNA methylation patterns. *Development* 2007; 134:3959–3965.

17 Kimmel A, Oliver B. DNA Microarrays. Academic Press 2006.

18 Ruvkun G. The perfect storm of tiny RNAs. *Nat. Med.* 2008; 14:1041–1045.

19 Stefani G, Slack FJ. Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell. Biol.* 2008; 9:219–230.

20 Esquela-Kerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* 2006; 6:259–269.

21 Fabbri M, *et al.*. MicroRNAs. *Cancer J.* 2008; 14:1–6.

22 Blenkiron C, Miska EA. miRNAs in cancer: approaches, aetiology, diagnostics and therapy. *Hum. Mol. Genet.* 2007; 16:R106–113.

23 Kong W, *et al.*. Strategies for profiling microRNA expression. *J. Cell. Physiol.* 2009; 218:22–25.

24 Yin JQ, *et al.*. Profiling microRNA expression with microarrays. *Trends Biotechnol.* 2008; 26:70–76.

25 Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* 2004;22:245–52.

26 Rochfort S. Metabolomics reviewed: a new ‘‘omics’’ platform technology for systems biology and implications for natural products research. *J. Nat. Prod.* 2005;68:1813–20.

27 Harrigan GG, Goodacre R. Introduction. In: Harrigan GG, Goodacre R, editors. Metabolic Profiling. Its Role in Biomarker Discovery and Gene Function Analysis. Norwell, MA: Kluwer Academic Publishers 2003; 1–8.

28 Goodacre R, Timmins EM, Burton R, Kaderbhai N, Woodward AM, Kell DB, *et al.*. Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology* 1998;144:1157–70.

29 Clarke S, Goodacre R. Raman spectroscopy for whole organism and tissue profiling. In: Harrigan GG, Goodacre R, editors. Metabolic Profiling. Its Role in Biomarker Discovery and Gene Function Analysis. Norwell, MA: Kluwer Academic Publishers 2003; 95–110.

30 Zahn JA, Higgs RE, Hilton MD. Use of direct-infusion electrospray mass spectrometry to guide empirical development of improved conditions for expression of secondary metabolites from actinomycetes. *Appl. Environ. Microbiol.* 2001;67:377–86.

31 Tyers M, Mann M. From genomics to proteomics. *Nature* 2003; 422(6928):193-7.

32 Issaq HJ, Veenstra TD. The role of electrophoresis in disease biomarker discovery. *Electrophoresis* 2007;28(12):1980-8.

33 Issaq HJ, Blonder J. Electrophoresis and liquid chromatography/tandem mass spectrometry in disease biomarker discovery. *Journal of Chromatography B* 2009; 877:1222–1228.

34 O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 1975; *250*:4007–4021.

35 Bellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W. Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem. Biophys. Methods* 1982; 6, 317–339.

36 Ünlü M, Mary E, Morgan M E, Minden JS. Difference gel electrophoresis: a single method for detecting changes in protein extracts. *Electrophoresis* 1997; *18*:2071-2077.

37 Righetti PG, Castagna A, Antonucci F, Piubelli C, *et al.*. Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches. *J. Chromatogr. A* 2004; 1051:3–17.

38 Lilley KS, Friedman DB. All about DIGE: quantification technology for differential-display 2D-gel proteomics. *Expert Rev. Proteomics* 2004; 1:401–409.

39 Zhou G, Li H, DeCamp D, Chen S, *et al..* 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol. Cell. Proteomics* 2002; 1:117–124.

40 Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 1996; 379:466–469.

41 Schirle M, Heurtier MA, Kuster B. Profiling Core Proteomes of Human Cell Lines by One-dimensional PAGE and Liquid Chromatography-Tandem Mass Spectrometry. *Molecular & Cellular Proteomics* 2003; *2:1297–1305.*

42 Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW, Eling WM, Hall N, Waters A P, Stunnenberg H G, Mann M. Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature* 2002; 419:537–542.

43 Alldridge L, Metodieva G, Greenwood C, Al-Janabi K, Thwaites L, Sauven P, Metodiev M. Proteome Profiling of Breast Tumors by Gel Electrophoresis and Nanoscale Electrospray Ionization Mass Spectrometry. *J. Proteome Res.* 2008; 7 (4):1458-1469.

44 Washburn MP, Yates JR. New methods for proteome analysis: multidimensional chromatography and mass spectrometry. *Trends Biotechnol.* 2001; 19:27–30.

45 Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* 2001;19(3):242-7.

46 Motoyama A, Yates JR 3rd. Multidimensional LC separations in shotgun proteomics. *Anal Chem.* 2008; 80(19):7187-93.

47 Ahmed FE. The role of capillary electrophoresis–mass spectrometry to proteome analysis and biomarker discovery. *Journal of Chromatography B* 2009; 877:1963–1981.

48 Sniehotta M, Schiffer E, Zürbig P, Novak J, Mischak H. CE - a multifunctional application for clinical diagnosis. *Electrophoresis* 2007; 28:1407.

49 Fenselau C. A review of quantitative methods for proteomic studies. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2007; 855(1):14-20.

50 Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 2005;1(5):252-62.

51 America AH, Cordewener JH. Comparative LC-MS: a landscape of peaks and valleys. *Proteomics* 2008; 8(4):731-49.

52 Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem.* 2004;76(14):4193-201.

53 Martinkova J, Gadher SJ, Hajduch M, Kovarova H. Challenges in cancer research and multifaceted approaches for cancer biomarker quest. *FEBS Letters* 2009; 583:1772–1784.

54 Deininger SO, Ebert MP, Fütterer A, Gerhard M, Röcken C. MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *J. Proteome Res.* 2008; 7(12):5230-6.

55 Issaq HJ, Conrads TP, Prieto DA, Tirumalai R, Veenstra TD. SELDI-TOF MS for diagnostic proteomics. *Anal. Chem.* 2003; 75(7):148A-155A.

56 Xiao Z, Prieto D, Conrads TP, Veenstra TD, Issaq HJ. Proteomic patterns: their potential for disease diagnosis. *Mol. Cell. Endocrinol.* 2005; 230(1-2):95-106.

57 Kussmann M, Raymond F, Affolter M. OMICS-driven biomarker discovery in nutrition and health. *Journal of Biotechnology* 2006; 124:758–787.

58 Zhang X, Wei D, Yap Y, Li L, Guo S, Chen F. Mass spectrometry-based ‘‘omics’’ technologies

in cancer diagnostics. *Mass Spectrometry Reviews* 2007; 26:403– 431.

59 Fox CH, Johnson FB, Whiting J, Roller PP. Formaldehyde fixation. *J. Histochem. Cytochem.* 1985; 33:845-853.

60 Metz B, Kersten GFA, Hoogerhout P, Brugghe HF, *et al.*. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J. Biol. Chem.* 2004; 279:6235-6243.

61 Metz B, Kersten GFA, Baart GJ, De Joong A, *et al.*. Identification of formaldehyde-induced modifications in proteins: reactions with insulin. *Bioconjug. Chem.* 2006; 17:815-822.

62 Toews J, Rogalski JC, Clark TJ, Kast J. Mass spectrometric identification of formaldehyde- induced peptide modifications under *in vivo* protein cross-linking conditions. *Analytica chimica acta* 2008; 618: 168–183.

63 Conti CJ, Larcher F, Chesner J, Aldaz CM. Polyacrylamide gel electrophoresis and immunoblotting of proteins from paraffin-embedded tissue sections. *J. Histochem. Cytochem.* 1988; 36:547-550.

64 Shi SR, Key ME, Kalra KL. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J Histochem Cytochem.* 1991; 39:741-748.

65 Shi SR, Cote RJ, Taylor CR. Antigen Retrieval Techniques: Current Perspectives. *J. Histochem. Cytochem.* 2001; 49(8):931–937.

66 Sompuram SR, Vani K, Messana E, Bogen SA. A Molecular Mechanism of Formalin Fixation and Antigen Retrieval. *Am. J. Clin. Pathol.* 2004; 121:190-199.

67 Rait VK, O'Leary TJ, Mason JT. Modeling formalin fixation and antigen retrieval with bovine pancreatic ribonuclease A: I—Structural and functional alterations. *Laboratory Investigation* 2004; 84:292–299.

68 Rait VK, Xu L, O'Leary TJ, Mason JT. Modeling formalin fixation and antigen retrieval with bovine pancreatic RNase A II. Interrelationship of cross-linking, immunoreactivity, and heat treatment. *Laboratory Investigation* 2004; 84:300–306.

69 Yamashita S. Heat-induced antigen retrieval: Mechanisms and application to histochemistry. *Progress in Histochemistry and Cytochemistry* 2007; 41:141–200.

70 Bogen SA, Vani K, Sompuram SR. Molecular mechanisms of antigen retrieval: antigen retrieval reverses steric interference caused by formalin-induced cross-links. *Biotech. Histochem.* 2009; 84(5):207-15.

71 Ikeda K, Monden T, Kanoh T, Tsujie M, *et al*.. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J. Histochem. Cytochem.* 1998; 46:397-403.

72 Yamashita S, Okada Y. Mechanisms of heat-induced antigen retrieval: analyses in vitro employing SDS-PAGE and immunohistochemistry. *J. Histochem. Cytochem.* 2005; 53:13-21.

73 Chu WS, Liang Q, Liu J, Wei MQ, *et al*.. A nondestructive molecule extraction method allowing morphological and molecular analyses using a single tissue section. *Lab. Invest.* 2005; 85:1416-1428.

74 Prieto DA, Hood BL, Darfler MM, Guiel TG, *et al.*. Liquid tissue: proteomic profiling of formalin-fixed tissues. *Biotechniques* 2005; June(suppl):32–35.

75 Shi SR, Liu C, Balgley BM, Lee C, Taylor CR. Protein extraction from formalin-fixed, paraffin embedded tissue sections: quality evaluation by mass spectrometry. *J. Histochem. Cytochem.* 2006; 54:739-743.

76 Rahimi F, Shepherd CE, Hallyday GM, Geczy CL, Raftery MJ. Antigen-epitope retrieval to facilitate proteomic analysis of formalin-fixed archival brain tissue. *Anal. Chem.* 2006; 78:7216-7221.

77 Jiang X, Jiang X, Feng S, Tian R, *et al.*. Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues. *J. Prot. Res.* 2007; 6:1038-1047.

78 Hwang SI, Thumar J, Lundgren DH, Rezaul K, *et al*.. Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. *Oncogene* 2007; 26:65-76.

79 Nirmalan NJ, Harnden P, Selby PJ, Banks RE. Development and validation of a novel protein extraction methodology for quantitation of protein expression in

formalin-fixed paraffin-embedded tissues using western blotting. *J. Pathol.* 2009; 217(4):497-506.

80 Becker KF. Quantitative protein analysis from formalin-fixed tissues: implications for translational clinical research and nanoscale molecular diagnosis. *J. Pathol.* 2007; 211:370-378.

81 Hipp S, Walch A, Schuster T, Höfler H, Becker KF. Precise measurement of the E-cadherin repressor Snail in formalin-fixed endometrial carcinoma using protein lysate microarrays. *Clin. Exp. Metastasis* 2008; 25(6):679-83.

82 Becker KF, Mack H, Schott C, Hipp S, Rappl A, Piontek G, Hofler H. Extraction of Phosphorylated Proteins from Formalin-Fixed Cancer Cells and Tissues. *The Open Pathology Journal* 2008; 2:46-52

83 Becker KF, Schitt C, Bercker I, Hofler H. Guided protein extraction from formalin-fixed tissues for quantitative multiplex analysis avoids detrimental effects of histological stains. *Proteomics Clin. Appl.* 2008; 2:737-743.

84 Ahram M, Flaig MJ, Gillespie JW, Duray PH, *et al.*. Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications. *Proteomics* 2003; 3:413-421.

85 Ono A, Kumai T, Koizumi H, Nishikawa, *et al.*. Overexpression of heat shock protein 27 in squamous cell carcinoma of the uterine cervix: a proteomic analysis using archival formalin-fixed, paraffin-embedded tissues. *Hum. Pathol.* 2009; 40:41-49.

86 Crockett DK, Lin Z, Vaughn CP, Lim MS, Elenitoba-Johnson KS. Identification of proteins from formalin-fixed, paraffin embedded cells by LC-MS/MS. *Lab. Invest.* 2005; 85:1405-1415.

87 Hood BL, Darfler MM, Guiel TG, Furusato B, *et al*.. Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol. Cell. Proteomics* 2005; 4:1741-1753.

88 Palmer-Toy DE, Krastins B, Sarracino DA, Nadol JB Jr, Merchant SN. Efficient method for the proteomic analysis of fixed and embedded tissues. *J. Prot. Res.* 2005; 4:2404-2411.

89 Guo T, Wang W, Rudnick PA, Song T, *et al*.. Proteome analysis of microdissected formalin-fixed and paraffin-embedded tissue specimens. *J. Histochem. Cytochem.* 2007; 55:763-772.

90 Jain MR, Liu T, Hu J, Darfler M, Fitzhugh V, Rinaggio J, Li H. Quantitative Proteomic Analysis of Formalin Fixed Paraffin Embedded Oral HPV Lesions from HIV Patients. *Open Proteomics J.* 2008; 1:40-45.

91 Patel V, Hood BL, Molinolo AA, Lee NH, Conrads TP, Braisted JC, Krizman DB, Veenstra TD, Gutkind JS. Proteomic analysis of laser-captured paraffin-embedded tissues: a molecular portrait of head and neck cancer progression. *Cancer Res.* 2008; 14(4):1002-14.

92 Rodriguez FJ, Gamez JD, Vrana JA, Theis JD, Giannini C, Scheithauer BW, Parisi JE, Lucchinetti CF, Pendlebury WW, Bergen HR 3rd, Dogan A. Immunoglobulin derived depositions in the nervous system: novel mass spectrometry application for protein characterization in formalin-fixed tissues. *Lab. Invest.* 2008; 88(10):1024-37.

93 Xu H, Yiang L, Wang W, Shi SR, *et al.*. Antigen retrieval for proteomic characterisation of formalin-fixed and paraffin-embedded tissues. *J. Prot. Res.* 2008; 7:1098-1108.

94 Cheung W, Darfler MM, Alvarez H, Hood BL, Conrads TP, Habbe N, Krizman DB, Mollenhauer J, Feldmann G, Maitra A. Application of a global proteomic approach to archival precursor lesions: deleted in malignant brain tumors 1 and tissue transglutaminase 2 are upregulated in pancreatic cancer precursors. *Pancreatology* 2008; 8(6):608-16.

95 Huang SK, Darfler MM, Nicholl MB, You J, Bemis KG, Tegeler TJ, Wang M, Wery JP, Chong KK, Nguyen L, Scolyer RA, Hoon DS. LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis. *PLoS One* 2009; 4(2):e4430.

96 Balgley BM, Guo T, Zhao K, Fang X, Tavassoli FA, Lee CS. Evaluation of archival time on shotgun proteomics of formalin-fixed and paraffin-embedded tissues. *J. Proteome Res.* 2009; 8(2):917-25.

97 Sprung RW Jr, Brock JW, Tanksley JP, Li M, Washington MK, Slebos RJ, Liebler DC. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol Cell Proteomics* 2009; 8(8):1988-98.

98 Lemaire R, Desmons A, Tabet JC, Day R, Salzet M, Fournier I. Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections. *J. Proteome Res.* 2007; 6(4):1295-305.

99 Ronci M, Bonanno E, Colantoni A, Pieroni L, Di Ilio C, Spagnoli LG, Federici G, Urbani A. Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations. *Proteomics* 2008; 8(18):3702-14.

100 Groseclose MR, Massion PP, Chaurand P, Caprioli RM. High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. *Proteomics* 2008; 8(18):3715-24.

101 Stauber J, Lemaire R, Franck J, Bonnel D, Croix D, Day R, Wisztorski M, Fournier I, Salzet M. MALDI imaging of formalin-fixed paraffin-embedded tissues: application to model animals of Parkinson disease for biomarker hunting. *J. Proteome Res.* 2008; 7(3):969-78.

102 Aerni HR, Cornett DS, Caprioli RM. High-throughput profiling of formalin-fixed paraffin-embedded tissue using parallel electrophoresis and matrix-assisted laser desorption ionization mass spectrometry. *Anal. Chem.* 2009; 81(17):7490-5.

103 Djidja MC, Francese S, Loadman PM, Sutton CW, Scriven P, Claude E, Snel MF, Franck J, Salzet M, Clench MR. Detergent addition to tryptic digests and ion mobility separation prior to MS/MS improves peptide yield and protein identification for in situ proteomic investigation of frozen and formalin-fixed paraffin-embedded adenocarcinoma tissue sections. *Proteomics* 2009; 9(10):2750-63.

104 Bradford M. A rapid and sensitive method for the quantification of mg quantities of protein. *Anal. Chem.* 1976; 72:248-254.

105 Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970; 227:680-685.

106 Westermeier R. Sensitive, quantitative and fast modifications for Coomassie blue staining of polyacrilamide gels. *Proteomics* 2006; 6:63-64.

107 Chevallet M, Luche S, Rabilloud T. Silver staining of proteins in polyacrylamide gels. *Nat. Protoc.* 2006; 1:1852-1858.

