Susana Patrícia Guerreiro Jorge Penedo

Mestre em Biotecnologia

Deciphering the interplay of molecular alterations underpinning renal cell carcinoma by label-free mass spectrometry and clinical proteomics:

A systems medicine approach for precision diagnosis

Dissertação para obtenção do Grau de Doutor em

Química

Orientador: Doutor Hugo Miguel Baptista Carreira dos Santos, Investigador Auxiliar, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa

Co-orientadores: Doutor William A. LaFramboise, Professor e director executivo, Drexel University, USA

Doutor Carlos Lodeiro Espiño, Professor Associado com agregação, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa

Júri:

Presidente: Professor Doutor João Paulo Goulão Crespo

Arguentes: Doutora Deborah Penque

Professor Doutor José Ricardo Ramos Franco Tavares

Vogais: Professor Doutor Rajiv Dhir

Doutor Jacek Ramon Wiśniewski

Professor Doutor José Luis Capelo Martinez

Doutor Hugo Miguel Baptista Carreira dos Santos

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro, 2020

Susana Patrícia Guerreiro Jorge Penedo

Mestre em Biotecnologia

Deciphering the interplay of molecular alterations underpinning renal cell carcinoma by label-free mass spectrometry and clinical proteomics:

A systems medicine approach for precision diagnosis

Dissertação para obtenção do Grau de Doutor em
Química

This thesis was supported by the Foundation for Science and Technology,

with the fellowship, reference number SFRH/BD/120537/2016

and by the PROTEOMASS Scientific Society

in collaboration of University of Pittsburgh Medical Center

Deciphering the interplay of molecular alterations underpinning renal cell carcinoma by label-free mass spectrometry and clinical proteomics: A systems medicine approach for precision diagnosis

To my family…

ACKNOWLEDGMENTS

"*Great things are not done by one person. They're done by a team of people"*

– Steve Jobs

After a long journey, with a variety of professional and personal challenges, I would like to express my gratitude to those who supported me, who taught me, who gave me their hand, who made me grow and accomplish this academic achievement.

First, I would like to thank my supervisor, Dr. Hugo Santos, who has always supported me throughout my thesis with his patience and knowledge. Thanks for saving me so many times when I was afraid of the big boss. I would like to express my gratitude to my co-supervisor, Professor Lodeiro, for all the support and help provided during this thesis, but also since I arrived at the BIOSCOPE group. A special thanks to Professor LaFramboise, who has embraced me so affectionally. Thank you for everything you have done for me, I will never forget all the nice moments shared in the USA, also with sweet Karin, lovely Leyland, and Leah. A big "HUGE". And of course, to Professor Capelo, for believing and giving me the opportunity to integrate the BIOSCOPE Research group back in 2014, after a singular e-mail subjected "*Candidatura expontânea".* Thank you for sharpening my critical thinking, for preparing me for the future, and for this doctoral. Without you, none of this would be possible. I extend my words to Doctor Rajiv Dhir for all the support and to all the BIOCOPE group, Dr. Elisabete Oliveira, Dr. Javier Lodeiro, Dr. Adrian Lodeiro, Gonçalo Marcelo, Silvia Nutti, and, obviously, my two war companions Gonçalo Martins e Luis Carvalho. I have also to thank my former colleague and friend Eduardo Araújo, this journey has started with you and was somehow linked until the end. Three other girls have marked my life during this period, Joana Flores with her contagious smile and happiness, and the "proteomassinhas" Marta Chaves and Ana Laço. Our friendship started inside of that green room I will always remember our moments, I deeply miss you girls.

To my longtime friends, Eric, Nuno, Hélder, Daniel, Miguel, and my "Marias" Rita, Marilia, Alexandra, and now also the little Eva. Thank you, for all the craziest moments and friendship.

I am also deeply thankful to my beautiful family. To my mom Lina and dad Justino, no words can thank you. You always believed in me; you always have encouraged me. You gave me everything. And to you, my dear sis Vânia, you always had my back. You always knew me better. Thank you "mana". To my little cheeky nephews and godchildren Gustavo e Mateus who are the sweetest human beings apart of the energy, these two little creatures can have!!! And you too, my acquired brother Paulo, thank you for being part of this family. Also, I would like to thank all my closest family, my grandmother Celeste, my aunties Fernanda, Ana e Marlene, my

uncles Zé, Hilário and Miguel, and my cousins, Marília, Paulo, Gonçalo, Tomás, Vasco and Laura. Yes, I have to mention that all of them were important not only during this time but always. “Even Google can’t find a family like mine”. And also, to my newest “in-law” family, who became part of my life. Finally, to those who I keep in my heart, thank you!

“Estou também profundamente grata à minha família maravilhosa. À minha mãe Lina e ao meu Pai Justino, nem tenho palavras para poder descrever a minha gratidão. A vocês que sempre acreditaram em mim e que sempre me encorajaram. Vocês deram-me tudo! E tu minha irmã Vânia, que sempre me apoiaste e que me conheces como ninguém. Muito obrigada Mana! Aos meus sobrinhos e afilhados traquinas, Gustavo e Mateus, que são os seres humanos mais doces depois de toda a energia que duas criaturas pequenas podem ter!!! E tu também, meu mano apegadiço Paulo, obrigada por fazeres parte desta família. Também gostaria de agradecer aos meus familiares mais próximos, à minha avó Celeste, às minhas tias Fernanda, Ana e Marlene, aos meus tios Zé, Hilário e Miguel e aos meus primos Marília, Paulo, Gonçalo, Tomás, Vasco e Laura. E sim, tinha que dizer o nome de todos, porque todos vocês foram importantes não só nesta fase, mas na minha vida inteira. “Uma família como a minha, nem o Google encontra”. E ainda à minha mais recente família, que passou a fazer parte da minha vida também. Por fim, a todos os que guardo no meu coração o meu muito obrigada.”

And you Pedro. You are the most important person in this thesis. You were the one who firstly cheered me in every victory and the one who wiped my tears out in every failure. You always had faith in me even when I had so many doubts. I only was able to start and finish this thesis because of you. Thank you very much for being part of my life. For being a reasonable husband (just kidding, you are The best) and letting me keep dreaming. You are my feet on the ground. And last but not least, to our Kaya! She spent so much time writing this thesis with me or should say sleeping on my lap while I was writing. Even though, she warmed my heart every time I got tired.

Thank you all, this thesis is for all of you!

ABSTRACT

Renal neoplasia is the 14th most common tumor type diagnosed worldwide. With a vast heterogeneity, renal neoplasia encompasses different subtypes. 90% of the neoplasms arise from the epithelial layer of the nephron and vary from benign renal masses (renal oncocytoma, RO) to more indolent or aggressive cancers (renal cell carcinomas, RCC). As RCC subtypes, clear cell (ccRCC) subtype is the most predominant subtype, followed by papillary (pRCC) and chromophobe (chRCC). Despite the different outcomes, some overlapped histological and morphological features difficult their differentiation and diagnosis. Therefore, new approaches for a clear and accurate diagnosis are still needed.

To achieve this goal, renal tissue biopsies diagnosed with ccRCC (n = 7), pRCC (n = 5), chRCC (n = 5), RO (n = 5) and normal adjacent tissue (NAT, n= 5) were enrolled in this study. As a very resourceful tool for proteome analysis and biomarker discovery, mass spectrometry (MS)-based methods were used to interrogate the proteome of each tumor in order to undisclosed differences trough which to develop faster and accurate diagnostics.

The results achieved with this doctoral thesis include i) the accomplishment of an effective ultrasonic workflow to recover the proteome of optimal cutting temperature (OCT)-embedded tissues, ii) a novel analytical approach based on MALDI-MS profiling to distinguish chRCC from RO, iii) a 109-protein panel to discriminate between chRCC and RO and NAT, iv) a top 24-protein panel to diagnose ccRCC, pRCC, chRCC and RO based on absolute concentration values, v) the translation and validation of three promising biomarkers by immunohistochemical analysis, and vi) an approach for phosphopeptide enrichment.

This work brings new insights into the different mechanisms underlying formation of these tumors as well as it provides valuable information to improve clinical diagnosis by opening new avenues for immunohistochemistry and mass spectrometry-based approaches.

Keywords: Renal tumors, Differential diagnosis, Mass spectrometry, Proteome analysis.

RESUMO

A neoplasia renal é o 14º tumor mais diagnosticado no mundo. Com uma elevada heterogeneidade, a neoplasia renal abrange diferentes subtipos. 90% das neoplasias são provenientes da camada epitelial do nefrónio, variando desde massas benignas (oncocitoma renal - OR) - a cancros mais indolentes ou agressivos (carcinomas de células renais - CCR). Dentro dos subtipos de CCR, o subtipo de células claras (CCRcc) é considerado o mais predominante, seguido pelo papilar (CCRp) e pelo cromófobo (CCRcr). Contudo, a similaridade de algumas características histológicas e morfológicas dificultam a sua caracterização, e consequentemente, o seu diagnóstico pelo que são necessárias novas abordagens para um diagnóstico inequívoco.

Para atingir este objetivo, biópsias de tecido renal diagnosticadas com CCRcc (n = 7), CCRp, (n = 5), CCRcr (n = 5), OR (n = 5) e tecido adjacente normal (TAN, n = 5) foram incluídos neste estudo. Como ferramenta muito útil para análise do proteoma e descoberta de biomarcadores, a espectrometria de massa foi utilizada para analisar os proteomas de cada tumor em estudo.

Os resultados alcançados incluem: i) o desenvolvimento de um procedimento ultrassónico eficaz para a recuperação do proteoma de tecidos preservados em OCT (*optimal cutting temperature medium)*, ii) uma nova abordagem analítica baseada no perfil de MALDI para distinguir CCRcr de OR iii) um painel de 109 proteínas para discriminar entre CCRcr, OR e TAN, iv) um painel de 24 proteínas principais para diagnosticar CCRcc, CCRp, CCRcr e OR com base em valores de concentração absoluta v) aplicação e validação, por imunohistoquímica, de três potenciais biomarcadores e vi) uma abordagem quantitativa para enriquecimento de fosfopeptidos.

Este trabalho revela novas perspetivas sobre os diferentes mecanismos subjacentes à formação desses tumores. Além disso, estas descobertas fornecem informações valiosas para melhorar o diagnóstico clínico, abrindo novos caminhos para a imunohistoquímica e sua implementação na medicina.

Palavras-chave: Tumores renais, Diagnostico diferencial, Espectrometria de massa, Análise proteómica.

TABLE OF CONTENTS

FIGURES INDEX

Supplementary figures

TABLES INDEX

Supplementary Tables

Electronic supplementary Tables

ABBREVIATIONS

A

AACT	Amino acid-coded mass tagging
ACN	Acetonitrile
AIBN	2,2'-Azobis2-methyl-propionitrile
AJCC	American Joint Committee on Cancer
Ambic	Ammonium bicarbonate
AMCAR	Alpha-methylacyl-coa racemase
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
APEX	Absolute protein expression
APPI	Atmospheric pressure photoionization
APS	Ammonium persulphate
AQUA	Synthetic peptides
ATP6V0A4	V-type ATPase 166 kDa subunit a isoform 4

B

BCA	Bicinchoninic acid
BSA	Bovine serum albumin

C

CA9	Carbonic anhydrase IX
ccRCC	Clear cell renal cell carcinoma
ccRCC	Clear cell cell carcinoma
CDH1	Cadherins 1
CDH16	Cadherins 1
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
chRCC	Chromophobe cell carcinoma
CI	Chemical ionization
CID	Collision-induced dissociation
CKB	Creatine kinase B
CKMT2	Creatine kinase S
CLDN7	Claudins 7
CLDN8	Claudins 8
CT	Computed tomography

D

DART	Direct analysis in real-time
DC	Direct current
DDA.	Data-dependent acquisition
DESI	Surface ionization methods, including the desorption electrospray ionization
DHB	2,5-dihydrobenzoic acid
DIA.	Data-independent acquisition
DLAT	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex
DTT	Dithiodeithrol

E

ECD	Electron capture dissociation
EI	Electron ionization
ELISA	Enzyme-linked immunosorbent assay
emPAI	Exponentially modified protein abundance index
ESI	Electrospray ionization
ETD	Electron transfer dissociation
EAU	European Association of Urology

F

FA	Formic acid
FAB	Fast atom bombardment
FASP	Filter-aided sample preparation
FD	Field desorption
FDG	18F-2-fluoro-2-deoxyglucose
FDR	False discovery rate
FFPE	Formalin-fixed and paraffin-embedded
FI	Field ionization
FNA	Fine needle aspiration
FT	Fourier transform

G

GC	Gas chromatography
GMA	Glycidyl methacrylate
GO	Gene ontology

H

HILIC	Hydrophilic interaction liquid chromatography
HK1	Hexokinase 1
HPLC	High-performance LC

I

IAA	Iodoacetamide
ICAT	Isotope-coded affinity tag
ICR	Ion cyclotron resonance
IDH3A	Isocitrate dehydrogenase [NAD] subunit alpha
IDH3B	Isocitrate dehydrogenase [NAD] subunit beta
IEX	Ion-exchange chromatography
IHC	Immunohistochemical
IHC	Immunohistochemistry
IMAC	Immobilized metal affinity chromatography
IR	Infrared
IRMPD	Infrared multiphoton dissociation
ISUP	International Society of Urological Pathology
IT	Ion trap
iTRAQ	Isobaric tags for relative and absolute quantification

K

KEGG	Kyoto encyclopedia of genes and genomes
KIT	Transcription factors mast/stem cells grow factor receptor
KRT7	Keratin type II cytoskeletal 7

L

LAMP1	Lysosome-associated membrane glycoprotein 1
LC	Liquid chromatography
LFQ	Label-free quantification
LGALS3	Galectin-3
LIFDI	Liquid introduction field desorption ionization
LIT	Linear ion trap
LLE	Liquid-liquid extraction
LOH	Loss of heterozygosity

M

M	Metastases
m	Mass of the ion
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MAPs	Most abundant proteins
MCP.	Multichannel plate detector
MME	Neprilysin
MOAC	Metal oxide affinity chromatography
MRI	Magnetic resonance imaging
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS1	First scan
MS2	Second scan
MWCO	Molecular weight cutoff

N

N	Regional lymph nodes involvement
NAT	Normal adjacent tissues

O

OCT	Optimal cutting temperature
OXPHOS	Oxidative phosphorylation

P

PAX2	Paired box protein 2
PAX8	Paired box protein 8
PCA	Principal component analysis
PDI	Polydispersity
PEG	Polyethylene glycol
PET	Positron emission tomography
PLCG2	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2
PLIN2	Perilipin-2
PMF	Peptide mass fingerprinting
PP	Protein precipitation
PPIs	Protein-protein interactions

ppm	Parts per million
pRCC	Papillary renal cell carcinoma
PRM	Parallel reaction monitoring
PSAQ	Protein standard absolute quantification
PSM	Peptide-to-spectrum match
PSMA	Prostate-specific membrane antigen
Ps-NPs	Polystyrene nanoparticles
PTMs	Post-translational modification
PVA	Polyvinyl alcohol
PVALB	Parvalbumin

Q

Q	Quadrupole
QCAT	Artificial concatemer of standard peptides
QIT	Quadrupole ion trap
QqQ	Triple quadrupole
RCC	Renal cell carcinomas
RF	Radio frequency
RO	Renal oncocytoma
ROC	Receiver operating characteristics
RPC	Reverse-phase chromatography

S

SAX	Strong anion exchange chromatography
SC	Spectral counting
SCX	Strong cation exchange chromatography
SDC	Sodium deoxycholate
SDH	Succinate dehydrogenase
SDS,	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling by amino acids in cell culture
SN	Supernatant
SPE	Solid-phase extraction
SRM	Selected reaction monitoring

T

T	Tumor extension

TCA	Trichloroacetic acid
TEMED	N,N,N',N' – tetramethylethylene diamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIS	Timed ion selector
TMAs	Tissue micro arrays
TMPTMA	Trimethylolpropane trimethacrylate
TMT	Tandem mass tags
TOF	Time of flight
TPA	Total protein approach

U

UBLs	Ubiquitin-like proteins
UE	Ultrasound energy
UHPLC	Ultra-high-performance LC
UniProt	Universal protein resource
UPMC	University of Pittsburgh Medical Center
US	Ultrasound
UV	Ultraviolet

V

v/v	Volume per volume
VHL	Von Hippel–Lindau
VIM	Vimentin

W

w/v	Weight per volume
w/w	Weight per weight
WHO	World health organization

X

XIC	Extracted ion chromatogram

Z

z	Electrical charge
α-CHCA	α -Cyano-4-hydroxy-cinnamic acid

1D	One dimensional
2D	Two dimensional
3D	Tree-dimensional

CHAPTER I.