108 Zeeberg BR, Feng W, Wang G, Wang MD. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* 2003; 4:R28.

109 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc. 2009;* 4(1*):44-57.*

110 Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis *Nucleic Acids Res*. 2003; 31:3784-3788.

111 Grissom RJ. Heterogeneity of variance in clinical data. *J. Consult. Clin. Psychol.* 2000; 68(1):155-65.

112 Sompuram SR, Vani K, Hafer LJ, Bogen SA. Antibodies immunoreactive with formalin-fixed tissue antigens recognize linear protein epitopes. *Am. J. Clin. Pathol.* 2006; 125:82-90.

113 Ponten F, Jirström K, Uhlen M. The Human Protein Atlas – A tool for pathology. *J. Pathol.* 2008; 4:387-393.

114 Rabek JP, Boylston WH 3rd, Papaconstantinou J. Carbonylation of ER chaperone proteins in aged mouse liver. *Biochem. Biophys. Res. Commun.* 2003; 305(3):566-72.

115 Song J, Finnerty CC, Herndon DN, Boehning D, Jeschke MG. Severe burn-induced endoplasmic reticulum stress and hepatic damage in mice. *Mol. Med.* 2009; 15(9-10):316-20.

116 Alanen HI, Williamson RA, Howard MJ, Hatahet FS, Salo KE, Kauppila A, Kellokumpu S, Ruddock LW. ERp27, a new non-catalytic endoplasmic reticulum-located human protein disulfide isomerase family member, interacts with ERp57. *J. Biol. Chem.* 2006; 281(44):33727-38.

117 Liu F, Inageda K, Nishitai G, Matsuoka M. Cadmium induces the expression of Grp78, an endoplasmic reticulum molecular chaperone, in LLC-PK1 renal epithelial cells. *Environ. Health Perspect.* 2006; 114(6):859-64.

118 Lou LX, Geng B, Chen Y, Yu F, Zhao J, Tang CS. Endoplasmic reticulum stress involved in heart and liver injury in iron-loaded rats. *Clin. Exp. Pharmacol. Physiol.* 2009; 36(7):612-8.

119 Qian Y, Zheng Y, Ramos KS, Tiffany-Castiglioni E. GRP78 compartmentalized redistribution in Pb-treated glia: role of GRP78 in lead-induced oxidative stress. *Neurotoxicology* 2005; 26(2):267-75.

120 Morrison DK. The 14-3-3 proteins: integrators of diverse signaling cues that impact cell fate and cancer development. *Trends Cell. Biol.* 2009; 19(1):16-23.

121 Lee IN, Chen CH, Sheu JC, Lee HS, Huang GT, Yu CY, Lu FJ, Chow LP. Identification of human hepatocellular carcinoma-related biomarkers by two-dimensional difference gel electrophoresis and mass spectrometry. *J. Proteome Res.* 2005; 4(6):2062-9.

122 Grune T, Reinheckel T, Li R, North JA, Davies KJ. Proteasome-dependent turnover of protein disulfide isomerase in oxidatively stressed cells. *Arch. Biochem. Biophys.* 2002; 397(2):407-13.

123 Kajiya H, Takekoshi S, Takei M, Egashira N, Miyakoshi T, Serizawa A, Teramoto A, Osamura RY. Selection of buffer pH by the isoelectric point of the antigen for the efficient heat-induced epitope retrieval: re-appraisal for nuclear protein pathobiology. *Histochem. Cell. Biol.* 2009 (ahead of print).

124 Gustafsson BI, Kidd M, Chan A, Malfertheiner MV, Modlin IM. Bronchopulmonary neuroendocrine tumors. *Cancer* 2008; 113(1):5-21.

125 Pelosi G, Rodriguez J, Viale G, Rosai J. Typical and atypical pulmonary carcinoid tumor overdiagnosed as small-cell carcinoma on biopsy specimens: a major pitfall in the management of lung cancer patients. *Am. J. Surg. Pathol.* 2005; 29(2):179-87.

126 Anbazhagan R, Tihan T, Bornman DM, Johnston JC, Saltz JH, Weigering A, Piantadosi S, Gabrielson E. Classification of small cell lung cancer and pulmonary carcinoid by gene expression profiles. *Cancer Res.* 1999; 59(20):5119-22.

127 Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, Meyerson M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* 2001; 98(24):13790-5.

128 Virtanen C, Ishikawa Y, Honjoh D, Kimura M, Shimane M, Miyoshi T, Nomura H, Jones MH. Integrated classification of lung tumors and cell lines by expression profiling. *Proc. Natl. Acad. Sci. USA* 2002; 99(19):12357-62.

129 Pedersen N, Mortensen S, Sørensen SB, Pedersen MW, Rieneck K, Bovin LF, Poulsen HS. Transcriptional gene expression profiling of small cell lung cancer cells. *Cancer Res.* 2003; 63(8):1943-53.

130 Jones MH, Virtanen C, Honjoh D, Miyoshi T, Satoh Y, Okumura S, Nakagawa K, Nomura H, Ishikawa Y. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *Lancet* 2004; 363(9411):775-81.

131 Taniwaki M, Daigo Y, Ishikawa N, Takano A, Tsunoda T, Yasui W, Inai K, Kohno N, Nakamura Y. Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer. *Int. J. Oncol.* 2006; 29(3):567-75.

132 Okubo C, Minami Y, Tanaka R, Uchihara T, Anami Y, Furuya S, Morishita Y, Iijima T, Noguchi M. Analysis of differentially expressed genes in neuroendocrine carcinomas of the lung. *J. Thorac. Oncol.* 2006; 1(8):780-6.

133 Cho NH, Koh ES, Lee DW, Kim H, Choi YP, Cho SH, Kim DS. Comparative proteomics of pulmonary tumors with neuroendocrine differentiation. *J. Proteome Res.* 2006; 5(3):643-50.

134 Gustafsson BI, Kidd M, Modlin IM. Neuroendocrine tumors of the diffuse neuroendocrine system. *Curr. Opin. Oncol.* 2008; 20(1):1-12.

135 Fay J, Kelehan P, Lambkin H, Schwartz S. Increased Expression of Cellular RNA-Binding Proteins in HPV-Induced Neoplasia and Cervical Cancer. *Journal of Medical Virology* 2009; 81:897-907.

136 Zech VFE, Dlaska M, Tzankov A, Hilbe W. Prognostic and diagnostic relevance of hnRNP A2/B1, hnRNP B1 and S100 A2 in non-small cell lung cancer. *Cancer Detection and Prevention* 2006; 30:395–402.

137 Yang Z, Chang YJ, Miyamoto H, Yeh S, Yao JL, Di Sant'Agnese PA, Tsai MY, Chang C. Suppression of Androgen Receptor Transactivation and Prostate Cancer Cell Growth by Heterogeneous Nuclear Ribonucleoprotein A1 via Interaction with Androgen Receptor Coregulator ARA54. *Endocrinology* 2007; 148(3):1340–1349.

138 Pino I, Pio R, Toledo G, Zabalegui N, Vicent S, Rey N, Lozano MD, Torre W, Garcìa-Foncillas J, Montuenga LM. Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer. *Lung Cancer* 2003; 41:131-143.

139 He Y, Brown MA, Rothnagel JA, Saunders NA, Smith R. Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation. *Journal of Cell Science* 2006; 118:3173-3183.

140 Hatakeyama S, Sugihara K, Nakayama J, Akama TO, Wong SMA, Kawashima H, Zhang J, Smith DF, Ohyama C, Fukuda M, Fukuda MN. Identification of mRNA splicing factors as the endothelial receptor for carbohydrate-dependent lung colonization of cancer cells. *PNAS* 2009; 106(9):3095–3100.

141 Sadow PM, Rumilla KM, Erickson LA, Lloyd RV. Stathmin Expression in Pheochromocytomas, Paragangliomas, and in other Endocrine Tumors. *Endocr. Pathol.* 2008; 19:97–103.

142 Iancu-Rubin C, Atweh GF. p27Kip1 and stathmin share the stage for the first time. *TRENDS in Cell Biology* 2005; 15(7):346-348.

143 Mussunoor S, Murray GI. The role of annexins in tumour development and progression. *J. Pathol.* 2008; 216:131–140.

144 Zhang B, Wang Y, Su Y. Peroxiredoxins, a novel target in cancer radiotherapy. *Cancer Letters* 2009; 286:154–160.

145 Kinnula VL, Paakko P, Soini Y. Antioxidant enzymes and redox regulating thiol proteins in malignancies of human lung. *FEBS Letters* 2004; 569:1–6.

146 Rhee SG, Chae HZ, Kim K. Peroxiredoxins: A historical overview and speculative preview of novel mechanisms and emerging concepts in cell signaling. *Free Radical Biology and Medicine* 2005; 38:1543–1552.

147 Manevich Y, Fisher AB. Peroxiredoxin 6, a 1-Cys peroxiredoxin, functions in antioxidant defense and lung phospholipid metabolism. *Free Radical Biology and Medicine* 2005; 38:1422–1432.

148 Svensk AM, Soini Y, Pääkkö P, Hirvikoski P, Kinnula VL. Differential Expression of Superoxide Dismutases in Lung Cancer. *Am. J. Clin. Pathol.* 2004; 122:395-404.

149 Kinnula VL, Crapo JD. Superoxide Dismutases in the Lung and Human Lung Diseases. *Am. J. Respir. Crit. Care Med.* 2003; 167:1600–1619.

150 Miwa N, Uebi T, Kawamura S. S100–annexin complexes – biology of conditional association. *FEBS Journal* 2008; 275:4945–4955.

151 Salama I, Malone PS, Mihaimeed F, Jones JL. A review of the S100 proteins in cancer. EJSO 2008; 34:357e364.

152 Buxbaum JN, Reixach N. Transthyretin: the servant of many masters. *Cell. Mol. Life Sci.* 2009; 66:3095–3101.

153 Liang D, Carvalho S, Bon S, Massoulie J. Unusual transfer of CutA into the secretory pathway, evidenced by fusion proteins with acetylcholinesterase. *FEBS Journal* 2009; 276:4473–4482.

154 Rho JH, Roehrl MHA, Wang JY. Tissue Proteomics Reveals Differential and Compartment-Specific Expression of the Homologs Transgelin and Transgelin-2 in Lung Adenocarcinoma and Its Stroma. *Journal of Proteome Research* 2009 (ahead of print).

155 Assinder SJ, Stanton JAL, Prasad PD. Transgelin: An actin-binding protein and tumour suppressor. *The International Journal of Biochemistry & Cell Biology* 2009; 41:482–486.

156 Huang Q, Huang Q, Chen W, Wang L, Lin W, Lin J, Lin X. Identification of transgelin as a potential novel biomarker for gastric adenocarcinoma based on proteomics technology. *J. Cancer Res. Clin. Oncol.* 2008; 134:1219–1227.

APPENDICES

Appendix 1

List of fresh-frozen skeletal muscle proteins identified by 1D-GeLC-MS/MS, ordered according to the number of queries matched. For "number of queries matched" is intended the total number of peptide hits assigned to a protein, as a sum of data obtained from *all* 21 cut *gel bands*. Sequence coverage instead refers to the best percentage reached for a *single gel band*. Shaded lines indicate proteins identified also in FFPE extracts.

Protein	*Queries matched*	*Sequence Coverage*	*Mowse Score*	*Nominal mass*	*Calculated pI value*	*Swiss-Prot acc. no.*
Myosin heavy chain (several isoforms)	281	36	5519	223764	5.57	Q9BE40
Actin, alpha skeletal muscle	171	47	2947	42366	5.23	P68138
Creatine kinase M-type	163	37	3353	43190	6.63	Q9XSC6
Fructose-bisphosphate aldolase A	78	42	1754	39774	8.31	P00883
Myoglobin	77	43	2686	17044	6.87	P02190
Myosin light chain 1, skeletal muscle isoform	67	35	635	20994	4.97	P02602
Pyruvate kinase muscle isozyme	52	25	997	58522	7.24	P11979
Alpha-actinin-2	47	35	1998	104284	5.31	Q3ZC55
Tropomyosin-1 alpha chain	44	54	1531	32718	4.69	P58771
Tropomyosin beta chain	43	39	1041	32931	4.66	P58774
Triosephosphate isomerase	41	53	1343	26901	6.45	Q5E956
Troponin C, skeletal muscle	40	65	1116	18141	4.06	P02586
Glycogen phosphorylase, muscle form	40	26	725	97702	6.65	O18751
Myosin regulatory light chain 2, skeletal muscle isoform	39	60	1310	19057	4.82	P97457
Troponin I, fast skeletal muscle	37	31	895	21496	8.87	P48788
Glyceraldehyde-3-phosphate dehydrogenase	34	31	1386	36073	8.50	P04406
Serum albumin precursor	31	23	624	71139	5.80	P14639
Four and a half LIM domains protein 1	30	26	615	33821	8.76	Q9WUH4
Alpha-actinin-3	26	23	975	103605	5.31	O88990
ADP/ATP translocase 1	25	27	320	33174	9.84	P02722
Phosphoglucomutase-1	25	16	654	61805	6.58	P00949
Carbonic anhydrase 3	24	32	682	29637	7.17	Q3SZX4

Beta-enolase	23	19	667	47409	7.6	Q3ZC09
ATP synthase sub. alpha, mit. prec.	19	15	614	59828	9.16	P25705
ATP synthase sub. beta, mit. prec.	17	24	784	56249	5.15	P00829
L-lactate dehydrogenase A chain	16	23	445	36947	8.16	P19858
Heat shock cognate 71 kDa	16	22	577	71424	5.49	P19120
Desmin	15	28	400	53556	5.21	O62654
Heat shock 70 kDa protein 1A/1B	15	19	604	70495	5.54	Q27975
Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	15	5	322	110532	2.17	Q0VCY0
Creatine kinase, sarcomeric mitoch. prec.	14	15	375	47899	8.64	Q6P8J7
Glutathione S-transferase Mu 1	13	21	457	25789	6.90	Q9N0V4
Aconitate hydratase, mitoch. prec	13	12	186	86151	8.08	Q99KI0
6-phosphofructokinase	13	10	353	86083	8.48	Q867C9
Alpha crystallin B chain	12	28	202	20024	6.76	P02510
Troponin T, fast skeletal muscle	11	14	100	31805	5.71	P45378
Fructose-bisphosphate aldolase C	11	9	933	39830	6.41	P09972
Adenylate kinase isoenzyme 1	10	26	270	21764	8.40	P00570
Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic	9	21	123	38194	6.42	Q5EA88
Phosphoglycerate mutase 2	9	12	194	28919	8.99	P15259
Glucose-6-phosphate isomerase	9	8	239	63059	7.33	Q3ZBD7
Troponin T, slow skeletal muscle	9	7	199	31265	5.71	Q8MKH6
Troponin I, slow skeletal muscle	7	24	181	21856	9.61	Q9WUZ5
Malate dehydrogenase, mitoch. prec.	7	21	118	36102	8.82	Q32LG3
Voltage-dependent anion-selective channel protein 1	7	17	292	30836	8.62	P45879
Phosphatidylethanolamine-BP 1	6	21	174	21087	6.96	P13696
PDZ and LIM domain protein 7	6	7	142	47486	8.69	Q3SX40
Troponin C, slow skel. and card. muscles	5	24	294	18519	4.05	P63315
Glutathione S-transferase A1	5	21	70	25549	8.66	Q28035
14-3-3 protein gamma	5	14	68	28456	4.80	P61981
Phosphoglycerate kinase 1	5	8	146	44985	8.30	P00558
Voltage-dependent anion-selective channel protein 3	5	7	197	31062	8.95	Q9MZ13
Myotilin	5	4	125	55738	9.02	Q9JIF9
Heat-shock protein beta-6	4	28	142	17515	5.95	Q148F8
Histone H2A type 1-B	4	21	64	14127	11.05	P04908
NADH dehydrogenase iron-sulfur protein 3, mitoch. prec.	4	12	201	30473	6.54	P17694

Elongation factor 1-alpha 2	4	7	93	50780	9.11	Q32PH8
Very-long-chain specific acyl-CoA dehydrogenase, mitoch. prec.	4	5	78	70745	8.92	P48818
AMP deaminase 1	4	4	134	87118	6.43	P23109
AP-1 complex subunit mu-1	4	1	31	48727	6.82	Q2KJ81
Calsequestrin-1	3	25	97	6610	4.61	P31236
14-3-3 protein epsilon	3	7	70	29326	4.63	P62261
Aldehyde dehydrogenase, mitochondrial precursor	3	6	104	57129	7.55	P20000
Malate dehydrogenase, cytoplasmic	3	6	94	36700	6.16	Q3T145
Aspartate aminotransferase, mitochondrial	3	4	98	44938	9.04	P08907
Myosin-binding protein C, slow-type	3	3	111	129240	5.78	Q00872
Carbonic anhydrase 2	2	32	682	29637	7.71	P00921
Protein DJ-1	2	12	94	20194	6.84	Q5E946
Alpha-S1-casein precursor	2	10	55	24570	4.98	P02662
Heat-shock protein beta-1	2	8	127	22436	5.98	Q3T149
ATP synthase gamma chain, mitochondrial precursor	2	7	76	33108	9.34	Q4LDE7
Myozenin-1	2	6	75	31654	9.17	Q8SQ24
Kappa-casein precursor	2	5	64	21370	6,30	P02668
Dihydrolipoyl dehydrogenase, mitochondrial precursor	2	4	90	54689	7.95	P49819
PDZ and LIM domain protein 5	2	4	52	64544	8.61	Q8CI51
Four and a half LIM domains protein 3	2	3	62	33792	5.75	Q9R059
Heat shock protein HSP 90-alpha	2	3	70	85077	4.93	Q76LV2
LIM domain-binding protein 3	2	1	53	78226	8.47	O75112
Cytochrome c1 heme protein, mitochondrial precursor	1	4	45	35616	9.14	P00125
Cytochrome c oxidase subunit 2	1	3	48	25836	4.74	P00405
Transmembrane protein 38A	1	3	47	33694	8.56	Q9H6F2
Succinate dehydrogenase iron-sulfur protein, mitoch. prec.	1	3	32	32296	8.91	Q3T189
Myosin-binding protein H	1	2	85	52531	6.30	Q13203
Myc box-dependent-interact. prot. 1	1	1	39	64887	4.97	O00499
SET and MYND domain-containing protein 1	1	1	46	57549	6.66	Q8NB12
Total: 85 proteins identified in fresh-frozen skeletal muscle tissue extracts						

Appendix 2

List of FFPE skeletal muscle proteins identified by 1D-GeLC-MS/MS, ordered according to the number of queries matched. For "number of queries matched" is intended the total number of peptide hits assigned to a protein, as a sum of data obtained from *all* 21 cut *gel bands*. Sequence coverage instead refers to the best percentage reached for a *single gel band*. Shaded lines indicate proteins identified also in fresh-frozen extracts.