General Introduction

I.1 KIDNEY CANCER

The biological homeostasis is fundamental for organisms where the body's cells are programmed to divide or die to maintain a balanced and controlled organic system. When cells stop responding appropriately to cell regulatory system signals it may lead to uncontrolled cell growth and division [1]. The continual unregulated proliferation of cells drives the development of tumoral masses, which can be benign neoplasms if they are limited and confined to its original location, or malignant if the abnormal cellular growth is capable of invading surrounding normal tissues and ultimately spread through the body. Malignant tumors are cancerous due to their ability to invade and metastasize making them so harmful [2]. Apart from the essence of the tumor cells, neoplasms can be classified accordingly to the type of cells from which they arise. Solid tumors in epithelial cells are called carcinomas, while in connective tissues sarcomas [2]. For the malignancies in cells from the blood and immune systems, they are named as leukemias and lymphomas, respectively. On the other hand, tumors can also be classified by the tissue/organ of origin, such as lung, breast, kidney, among others.

Kidneys are a vital organ for body homeostasis through control of blood pressure and blood composition [3]. They play a role in hormone production that stimulates erythropoiesis, the blood's extraneous materials filtration, electrolyte balance maintenance, and blood pressure regulation [4]. When abnormal pathologies arise, such as kidney cancer, integrity and organ function is compromised, resulting in detrimental outcomes. Kidney cancer, also known as renal cancer, is the 14th most common cancer worldwide which accounts for over 400,000 new cases every year [5]. However, a particular challenge in renal neoplasia is the vast heterogeneity of this cancer type. Different subtypes of renal cancer can be classified according to histological origins, genetic alterations, and clinical course.

I.1.1 RENAL NEOPLASMS: CLASSIFICATION AND CHARACTERISTICS

The classification of renal neoplasia is based on morphology, immunohistochemistry, cytogenetics, and molecular pathology of each tumor subtype [6,7]. In 2004, the World Health Organization (WHO) published the histologic classification of renal tumors [8]. Over the past decade, new tumor types have been discovered and recognized as distinct tumor entities according to their clinicopathological features, and the WHO classification was revised in 2016 (Table I.1) [7].

Table I.1- 2016 World Health Organization classification of renal neoplasms [7].

Renal cell tumors
Clear cell renal cell carcinoma
Multilocular cystic renal neoplasm of low malignant potential
Papillary renal cell carcinoma
Hereditary leiomyomatosis renal cell carcinoma (HLRCC)-associated renal carcinoma
Chromophobe renal cell carcinoma
Collecting duct carcinoma
Renal medullary carcinoma
MiT Family translocation carcinomas
Succinate dehydrogenase (SDH) deficient renal carcinoma
Mucinous tubular and spindle cell carcinoma
Tubulocystic renal cell carcinoma
Acquired cystic disease associated with renal cell carcinoma
Clear cell papillary renal cell carcinoma
Renal cell carcinoma, unclassified
Papillary adenoma
Oncocytoma
Metanephric tumors
Metanephric adenoma
Metanephric adenofibroma
Metanephric stromal tumor
Nephroblastic and cystic tumors occurring mainly in children
Nephrogenic rests
Nephroblastoma
Cystic partially differentiated nephroblastoma
Cystic nephroma, pediatric type
Mesenchymal tumors
Mesenchymal tumors occurring mainly in children
Clear cell sarcoma
Rhabdoid tumor
Congenital mesoblastic nephroma
Ossifying renal tumor of infants
Mesenchymal tumors occurring mainly in adults
Leiomyosarcoma (including renal vein)
Angiosarcoma
Rhabdomyosarcoma
Haemangiopericytoma
Undifferentiated pleomorphic sarcoma
Osteosarcoma

Table I.1 (cont.)

Synovial sarcoma
Primitive neuroectodermal tumor (Ewing sarcoma)
Angiomyolipoma
Epithelioid angiomyolipoma
Leiomyoma
Haemangioma
Lymphangioma
Haemangioblastoma
Juxtaglomerular cell tumor
Renomedullary interstitial tumor
Schwannoma
Solitary fibrous tumor
Mixed epithelial-stromal tumor family
Adult cystic nephroma
Mixed epithelial and stromal tumor

Renal cell tumors are the most dominant forms of neoplasms, accounting for 90% of all kidney cancers [9]. It is most commonly diagnosed in patients in their early sixties afflicting males about twice as often as females [10]. As a heterogeneous disease, renal cell tumors can vary from benign renal masses [e.g. renal oncocytoma (RO)] to more indolent or aggressive tumors such as the renal cell carcinomas (RCC) subtypes [e.g. chromophobe RCC (chRCC), papillary RCC (pRCC) or clear cell RCC (ccRCC)].

Arising from the epithelial layer of the nephron, these tumor entities display traits associated with particular cell types, as seen in Figure 1 [11]. While ccRCC and pRCC display proximal tubule traits, the cellular origin of chRCC and RO are associated with the distal nephron, in particular, in the intercalated cells of the connecting tubules and collecting duct system [11].

ccRCC is considered the most aggressive and common RCC variant encompassing 75% of all RCC neoplasms [12]. Histologically, is defined by the clear cytoplasm due to their lipid and glycogen-rich content (Figure I.2) [13]. Although, 95% of the cases are sporadic, the remaining 5% are associated with hereditary syndromes, such as von Hippel-Lindau disease or more rarely with Cowden's syndrome, Birt-Hogg-Dubé syndrome, tuberous sclerosis complex, and succinate dehydrogenase-deficient RCC [14]. The molecular genetic hallmarks of ccRCC are mutations in VHL gene mutations, hypermethylation of VHL gene promoter, and loss of heterozygosity (LOH) at chromosome arm 3p. Loss of 14q, and gains of 5q and 7q, are also commonly observed in this variant [11]. Apart from the functional inactivation of the von Hippel–Lindau (VHL) tumor-suppressor gene, located on 3p21, this region encompasses three

additional genes, *PBRM1*, *BAP1*, *SETD2*, and prevalent mutations in these chromatin- and histone-regulating genes have also been reported [15].

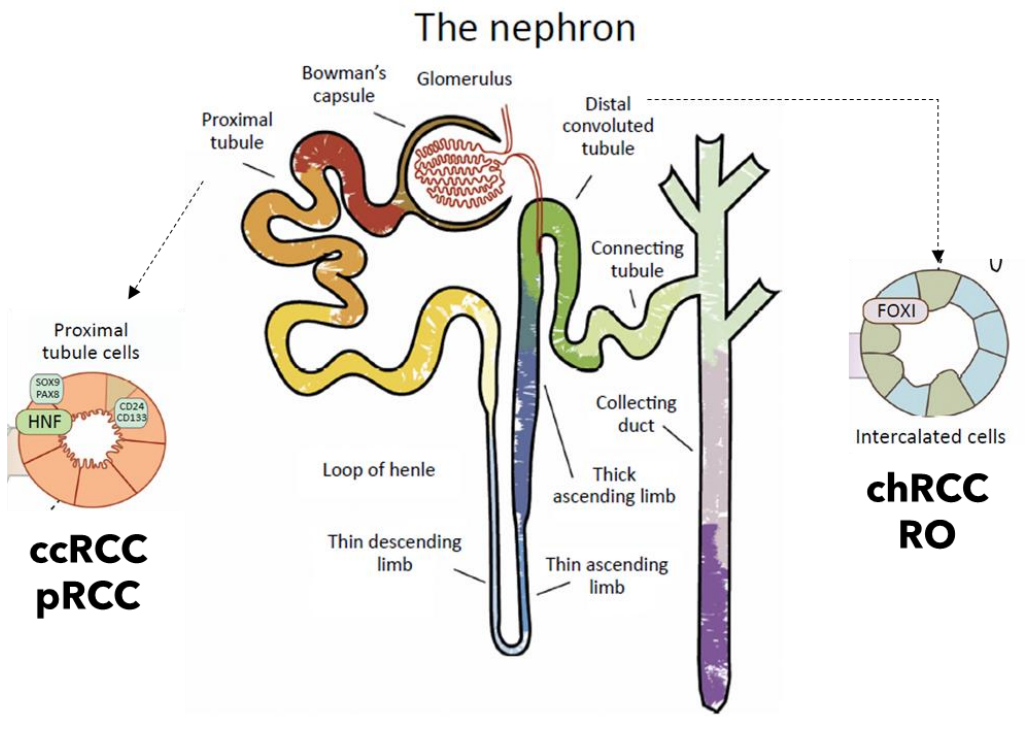

Figure I.1 – Cellular origin of RCC subtypes in the renal nephron. ccRCCs and pRCCs originate from proximal tubule or parietal cells (bottom left), and chRCCs and RO stem from intercalated cells of the distal nephron and collecting ducts (bottom right). Adapted from Lindgren et al. [11].

The second most common type of RCC is the pRCC, which accounts for approximately 10% of the cases [12]. Although pRCC is characterized by papillary structures with a spindle-shaped pattern, this variant presents two subtypes, type 1 and type 2, based on histological appearance and biological behavior. Generally, type 1 pRCC is less aggressive and displays a single layer of basophilic cells with scarce clear cytoplasm and hyperchromatic nuclei surrounding the basal membrane. In its turn, the more aggressive histological subtype 2 is more heterogeneous and is characterized by papillae covered by cells with abundant granular eosinophilic cytoplasm, with prominent nucleoli associated with areas of necrosis (Figure I.2) [13]. Genetic profiles also reveal some differences between the two types. Gains of 7p and 17p and mutations in the *MET* gene are typically shown in Type 1 tumors. While cytogenetic aberrations of type 2 tumors are associated with gains of 8q, loss of heterozygosity of chromosomes 1p and 9p, and mutation in *CDKN2A*, *SETD2*, and *NRF2* genes [16,17].

As a more indolent behavior, but still with the ability to metastasize, the chRCC is the third most common RCC subtype with an incidence of approximately 5% [12]. Large pale cells with reticulated cytoplasm and peniculear halos are characteristic of these tumors (Figure I.2).

Cytogenetic alterations account for losses in chromosomes 1, 2, 6, 10, 13, 17, and 20. At a molecular level, mutations in *TP53* and *PTEN* genes have also been reported [15].

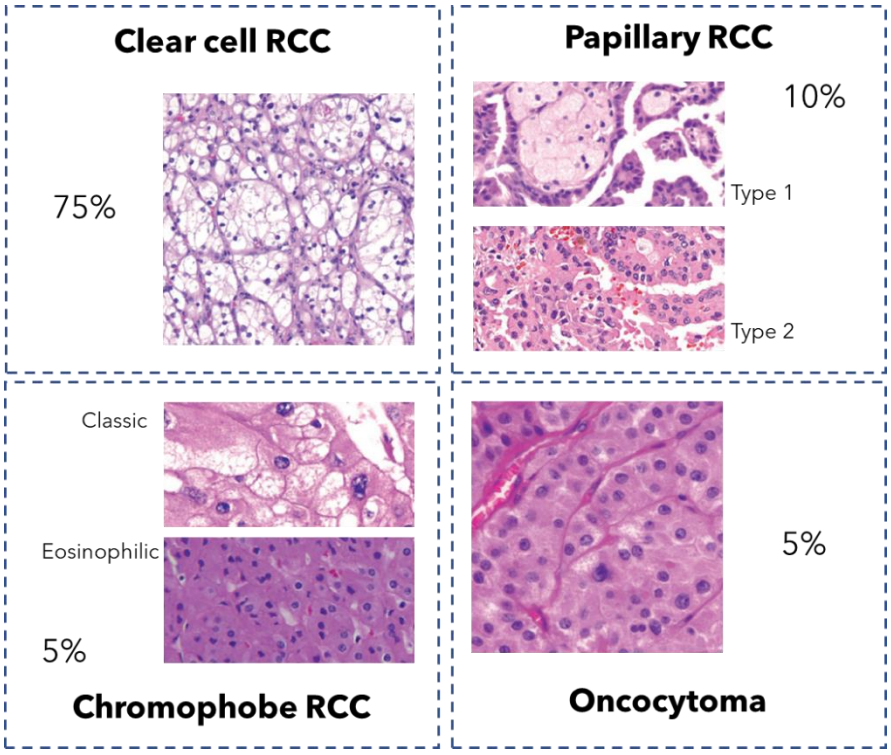

Figure I.2 – Histologic subtypes of renal tumors [18].

Lastly, renal oncocytoma is a benign renal tumor that represents 5% of renal tumors' diagnosis. The appearance of RO is of round to polygonal cells with densely granular eosinophilic cytoplasm, and round uniform nuclei (Figure I.2) [19]. Genetic factors associated with these tumor types include losses of chromosome 1 and Y, rearrangement of the *CCND1* gene, as well as mutations in mitochondrial genes (*COX1, COX2, MTND4,* and *MTCYB*) [15].

I.1.2 Diagnosis, Staging, and Grading systems

Renal cell tumors are often diagnosed incidentally during thoracoabdominal imaging ordered for unrelated complains due to their asymptomatic profile especially in the early stages [20]. Symptoms tend to appear with the progression of the disease and can be the result of local tumor growth, hemorrhage, paraneoplastic syndromes or metastatic evolution [20]. Usually, the ureteral, vascular, or thromboembolic obstruction cause flank pain which can develop into hematuria [20]. The physical examination of palpable abdominal masses is then complemented by non-invasive imaging techniques. According to the latest guidelines prepared by the European Association of Urology (EUA) in 2019 [9], computed tomography (CT), abdominal

ultrasound, and magnetic resonance imaging (MRI) are the imaging modalities used to detect and characterize renal masses. On the basis of imaging findings, renal masses can be classified as solid masses or in more atypical forms as cystic lesions. In the case of solid masses, the most important criterion used for differentiating malignant lesions in the presence of enhancement in imaging scans [21]. The diffusion of contrast agents in pathologies, like cancer, leans towards a bigger enhanced area due to the formation of new blood vessels of tumor angiogenesis [22]. On the other hand, Bosniak classification is used to classify cystic masses, where renal lesions are classified into five categories to predict malignancy risk [23,24].

I.1.2.1 TNM classification system

In renal masses, the most important factors in predicting the clinical behavior and outcome are based on tumor size and extent of invasion. Over the years, the TNM staging systems, which reflect the evaluation of primary tumor extension (T), regional lymph nodes involvement (N), and presence of distant metastases (M) have become the predominant grading system for RCC. In Table I.2 is represented the TMN staging system for renal cancer according to the American Joint Committee on Cancer (AJCC) Staging guidelines [25].

Tumors with 4 cm or less are classified as T1a, over 4 cm but less than 7 cm are designed as T1b, between 7 cm and 10 cm are classified as T2a and more than 10 cm are designed as T2b. In T1 and T2 stages, the tumor is confined within the kidney. When tumors extended beyond the renal parenchyma to involve major veins, the pelvicalyceal system, or perinephric tissues, they are considered as T3. Staging T3 is subdivided into: T3a if tumors do not extend into vena cava, T3b when the extension reaches the vena cava below the diaphragm, and T3c when the extension reaches the vena cava above the diaphragm or invades the wall of the vena cava. Lastly, T4 is assessed when tumors directly invade the ipsilateral adrenal gland or invade beyond Gerota fascia. Stage N0 state for no regional lymph nodes and N1 for the involvement of regional lymph nodes. NX denotes that regional lymph nodes cannot be assessed. The presence of distant metastases is classified as M0 when they are absent or M1 for their presence. Clinical staging is then assigned to the TNM categories as presented in Table I.3 and Figure I.3.

I.1.2.2 Histopathological ascertainment of tissue biopsies

Imaging techniques are frequently used in routine clinical practice, resulting in accurate diagnosis in most cases. However, some pitfalls arise in the characterizations of renal masses through these techniques. For instance, RO cannot reliably be distinguished from malignant renal neoplasms through imaging approaches. In addition, inconsistencies of CT Hounsfield unit measurements, pseudoenhancement of simple renal cysts at CT, and the lack of a standardized approach for evaluating renal mass enhancement contribute to the potential misdiagnosis. In this case, histological diagnosis is recommended to avoid unnecessary surgery in the event of

benign lesions, to select patients for surveillance, and to obtain histology before ablative treatment [9].

Table I.2 – TNM classification system [25].

	Definition	Subdivision	Extension
Tumor stage (T)			
T1	Tumor < 7 cm in greatest dimension	T1a: tumor ≤ 4 cm	Limited to the kidney
		T1b: tumor > 4 cm but ≤ 7 cm	
T2	Tumor > 7 cm in greatest dimension	T2a: tumor > 7 cm but ≤ 10 cm	Limited to the kidney
		T2b: tumor > 10 cm	
T3	Tumor extends into major veins or perinephric tissues	T3a: Tumor in renal vein or renal sinus fat	Not beyond Gerota's fascia
		T3b: Tumor extends into the vena cava	Not beyond Gerota's fascia, below the diaphragm
		T3c: Tumor extends into the vena cava	Not beyond Gerota's fascia, above the diaphragm
T4	Tumor invades beyond Gerota's fascia		Including contiguous extension into the ipsilateral adrenal gland
Regional lymph nodes (N)			
N0	No regional lymph node metastasis		
N1	Metastasis in regional lymph node(s)		
NX	Regional lymph nodes cannot be assessed		
Distant metastasis (M)			
M0	No distant metastasis		
M1	Distant metastasis		

For histopathological analysis, it is necessary the collection of renal tissue. Percutaneous renal tumor biopsies are minimal invasive medical procedures, with a low risk of complications, where the renal mass is accessed via needle-puncture of the skin [26]. Commonly, the needle core biopsy and the fine needle aspiration (FNA) biopsy are the most used methods for renal tumor tissue, however, the core biopsies are preferable over FNA biopsies due to their superior diagnostic yield [27]. Providing additional information, the

histochemical analysis performed on tissue biopsies can predict clinical outcomes based on numerous prognostic factors including the evaluation of i) nuclear grade, ii) sarcomatoid features, iii) tumor necrosis, and iv) microvascular invasion [9,28]. Grading renal tumors based on nuclear morphology has been recognized as a prognostic factor. Over the past years, the Fuhrman grading system [29] has been massively applied in pathology practice for the RCC grading system. However, its prognostic value has been questioned [30]. To overcome the difficulties in associating the former grading system, the WHO and the International Society of Urological Pathology (ISUP) have reached a consensus and recommended the use of the four-tiered WHO/ISUP grading system [31]. Although the novel ISUP grading system has been implemented by the WHO for ccRCC and pRCC subtypes, this grading evaluation has not been recommended for other variants such as chRCC [32].