Protein	*Queries matched*	*Sequence Coverage*	*Mowse Score*	*Nominal mass*	*Calculated pI value*	*Swiss-Prot acc. No.*
Actin, alpha skeletal muscle	456	49	3075	42381	5.23	P68138
Creatine kinase M-type	153	26	856	43190	6.63	Q9XSC6
Myosin light chain 1, skeletal muscle isoform	124	39	1547	20993	4.97	P02602
Myoglobin	121	46	1642	17044	6.87	P02190
Fructose-bisphosphate aldolase A	92	39	1343	39774	8.31	P00883
Triosephosphate isomerase	69	47	1676	26901	6.45	Q5E956
Heat-shock protein beta-1	60	63	1181	22436	5.98	Q3T149
Beta-enolase	54	35	1676	47409	7.6	Q3ZC09
Carbonic anhydrase 3	50	16	155	29637	7.71	Q3SZX4
Tropomyosin-1 alpha chain	48	21	441	32746	4.69	P09493
Tropomyosin beta chain	41	30	389	32945	4.66	P07951
Myosin regulatory light chain 2, skeletal muscle isoform	40	56	1235	19057	4.82	P97457
Serum albumin precursor	38	18	319	71139	5.80	P14639
Glyceraldehyde-3-phosphate dehydrogenase	29	11	379	36073	8.50	P04406
Pyruvate kinase muscle isozyme	22	21	378	58522	7.24	P11979
Myosin heavy chain (several isoforms)	22	5	485	223946	5.59	Q9BE40
Adenylate kinase isoenzyme 1	19	29	210	21764	8.40	P00570
Alpha-actinin-2	19	16	953	104284	5.31	Q3ZC55
ATP synthase sub. alpha, mit. prec.	16	20	708	59797	9.21	P25705
Fructose-bisphosphate aldolase C	14	39	1343	39774	8.31	P09972
Phosphoglycerate mutase 2	13	16	205	28919	8.99	P15259
Hemoglobin subunit alpha-1/2	12	25	153	15212	8,72	P68240
Alpha crystallin B chain	11	32	156	20024	6.76	P02510
Glycogen phosphorylase, muscle form	11	14	297	97702	6.65	O18751
Alpha-actinin-3	11	11	449	103605	5.31	O88990
Phosphatidylethanolamine-binding protein 1	10	25	147	21087	6.96	P13696

ATP synthase sub. beta, mit. prec.	10	14	571	56249	5.15	P00829
Hemoglobin subunit beta-C	9	16	130	15680	11,57	P68056
Glutathione S-transferase Mu 1	9	16	261	25789	6.90	Q9N0V4
Creatine kinase, sarcomeric mitochondrial precursor	9	5	160	47714	8.45	Q3ZBP1
Troponin I, slow skeletal muscle	8	20	206	21850	9.61	Q9WUZ5
Heat shock 70 kDa protein 1A/1B	8	12	311	70500	5.68	Q27975
Troponin I, fast skeletal muscle	7	26	129	21496	8.87	P48788
Phosphoglucomutase-1	6	9	127	71764	6.30	P00949
Troponin C, skeletal muscle	5	35	206	18167	4.06	P02586
Calsequestrin-1	5	25	176	6610	4.71	P31236
Glutathione S-transferase A1	5	12	61	25549	8.66	Q28035
Malate dehydrogenase, mitoch. prec.	5	7	41	36102	8.82	Q32LG3
Heat shock cognate 71 kDa protein	5	5	245	71423	5.49	P19120
L-lactate dehydrogenase A chain	5	3	151	36947	8.16	P19858
14-3-3 protein gamma	4	14	67	28456	4.80	P61981
Malate dehydrogenase, cytoplasmic	4	7	116	36700	6.16	Q3T145
PDZ and LIM domain protein 7	4	5	67	47486	8.69	Q3SX40
Peroxiredoxin-6	3	13	115	25108	6.00	O77834
Troponin T, fast skeletal muscle	3	10	81	31805	5.71	P45378
Phosphoglycerate kinase 1	3	6	63	44908	8.48	P00558
Decorin precursor	3	5	48	40196	8.72	P21793
Troponin T, slow skeletal muscle	3	4	78	31265	5.71	Q8MKH6
Heat-shock protein beta-6	2	14	65	17515	5.95	Q148F8
Telethonin	2	13	93	19204	5.50	Q6T8D8
Histone H2B type 1-A	2	12	32	14159	10.31	Q96A08
Thioredoxin-dependent peroxide reductase, mitoch. prec.	2	10	74	28406	7.15	P35705
NADH dehydrogenase flavoprotein 2, mitoch. prec.	2	9	91	27575	8.21	P25708
NADH dehydrogenase iron-sulfur protein 3, mitoch. prec.	2	9	115	30437	6.54	P17694
Histone H2A type 1-A	2	6	42	14225	10.86	P04908
Kappa-casein precursor	2	5	69	21370	6,30	P02668
Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic	2	2	45	38194	6.42	Q5EA88
Sarcalumenin precursor	2	1	89	101069	4.34	Q86TD4
Superoxide dismutase [Mn], mitochondrial	1	4	52	22348	6.86	Q8HXP0
Carbonic anhydrase 2	1	3	35	29193	6.41	P00921
Cytochrome c oxidase subunit 2	1	3	53	25836	4.74	P00405
Voltage-dependent anion-selective channel protein 1	1	3	70	30836	8.62	P45879

Collagen alpha-2(I) chain precursor	1	2	40	53495	7.23	Q28668
Prolargin precursor	1	2	43	44054	9.59	Q9GKN8
Glucose-6-phosphate isomerase	1	1	36	63335	8.43	Q3ZBD7
Myc box-dependent-interacting prot. 1	1	1	51	64887	4.97	O00499
Total: 66 proteins identified in FFPE skeletal muscle tissue extracts						

Appendix 3

List of proteins comparatively identified in fresh-frozen and FFPE tissues by 2D-PAGE-MS. Queries matched value represents number of peptide hits assigned to a protein for a single spot.

Spot no.	*Protein*	*Swiss-Prot acc. no.*	*Theor. mol. mass*	*Theor. pI value*	*Queries matched*	*Sequence Coverage*	*Est. MW on gel*	*Est. pI on gel*	*Queries matched*	*Sequence Coverage*	*Est. MW on gel*	*Est. pI on gel*
					FRESH				FFPE			
1	Myosin heavy chain, skeletal muscle, adult 1	Q9BE40	223764	5.57	50	25	88612	5.44	11	27	83781	5.42
2	Serum albumin precursor	P14639	71139	5.80	19	38	71139	5.80	16	33	71139	5.80
3	Phosphoglucomutase-1	P00949	61805	6.58	31	23	64523	6.75	11	10	63677	6.54
4	Myosin heavy chain, skeletal muscle, adult 1	Q9BE40	223764	5.57	18	10	62967	5.72	3	20	59770	5.79
5	Pyruvate kinase isozyme M1	P11979	58522	7.24	13	30	58522	7.24	31	22	58522	7.24
6	Desmin	O62654	53556	5.21	27	65	56967	5.30	4	9	56411	5.41
7	ATP synthase subunit beta, mitochondrial precursor	P00829	56249	5.15	21	53	53835	5.02	3	6	53453	5.04
8	Beta-enolase	Q3ZC09	47409	7.60	13	41	49258	7.42	11	37	48823	7.25
9	Beta-enolase	Q3ZC09	47409	7.60	13	41	49125	7.15	11	37	48584	7.08
10	Beta-enolase	Q3ZC09	47409	7.60	13	41	48730	7.00	11	37	48465	6.90
11	Beta-enolase	Q3ZC09	47409	7.60	15	39	48730	6.83	11	37	48111	6.71
12	Beta-enolase	Q3ZC09	47409	7.60	11	33	48862	6.64	11	37	47525	6.51

13	Actin, alpha skeletal muscle	P68138	42366	5.23	18	49	47054	5.56	6	16	45365	5.86
14	Actin, alpha skeletal muscle	P68138	42366	5.23	18	49	46801	5.70	6	16	45143	6.02
15	Actin, alpha skeletal muscle	P68138	42366	5.23	18	49	44828	5.20	15	48	45143	5.29
16	Creatine kinase M-type	Q9XSC6	43190	6.63	12	35	42136	6.94	10	30	44376	7.03
17	Fructose-bisphosphate aldolase A	P00883	39774	8.31	11	30	40034	8.65	6	24	40931	8.59
18	Troponin T, slow skeletal muscle	Q8MKH6	31265	5.71	2	9	39926	5.70	2	9	38784	5.71
19	Tropomyosin-1 alpha chain	P58771	32931	4.66	43	65	38243	4.89	43	58	38879	4.83
	Tropomyosin beta chain	P58774	32718	4.69	36	57	38243	4.89	14	47	38879	4.83
	Tropomyosin alpha-3 chain	P06753	32856	4.68	9	51	38243	4.89	10	51	38879	4.83
20	Glyceraldehyde-3-phosphate dehydrogenase	P10096	36073	8.50	14	49	35370	7.29	13	45	36930	7.34
	L-lactate dehydrogenase A chain	P19858	36947	8.16	9	33	35370	7.29	4	14	36930	7.34
21	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic	Q5EA88	38194	6.42	12	34	34338	6.76	4	8	35511	6.85
22	Carbonic anhydrase 3	Q3SZX4	29637	7.71	5	16	29611	7.32	9	35	28868	7.39
23	14-3-3 protein gamma	P61981	28456	4.80	11	21	29215	4.78	6	26	31573	5.02
	14-3-3 protein epsilon	P62261	29326	4.63	5	9	29215	4.78	4	15	31573	5.02
24	Phosphoglycerate mutase 2	O70250	28980	8.65	6	15	28908	8.65	4	14	28908	8.65
25	Triosephosphate isomerise	Q5E956	26901	6.45	6	31	26925	7.00	13	68	26094	6.99
26	Triosephosphate isomerise	Q5E956	26901	6.45	11	58	26698	6.54	7	36	26102	6.52
27	Triosephosphate isomerise	Q5E956	26901	6.45	6	31	26623	6.70	13	68	25593	6.61

28	Triosephosphate isomerise	Q5E956	26901	6.45	6	31	26623	6.86	13	68	25792	6.83
29	Heat-shock protein beta-1	Q3T149	22436	5.98	2	9	24526	5.41	8	47	23224	5.45
30	Heat-shock protein beta-1	Q3T149	22436	5.98	6	47	24182	5.56	8	47	23405	5.65
31	Heat-shock protein beta-1	Q3T149	22436	5.98	10	58	23507	5.86	8	47	22866	5.88
32	Heat-shock protein beta-1	Q3T149	22436	5.98	2	13	22917	6.15	7	37	22513	6.19
33	Myosin light chain 1, skeletal muscle isoform	P02602	20993	4.97	28	58	20993	4.97	30	60	20993	4.97
34	Adenylate kinase isoenzyme 1	P00570	21764	8.40	12	53	20433	8.56	2	11	20203	8.60
35	Myosin light chain 1, skeletal muscle isoform	P02602	20993	4.97	10	28	20382	4.91	13	36	19785	4.88
36	Troponin C, skeletal muscle	P02586	18141	4.06	9	60	19951	4.45	9	60	19635	4.31
37	Alpha crystallin B chain	P23927	20056	6.76	15	62	19926	7.05	10	60	19681	7.43
38	Heat-shock protein beta-6	Q148F8	17515	5.95	6	44	19764	6.21	4	23	19510	6.54
39	Myosin regulatory light chain 2, skeletal muscle isoform	P02608	19128	4.82	16	78	19128	4.82	16	78	19128	4.82

Appendix 4

Analysis of spot intensity reproducibility among DIGE maps. A well-preserved spot was chosen as standard, and volumes of ten selected spots per tissue were normalised, according to standard volume. Means, standard deviations (SD), and coefficients of variation (CV) are reported for each spot, and for the whole spot panel as average.

A - Skeletal muscle

	Master no.	Frfr1 Volume	Frfr2 Volume	Frfr3 Volume	Frfr mean	Frfr1 Norm. vol	Frfr2 Norm. vol	Frfr3 Norm. vol	Frfr Norm. mean	Frfr SD	Frfr CV	FFPE1 Volume	FFPE2 Volume	FFPE3 Volume	FFPE mean	FFPE1 Norm. vol	FFPE2 Norm. vol	FFPE3 Norm. vol	FFPE Norm. mean	FFPE SD	FFPE CV
std	2447	438173	134938	2449354	1007488	1,0000	1,0000	1,0000	1,0000	0,0000	0,00	59220	55050	343578	152616	1,0000	1,0000	1,0000	1,0000	0,0000	0,00
1	994	1085070	321916	6203006	2536664	2,4764	2,3857	2,5325	2,4648	0,0741	0,03	144951	125804	987848	419534	2,4477	2,2853	2,8752	2,5360	0,3047	0,12
2	1323	2266008	800983	7110388	3392460	5,1715	5,9359	2,9030	4,6701	1,5774	0,34	215868	192874	536535	315092	3,6452	3,5036	1,5616	2,9035	1,1642	0,40
3	1552	946710	314690	6066125	2442508	2,1606	2,3321	2,4766	2,3231	0,1582	0,07	38438	22281	345383	135367	0,6491	0,4047	1,0053	0,6864	0,3020	0,44
4	2007	986925	303342	5689669	2326645	2,2524	2,2480	2,3229	2,2744	0,0421	0,02	266815	138370	1125603	510263	4,5055	2,5135	3,2761	3,4317	1,0051	0,29
5	2108	370933	421836	6993028	2595266	0,8465	3,1261	2,8550	2,2759	1,2453	0,55	109641	75872	630445	271986	1,8514	1,3782	1,8349	1,6882	0,2686	0,16
6	2197	1393018	665216	7397226	3151820	3,1792	4,9298	3,0201	3,7097	1,0596	0,29	183619	401247	1750714	778527	3,1006	7,2888	5,0955	5,1616	2,0949	0,41
7	2204	826117	273674	4496053	1865281	1,8854	2,0281	1,8356	1,9164	0,0999	0,05	230114	201859	974661	468878	3,8857	3,6668	2,8368	3,4631	0,5534	0,16
8	4291	14756668	5882788	21622416	1007488	33,6777	43,5962	8,8278	28,7006	17,9106	0,62	3417189	4164877	5305322	152616	57,7033	75,6563	15,4414	49,6003	30,9144	0,62
9	3723	1093903	298664	3304156	1565574	2,4965	2,2133	1,3490	2,0196	0,5978	0,30	298350	235217	1202202	578590	5,0380	4,2728	3,4991	4,2699	0,7695	0,18
10	3745	461548	130329	3120149	1237342	1,0533	0,9658	1,2739	1,0977	0,1587	0,14	209291	171141	2038987	806473	3,5341	3,1088	5,9346	4,1925	1,5236	0,36
	mean				2212105				5,1452	**2,2924**	**0,24**				443733				7,7933	**3,8900**	**0,31**