Table I.3 – AJCC\ Stage groups [25].

Stage	Tumor stage	Lymph Nodes	Metastasis
I	T1	N0	M0
II	T2	N0	M0
III	T1	N1	M0
III	T2	N1	M0
III	T3	N0	M0
III	T3	N1	M0
IV	T4	Any N	M0
IV	Any T	Any N	M1

In this system, grades from 1 to 3 are based on nucleolar prominence. Like so, tumors having i) inconspicuous or absent nucleoli at x 400 magnification are classified as grade 1, ii) inconspicuous nucleoli at x 100 magnification but distinctly visible at x 400 are classified as grade 2, and iii) nucleoli distinctly visible at low-power magnification (x 100) are classified as grade 3. The fourth grade of the WHO/ISUP grading system encompasses tumors extreme pleomorphic giant cells or the presence of rhabdoid and/or sarcomatoid differentiation [31]. Characterized by atypical spindles cells and resemblance to any form of sarcoma, sarcomatoid morphology can be found in all RCC subtypes and represents tumor aggressiveness with a worse clinical outcome [33,34]. The presence of tumor necrosis has also been associated with a less favorable clinical outcome in RCC [35,36]. Microvascular invasion is defined as the presence of small vessels within the renal parenchyma. In several studies, this parameter has been found associated with prognostic value [28,37,38], however, fully recommendation to use the microvascular invasion to prognosis RCC tumors was not achieved in the consensus conference [31].

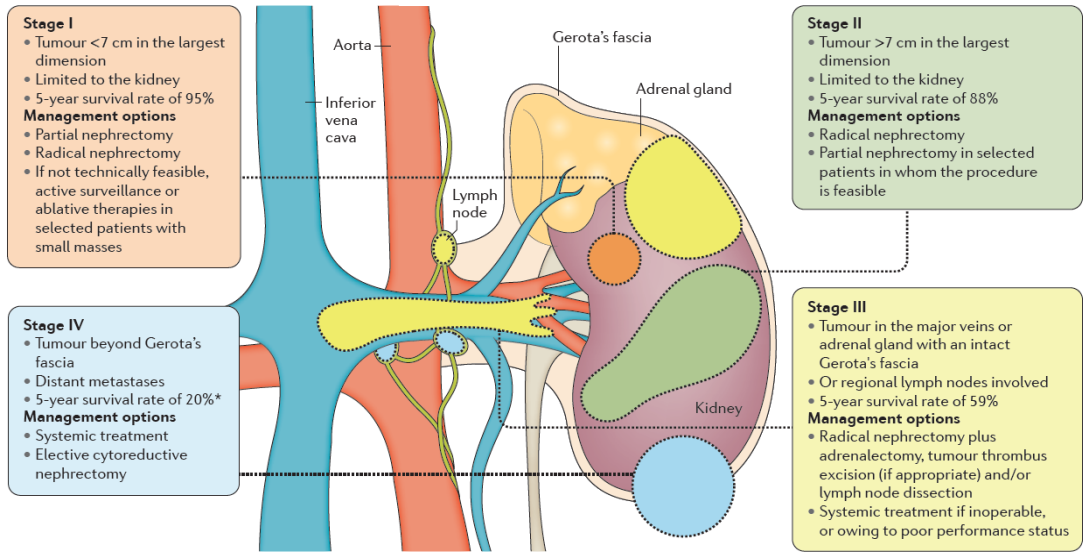

Figure I.3 - Kidney cancer stages. Adapted from [15].

I.1.3 BIOMARKERS IN KIDNEY NEOPLASMS

A specific challenge in diagnosing and prognosis renal cell tumors is the wide spectrum of histological subtypes with different outcomes. The proper identification of the histologic tumor type is crucial in clinical practice to avoid undertreatment of most aggressive forms or overtreatment of more indolent or even benign subtypes. Therefore, when the analysis of imaging scans *per se* reveals to be inconclusive, the use of specific markers represents a support in renal cell tumors' diagnosis and further clinical decision making [39].

I.1.3.1 Imaging markers

On par with the detection of tumoral masses, imaging techniques can be used to the identification of molecular features specific for each tumor, in a non-invasive manner and through the mean of radioactive tracers [39]. Resulting from the combination of a chemical compound and a radioactive element, the radiotracers are used to track pathological conditions such as cancer. The most common radiotracer is the ^{18}F-2-fluoro-2-deoxyglucose (FDG) which take the advantage of the high glucose metabolism rates of cancer to flag those regions, particularly in advance stages with metastasis [40]. However, the use of FDG on positron emission tomography (PET)/CT is not recommended for RCC diagnosis due to their low sensitivity - only 22% - for localized RCC [41]. Other explored molecules are the ^{124}Iodine cG250 [42] and ^{18}F-VM4-037 [43]. These radiotracers have been used to trace the carbonic anhydrase IX (CA9), a protein that has been verified to be highly and homogenously expressed in ccRCC [44]. However, these radioactive markers have some disadvantages such as a proper contrast between tumor and normal parenchyma or prolonged waiting times for an adequate

background visualization [45]. The molecular imaging biomarker prostate-specific membrane antigen (PSMA)-targeted ^{18}F-DCFPyL is also been used to detect metastasis in RCC patients [46]. However, the small cohort of patients and this preliminary study, further evaluation of the sensitivity is required to fully image RCC with PSMA-based PET techniques. Molecular imaging with PET has been proven as a powerful tool to trace cancerous processes particularly in more aggressive forms as ccRCC or metastatic stages. Unfortunately, this technique still presents some limitations and a reduced capacity to stratify the wide range of RCC subtypes.

MRI is another non-invasive technique that can be applied to characterize and assess renal masses [39]. However, the major obstacle of MRI implementation in clinical practice is the technical aspects and expertise required to consistently obtain high-quality images of kidney tumors [47].

I.1.3.2 Blood-based and urine protein markers

The initial diagnosis of renal masses is frequently based on noninvasive imaging techniques [12]. Nevertheless, these techniques often lack in effectively detect them in the early stages of the disease. Liquid biopsies such as serum and urine, also represent valuable sources of pathological information, being routinely implemented in the clinical practice [48]. The evaluation of potential biomarkers in these biological samples has been widely approached in biomarker discovery for many pathologies including renal cell tumors. Although the discovery and evaluation of potential biomarkers in these biological fluids are often challenging process due to the frequent very low concentration of the proteins of interest or the camouflage of such proteins by the high-abundant proteins [49], some efforts have been done in this field. Blood-derivative specimens, such as serum and plasma, are very accessible clinical samples and due to the direct contact with disease-affected tissues, specific proteins can be secreted into the bloodstream, making them a valuable source for biomarker discovery. Several studies have been reporting potential serological biomarkers for the diagnosis of RCC [50]. To a less extent, some efforts have also been done in plasma samples. In Table I.4 are summarized the most promising protein biomarker for renal cell tumors diagnosis.

However, none of the proposed biomarkers has been proven to be reliable and specific for RCC differentiation.

The urinary fluid is also considered an attractive biological matrix for candidate soluble protein biomarkers, especially in renal cell tumors which take the advantage of the close contact and communication of the tumoral cells with the urine-forming tubular elements of the kidney [50]. Potential urinary biomarkers are presented in Table I.5

Unfortunately, some technical aspects, such as the lack of reproducibility and the complexity of the urine, represent the major shortcomings in this fluid for biomarker discovery [49]. Additionally, some reported markers have also been used to evaluate other cancer types or kidney injuries [51,52], compromising their specificity as a diagnostic marker for RCC.

Table I.4 - RCC-associated protein biomarkers in blood-derivative samples.