B - Liver

	Master no.	Frfr1 Volume	Frfr2 Volume	Frfr3 Volume	Frfr mean	Frfr1 Norm. vol	Frfr2 Norm. vol	Frfr3 Norm. vol	Frfr Norm. mean	Frfr SD	Frfr CV	FFPE1 Volume	FFPE2 Volume	FFPE3 Volume	FFPE mean	FFPE1 Norm. vol	FFPE2 Norm. vol	FFPE3 Norm. vol	FFPE Norm. mean	FFPE SD	FFPE CV
std	3482	454360	483503	1629488	855784	1,0000	1,0000	1,0000	1,0000	0,0000	0,00	155925	356753	493121	335266	1,0000	1,0000	1,0000	1,0000	0,0000	0,00
1	2130	860784	719337	2051218	1210446	1,8945	1,4878	1,2588	1,5470	0,3220	0,21	280041	908580	823355	670659	1,7960	2,5468	1,6697	2,0042	0,4742	0,24
2	2258	1336657	1141394	3029084	1835712	2,9418	2,3607	1,8589	2,3871	0,5419	0,23	286039	1730284	824352	946892	1,8345	4,8501	1,6717	2,7854	1,7899	0,64
3	2398	625572	620515	1077762	774616	1,3768	1,2834	0,6614	1,1072	0,3889	0,35	38674	196676	80244	105198	0,2480	0,5513	0,1627	0,3207	0,2042	0,64
4	3014	1645170	1198431	4137239	2326947	3,6209	2,4786	2,5390	2,8795	0,6427	0,22	111562	448166	459739	339822	0,7155	1,2562	0,9323	0,9680	0,2721	0,28
5	3269	971706	740817	2778897	1497140	2,1386	1,5322	1,7054	1,7921	0,3124	0,17	124982	252231	652779	343331	0,8016	0,7070	1,3238	0,9441	0,3322	0,35
6	3435	913120	953507	2168567	1345065	2,0097	1,9721	1,3308	1,7709	0,3815	0,22	219560	1225616	1057543	834240	1,4081	3,4355	2,1446	2,3294	1,0262	0,44
7	5428	10370122	5401476	22270230	855784	22,8236	11,1715	13,6670	15,8874	6,1351	0,39	1290008	2358838	3700118	335266	8,2733	6,6120	7,5035	7,4629	0,8314	0,11
8	3498	2062528	1158786	4106799	2442704	4,5394	2,3966	2,5203	3,1521	1,2030	0,38	175199	311320	319744	268754	1,1236	0,8726	0,6484	0,8816	0,2377	0,27
9	4285	1642346	813342	3372819	1942836	3,6146	1,6822	2,0699	2,4556	1,0223	0,42	572762	861700	1825338	1086600	3,6733	2,4154	3,7016	3,2634	0,7346	0,23
10	4723	1151510	656323	2908894	1572242	2,5344	1,3574	1,7852	1,8923	0,5957	0,31	383355	863685	1703133	983391	2,4586	2,4210	3,4538	2,7778	0,5857	0,21
	mean				1580349				3,4871	**1,1546**	**0,29**				591415				2,3737	**0,6488**	**0,34**

Appendix 5

List of proteins differentially represented in fresh-frozen skeletal muscle and liver 2D-DIGE maps. Percentage of lysine, and arginine residues, and lysine-to-arginine ratio (K/R) for all proteins in list are indicated.

Skeletal muscle

Spot n.	ID	% Lys	% Arg	K/R
1	Myosin-1	10,8%	5,3%	2,0
2	Serum albumin	10,3%	3,8%	2,7
3	ATP-synthase subunit beta, mitochondrial	4,6%	4,6%	1,0
4	Tropomyosin alpha-1 chain	13,7%	4,9%	2,8
5	Pyruvate kinase isozymes M1/M2	7,0%	5,8%	1,2
6	Phosphoglycerate kinase 1	10,3%	2,6%	4,0
7	Glyceraldehyde-3-phosphate dehydrogenase	7,8%	3,0%	2,6
8	Fructose-bisphosphate aldolase A	7,2%	4,1%	1,8
9	Beta-enolase	8,5%	4,2%	2,0
10	Troponin T, fast skeletal muscle	14,1%	9,6%	1,5
	mean	9,4%	4,8%	2,2

Liver

Spot n.	ID	% Lys	% Arg	K/R
1	78 kDa glucose regulated protein	9,4%	4,2%	2,2
2	ATP-synthase subunit beta, mitochondrial	4,6%	4,6%	1,0
3	Protein disulfide-isomerase (A1)	10,4%	2,0%	5,2
4	Calreticulin	10,2%	2,0%	5,1
5	60 kDa heat shock protein, mitochondrial	9,7%	3,1%	3,1
6	Alpha-enolase	8,5%	4,2%	2,0
7	Endoplasmin	10,0%	4,6%	2,2
8	Serum albumin	10,3%	3,8%	2,7
9	Heat shock cognate 71 kDa protein	8,3%	4,3%	1,9
10	Aldose 1-epimerase	5,6%	3,8%	1,5
	mean	8,7%	3,7%	2,7

Appendix 6

List of proteins differentially represented in FFPE skeletal muscle and liver 2D-DIGE maps. Percentage of lysine, and arginine residues, and lysine-to-arginine ratio (K/R) for all proteins in list are indicated.

Skeletal muscle

Spot n.	ID	% Lys	% Arg	K/R
1	Actin, cytoplasmic 1	5,1%	4,8%	1,1
2	Heat shock protein beta-1	3,5%	7,5%	0,5
3	Myosin light chain 1, skeletal muscle isoform	9,9%	2,1%	4,7
4	Triosephosphate isomerase	8,5%	3,2%	2,7
5	Myosin regulatory light chain 2, skeletal muscle isoform type 2	9,5%	3,6%	2,6
6	Cytochrome c oxidase subunit 5°	6,4%	6,4%	1,0
7	Troponin C, skeletal muscle	5,7%	4,4%	1,3
8	Myosin light chain 1, skeletal muscle isoform (truncated form 55-192)	7,2%	2,9%	2,5
9	Actin, cytoplasmic 1 (truncated form 2-245)	4,5%	4,9%	0,9
10	Troponin C, skeletal muscle (truncated form 2-143)	5,6%	4,2%	1,3
	mean	6,6%	4,4%	1,9

Liver

Spot n.	ID	% Lys	% Arg	K/R
1	3-hydroxyacyl-CoA dehydrogenase type-2	4,2%	5,0%	0,8
2	Actin, cytoplasmic 1	5,1%	4,8%	1,1
3	Ferritin light chain	5,7%	7,5%	0,8
4	Superoxide dismutase [Cu-Zn]	7,3%	3,3%	2,2
5	Catalase	5,3%	6,1%	0,9
6	Sorbitol dehydrogenase	6,5%	3,7%	1,8
7	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	6,2%	4,6%	1,3
8	Proteasome subunit beta type-6	2,0%	5,4%	0,4
9	Enoyl-CoA hydratase, mitochondrial	8,0%	3,0%	2,7
10	Heat shock protein beta-1	3,5%	7,5%	0,5
	mean	5,4%	5,1%	1,2

Appendix 7

List of proteins identified by GeLC-MS/MS in sheep FFPE pathological liver extracts (samples A and B).

Band	AN	Protein	MW	pI	Score	QM	UP	SC
A1	Q0VCX2	78 kDa glucose-regulated protein	72470	5,07	376	11	10	19%
A1	P14639	Serum albumin	71139	5,80	226	9	8	13%
A1	Q13087	Protein disulfide-isomerase A2	58512	4,89	133	9	6	8%
A1	P19483	ATP synthase subunit alpha, mitochondrial	59797	9,21	98	1	1	2%
A1	P19120	Heat shock cognate 71 kDa protein	71424	5,37	94	2	2	3%
A1	P62803	Histone H4	11360	11,36	92	2	2	17%
A1	P30122	Bile salt-activated lipase (Fragment)	65463	5,32	63	1	1	2%
A1	P05307	Protein disulfide-isomerase	57629	4,80	60	4	4	7%
A1	Q5I597	Betaine--homocysteine S-methyltransferase 1	45306	6,40	57	1	1	2%
A1	Q95M18	Endoplasmin	92654	4,76	41	1	1	1%
A1	P61284	60S ribosomal protein L12	17979	9,48	40	1	1	5%
A1	Q8NHW5	60S acidic ribosomal protein P0-like	34514	5,41	38	1	1	3%
A1	Q8NF86	Serine protease 33	30395	10,03	35	1	1	2%
A1	P68103	Elongation factor 1-alpha 1	50451	9,10	34	1	1	1%
A2	P05307	Protein disulfide-isomerase	57629	4,80	487	19	14	27%
A2	Q95140	60S acidic ribosomal protein P0	34520	5,71	225	7	6	30%
A2	Q5EAD2	D-3-phosphoglycerate dehydrogenase	57327	6,47	215	8	5	10%
A2	P19483	ATP synthase subunit alpha, mitochondrial	59797	9,21	212	5	4	8%
A2	P38657	Protein disulfide-isomerase A3	57293	6,23	177	6	5	11%
A2	P17248	Tryptophanyl-tRNA synthetase, cytoplasmic	54177	5,49	172	5	5	10%
A2	P00829	ATP synthase subunit beta, mitochondrial	56249	5,15	93	2	2	5%
A2	Q2KJ32	Selenium-binding protein 1	53092	6,03	89	1	1	2%
A2	P52193	Calreticulin	48180	4,31	88	5	5	10%
A2	P62803	Histone H4	11360	11,36	85	3	3	27%
A2	Q5I597	Betaine--homocysteine S-methyltransferase 1	45306	6,40	81	4	4	8%
A2	Q2NL29	Inositol-3-phosphate synthase 1	61293	5,68	81	2	2	4%
A2	Q3SYU2	Elongation factor 2	96276	6,41	74	2	2	2%
A2	Q07536	Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	58482	8,29	72	2	2	3%
A2	Q5E995	40S ribosomal protein S6	28820	10,85	70	3	3	13%
A2	A5PK51	Nicotinate phosphoribosyltransferase	58291	6,69	69	2	2	4%
A2	P68103	Elongation factor 1-alpha 1	50451	9,10	68	3	3	5%
A2	Q13087	Protein disulfide-isomerase A2	58512	4,89	64	2	2	4%
A2	P63243	Guanine nucleotide-binding protein subunit beta-2-like 1	35511	7,60	56	2	2	6%
A2	P26452	40S ribosomal protein SA	32977	4,80	55	1	1	4%
A2	Q32L47	Endoplasmic reticulum resident protein ERp27	23072	4,81	50	1	1	5%
A2	P00766	Chymotrypsinogen A	26220	8,52	49	1	1	3%
A2	P02075	Hemoglobin subunit beta	16120	6,75	49	1	1	6%
A2	P31081	60 kDa heat shock protein, mitochondrial	61240	5,60	48	1	1	1%
A2	P04908	Histone H2A type 1-B/E	14127	11,05	43	2	2	21%
A2	Q96A08	Histone H2B type 1-A	14159	10,31	41	3	2	12%

A2	Q64316	Actin, cytoplasmic 1	42052	5,29	38	1	1	4%
A2	Q96L12	Calreticulin-3	45195	6,19	35	2	2	3%
A2	Q3T169	40S ribosomal protein S3	26842	9,68	34	1	1	3%
A2	P68250	14-3-3 protein beta/alpha	28178	4,80	31	1	1	3%
A2	P30122	Bile salt-activated lipase (Fragment)	65463	5,32	31	1	1	1%
A3	Q3T169	40S ribosomal protein S3	26842	9,68	272	8	7	30%
A3	P26452	40S ribosomal protein SA	32977	4,80	148	3	3	12%
A3	Q32L47	Endoplasmic reticulum resident protein ERp27	23072	4,81	139	3	2	7%
A3	Q64475	Histone H2B type 1-B	13944	10,31	132	8	5	40%
A3	P63243	Guanine nucleotide-binding protein subunit beta-2-like 1	35511	7,60	126	4	4	11%
A3	P62261	14-3-3 protein epsilon	29326	4,63	112	3	3	12%
A3	Q8NHW5	60S acidic ribosomal protein P0-like	34514	5,41	111	4	4	17%
A3	Q5E983	Elongation factor 1-beta	24960	4,51	103	1	1	6%
A3	Q5E973	60S ribosomal protein L1	21636	11,69	100	1	1	6%
A3	O18789	40S ribosomal protein S2	31501	10,25	94	2	2	7%
A3	Q2KJE4	Electron transfer flavoprotein subunit alpha, mitochondrial	35282	8,77	93	2	2	8%
A3	Q5E995	40S ribosomal protein S6	28820	10,85	93	2	2	8%
A3	Q76N24	40S ribosomal protein S4, X isoform	29807	10,16	92	4	4	9%
A3	P62803	Histone H4	11360	11,36	88	4	4	38%
A3	P62265	40S ribosomal protein S14	16434	10,07	80	1	1	7%
A3	P04908	Histone H2A type 1-B/E	14127	11,05	78	4	2	21%
A3	Q3T165	Prohibitin	29843	5,57	66	1	1	4%
A3	P68103	Elongation factor 1-alpha 1	50451	9,10	65	2	2	4%
A3	Q9N1F5	Glutathione S-transferase omega-1	27743	6,84	63	1	1	5%
A3	Q64316	Actin, cytoplasmic 1	42052	5,29	61	2	2	7%
A3	Q3T0S6	60S ribosomal protein L8	28235	11,03	59	2	2	6%
A3	Q56JX6	40S ribosomal protein S28	7893	10,70	59	1	1	17%
A3	P02253	Histone H1.1 (Fragment)	10359	10,49	57	1	1	11%
A3	P02075	Hemoglobin subunit beta	16120	6,75	54	1	1	6%
A3	Q56K03	60S ribosomal protein L27a	16681	11,12	52	1	1	7%
A3	P00766	Chymotrypsinogen A	26220	8,52	50	1	1	3%
A3	P81623	Endoplasmic reticulum protein ERp29	28845	5,63	50	1	1	4%
A3	Q56JX5	40S ribosomal protein S25	13791	10,12	50	1	1	8%
A3	P68002	Voltage-dependent anion-selective channel protein 2	32113	7,48	47	1	1	4%
A3	Q28153	Chymotrypsin-like elastase family member 1	28956	8,89	44	2	2	7%
A3	Q28960	Carbonyl reductase [NADPH] 1	32015	7,60	44	1	1	3%
A3	Q3T0X6	40S ribosomal protein S16	16549	10,21	44	1	1	6%
A3	Q6NXT2	Histone H3-like	15318	11,10	41	1	1	5%
A3	Q9MZS8	Cathepsin D (Fragment)	40245	6,54	40	1	1	2%
A3	Q9JKY1	Peroxiredoxin-1	22533	8,22	39	2	2	9%
A3	Q32PD5	40S ribosomal protein S19	16051	10,31	38	1	1	6%
A3	P05631	ATP synthase subunit gamma, mitochondrial	33108	9,34	37	1	1	3%
A3	P68250	14-3-3 protein beta/alpha	28178	4,80	36	2	2	6%
A3	Q0VC36	14-3-3 protein sigma	27946	4,65	36	2	2	6%
A3	Q2TBQ5	60S ribosomal protein L7a	30178	10,61	32	3	2	7%
A3	Q3T0W9	60S ribosomal protein L19	23565	11,48	31	1	1	4%
A4	O02691	3-hydroxyacyl-CoA dehydrogenase type-2	27294	8,45	410	8	7	39%
A4	Q9BGI2	Peroxiredoxin-4	30950	6,01	396	10	7	33%

A4	Q28153	Chymotrypsin-like elastase family member 1	28956	8,89	177	6	5	28%
A4	Q56JZ1	60S ribosomal protein L13	24391	11,59	169	5	5	25%
A4	Q9N0V4	Glutathione S-transferase Mu 1	25789	6,90	168	6	6	28%
A4	P02075	Hemoglobin subunit beta	16120	6,75	157	5	4	30%
A4	P00766	Chymotrypsinogen A	26220	8,52	155	4	3	15%
A4	Q5E956	Triosephosphate isomerase	26901	6,45	143	4	4	21%
A4	Q5E958	40S ribosomal protein S8	24475	10,32	140	4	4	22%
A4	P62803	Histone H4	11360	11,36	128	5	4	38%
A4	Q3T114	Ribonuclease UK114	14321	6,18	128	3	3	32%
A4	Q5E947	Peroxiredoxin-1	22423	8,59	124	4	3	15%
A4	P33778	Histone H2B type 1-B	13942	10,31	109	3	3	24%
A4	P05805	Proproteinase E	27890	5,11	108	5	4	16%
A4	P68240	Hemoglobin subunit alpha-1/2	15212	8,72	103	2	2	16%
A4	Q8NHW5	60S acidic ribosomal protein P0-like	34514	5,41	91	3	3	11%
A4	Q32L47	Endoplasmic reticulum resident protein ERp27	23072	4,81	89	1	1	5%
A4	O46415	Ferritin light chain	20032	5,89	84	1	1	8%
A4	Q3T149	Heat shock protein beta-1	22436	5,98	82	3	3	21%
A4	Q3T0U2	60S ribosomal protein L14	23535	10,79	80	1	1	5%
A4	Q3SZ18	Hypoxanthine-guanine phosphoribosyltransferase	24711	7,66	79	2	2	11%
A4	P26452	40S ribosomal protein SA	32977	4,80	73	3	3	14%
A4	O15173	Membrane-associated progesterone receptor component 2	23861	4,76	73	1	1	6%
A4	Q99895	Chymotrypsin-C	30092	7,53	66	1	1	4%
A4	O77834	Peroxiredoxin-6	25108	6,00	65	2	2	10%
A4	Q91WS4	Betaine--homocysteine S-methyltransferase 2	40416	6,10	59	2	2	4%
A4	P19803	Rho GDP-dissociation inhibitor 1	23464	5,12	58	1	1	7%
A4	Q3T0W9	60S ribosomal protein L19	23565	11,48	55	1	1	4%
A4	Q56K03	60S ribosomal protein L27a	16681	11,12	54	1	1	7%
A4	Q3T0S6	60S ribosomal protein L8	28235	11,03	54	1	1	4%
A4	P00767	Chymotrypsinogen B	26309	4,99	53	1	1	6%
A4	Q64316	Actin, cytoplasmic 1	42052	5,29	51	1	1	2%
A4	P62265	40S ribosomal protein S14	16434	10,07	51	1	1	7%
A4	P15246	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	24664	7,05	50	1	1	4%
A4	Q5E988	40S ribosomal protein S5	23033	9,73	48	1	1	7%
A4	Q5E9E6	60S ribosomal protein L10a	24987	9,94	46	1	1	3%
A4	P04908	Histone H2A type 1-B/E	14127	11,05	38	3	3	26%
A4	Q2TBV3	Electron transfer flavoprotein subunit beta	27910	8,24	36	2	2	8%
A4	Q2YDE4	Proteasome subunit alpha type-6	27838	6,34	33	1	1	5%
A4	Q3T0X6	40S ribosomal protein S16	16549	10,21	33	1	1	6%
A4	Q2TBQ3	Guanidinoacetate N-methyltransferase	26821	5,70	29	2	2	8%
B1	P00432	Catalase	60106	6,78	391	14	10	19%
B1	P02769	Serum albumin	71244	5,82	322	18	17	31%
B1	P00367	Glutamate dehydrogenase 1, mitochondrial	61701	7,66	203	11	9	17%
B1	P68240	Hemoglobin subunit alpha-1/2	15212	8,72	156	5	3	23%
B1	P00829	ATP synthase subunit beta, mitochondrial	56249	5,15	129	3	3	8%
B1	Q9N0V4	Glutathione S-transferase Mu 1	25789	6,90	117	3	3	13%
B1	P19483	ATP synthase subunit alpha, mitochondrial	59797	9,21	110	3	3	7%
B1	O02691	3-hydroxyacyl-CoA dehydrogenase type-2	27294	8,45	105	6	6	37%
B1	Q64316	Actin, cytoplasmic 1	42052	5,29	100	4	4	13%

B1	P02075	Hemoglobin subunit beta	16120	6,75	93	4	3	21%
B1	Q27975	Heat shock 70 kDa protein 1A	70500	5,68	87	2	2	3%
B1	P62803	Histone H4	11360	11,36	77	3	3	27%
B1	P31327	Carbamoyl-phosphate synthase [ammonia], mitochondrial	165975	6,30	69	2	2	1%
B1	Q2KJH9	4-trimethylaminobutyraldehyde dehydrogenase	54854	5,84	57	1	1	2%
B1	P13745	Glutathione S-transferase A1	25706	9,04	54	2	2	6%
B1	P30115	Glutathione S-transferase A3	25401	8,76	54	2	2	7%
B1	P79381	Epoxide hydrolase 1	52646	7,25	51	1	1	1%
B1	Q9UC56	Stress-70 protein, mitochondrial	73920	5,87	50	2	2	3%
B1	P52210	Fructose-bisphosphate aldolase B	39948	8,58	45	1	1	3%
B1	P15392	Cytochrome P450 2A4	56672	9,15	37	2	2	3%
B1	Q9TTJ5	Regucalcin	33857	5,54	33	2	2	8%
B2	P00432	Catalase	60106	6,78	872	26	15	35%
B2	O02691	3-hydroxyacyl-CoA dehydrogenase type-2	27294	8,45	322	4	3	18%
B2	P00367	Glutamate dehydrogenase 1, mitochondrial	61701	7,66	277	11	11	23%
B2	Q07536	Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	58482	8,29	269	8	6	11%
B2	P19483	ATP synthase subunit alpha, mitochondrial	59797	9,21	268	5	5	11%
B2	P31327	Carbamoyl-phosphate synthase [ammonia], mitochondrial	165975	6,30	188	5	4	2%
B2	P00829	ATP synthase subunit beta, mitochondrial	56249	5,15	122	2	2	5%
B2	P51977	Retinal dehydrogenase 1	55417	6,37	96	3	3	5%
B2	Q6UWU3	Dehydrogenase/reductase SDR family member 4	27782	7,66	92	2	2	8%
B2	P09488	Glutathione S-transferase Mu 1	25923	6,24	87	3	2	14%
B2	Q58DM8	Enoyl-CoA hydratase, mitochondrial	31565	8,82	78	1	1	7%
B2	P02075	Hemoglobin subunit beta	16120	6,75	76	3	3	21%
B2	Q16788	Liver carboxylesterase 1	62766	6,15	72	2	2	4%
B2	Q2KJH9	4-trimethylaminobutyraldehyde dehydrogenase	54854	5,84	69	1	1	2%
B2	P20000	Aldehyde dehydrogenase, mitochondrial	57073	7,55	68	2	2	3%
B2	Q6Q2C2	Epoxide hydrolase 2	63354	6,06	62	1	1	1%
B2	P68240	Hemoglobin subunit alpha-1/2	15212	8,72	61	3	2	16%
B2	P31081	60 kDa heat shock protein, mitochondrial	61240	5,60	60	1	1	1%
B2	Q05555	Cytochrome P450 2A10	57325	9,24	54	3	3	5%
B2	A7YWE4	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	62084	8,37	50	2	2	3%
B2	Q2KJ32	Selenium-binding protein 1	53092	6,03	47	1	1	2%
B2	Q64316	Actin, cytoplasmic 1	42052	5,29	44	2	2	7%
B2	Q2TBR0	Propionyl-CoA carboxylase beta chain, mitochondrial	58901	7,14	41	2	2	3%
B2	P12378	UDP-glucose 6-dehydrogenase	55786	7,51	37	2	2	4%
B2	Q5E956	Triosephosphate isomerase	26901	6,45	35	1	1	5%
B2	P10096	Glyceraldehyde-3-phosphate dehydrogenase	36073	8,50	33	1	1	4%
B3	Q9TTJ5	Regucalcin	33857	5,54	204	3	3	14%
B3	P00432	Catalase	60106	6,78	154	5	3	9%
B3	Q2KJE4	Electron transfer flavoprotein subunit alpha, mitochondrial	35282	8,77	125	4	4	14%
B3	Q16698	2,4-dienoyl-CoA reductase, mitochondrial	36330	9,35	116	3	3	10%
B3	O02691	3-hydroxyacyl-CoA dehydrogenase type-2	27294	8,45	105	1	1	7%
B3	P31327	Carbamoyl-phosphate synthase	165975	6,30	98	2	2	1%

		[ammonia], mitochondrial						
B3	Q3T0Z7	Dihydropteridine reductase	25716	6,90	97	3	3	14%
B3	Q58DK5	Delta-aminolevulinic acid dehydratase	36514	6,51	96	1	1	4%
B3	P00829	ATP synthase subunit beta, mitochondrial	56249	5,15	93	1	1	2%
B3	O35459	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	36437	7,60	92	2	2	7%
B3	P25325	3-mercaptopyruvate sulfurtransferase	33443	6,13	90	1	1	4%
B3	P68240	Hemoglobin subunit alpha-1/2	15212	8,72	89	3	3	23%
B3	Q9D7V9	N-acylethanolamine-hydrolyzing acid amidase	40221	5,99	86	1	1	4%
B3	P02080	Hemoglobin subunit beta-C(NA)	15680	11,57	85	4	3	14%
B3	P05631	ATP synthase subunit gamma, mitochondrial	33108	9,34	83	2	2	7%
B3	P00366	Glutamate dehydrogenase 1, mitochondrial	61815	7,25	80	2	2	3%
B3	P45879	Voltage-dependent anion-selective channel protein 1	30836	8,62	71	2	2	7%
B3	Q3T001	Hydroxysteroid 17-beta dehydrogenase 6	36517	8,65	64	2	2	6%
B3	Q2KJ64	Arginase-1	35272	6,09	63	1	1	3%
B3	Q9N1F5	Glutathione S-transferase omega-1	27743	6,84	52	1	1	5%
B3	Q32LG3	Malate dehydrogenase, mitochondrial	36102	8,82	51	2	2	6%
B3	Q96CN7	Isochorismatase domain-containing protein 1	32501	6,96	51	1	1	4%
B3	P61603	10 kDa heat shock protein, mitochondrial	10925	8,89	44	1	1	9%
B3	Q28035	Glutathione S-transferase A1	25549	8,66	41	2	2	6%
B3	P30115	Glutathione S-transferase A3	25401	8,76	41	2	2	7%
B3	Q64316	Actin, cytoplasmic 1	42052	5,29	38	1	1	2%
B3	Q96QV6	Histone H2A type 1-A	14225	10,86	38	1	1	6%
B3	O97764	Zeta-crystallin	35532	8,29	38	1	1	3%
B3	P04908	Histone H2A type 1-B/E	14127	11,05	37	2	2	21%
B3	Q3T165	Prohibitin	29843	5,57	34	1	1	4%
B3	P09670	Superoxide dismutase [Cu-Zn]	15856	6,14	33	1	1	5%
B3	Q96A08	Histone H2B type 1-A	14159	10,31	30	2	2	12%
B4	O02691	3-hydroxyacyl-CoA dehydrogenase type-2	27294	8,45	605	13	8	43%
B4	Q58DM8	Enoyl-CoA hydratase, mitochondrial	31565	8,82	318	7	6	28%
B4	Q5E956	Triosephosphate isomerase	26901	6,45	282	6	6	30%
B4	Q9TSM5	Glutathione S-transferase Mu 1	25732	6,09	214	6	5	16%
B4	P28161	Glutathione S-transferase Mu 2	25899	6,00	207	6	5	16%
B4	P68240	Hemoglobin subunit alpha-1/2	15212	8,72	175	5	4	30%
B4	P02075	Hemoglobin subunit beta	16120	6,75	166	7	4	22%
B4	P04040	Catalase	59947	6,90	123	3	3	6%
B4	Q3T094	Protein ETHE1, mitochondrial	28395	6,25	119	3	3	14%
B4	P08760	GTP:AMP phosphotransferase mitochondrial	25656	9,02	109	4	4	17%
B4	Q2TBV3	Electron transfer flavoprotein subunit beta	27910	8,24	96	5	4	14%
B4	O77834	Peroxiredoxin-6	25108	6,00	67	1	1	5%
B4	Q28035	Glutathione S-transferase A1	25549	8,66	61	5	4	17%
B4	P00366	Glutamate dehydrogenase 1, mitochondrial	61815	7,25	56	3	3	5%
B4	P51781	Glutathione S-transferase alpha M14	25378	8,86	53	4	4	15%
B4	P60713	Actin, cytoplasmic 1	42052	5,29	53	1	1	2%
B4	P42126	3,2-trans-enoyl-CoA isomerase, mitochondrial	33080	8,80	52	1	1	2%
B4	Q9N1F5	Glutathione S-transferase omega-1	27743	6,84	52	1	1	5%
B4	P62803	Histone H4	11360	11,36	44	2	2	17%
B4	P30115	Glutathione S-transferase A3	25401	8,76	38	2	2	7%

B4	Q3T0S5	Fructose-bisphosphate aldolase B	39917	8,69	31	1	1	3%
B4 bis	O02691	3-hydroxyacyl-CoA dehydrogenase type-2	27294	8,45	2360	38	10	44%
B4 bis	Q5E956	Triosephosphate isomerase	26901	6,45	290	6	6	30%
B4 bis	P68240	Hemoglobin subunit alpha-1/2	15212	8,72	171	6	4	23%
B4 bis	O97492	Catalase	60102	6,88	166	6	5	10%
B4 bis	Q9N0V4	Glutathione S-transferase Mu 1	25789	6,90	166	5	5	21%
B4 bis	P02080	Hemoglobin subunit beta-C(NA)	15680	11,57	131	5	4	16%
B4 bis	Q28035	Glutathione S-transferase A1	25549	8,66	128	11	7	26%
B4 bis	Q3T149	Heat shock protein beta-1	22436	5,98	117	5	4	25%
B4 bis	P51781	Glutathione S-transferase alpha M14	25378	8,86	108	8	4	15%
B4 bis	Q3T094	Protein ETHE1, mitochondrial	28395	6,25	85	1	1	5%
B4 bis	O18879	Glutathione S-transferase A2	25814	8,66	81	6	3	11%
B4 bis	P30115	Glutathione S-transferase A3	25401	8,76	69	6	2	7%
B4 bis	P19803	Rho GDP-dissociation inhibitor 1	23464	5,12	56	1	1	7%
B4 bis	P35705	Thioredoxin-dependent peroxide reductase, mitochondrial	28406	7,15	53	3	3	12%
B4 bis	P62803	Histone H4	11360	11,36	51	1	1	7%
B4 bis	Q9N1F5	Glutathione S-transferase omega-1	27743	6,84	43	1	1	5%

AN, Swiss-Prot Accession number; QM, Queries Matched; UP, Unique Peptides; SC, Sequence Coverage.

Appendix 8

List of typical carcinoid protein identifications, ordered by Swiss-Prot accession number.

Accession no.	Protein	Total peptide hits	Unique peptides
A2RUB1	Uncharacterized protein FLJ35848	1	1
A6NHG4	D-dopachrome decarboxylase-like protein	1	1
A8MT79	Putative zinc-alpha-2-glycoprotein-like 1	5	1
O00299	Chloride intracellular channel protein 1	1	1
O00483	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 4	3	3
O14558	Heat shock protein beta-6	2	1
O14773	Tripeptidyl-peptidase 1	4	1
O15049	NEDD4-binding protein 3	1	1
O15145	Actin-related protein 2/3 complex subunit 3	3	2
O15511	Actin-related protein 2/3 complex subunit 5	1	1
O43488	Aflatoxin B1 aldehyde reductase member 2	6	5
O43598	Deoxyribonucleoside 5'-monophosphate N-glycosidase	1	1
O43678	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 2	1	1
O60888	Protein CutA	7	1
O75340	Programmed cell death protein 6	1	1
O75367	Core histone macro-H2A.1	3	2
O75368	SH3 domain-binding glutamic acid-rich-like protein	1	1
O75964	ATP synthase subunit g, mitochondrial	1	1
O94772	Lymphocyte antigen 6H	1	1
O95399	Urotensin-2	1	1
O95716	Ras-related protein Rab-3D	1	1
O95994	Anterior gradient protein 2 homolog	5	2
O96009	Napsin-A	4	2
P00167	Cytochrome b5	2	2
P00338	L-lactate dehydrogenase A chain	8	2
P00352	Retinal dehydrogenase 1	6	3
P00387	NADH-cytochrome b5 reductase 3	3	3
P00403	Cytochrome c oxidase subunit 2	11	2
P00441	Superoxide dismutase [Cu-Zn]	9	2
P00558	Phosphoglycerate kinase 1	1	1
P00568	Adenylate kinase isoenzyme 1	6	4
P00738	Haptoglobin	13	4
P00739	Haptoglobin-related protein	1	1
P00915	Carbonic anhydrase 1	54	6
P00918	Carbonic anhydrase 2	3	3
P01009	Alpha-1-antitrypsin	18	4
P01011	Alpha-1-antichymotrypsin	9	4
P01591	Immunoglobulin J chain	2	1
P01605	Ig kappa chain V-I region Lay	2	1
P01614	Ig kappa chain V-II region Cum	2	1
P01620	Ig kappa chain V-III region SIE	3	2
P01621	Ig kappa chain V-III region NG9	1	1
P01700	Ig lambda chain V-I region HA	1	1