Protein marker	Profile	Ref
Serum fluid		
20S proteasome	↑ ccRCC vs HC	[53]
72 kDa type IV collagenase (MMP2)	↑ RCC vs HC	[54]
Arrestin-1 (ARRB1) autoantibody	↑ RCC and RO vs HC	[55]
Baculoviral IAP repeat-containing protein 5 (BIRC5) also known as survivin	↑ RCC vs HC	[56]
CD276 antigen (CD276) also known as B7H3	↓ ccRCC vs HC	[57]
Ceruloplasmin (CP)	↑ ccRCC vs HC	[58]
Complement C1q subcomponent subunit B (C1QB)	↑ ccRCC vs HC	[58]
Complement C1q subcomponent subunit C (C1QC)	↑ ccRCC vs HC	[58]
C-X-C motif chemokine 13 (CXCL13)	↑ ccRCC vs HC	[59]
Cytotoxic T-lymphocyte protein 4 (CTLA4)	↓ ccRCC vs HC	[57]
DnaJ homolog subfamily C member 7 (DNAJC7)	↑ RCC vs OKI and HC	[60]
Endothelial cell-specific molecule 1 (ESM1)	↑ RCC vs HC	[61]
Heat shock cognate 71 kDa protein (HSPA8)	↑ ccRCC vs HC	[62]
Heat shock protein beta-1 (HSPB1)	↑ ccRCC vs HC	[63]
High mobility group protein B1 (HMGB1)	↑ mRCC vs HC	[64]
Hypoxia-inducible factor propyl hydroxylase-3 (PHD3)	↑ RCC vs HC	[65,66]
Intact form KRT18 (M65)	↑ mRCC vs HC	[67]
Interleukin-2 receptor subunit alpha (IL2RA) also known as CD25	↑ ccRCC vs HC	[57]
Kallistatin (SERPINA4)	↓ ccRCC vs HC	[58]
Lumican (LUM)	↓ ccRCC vs HC	[58]
Matrix metalloproteinase-9 (MMP9)	↑ RCC vs HC	[54]
Osteopontin (SPP1)	↑ pRCC vs ccRCC and chRCC	[68]
Protein S100-A8 (S100A8)	↑ ccRCC vs HC	[58]
Protein S100-A9 (S100A9)	↑ ccRCC vs HC	[58]
Tumor necrosis factor receptor-associated factor 1 (TRAF1)	↑ ccRCC vs HC	[69]
Vascular endothelial growth factor A (VEGF)	↑ RCC vs HC	[54]
Zymogen granule protein 16 homolog b (ZG16B)	↑ ccRCC vs HC	[58]
Plasma fluid		
Carbonic anhydrase IX (CA9)	↑ ccRCC vs HC and BT	[70]
Fibronectin 1 (FN1)	↑ RCC vs HC	[71]
Hepatitis A virus cellular receptor 1 (HAVCR) also known as kidney injury molecule-1 (KIM-1)	↑ ccRCC vs HC and BT ↑ ccRCC vs HC	[72] [73]

RCC: renal cell carcinoma; ccRCC: clear cell RCC; mRCC: metastatic RCC; pRCC: papillary RCC; *chRCC: chromophobe RCC; RO: renal oncocytoma; HC: healthy controls; OKI: other kidney injuries; BT: benign tumors.*

I.1.3.3 Tissue and immunohistochemical marker

Proper differentiation of tumor subtypes is crucial for medical decision-making. As aforementioned in I.1.2.2. Histopathological ascertainment of tissue biopsies section, tumoral tissue biopsies are usually used to support tumor classification. Since tumor-derived biomarkers are expected to be at higher concentrations in tumoral tissues [9], the use of immunohistochemical (IHC) stains is often applied for the differential diagnosis [80]. In Table I.6 are presented the most common markers used in immunohistochemical analysis.

Table I.5 – RCC-associated protein biomarkers in urinary samples.

Protein marker	Profile	Ref
Aquaporin-1 (AQP1)	↑ ccRCC and p RCC vs chRCC, RO and HC	[74]
Cystatin-C (CST3)	↑ ccRCC vs HC	[75]
Glutaredoxin-1 (GLRX)	↑ ccRCC vs HC	[75]
Hepatitis A virus cellular receptor 1 (HAVCR) also known as kidney injury molecule-1 (KIM-1)	↑ ccRCC vs HC	[76,77]
Nuclear mitotic apparatus protein 1 (NUMA1) also known as nuclear matrix protein 22 (NMP22)	↑ RCC vs HC	[78,79]
Perilipin-2 (PLIN2)	↑ ccRCC and p RCC vs chRCC, RO and HC	[74]

RCC: renal cell carcinoma; ccRCC: clear cell RC; pRCC: papillary RCC; chRCC: chromophobe RCC; HC: healthy controls; RO: renal oncocytoma.

Currently, the immuneprofile-based classification of RCC is based on (i) structural molecules, such as keratin type II cytoskeletal 7 (KRT7), vimentin (VIM), cadherins 1 and 16 (CDH1 and CDH16), and claudins 7 and 8 (CLDN7 and CLDN8); (ii) proteins related to metabolic pathways, such as alpha-methylacyl-CoA racemase (AMCAR), CA9 and parvalbumin (PVALB), (iii) the transcription factors mast/stem cells grow factor receptor (KIT), and (iv) other molecules including neprilysin (MME), paired box proteins (PAX2 and PAX8), protein S100A1 and RCC marker [49,80,81]. Despite the IHC remains a widely used tool for renal tumors diagnosis, the precision of the classifications is frequently compromised by the ambiguous immunohistochemical results. For instance, the immune marker KRT7 stains positively for pRCC and generally negative for the other subtypes. However, occasional positivity has been also observed in ccRCC and RO entities [15]. Recently, Kim *et al.* [82] proposed an immune algorithm based on the canonical KRT7/CA9/AMACR triple panel, even so, an adjunctive panel

is necessary in equivocal cases. Therefore, the IHC triage in renal cell tumors is still vulnerable to misdiagnosis and a precise classification is still necessary.

The unmet demand to identify molecular changes associated with the different phenotypes has rendered numerous proteomic studies through the research community. Table I.7 summarizes the more relevant ones performed in the latest years for RCC tissue biomarker discovery.

I.1.4 MEDICAL CARE

Once the renal mass is diagnosed, medical decision-making is based on the histotypes and progression of the disease. Currently, surgery is the only curative treatment for localized renal tumors. However, due to the negative consequences often allied to surgical excision, active surveillance would represent a better approach when tumoral masses are limited to the kidney and with a threshold size of 3 cm [83]. Active surveillance is also applied when benign subtypes, like RO, are confirmed by IHC. In this way, with the monitorization of the tumor size and the IHC analysis, the urologist may decide whether to keep the surveillance or to perform surgical procedures instead, Figure I.4.

Table I.6 – Selected immunohistochemical staining panel for renal neoplasms [80].

Protein marker	ccRCC	pRCC	chRCC	RO
Alpha-methylacyl-CoA racemase (AMACR)	-/+	+	-	-/+
Cadherin-16 (CDH16) also known as kidney-specific cadherin (Ksp-cadherin)	-/+	-/+	+	+/-
Cadherin-1 (CDH1) also known as epithelial cadherin (e-cadherin)	-/+	-/+	+/-	+/-
Carbonic anhydrase IX (CA9)	+	+/-	-/+	-/+
Claudin-7 and -8 (CLDN7 and CLDN8)	-	-/+	+	+/-
Keratin, type II cytoskeletal 7 (KRT7) also known as CK7	-/+	+[a]	+	Focal[b]
Mast/stem cell growth factor receptor Kit (KIT) also known as CD117	-/+	-/+	+	-/+
Neprilysin (MME) also known as CD10	+	+	-/+	-/+
Paired box protein PAX-2 and PAX-8 (PAX2 and PAX8)	+	+	+/-	+
Parvalbumin (PVALB)	-/+	-/+	+	+
Protein S100-A1 (S100A1)	+/-	+/-	-/+	+
RCC-Marker[c]	+	+	+/-	-
Vimentin (VIM)	+	+	-/+	-/+

(+) usually positive; (-) usually negative; (+/-) frequently positive; (-/+) occasionally positive, [a] maybe negative in type II pRCC, [b] in contract with diffuse and strong immunoreactivity in chRCC, RO typically shows scattered focal positivity (< 5% of tumor cells staining, [c] monoclonal antibody against a 200 kDa glycoprotein expressed on the brush boarder of proximal tubules cells.

Table I.7 – Tissue proteomic studies in renal tumor masses.