P01763	Ig heavy chain V-III region WEA	1	1
P01834	Ig kappa chain C region	87	5
P01842	Ig lambda chain C regions	42	2
P01857	Ig gamma-1 chain C region	11	4
P01859	Ig gamma-2 chain C region	25	4
P01860	Ig gamma-3 chain C region	1	1
P01861	Ig gamma-4 chain C region	1	1
P01876	Ig alpha-1 chain C region	60	6
P02008	Hemoglobin subunit zeta	3	2
P02042	Hemoglobin subunit delta	929	9
P02100	Hemoglobin subunit epsilon	7	2
P02452	Collagen alpha-1	3	1
P02647	Apolipoprotein A-I	16	7
P02649	Apolipoprotein E	1	1
P02743	Serum amyloid P-component	9	4
P02753	Retinol-binding protein 4	3	3
P02763	Alpha-1-acid glycoprotein 1	2	1
P02766	Transthyretin	6	2
P02768	Serum albumin	1287	27
P02787	Serotransferrin	16	4
P02790	Hemopexin	11	3
P02792	Ferritin light chain	55	6
P02794	Ferritin heavy chain	7	2
P04004	Vitronectin	1	1
P04040	Catalase	4	2
P04075	Fructose-bisphosphate aldolase A	25	8
P04083	Annexin A1	1	1
P04179	Superoxide dismutase [Mn], mitochondrial	5	3
P04406	Glyceraldehyde-3-phosphate dehydrogenase	59	8
P04632	Calpain small subunit 1	1	1
P04732	Metallothionein-1E	8	1
P04792	Heat shock protein beta-1	54	6
P04908	Histone H2A type 1-B/E	737	6
P05060	Secretogranin-1	51	5
P05109	Protein S100-A8	15	5
P05164	Myeloperoxidase	1	1
P05387	60S acidic ribosomal protein P2	2	1
P05408	Neuroendocrine protein 7B2	1	1
P06576	ATP synthase subunit beta, mitochondrial	33	5
P06702	Protein S100-A9	11	5
P06703	Protein S100-A6	39	3
P06733	Alpha-enolase	20	6
P06748	Nucleophosmin	1	1
P07203	Glutathione peroxidase 1	1	1
P07305	Histone H1.0	4	2
P07339	Cathepsin D	23	6
P07355	Annexin A2	16	8
P07437	Tubulin beta chain	14	5
P07602	Proactivator polypeptide	6	3
P07737	Profilin-1	2	1
P07741	Adenine phosphoribosyltransferase	8	3
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	13	2
P07988	Pulmonary surfactant-associated protein B	5	2

P08107	Heat shock 70 kDa protein 1	9	5
P08123	Collagen alpha-2	13	3
P08238	Heat shock protein HSP 90-beta	3	3
P08294	Extracellular superoxide dismutase [Cu-Zn]	8	2
P08311	Cathepsin G	3	1
P08670	Vimentin	52	14
P08758	Annexin A5	49	6
P08865	40S ribosomal protein SA	15	4
P09104	Gamma-enolase	2	2
P09211	Glutathione S-transferase P	4	2
P09382	Galectin-1	1	1
P09417	Dihydropteridine reductase	4	1
P09471	Guanine nucleotide-binding protein G(o) subunit alpha	2	2
P09525	Annexin A4	1	1
P09972	Fructose-bisphosphate aldolase C	1	1
P0C7P4	Cytochrome b-c1 complex subunit Rieske-like protein 1	2	2
P10253	Lysosomal alpha-glucosidase	2	2
P10412	Histone H1.4	2	2
P10599	Thioredoxin	1	1
P10645	Chromogranin-A	66	5
P10909	Clusterin	3	2
P11021	78 kDa glucose-regulated protein	5	5
P11142	Heat shock cognate 71 kDa protein	13	4
P11226	Mannose-binding protein C	1	1
P11684	Uteroglobin	1	1
P12277	Creatine kinase B-type	16	3
P12724	Eosinophil cationic protein	2	1
P13521	Secretogranin-2	3	1
P13929	Beta-enolase	2	2
P13987	CD59 glycoprotein	2	1
P14174	Macrophage migration inhibitory factor	11	2
P14314	Glucosidase 2 subunit beta	2	1
P14649	Myosin light chain 6B	2	1
P14678	Small nuclear ribonucleoprotein-associated proteins B and B'	1	1
P14854	Cytochrome c oxidase subunit 6B1	3	1
P15151	Poliovirus receptor	1	1
P15157	Tryptase alpha-1	2	2
P15309	Prostatic acid phosphatase	3	1
P15531	Nucleoside diphosphate kinase A	5	3
P15924	Desmoplakin	1	1
P16070	CD44 antigen	2	1
P16104	Histone H2A.x	13	3
P16403	Histone H1.2	18	3
P16870	Carboxypeptidase E	3	1
P16949	Stathmin	1	1
P17066	Heat shock 70 kDa protein 6	6	3
P17661	Desmin	1	1
P17900	Ganglioside GM2 activator	2	1
P18085	ADP-ribosylation factor 4	2	1
P18124	60S ribosomal protein L7	1	1
P18621	60S ribosomal protein L17	3	2
P18669	Phosphoglycerate mutase 1	4	2

P19105	Myosin regulatory light chain 12A	15	4
P19484	Transcription factor EB	2	2
P20336	Ras-related protein Rab-3A	1	1
P20337	Ras-related protein Rab-3B	1	1
P20592	Interferon-induced GTP-binding protein Mx2	2	1
P20671	Histone H2A type 1-D	138	6
P20711	Aromatic-L-amino-acid decarboxylase	2	1
P20774	Mimecan	5	2
P20962	Parathymosin	12	3
P21291	Cysteine and glycine-rich protein 1	2	1
P21810	Biglycan	1	1
P22061	Protein-L-isoaspartate	1	1
P22090	40S ribosomal protein S4, Y isoform 1	1	1
P22392	Nucleoside diphosphate kinase B	3	3
P22492	Histone H1t	15	1
P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	62	7
P23284	Peptidyl-prolyl cis-trans isomerase B	29	7
P23396	40S ribosomal protein S3	2	2
P23528	Cofilin-1	13	3
P23763	Vesicle-associated membrane protein 1	1	1
P24666	Low molecular weight phosphotyrosine protein phosphatase	1	1
P24844	Myosin regulatory light polypeptide 9	7	4
P25705	ATP synthase subunit alpha, mitochondrial	22	6
P25786	Proteasome subunit alpha type-1	1	1
P26447	Protein S100-A4	2	2
P26599	Polypyrimidine tract-binding protein 1	1	1
P27348	14-3-3 protein theta	15	5
P27635	60S ribosomal protein L10	1	1
P27695	DNA-(apurinic or apyrimidinic site) lyase	1	1
P28065	Proteasome subunit beta type-9	3	1
P28070	Proteasome subunit beta type-4	1	1
P28072	Proteasome subunit beta type-6	6	2
P28074	Proteasome subunit beta type-5	2	1
P29401	Transketolase	2	2
P30041	Peroxiredoxin-6	19	3
P30043	Flavin reductase	8	3
P30044	Peroxiredoxin-5, mitochondrial	6	1
P30048	Thioredoxin-dependent peroxide reductase, mitochondrial	10	3
P30049	ATP synthase subunit delta, mitochondrial	3	2
P30050	60S ribosomal protein L12	24	3
P30086	Phosphatidylethanolamine-binding protein 1	57	5
P30101	Protein disulfide-isomerase A3	7	1
P30626	Sorcin	1	1
P31151	Protein S100-A7	2	2
P31942	Heterogeneous nuclear ribonucleoprotein H3	1	1
P31946	14-3-3 protein beta/alpha	35	5
P31947	14-3-3 protein sigma	30	3
P31949	Protein S100-A11	10	2
P32119	Peroxiredoxin-2	191	11
P32969	60S ribosomal protein L9	3	2
P33778	Histone H2B type 1-B	334	6
P35232	Prohibitin	1	1

P35237	Serpin B6	1	1
P35268	60S ribosomal protein L22	2	1
P37802	Transgelin-2	46	8
P37837	Transaldolase	16	3
P38159	Heterogeneous nuclear ribonucleoprotein G	1	1
P39019	40S ribosomal protein S19	24	5
P40121	Macrophage-capping protein	2	1
P40925	Malate dehydrogenase, cytoplasmic	5	2
P40926	Malate dehydrogenase, mitochondrial	9	3
P42126	3,2-trans-enoyl-CoA isomerase, mitochondrial	2	1
P42345	Serine/threonine-protein kinase mTOR	1	1
P46776	60S ribosomal protein L27a	4	1
P46781	40S ribosomal protein S9	4	4
P46782	40S ribosomal protein S5	18	3
P47929	Galectin-7	2	2
P48047	ATP synthase subunit O, mitochondrial	3	2
P48556	26S proteasome non-ATPase regulatory subunit 8	1	1
P49189	4-trimethylaminobutyraldehyde dehydrogenase	6	3
P49458	Signal recognition particle 9 kDa protein	1	1
P49721	Proteasome subunit beta type-2	2	1
P49755	Transmembrane emp24 domain-containing protein 10	1	1
P50395	Rab GDP dissociation inhibitor beta	2	2
P51148	Ras-related protein Rab-5C	2	1
P51153	Ras-related protein Rab-13	2	2
P51157	Ras-related protein Rab-28	1	1
P51571	Translocon-associated protein subunit delta	1	1
P51857	3-oxo-5-beta-steroid 4-dehydrogenase	1	1
P51884	Lumican	12	3
P51888	Prolargin	18	4
P51911	Calponin-1	1	1
P54578	Ubiquitin carboxyl-terminal hydrolase 14	1	1
P56385	ATP synthase subunit e, mitochondrial	1	1
P57721	Poly(rC)-binding protein 3	2	2
P59665	Neutrophil defensin 1	18	1
P59998	Actin-related protein 2/3 complex subunit 4	8	4
P60174	Triosephosphate isomerase	36	6
P60660	Myosin light polypeptide 6	53	5
P60709	Actin, cytoplasmic 1	896	15
P60866	40S ribosomal protein S20	11	4
P60903	Protein S100-A10	1	1
P61019	Ras-related protein Rab-2A	5	2
P61088	Ubiquitin-conjugating enzyme E2 N	10	3
P61160	Actin-related protein 2	2	2
P61163	Alpha-centractin	2	1
P61457	Pterin-4-alpha-carbinolamine dehydratase	3	2
P61586	Transforming protein RhoA	2	1
P61604	10 kDa heat shock protein, mitochondrial	7	2
P61626	Lysozyme C	8	2
P61769	Beta-2-microglobulin	2	2
P61978	Heterogeneous nuclear ribonucleoprotein K	6	3
P61981	14-3-3 protein gamma	12	4
P62081	40S ribosomal protein S7	2	2
P62158	Calmodulin	23	4

P62244	40S ribosomal protein S15a	1	1
P62249	40S ribosomal protein S16	12	4
P62258	14-3-3 protein epsilon	24	6
P62263	40S ribosomal protein S14	13	2
P62269	40S ribosomal protein S18	17	4
P62277	40S ribosomal protein S13	4	2
P62280	40S ribosomal protein S11	5	3
P62304	Small nuclear ribonucleoprotein E	4	2
P62306	Small nuclear ribonucleoprotein F	1	1
P62310	U6 snRNA-associated Sm-like protein LSm3	4	1
P62318	Small nuclear ribonucleoprotein Sm D3	2	1
P62491	Ras-related protein Rab-11A	5	4
P62701	40S ribosomal protein S4, X isoform	1	1
P62736	Actin, aortic smooth muscle	284	13
P62750	60S ribosomal protein L23a	1	1
P62805	Histone H4	725	8
P62807	Histone H2B type 1-C/E/F/G/I	348	7
P62826	GTP-binding nuclear protein Ran	2	1
P62829	60S ribosomal protein L23	4	1
P62834	Ras-related protein Rap-1A	1	1
P62851	40S ribosomal protein S25	8	3
P62857	40S ribosomal protein S28	7	2
P62873	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	6	3
P62899	60S ribosomal protein L31	17	3
P62913	60S ribosomal protein L11	14	2
P62917	60S ribosomal protein L8	6	2
P62937	Peptidyl-prolyl cis-trans isomerase A	45	7
P62988	Ubiquitin	95	5
P63104	14-3-3 protein zeta/delta	49	8
P63208	S-phase kinase-associated protein 1	1	1
P63220	40S ribosomal protein S21	3	2
P63241	Eukaryotic translation initiation factor 5A-1	2	2
P68104	Elongation factor 1-alpha 1	4	3
P68363	Tubulin alpha-1B chain	2	2
P68371	Tubulin beta-2C chain	10	5
P68871	Hemoglobin subunit beta	3459	13
P69891	Hemoglobin subunit gamma-1	15	2
P69905	Hemoglobin subunit alpha	2580	10
P80294	Metallothionein-1H	1	1
P81605	Dermcidin	1	1
P84074	Neuron-specific calcium-binding protein hippocalcin	1	1
P84077	ADP-ribosylation factor 1	14	3
P84090	Enhancer of rudimentary homolog	2	1
P84103	Splicing factor, arginine/serine-rich 3	4	2
Q01105	Protein SET	2	1
Q01469	Fatty acid-binding protein, epidermal	6	2
Q01995	Transgelin	41	10
Q02543	60S ribosomal protein L18a	2	2
Q03113	Guanine nucleotide-binding protein subunit alpha-12	2	1
Q04837	Single-stranded DNA-binding protein, mitochondrial	7	1
Q04917	14-3-3 protein eta	8	4
Q06323	Proteasome activator complex subunit 1	2	1

Q06830	Peroxiredoxin-1	82	7
Q07020	60S ribosomal protein L18	2	1
Q07507	Dermatopontin	10	2
Q07617	Sperm-associated antigen 1	1	1
Q07955	Splicing factor, arginine/serine-rich 1	1	1
Q08380	Galectin-3-binding protein	2	1
Q12765	Secernin-1	1	1
Q12905	Interleukin enhancer-binding factor 2	2	1
Q12929	Epidermal growth factor receptor kinase substrate 8	2	1
Q13011	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	3	1
Q13017	Rho GTPase-activating protein 5	1	1
Q13228	Selenium-binding protein 1	1	1
Q13404	Ubiquitin-conjugating enzyme E2 variant 1	4	2
Q13510	Acid ceramidase	14	3
Q13885	Tubulin beta-2A chain	20	4
Q13938	Calcyphosin	5	5
Q14011	Cold-inducible RNA-binding protein	9	4
Q14789	Golgin subfamily B member 1	3	3
Q14894	Mu-crystallin homolog	9	4
Q15084	Protein disulfide-isomerase A6	1	1
Q15181	Inorganic pyrophosphatase	5	2
Q15185	Prostaglandin E synthase 3	1	1
Q15286	Ras-related protein Rab-35	1	1
Q15365	Poly(rC)-binding protein 1	2	1
Q15661	Tryptase beta-1	10	3
Q15717	ELAV-like protein 1	1	1
Q15772	Striated muscle preferentially expressed protein kinase	1	1
Q15878	Voltage-dependent R-type calcium channel subunit alpha-1E	2	2
Q16568	Cocaine- and amphetamine-regulated transcript protein	2	1
Q16695	Histone H3.1t	104	6
Q2LD37	Uncharacterized protein KIAA1109	1	1
Q32P51	Heterogeneous nuclear ribonucleoprotein A1-like protein 2	3	2
Q59H18	Serine/threonine-protein kinase TNNI3K	1	1
Q5D862	Filaggrin-2	1	1
Q5JNZ5	Putative 40S ribosomal protein S26-like 1	1	1
Q5JSL3	Dedicator of cytokinesis protein 11	3	1
Q5T7P6	Transmembrane protein 78	9	1
Q5VV41	Rho guanine nucleotide exchange factor 16	1	1
Q6AWC2	Protein WWC2	1	1
Q6NXG1	Epithelial splicing regulatory protein 1	1	1
Q6NXT2	Histone H3-like	50	2
Q6PRD1	Probable G-protein coupled receptor 179	1	1
Q6ZMR3	L-lactate dehydrogenase A-like 6A	2	2
Q71U36	Tubulin alpha-1A chain	37	6
Q71UI9	Histone H2A.V	14	3
Q71UM5	40S ribosomal protein S27-like protein	2	1
Q7L0Q8	Rho-related GTP-binding protein RhoU	1	1
Q86V86	Serine/threonine-protein kinase Pim-3	1	1
Q86XF0	Dihydrofolate reductase-like protein 1	2	2
Q8IUE6	Histone H2A type 2-B	73	2
Q8IWL2	Pulmonary surfactant-associated protein A1	15	3
Q8IZX4	Transcription initiation factor TFIID 210 kDa subunit	1	1

Q8N257	Histone H2B type 3-B	3	3
Q8N729	Neuropeptide W	1	1
Q8N957	Ankyrin repeat and fibronectin type-III domain-containing protein 1	2	1
Q8NAT1	Uncharacterized glycosyltransferase AGO61	2	1
Q8NEZ4	Histone-lysine N-methyltransferase MLL3	3	2
Q8NFU3	Putative thiosulfate sulfurtransferase KAT	1	1
Q8NHW5	60S acidic ribosomal protein P0-like	4	2
Q8TDP1	Ribonuclease H2 subunit C	1	1
Q8TEL6	Trpc4-associated protein	1	1
Q8WU39	Proapoptotic caspase adapter protein	2	1
Q8WY91	THAP domain-containing protein 4	1	1
Q8WZ42	Titin	6	6
Q8WZ73	E3 ubiquitin-protein ligase rififylin	1	1
Q92522	Histone H1x	1	1
Q92558	Wiskott-Aldrich syndrome protein family member 1	2	1
Q92614	Myosin-XVIIIa	1	1
Q92878	DNA repair protein RAD50	1	1
Q92930	Ras-related protein Rab-8B	2	2
Q969M7	NEDD8-conjugating enzyme UBE2F	1	1
Q96A08	Histone H2B type 1-A	357	5
Q96CX2	BTB/POZ domain-containing protein KCTD12	5	3
Q96FQ6	Protein S100-A16	1	1
Q96HE7	ERO1-like protein alpha	1	1
Q96IU4	Abhydrolase domain-containing protein 14B	5	1
Q96MF2	SH3 and cysteine-rich domain-containing protein 3	1	1
Q96PC2	Inositol hexakisphosphate kinase 3	1	1
Q96PS8	Aquaporin-10	1	1
Q96PX9	Pleckstrin homology domain-containing family G member 4B	2	1
Q96QV6	Histone H2A type 1-A	101	2
Q96T51	RUN and FYVE domain-containing protein 1	1	1
Q99497	Protein DJ-1	4	2
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2	2	1
Q99758	ATP-binding cassette sub-family A member 3	1	1
Q99943	1-acyl-sn-glycerol-3-phosphate acyltransferase alpha	11	1
Q9BPX5	Actin-related protein 2/3 complex subunit 5-like protein	2	1
Q9BTD8	RNA-binding protein 42	2	1
Q9BVC6	Transmembrane protein 109	6	1
Q9BXL6	Caspase recruitment domain-containing protein 14	1	1
Q9BZJ3	Tryptase delta	4	2
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3	5	1
Q9H2A2	Aldehyde dehydrogenase family 8 member A1	1	1
Q9H3E2	Sorting nexin-25	3	2
Q9H853	Putative tubulin-like protein alpha-4B	2	1
Q9NP55	Protein Plunc	2	2
Q9NP80	Calcium-independent phospholipase A2-gamma	1	1
Q9NP97	Dynein light chain roadblock-type 1	1	1
Q9NRJ7	Protocadherin beta-16	3	1
Q9NUJ1	Abhydrolase domain-containing protein 10, mitochondrial	1	1
Q9NV79	Protein-L-isoaspartate O-methyltransferase domain-containing protein 2	1	1
Q9NX65	Zinc finger protein 434	1	1
Q9NY65	Tubulin alpha-8 chain	10	4

Q9P0M6	Core histone macro-H2A.2	2	1
Q9P0M9	39S ribosomal protein L27, mitochondrial	1	1
Q9UBR2	Cathepsin Z	2	1
Q9UGV6	High mobility group protein 1-like 10	5	3
Q9UHG2	ProSAAS	4	4
Q9UHL4	Dipeptidyl-peptidase 2	3	2
Q9UI15	Transgelin-3	8	3
Q9UIF9	Bromodomain adjacent to zinc finger domain protein 2A	1	1
Q9Y281	Cofilin-2	1	1
Q9Y2V2	Calcium-regulated heat stable protein 1	2	1
Q9Y5S9	RNA-binding protein 8A	1	1
Q9Y6K5	2'-5'-oligoadenylate synthetase 3	2	2

Appendix 9

List of small cell lung carcinoma protein identifications, ordered by Swiss-Prot accession number.

Accession no.	Protein	Total peptide hits	Unique peptides
A1L020	RNA-binding protein MEX3A	1	1
A6NHQ2	rRNA/tRNA 2'-O-methyltransferase fibrillarin-like protein 1	1	1
B0I1T2	Myosin-Ig	3	1
O00148	ATP-dependent RNA helicase DDX39	1	1
O00422	Histone deacetylase complex subunit SAP18	2	1
O14569	Cytochrome b561 domain-containing protein 2	2	1
O14672	Disintegrin and metalloproteinase domain-containing protein 10	1	1
O14737	Programmed cell death protein 5	1	1
O15020	Spectrin beta chain, brain 2	1	1
O15131	Importin subunit alpha-6	1	1
O15145	Actin-related protein 2/3 complex subunit 3	3	1
O15511	Actin-related protein 2/3 complex subunit 5	2	1
O43390	Heterogeneous nuclear ribonucleoprotein R	2	2
O43617	Trafficking protein particle complex subunit 3	1	1
O43768	Alpha-endosulfine	1	1
O43808	Peroxisomal membrane protein PMP34	2	1
O75071	EF-hand domain-containing protein KIAA0494	2	1
O75340	Programmed cell death protein 6	3	3
O75347	Tubulin-specific chaperone A	1	1
O75390	Citrate synthase, mitochondrial	1	1
O75964	ATP synthase subunit g, mitochondrial	2	2
O95399	Urotensin-2	1	1
O95873	Uncharacterized protein C6orf47	1	1
O95881	Thioredoxin domain-containing protein 12	1	1
O95968	Secretoglobin family 1D member 1	1	1
P00167	Cytochrome b5	3	2
P00338	L-lactate dehydrogenase A chain	15	4
P00403	Cytochrome c oxidase subunit 2	14	2
P00558	Phosphoglycerate kinase 1	1	1
P00738	Haptoglobin	12	2
P00739	Haptoglobin-related protein	6	2
P00915	Carbonic anhydrase 1	15	4
P01009	Alpha-1-antitrypsin	10	4
P01011	Alpha-1-antichymotrypsin	39	4
P01024	Complement C3	1	1
P01591	Immunoglobulin J chain	1	1
P01598	Ig kappa chain V-I region EU	3	1
P01605	Ig kappa chain V-I region Lay	4	1
P01610	Ig kappa chain V-I region WEA	1	1
P01614	Ig kappa chain V-II region Cum	1	1
P01620	Ig kappa chain V-III region SIE	5	2
P01621	Ig kappa chain V-III region NG9	4	1
P01699	Ig lambda chain V-I region VOR	1	1

P01700	Ig lambda chain V-I region HA	4	2
P01715	Ig lambda chain V-IV region Bau	1	1
P01763	Ig heavy chain V-III region WEA	4	2
P01834	Ig kappa chain C region	54	4
P01842	Ig lambda chain C regions	32	3
P01857	Ig gamma-1 chain C region	70	7
P01859	Ig gamma-2 chain C region	28	5
P01860	Ig gamma-3 chain C region	26	5
P01876	Ig alpha-1 chain C region	19	4
P01889	HLA class I histocompatibility antigen, B-7 alpha chain	1	1
P02042	Hemoglobin subunit delta	142	7
P02100	Hemoglobin subunit epsilon	10	2
P02452	Collagen alpha-1(I) chain	1	1
P02671	Fibrinogen alpha chain	25	7
P02675	Fibrinogen beta chain	25	7
P02743	Serum amyloid P-component	3	2
P02751	Fibronectin	3	2
P02768	Serum albumin	524	19
P02787	Serotransferrin	4	3
P02790	Hemopexin	2	1
P02792	Ferritin light chain	10	3
P04004	Vitronectin	2	1
P04075	Fructose-bisphosphate aldolase A	34	8
P04406	Glyceraldehyde-3-phosphate dehydrogenase	56	4
P04792	Heat shock protein beta-1	56	8
P04908	Histone H2A type 1-B/E	612	5
P05109	Protein S100-A8	1	1
P05114	Non-histone chromosomal protein HMG-14	1	1
P05141	ADP/ATP translocase 2	3	3
P05204	Non-histone chromosomal protein HMG-17	5	1
P05387	60S acidic ribosomal protein P2	5	1
P06576	ATP synthase subunit beta, mitochondrial	24	5
P06702	Protein S100-A9	3	2
P06703	Protein S100-A6	16	2
P06731	Carcinoembryonic antigen-related cell adhesion molecule 5	4	3
P06733	Alpha-enolase	34	9
P06748	Nucleophosmin	1	1
P07195	L-lactate dehydrogenase B chain	2	2
P07311	Acylphosphatase-1	1	1
P07339	Cathepsin D	3	2
P07355	Annexin A2	10	3
P07437	Tubulin beta chain	155	10
P07737	Profilin-1	5	4
P07741	Adenine phosphoribosyltransferase	2	2
P07900	Heat shock protein HSP 90-alpha	10	3
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	29	6
P07996	Thrombospondin-1	1	1
P08107	Heat shock 70 kDa protein 1	3	3
P08123	Collagen alpha-2(I) chain	7	3
P08238	Heat shock protein HSP 90-beta	11	4
P08514	Integrin alpha-IIb	1	1
P08670	Vimentin	48	18
P08708	40S ribosomal protein S17	2	2

P08758	Annexin A5	17	4
P08865	40S ribosomal protein SA	18	4
P09104	Gamma-enolase	11	3
P09211	Glutathione S-transferase P	21	4
P09382	Galectin-1	1	1
P09455	Retinol-binding protein 1	1	1
P09496	Clathrin light chain A	2	1
P09651	Heterogeneous nuclear ribonucleoprotein A1	27	6
P09936	Ubiquitin carboxyl-terminal hydrolase isozyme L1	21	6
P09972	Fructose-bisphosphate aldolase C	3	2
P0C0L4	Complement C4-A	2	2
P0C7M2	Putative heterogeneous nuclear ribonucleoprotein A1-like protein 3	2	2
P0C7P4	Cytochrome b-c1 complex subunit Rieske-like protein 1	6	2
P10412	Histone H1.4	2	2
P10809	60 kDa heat shock protein, mitochondrial	2	1
P10909	Clusterin	2	2
P10915	Hyaluronan and proteoglycan link protein 1	87	11
P11021	78 kDa glucose-regulated protein	8	4
P11142	Heat shock cognate 71 kDa protein	20	4
P11226	Mannose-binding protein C	1	1
P12004	Proliferating cell nuclear antigen	4	3
P12109	Collagen alpha-1(VI) chain	4	2
P12111	Collagen alpha-3	6	2
P12235	ADP/ATP translocase 1	1	1
P12277	Creatine kinase B-type	15	5
P12724	Eosinophil cationic protein	1	1
P13224	Platelet glycoprotein Ib beta chain	4	2
P13688	Carcinoembryonic antigen-related cell adhesion molecule 1	7	1
P13929	Beta-enolase	2	2
P14136	Glial fibrillary acidic protein	10	3
P14174	Macrophage migration inhibitory factor	7	2
P14314	Glucosidase 2 subunit beta	5	2
P14618	Pyruvate kinase isozymes M1/M2	4	2
P14625	Endoplasmin	1	1
P14649	Myosin light chain 6B	4	1
P14678	Small nuclear ribonucleoprotein-associated proteins B and B'	4	2
P14770	Platelet glycoprotein IX	1	1
P14854	Cytochrome c oxidase subunit 6B1	2	1
P14866	Heterogeneous nuclear ribonucleoprotein L	1	1
P15153	Ras-related C3 botulinum toxin substrate 2	2	2
P15311	Ezrin	3	2
P15531	Nucleoside diphosphate kinase A	5	3
P15880	40S ribosomal protein S2	4	2
P16401	Histone H1.5	39	2
P16402	Histone H1.3	3	3
P16403	Histone H1.2	135	5
P16442	Histo-blood group ABO system transferase	1	1
P16949	Stathmin	34	7
P17066	Heat shock 70 kDa protein 6	5	2
P18124	60S ribosomal protein L7	6	1
P18621	60S ribosomal protein L17	5	2

P19105	Myosin regulatory light chain 12A	10	3
P20039	HLA class II histocompatibility antigen, DRB1-11 beta chain	1	1
P20592	Interferon-induced GTP-binding protein Mx2	3	1
P20671	Histone H2A type 1-D	859	5
P20930	Filaggrin	1	1
P20962	Parathymosin	10	3
P21291	Cysteine and glycine-rich protein 1	1	1
P21796	Voltage-dependent anion-selective channel protein 1	1	1
P21810	Biglycan	2	1
P22061	Protein-L-isoaspartate	1	1
P22087	rRNA 2'-O-methyltransferase fibrillarin	1	1
P22532	Small proline-rich protein 2D	1	1
P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	76	8
P23246	Splicing factor, proline- and glutamine-rich	1	1
P23284	Peptidyl-prolyl cis-trans isomerase B	17	6
P23396	40S ribosomal protein S3	10	4
P23528	Cofilin-1	19	5
P24844	Myosin regulatory light polypeptide 9	4	2
P25705	ATP synthase subunit alpha, mitochondrial	22	5
P26373	60S ribosomal protein L13	1	1
P26447	Protein S100-A4	4	2
P26599	Polypyrimidine tract-binding protein 1	2	2
P27348	14-3-3 protein theta	23	6
P27695	DNA-(apurinic or apyrimidinic site) lyase	2	1
P28065	Proteasome subunit beta type-9	1	1
P28072	Proteasome subunit beta type-6	3	1
P29692	Elongation factor 1-delta	1	1
P30041	Peroxiredoxin-6	3	3
P30043	Flavin reductase	5	2
P30048	Thioredoxin-dependent peroxide reductase, mitochondrial	5	2
P30049	ATP synthase subunit delta, mitochondrial	2	2
P30050	60S ribosomal protein L12	27	4
P30086	Phosphatidylethanolamine-binding protein 1	21	5
P30101	Protein disulfide-isomerase A3	3	1
P30405	Peptidyl-prolyl cis-trans isomerase, mitochondrial	2	2
P30626	Sorcin	8	3
P31150	Rab GDP dissociation inhibitor alpha	5	4
P31942	Heterogeneous nuclear ribonucleoprotein H3	21	6
P31943	Heterogeneous nuclear ribonucleoprotein H	4	3
P31946	14-3-3 protein beta/alpha	13	2
P31947	14-3-3 protein sigma	25	3
P31949	Protein S100-A11	5	3
P32119	Peroxiredoxin-2	76	9
P32969	60S ribosomal protein L9	1	1
P33778	Histone H2B type 1-B	472	6
P35232	Prohibitin	5	3
P35241	Radixin	2	2
P35268	60S ribosomal protein L22	1	1
P35321	Cornifin-A	1	1
P37802	Transgelin-2	22	4
P37837	Transaldolase	5	2
P38159	Heterogeneous nuclear ribonucleoprotein G	8	3
P38405	Guanine nucleotide-binding protein G	2	1

P38919	Eukaryotic initiation factor 4A-III	2	2
P39019	40S ribosomal protein S19	9	5
P40121	Macrophage-capping protein	1	1
P40925	Malate dehydrogenase, cytoplasmic	3	3
P40926	Malate dehydrogenase, mitochondrial	11	4
P41219	Peripherin	2	1
P42766	60S ribosomal protein L35	4	2
P46776	60S ribosomal protein L27a	8	1
P46777	60S ribosomal protein L5	4	3
P46778	60S ribosomal protein L21	4	3
P46779	60S ribosomal protein L28	6	4
P46781	40S ribosomal protein S9	29	8
P46782	40S ribosomal protein S5	6	4
P47914	60S ribosomal protein L29	2	1
P48047	ATP synthase subunit O, mitochondrial	6	3
P49006	MARCKS-related protein	7	2
P49207	60S ribosomal protein L34	1	1
P49720	Proteasome subunit beta type-3	2	1
P50552	Vasodilator-stimulated phosphoprotein	1	1
P51148	Ras-related protein Rab-5C	2	2
P51149	Ras-related protein Rab-7a	12	3
P51157	Ras-related protein Rab-28	5	1
P51571	Translocon-associated protein subunit delta	2	1
P51857	3-oxo-5-beta-steroid 4-dehydrogenase	2	1
P51884	Lumican	1	1
P52566	Rho GDP-dissociation inhibitor 2	1	1
P52597	Heterogeneous nuclear ribonucleoprotein F	1	1
P52907	F-actin-capping protein subunit alpha-1	1	1
P53814	Smoothelin	4	2
P53999	Activated RNA polymerase II transcriptional coactivator p15	1	1
P56385	ATP synthase subunit e, mitochondrial	4	2
P59665	Neutrophil defensin 1	5	1
P59998	Actin-related protein 2/3 complex subunit 4	16	4
P60059	Protein transport protein Sec61 subunit gamma	1	1
P60174	Triosephosphate isomerase	42	8
P60660	Myosin light polypeptide 6	19	3
P60709	Actin, cytoplasmic 1	979	15
P60842	Eukaryotic initiation factor 4A-I	2	1
P60866	40S ribosomal protein S20	3	2
P60903	Protein S100-A10	1	1
P60981	Destrin	4	3
P61019	Ras-related protein Rab-2A	4	2
P61088	Ubiquitin-conjugating enzyme E2 N	4	1
P61160	Actin-related protein 2	2	1
P61163	Alpha-centractin	2	1
P61313	60S ribosomal protein L15	10	3
P61604	10 kDa heat shock protein, mitochondrial	12	4
P61626	Lysozyme C	1	1
P61769	Beta-2-microglobulin	2	1
P61978	Heterogeneous nuclear ribonucleoprotein K	24	6
P61981	14-3-3 protein gamma	6	4
P62081	40S ribosomal protein S7	2	2
P62158	Calmodulin	11	2