Protein marker	Profile	Ref
Renal cell carcinoma (RCC)		
Arrestin-1 (ARRB1)	↑ RCC vs NAT	[55]
Reticulocalbin 1 (RCN1)	↑ RCC vs NAT	[84]
Clear cell RCC (ccRCC)		
10 kDa heat shock protein, mitochondrial (HSPE1)	↓ ccRCC vs NAT	[63]
14-3-3 protein zeta/delta (YWHAZ)	↑ ccRCC vs NAT	[85]
Acetyl-CoA acetyltransferase, mitochondrial (ACAT1)	↓ ccRCC vs NAT	[86]
ADP/ATP translocase 3 (SLC25A6)	↓ ccRCC vs NAT	[87]
Aldo-keto reductase family 1 member A1 (AKR1A1)	↓ ccRCC vs NAT	[87]
Alpha-enolase (ENO1)	↑ ccRCC vs NAT	[63]
AMMECR1-like protein (AMMECR1L)	↓ ccRCC vs NAT	[87]
Annexin A2 (ANXA2)	↑ ccRCC vs NAT	[88]
Annexin A5 (ANXA5)	↑ ccRCC vs NAT	[87]
Apolipoprotein A1 precursor (APOA1)	↑ ccRCC vs NAT	[89]
Calbindin (CALB1)	↓ ccRCC vs NAT	[87,90]
CD2-associated protein (CD2AP)	↑ ccRCC vs NAT	[87]
Complement component 1 Q subcomponent-binding protein, mitochondrial (C1QBP)	↓ ccRCC vs NAT	[91]
Coronin-1A (CORO1A)	↑ ccRCC vs NAT	[92]
Electron transfer flavoprotein regulatory factor 1 (ETFRF1) also known as LYR motif-containing protein 5 (LYRM5)	↓ ccRCC vs NAT	[87]
Ester hydrolase C11orf54 (C11orf54)	↓ ccRCC vs NAT	[87]
Fatty acid-binding protein, brain (FABP7)	↑ ccRCC vs NAT	[88]
Fatty acid-binding protein, heart (FABP3)	↓ ccRCC vs NAT	[90]
Ferritin (FTL)	↑ ccRCC vs NAT	[89]
Galectin-1 (LGALS1)	↑ ccRCC vs NAT	[85,88]
Gamma-enolase (ENO2)	↑ ccRCC vs NAT	[87]
Gelsolin (GSN)	↑ ccRCC vs NAT	[90]
Glutathione S-transferase P (GSTP1)	↑ ccRCC vs NAT	[87]
Haptoglobin (HP)	↑ ccRCC vs NAT	[89]
Heat shock protein beta-1 (HSPB1) also known as heat shock 27 kDa protein (Hsp27)	↑ ccRCC vs NAT	[63,89]
Hemoglobin subunit beta (HBB)	↑ ccRCC vs NAT	[89]
HIG1 domain family member 1A, mitochondrial (HIGD1A)	↑ ccRCC vs NAT	[87]
L-lactate dehydrogenase A chain (LDHA)	↑ ccRCC vs NAT	[63]
Nicotinamide N-methyltransferase (NNMT)	↑ ccRCC vs NAT	[93]
Peptidyl-prolyl cis-trans isomerase A (PPIA) also known as cyclophilin A (CYPA)	↑ ccRCC vs NAT	[88]
Perilipin-2 (PLIN2) also known as adipose differentiation-related protein (ADFP)	↑ ccRCC vs NAT	[92]

Table I.7 (cont.)

Peroxiredoxin-2 (PRDX2)	↑ ccRCC vs NAT	[89]
Peroxiredoxin-6 (PRDX6)	↑ ccRCC vs NAT	[87]
Profilin-1 (PFN1)	↑ ccRCC vs NAT	[85]
Protein-glutamine gamma-glutamyltransferase 2 (TGM2)	↑ ccRCC vs NAT	[87]
Retinal dehydrogenase 1 (ALDH1A1)	↑ ccRCC vs NAT	[87]
Superoxide dismutase [Mn], mitochondrial (SOD2)	↑ ccRCC vs NAT	[86]
Thioredoxin-dependent peroxide reductase, mitochondrial (PRDX3)	↑ ccRCC vs NAT	[87]
Triosephosphate isomerase (TPI1)	↑ ccRCC vs NAT	[89]
Vimentin (VIM)	↑ ccRCC vs NAT	[87]
Y-box-binding protein 1 (YBX1)	↑ ccRCC vs NAT	[91]
Papillary RCC (pRCC)		
Ferritin (FTL)	↑ pRCC vs NAT	[94]
Nicotinamide N-methyltransferase (NNMT)	↑ pRCC vs NAT	[93]
Protein S100-A11 (S100A11)	↑ pRCC vs NAT	[94]
Chromophobe (chRCC)		
Caspase-3 (CASP3)	↑ chRCC vs RO	[95]
Alpha-1-antitrypsin (SERPINA1)	↑ chRCC vs NAT	[89]
Hepatocyte nuclear factor 1-beta (HNF1B)	↓ chRCC vs RO	[96]
Leptin (LEP)	↑ RO and chRCC and ccRCC vs HC	[97]
Nicotinamide N-methyltransferase (NNMT)	↑ chRCC vs NAT	[93]
Superoxide dismutase (SOD2)	↑ chRCC vs NAT	[89]
UV excision repair protein RAD23 homolog B (RAD23D) also known as p58	↑ chRCC vs NAT	[89]
Oncocytoma (RO)		
Caspase-3 (CASP3)	↓ RO vs chRCC	[95]
Alpha-enolase (ENO1)	↑ RO vs NAT	[89]
Arrestin-1 (ARRB1)	↑ RO vs HC	[55]
Forkhead box protein I1 (FOXI1)	↑ RO vs chRCC and pRCC and ccRCC	[98]
Hepatocyte nuclear factor 1-beta (HNF1B)	↑ RO vs chRCC	[96]
Leptin (LEP)	↑ RO and chRCC and ccRCC vs HC	[97]

The management of medical treatment is largely dependent on an efficient diagnosis. The inability to differentiation between aggressive from indolent tumors represents a substantial pitfall for patient care. Therefore, the discovery of effective molecular markers for diagnosis and prognosis, especially those that are early-onset, is of utmost importance in the medical environment.

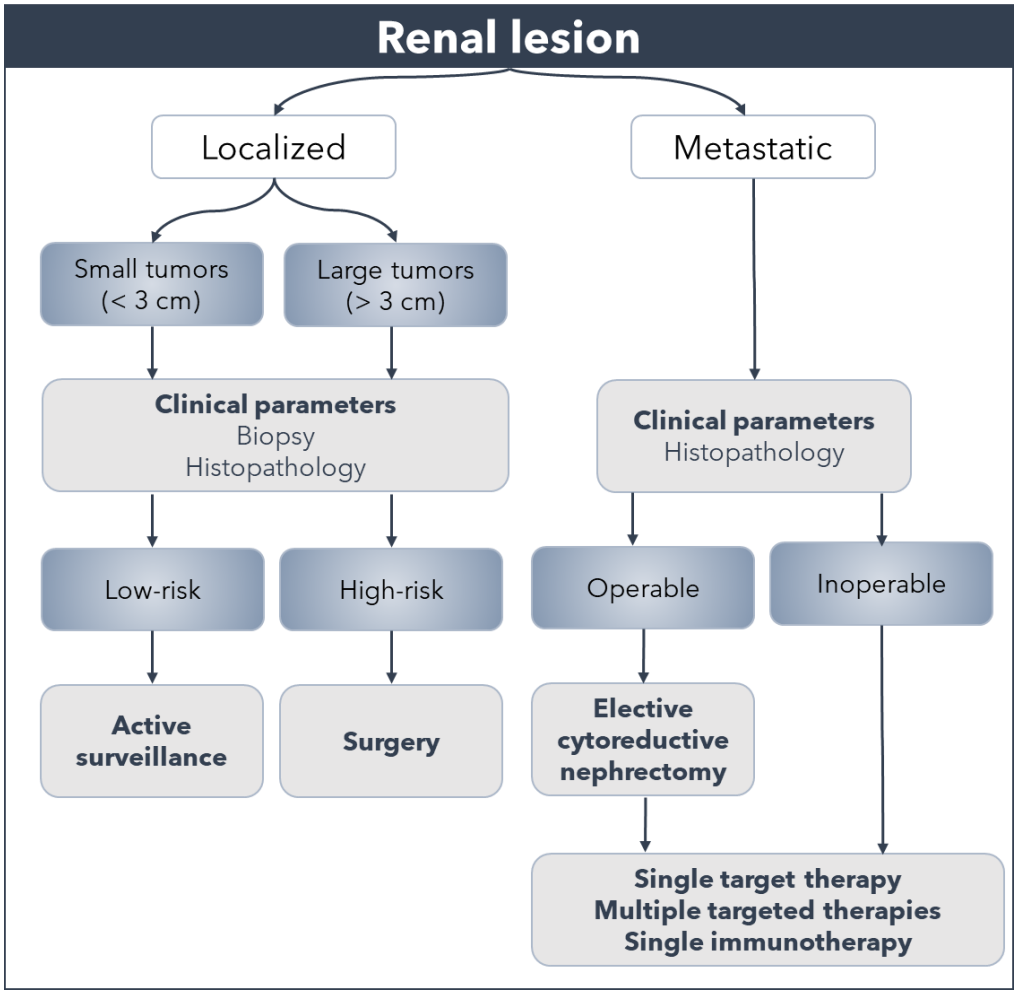

Figure I.4 - Medical decision-making algorithm for renal neoplasia. Adapted from [15].