P62241	40S ribosomal protein S8	5	3
P62244	40S ribosomal protein S15a	6	3
P62249	40S ribosomal protein S16	12	3
P62258	14-3-3 protein epsilon	7	3
P62263	40S ribosomal protein S14	10	2
P62266	40S ribosomal protein S23	2	2
P62269	40S ribosomal protein S18	19	7
P62277	40S ribosomal protein S13	7	5
P62280	40S ribosomal protein S11	19	5
P62304	Small nuclear ribonucleoprotein E	5	2
P62306	Small nuclear ribonucleoprotein F	5	2
P62310	U6 snRNA-associated Sm-like protein LSm3	1	1
P62318	Small nuclear ribonucleoprotein Sm D3	3	1
P62424	60S ribosomal protein L7a	1	1
P62491	Ras-related protein Rab-11A	7	3
P62701	40S ribosomal protein S4, X isoform	7	3
P62736	Actin, aortic smooth muscle	126	9
P62750	60S ribosomal protein L23a	13	3
P62753	40S ribosomal protein S6	1	1
P62805	Histone H4	1857	11
P62807	Histone H2B type 1-C/E/F/G/I	1208	7
P62820	Ras-related protein Rab-1A	8	2
P62826	GTP-binding nuclear protein Ran	13	3
P62829	60S ribosomal protein L23	2	1
P62834	Ras-related protein Rap-1A	3	1
P62847	40S ribosomal protein S24	1	1
P62851	40S ribosomal protein S25	4	3
P62857	40S ribosomal protein S28	6	2
P62873	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	8	3
P62899	60S ribosomal protein L31	9	1
P62913	60S ribosomal protein L11	28	7
P62917	60S ribosomal protein L8	14	3
P62937	Peptidyl-prolyl cis-trans isomerase A	19	6
P62988	Ubiquitin	178	5
P62995	Transformer-2 protein homolog beta	4	4
P63000	Ras-related C3 botulinum toxin substrate 1	2	2
P63104	14-3-3 protein zeta/delta	50	7
P63173	60S ribosomal protein L38	1	1
P63220	40S ribosomal protein S21	6	2
P63241	Eukaryotic translation initiation factor 5A-1	1	1
P63244	Guanine nucleotide-binding protein subunit beta-2-like 1	5	3
P68104	Elongation factor 1-alpha 1	40	5
P68363	Tubulin alpha-1B chain	32	8
P68371	Tubulin beta-2C chain	56	10
P68871	Hemoglobin subunit beta	1234	11
P69891	Hemoglobin subunit gamma-1	3	2
P69905	Hemoglobin subunit alpha	854	8
P80723	Brain acid soluble protein 1	1	1
P81605	Dermcidin	1	1
P83731	60S ribosomal protein L24	5	4
P84077	ADP-ribosylation factor 1	11	2
P84085	ADP-ribosylation factor 5	1	1

P84098	60S ribosomal protein L19	3	1
P84103	Splicing factor, arginine/serine-rich 3	26	3
Q01105	Protein SET	4	1
Q01130	Splicing factor, arginine/serine-rich 2	3	3
Q01469	Fatty acid-binding protein, epidermal	1	1
Q01995	Transgelin	13	4
Q02543	60S ribosomal protein L18a	9	5
Q03113	Guanine nucleotide-binding protein subunit alpha-12	2	1
Q04837	Single-stranded DNA-binding protein, mitochondrial	6	3
Q06830	Peroxiredoxin-1	56	11
Q07020	60S ribosomal protein L18	12	3
Q07507	Dermatopontin	4	2
Q07812	Apoptosis regulator BAX	1	1
Q07955	Splicing factor, arginine/serine-rich 1	19	5
Q12905	Interleukin enhancer-binding factor 2	15	4
Q12931	Heat shock protein 75 kDa, mitochondrial	1	1
Q13011	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	2	2
Q13162	Peroxiredoxin-4	1	1
Q13242	Splicing factor, arginine/serine-rich 9	9	3
Q13283	Ras GTPase-activating protein-binding protein 1	1	1
Q13404	Ubiquitin-conjugating enzyme E2 variant 1	1	1
Q13509	Tubulin beta-3 chain	50	6
Q13765	Nascent polypeptide-associated complex subunit alpha	2	1
Q13885	Tubulin beta-2A chain	28	7
Q14011	Cold-inducible RNA-binding protein	3	1
Q14019	Coactosin-like protein	4	2
Q14697	Neutral alpha-glucosidase AB	1	1
Q15084	Protein disulfide-isomerase A6	2	2
Q15102	Platelet-activating factor acetylhydrolase IB subunit gamma	2	1
Q15185	Prostaglandin E synthase 3	1	1
Q15233	Non-POU domain-containing octamer-binding protein	1	1
Q15286	Ras-related protein Rab-35	4	1
Q15365	Poly(rC)-binding protein 1	10	3
Q15366	Poly(rC)-binding protein 2	2	2
Q15661	Tryptase beta-1	1	1
Q15717	ELAV-like protein 1	3	2
Q16629	Splicing factor, arginine/serine-rich 7	3	2
Q16654	Pyruvate dehydrogenase [lipoamide] kinase isozyme 4, mitochondrial	1	1
Q16658	Fascin	3	3
Q16695	Histone H3.1t	404	7
Q16719	Kynureninase	1	1
Q460N5	Poly [ADP-ribose] polymerase 14	1	1
Q4G163	F-box only protein 43	1	1
Q59H18	Serine/threonine-protein kinase TNNI3K	4	1
Q5FWF4	Zinc finger Ran-binding domain-containing protein 3	2	1
Q5HYI8	Rab-like protein 3	3	1
Q5JNZ5	Putative 40S ribosomal protein S26-like 1	9	1
Q5SVQ8	Zinc finger and BTB domain-containing protein 41	1	1
Q5T2N8	ATPase family AAA domain-containing protein 3C	2	2
Q5T7P6	Transmembrane protein 78	5	1
Q5T8P6	RNA-binding protein 26	1	1
Q5TID7	Uncharacterized protein C1orf114	2	1

Q5VTU8	ATP synthase subunit epsilon-like protein, mitochondrial	3	3
Q5VV41	Rho guanine nucleotide exchange factor 16	2	1
Q5XG87	DNA polymerase sigma	1	1
Q63HR2	Tensin-like C1 domain-containing phosphatase	1	1
Q6IAA8	UPF0404 protein C11orf59	1	1
Q6IEE8	Schlafen family member 12-like	2	2
Q6NVV1	Putative 60S ribosomal protein L13a-like MGC87657	4	3
Q6NXG1	Epithelial splicing regulatory protein 1	2	1
Q6NXT2	Histone H3-like	80	3
Q6PRD1	Probable G-protein coupled receptor 179	2	1
Q6R327	Rapamycin-insensitive companion of mTOR	1	1
Q71U36	Tubulin alpha-1A chain	223	11
Q71UI9	Histone H2A.V	36	3
Q71UM5	40S ribosomal protein S27-like protein	3	2
Q7L2R6	Zinc finger protein 765	3	1
Q86SG6	Serine/threonine-protein kinase Nek8	1	1
Q86U44	N6-adenosine-methyltransferase 70 kDa subunit	1	1
Q86YZ3	Hornerin	1	1
Q8IUE6	Histone H2A type 2-B	91	3
Q8IY63	Angiomotin-like protein 1	1	1
Q8IYB3	Serine/arginine repetitive matrix protein 1	2	2
Q8N257	Histone H2B type 3-B	5	3
Q8N5V2	Ephexin-1	1	1
Q8N680	Zinc finger and BTB domain-containing protein 2	1	1
Q8NEZ4	Histone-lysine N-methyltransferase MLL3	2	1
Q8NHW5	60S acidic ribosomal protein P0-like	10	3
Q8TD57	Dynein heavy chain 3, axonemal	1	1
Q8WU39	Proapoptotic caspase adapter protein	11	4
Q8WWR8	Sialidase-4	1	1
Q92522	Histone H1x	7	2
Q92576	PHD finger protein 3	1	1
Q93045	Stathmin-2	6	2
Q93077	Histone H2A type 1-C	43	7
Q96A08	Histone H2B type 1-A	654	5
Q96DB2	Histone deacetylase 11	1	1
Q96KK5	Histone H2A type 1-H	48	8
Q96L21	60S ribosomal protein L10-like	3	2
Q96M91	Coiled-coil domain-containing protein 11	3	1
Q96PX9	Pleckstrin homology domain-containing family G member 4B	1	1
Q96Q83	Alpha-ketoglutarate-dependent dioxygenase alkB homolog 3	1	1
Q96QV6	Histone H2A type 1-A	92	2
Q96ST8	Coiled-coil domain-containing protein 123, mitochondrial	3	2
Q99623	Prohibitin-2	6	4
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2	1	1
Q99758	ATP-binding cassette sub-family A member 3	1	1
Q99880	Histone H2B type 1-L	7	3
Q9BPX5	Actin-related protein 2/3 complex subunit 5-like protein	1	1
Q9BRL6	Splicing factor, arginine/serine-rich 2B	11	2
Q9BRX8	Uncharacterized protein C10orf58	1	1
Q9BTC0	Death-inducer obliterator 1	1	1
Q9BUP3	Oxidoreductase HTATIP2	1	1
Q9BVA1	Tubulin beta-2B chain	15	6
Q9BVC6	Transmembrane protein 109	4	1

Q9BWT7	Caspase recruitment domain-containing protein 10	2	1
Q9BYT8	Neurolysin, mitochondrial	1	1
Q9H299	SH3 domain-binding glutamic acid-rich-like protein 3	2	1
Q9H3P7	Golgi resident protein GCP60	2	1
Q9H3Z7	Uncharacterized protein C20orf135	1	1
Q9H4B7	Tubulin beta-1 chain	8	4
Q9H853	Putative tubulin-like protein alpha-4B	5	1
Q9NPC6	Myozenin-2	1	1
Q9NQ38	Serine protease inhibitor Kazal-type 5	2	1
Q9NUJ7	PI-PLC X domain-containing protein 1	1	1
Q9NUQ9	Protein FAM49B	1	1
Q9NVH2	Integrator complex subunit 7	1	1
Q9NX63	Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial	1	1
Q9NY64	Solute carrier family 2, facilitated glucose transporter member 8	1	1
Q9NY65	Tubulin alpha-8 chain	70	5
Q9P2M7	Cingulin	1	1
Q9P2X0	Dolichol-phosphate mannosyltransferase subunit 3	1	1
Q9UBR2	Cathepsin Z	1	1
Q9UBZ4	DNA-(apurinic or apyrimidinic site) lyase 2	2	1
Q9UGV6	High mobility group protein 1-like 10	1	1
Q9UHB6	LIM domain and actin-binding protein 1	1	1
Q9UKL3	CASP8-associated protein 2	3	2
Q9UNX3	60S ribosomal protein L26-like 1	8	3
Q9UP52	Transferrin receptor protein 2	1	1
Q9Y277	Voltage-dependent anion-selective channel protein 3	1	1
Q9Y2H2	Phosphatidylinositide phosphatase SAC2	2	1
Q9Y2Q3	Glutathione S-transferase kappa 1	1	1
Q9Y2V2	Calcium-regulated heat stable protein 1	1	1
Q9Y3U8	60S ribosomal protein L36	1	1
Q9Y5G4	Protocadherin gamma-A9	1	1
Q9Y6H5	Synphilin-1	2	1
Q9Y6K5	2'-5'-oligoadenylate synthetase 3	5	2

Appendix 10

List of proteins shared by all TC and SCLC samples.

Accession no.	Protein	TC QM	SCLC QM
P31946	14-3-3 protein beta/alpha	35	13
P62736	Actin, aortic smooth muscle	284	126
P60709	Actin, cytoplasmic 1	896	979
P02768	Serum albumin	1287	524
P08758	Annexin A5	49	17
P25705	ATP synthase subunit alpha, mitochondrial	22	22
P06576	ATP synthase subunit beta, mitochondrial	33	24
P61604	10 kDa heat shock protein, mitochondrial	7	12
P23528	Cofilin-1	13	19
P00403	Cytochrome c oxidase subunit 2	11	14
P06733	Alpha-enolase	20	34
P04406	Glyceraldehyde-3-phosphate dehydrogenase	59	56
P62873	Guanine nucleotide-binding protein G	6	8
P16403	Histone H1.2	18	135
Q96QV6	Histone H2A type 1-A	101	92
P04908	Histone H2A type 1-B/E	737	612
P20671	Histone H2A type 1-D	138	859
Q8IUE6	Histone H2A type 2-B	73	91
Q96A08	Histone H2B type 1-A	357	654
P33778	Histone H2B type 1-B	334	472
P62807	Histone H2B type 1-C/E/F/G/I	348	1208
Q16695	Histone H3.1t	104	404
Q6NXT2	Histone H3-like	50	80
P62805	Histone H4	725	1857
P69905	Hemoglobin subunit alpha	2580	854
P68871	Hemoglobin subunit beta	3459	1234
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	13	29
P04792	Heat shock protein beta-1	54	56
P01876	Ig alpha-1 chain C region	60	19
P01859	Ig gamma-2 chain C region	25	28
P01834	Ig kappa chain C region	87	54
P01842	Ig lambda chain C regions	42	32
P14174	Macrophage migration inhibitory factor	11	7
P19105	Myosin regulatory light chain 12A	15	10
P60660	Myosin light polypeptide 6	53	19
P62937	Peptidyl-prolyl cis-trans isomerase A	45	19
P23284	Peptidyl-prolyl cis-trans isomerase B	29	17
Q06830	Peroxiredoxin-1	82	56
P32119	Peroxiredoxin-2	191	76
P62913	60S ribosomal protein L11	14	28
P30050	60S ribosomal protein L12	24	27
P62899	60S ribosomal protein L31	17	9
P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	62	76
P62263	40S ribosomal protein S14	13	10
P62269	40S ribosomal protein S18	17	19
Q04837	Single-stranded DNA-binding protein, mitochondrial	7	6
P37802	Transgelin-2	46	22

Q71U36	Tubulin alpha-1A chain	37	223
P07437	Tubulin beta chain	14	155
P60174	Triosephosphate isomerase	36	42
P62988	Ubiquitin	95	178

Appendix 11

Distribution of protein isoforms within principal protein families.

Protein family	AN	Protein member	TC TPH	SCLC TPH
hnRNPs	O43390	Heterogeneous nuclear ribonucleoprotein R	0	2
	P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	13	29
	P09651	Heterogeneous nuclear ribonucleoprotein A1	0	27
	P0C7M2	Putative heterog. nuclear ribonucleoprot. A1-like protein 3	0	2
	P14678	Small nuclear ribonucleoprotein-associated proteins B and B'	1	4
	P14866	Heterogeneous nuclear ribonucleoprotein L	0	1
	P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	62	76
	P31942	Heterogeneous nuclear ribonucleoprotein H3	1	21
	P31943	Heterogeneous nuclear ribonucleoprotein H	0	4
	P38159	Heterogeneous nuclear ribonucleoprotein G	1	8
	P52597	Heterogeneous nuclear ribonucleoprotein F	0	1
	P61978	Heterogeneous nuclear ribonucleoprotein K	6	24
	P62304	Small nuclear ribonucleoprotein E	4	5
	P62306	Small nuclear ribonucleoprotein F	1	5
	P62318	Small nuclear ribonucleoprotein Sm D3	2	3
	Q32P51	Heterogeneous nuclear ribonucleoprotein A1-like protein 2	3	0
Splicing factors	P23246	Splicing factor, proline- and glutamine-rich	0	1
	P84103	Splicing factor, arginine/serine-rich 3	4	26
	Q01130	Splicing factor, arginine/serine-rich 2	0	3
	Q07955	Splicing factor, arginine/serine-rich 1	1	19
	Q13242	Splicing factor, arginine/serine-rich 9	0	9
	Q16629	Splicing factor, arginine/serine-rich 7	0	3
	Q9BRL6	Splicing factor, arginine/serine-rich 2B	0	11
Tubulins	P07437	Tubulin beta chain	14	155
	P68363	Tubulin alpha-1B chain	2	32
	P68371	Tubulin beta-2C chain	10	56
	Q13509	Tubulin beta-3 chain	0	50
	Q13885	Tubulin beta-2A chain	20	28
	Q71U36	Tubulin alpha-1A chain	37	223
	Q9BVA1	Tubulin beta-2B chain	0	15
	Q9H4B7	Tubulin beta-1 chain	0	8
	Q9H853	Putative tubulin-like protein alpha-4B	2	5
	Q9NY65	Tubulin alpha-8 chain	10	70
Histones	O75367	Core histone macro-H2A.1	3	0
	P04908	Histone H2A type 1-B/E	737	612
	P07305	Histone H1.0	4	0
	P10412	Histone H1.4	2	2
	P16104	Histone H2A.x	13	0
	P16401	Histone H1.5	0	39
	P16402	Histone H1.3	0	3
	P16403	Histone H1.2	18	135
	P20671	Histone H2A type 1-D	138	859
	P22492	Histone H1t	15	0

P33778	Histone H2B type 1-B	334	472
P62805	Histone H4	725	1857
P62807	Histone H2B type 1-C/E/F/G/I	348	1208
Q16695	Histone H3.1t	104	404
Q6NXT2	Histone H3-like	50	80
Q71UI9	Histone H2A.V	14	36
Q8IUE6	Histone H2A type 2-B	73	91
Q8N257	Histone H2B type 3-B	3	5
Q92522	Histone H1x	1	7
Q93077	Histone H2A type 1-C	0	43
Q96A08	Histone H2B type 1-A	357	654
Q96KK5	Histone H2A type 1-H	0	48
Q96QV6	Histone H2A type 1-A	101	92
Q99880	Histone H2B type 1-L	0	7
Q9P0M6	Core histone macro-H2A.2	2	0