I.2 METHODS IN PROTEOMICS

Proteomics is the study of the proteome. The proteome is the entire set of proteins that is, or can be, expressed by a genome, cell, tissue, or organism at a certain time under defined conditions.

The proteome results from the post-transductional changes made on the mRNA and from the post-translational modifications made on the proteins [99]. With the sequence of the human genome in 2001 [100], approximately 25,000 protein-coding genes were revealed, however, the 'one gene, one protein, one function' paradigm does not explain the complex molecular biology and the functional phenotype of the organisms. Nowadays, it is estimated the existence of more than 1,000,000 individual protein species, suggesting that human biology relies on a degree of functional diversity and is not directly related to the number of protein-coding genes [101]. The splicing and alternative splicing during the transcription and the post-translational modifications following protein biosynthesis are the two major mechanisms of protein diversity [101].

Therefore, to unveil the intricate biology of an organism, or parts of an organism such as cells, tissues, or organs, proteomics has become an indispensable tool in biological and biomedical research representing a step towards a better understanding of the molecular biology events.

I.2.1 PROTEOMIC STRATEGIES TO UNVEIL INFORMATION FROM COMPLEX PROTEOMES

Nowadays, to analyze the entire or just a fraction of the proteome, two main strategies, the top-down and the bottom-up proteomics, can be followed.

I.2.1.1 Bottom-up proteomics

Bottom-up approaches entail protein cleavage into peptides, usually through enzymatic digestion, before their introduction to the mass spectrometer and subsequent analysis. This strategy involves the inference of the protein based on the proteolytic peptides detected in MS analysis [102].

I.2.1.2 Top-down proteomics

Emerging as an alternative to shotgun proteomics, the top-down approach analyzes intact proteins, instead of digested peptides. Acting at the protein level, this strategy involves the direct ionization of proteins and subsequent MS analysis. Only inside the MS analyzed, the protein ions are submitted to fragmentation, placing the entire sequence of the protein under examination [103].

If, on the one hand, the “Top Down” approach has arisen as a very promising method to analyses intact proteins through MS-based techniques, the proteolytic cleavage of the proteome through a “Bottom-Up” scheme still remains the most widely implemented through the proteomics community [104]. In the present work, the strategy carried out in the protein analysis was the bottom-up, therefore the following introduction topics will be mainly focused on this approach.

Typically, the proteomics pipeline includes the following steps: (i) sample collection and preservation, (ii) sample treatment, which includes the collection of the proteome from the sample, its purification, the reduction and alkylation of the purified proteome, the cleavage of the proteome and (iii) the proteome analysis. In the following sections, the sample collection, sample preservation and sample treatment will be discussed. The proteome analysis step will be covered in section I.4 Data systems and analysis.

I.2.2 SPECIMENS FOR PROTEOMICS AND THEIR PRESERVATION

In live organisms, the proteomes expressed vary with the different types of cells and organs. Although the genome is unique for the organism, only a fraction of its genes is being expressed in each cell type. This pattern of activated genes is the cause of the inherent biology

of each cell type and the differentiation of tissues and organs [101]. Thus, the selection of the correct sample type according to the research to be done, followed by appropriate collection and preservation, are the first issues to be considered in any experimental design.

I.2.2.1 Sample specimens for proteomics: from liquid to tissue biopsies

Sample specimens used in the proteomic analysis can vary from solid tissues from organs to body fluids such as blood or urine. The urine and the blood-driven specimens' serum and plasma, are the gold-standard specimens due to the information enclosed on them and also due to the minimal or non-invasive requirements for collecting them. Tissues are also a valuable source of etiological studies and provide distinct advantages over other biospecimens. For instance, no additional steps are required to achieve an operable range of protein concentration since their dynamic variety is more restricted than in serum and plasma. On the other hand, blood-driven samples have a wide dynamic range of protein concentrations [105] while the urine fluid may present problems due to the low protein content. Therefore, in these samples, additional steps to perform the proteomic analysis are required, in the case of serum and plasma, to deplete the most abundant proteins and in the case of urine, a step of protein enrichment can be required [106]. Although in a less extent, other specimens used include saliva, cerebrospinal fluid, tears, nasal secretion to name a few [107].

I.2.2.2 Preservation methods: optimal cutting temperature (OCT) compound

The integrity of the biospecimen needs to be maintained. The preservation and handling of samples are pivotal points to minimize the degradation of samples over time. In the case of tissues, one of the methods is the flash-freezing of the biological material. However, the morphological features of the tissues tend to be distorted with this approach. Therefore, some preservation methods such as (i) protein fixation with formaldehyde followed by paraffin embedding or (ii) unfixed tissues in polymers such as optimum cutting temperature compound (OCT) followed by freezing have been implemented in the collection and storage practice. In the past, the formalin-fixed and paraffin-embedded (FFPE) tissue blocks were widely used for immunohistochemical analysis, representing nowadays a massive portion of the sample repositories. In proteomic studies, the use of FFPE samples has been on the arena after the first report using shotgun-based FFPE proteome analysis [108]. However, a notable barrier for the investigation of those samples in proteomic analysis are the deleterious effects of formalin on protein structure resulting in an ineffective protein retrieval and also in the changes promoted by the fixation process on the protein structure [109]. Although several approaches have been suggested to overcome the low protein recovery from FFPE samples, the lower yields after extraction and the fewer proteins identified by tandem MS remains an inherent obstacle of this method of preservation [109]. As an alternative to FFPE the cryopreservation with OCT compound, which preserves the morphologic and immunobiological features of tissues for future studies, has been proposed [110]. In this case, and unlike the FFPE approach, the former

method does not involve covalent cross-linking of tissue proteins. However, the OCT compound is composed of polyethylene glycol (PEG) and polyvinyl alcohol (PVA) polymers which interfere with peptide analysis by LC-MS/MS by suppressing ion signal. Because PEG and PVA are soluble in aqueous solutions, the most used method to clean samples embedded in OCT include washing tissue biopsies with ethanol and water as reported by Loken *et al.* [110], and Zhang *et al.* [111]. Other alternatives comprising (i) protein precipitation with diethyl ether-methanol [112,113] or trichloroacetic acid (TCA) [114], (ii) filter-aided sample preparation (FASP) of tissues [112], (iii) sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) [112] and (iv) solid-phase extraction [115,116] have also been applied in proteomic studies, the water-based washing method remains the most common approach to remove the polymer contaminates from OTC-embedded samples [108,117,118].

I.2.3 SAMPLE PREPARATION AND TREATMENT

Nowadays, proteomic studies rely predominantly on mass spectrometry (MS) technology. However, the quality of the MS results is largely dependent on the sample preparation methods, since sample ionization is susceptible to buffers salts, polymers, and detergents [119].

I.2.3.1 Protein extraction and solubilization

The first step of a proteomic analysis consists of the extraction of the proteins present in the biological samples. Several strategies have been developed to maximize the highest protein yield including chemically promoted extraction, physical disruption, or the combination of both [120]. Different buffers can be used to cause cell disruption and protein denaturation/solubilization from a particular sample type. These buffers aim to break intra-molecular forces and their composition may include (i) chaotropic agents, with ammonium bicarbonate as the most frequently used, (ii) strong denaturants such as urea or guanidine, (iii) ionic detergents, which include sodium dodecyl sulfate (SDS) or deoxycholate (SDC), and (iv) non-ionic or zwitterionic detergents such as Triton X-100, NP-40, digitonin, or CHAPS which solubilize proteins and denature proteins [121]. Physical lysis equipment, such as homogenizers, bead beaters, and ultrasonication devices are also often used to aid cellular protein extraction.

I.2.3.1.1.1 Protein precipitation

In proteomic workflows a precipitation step is commonly used after protein extraction, to separate the proteins from other molecules and also as a way to concentrate the proteins. Protein precipitation (PP), is one of the methodologies frequently used to clean protein extracts [119]. PP occurs when the interactions between the protein and the aqueous media are altered due to a change of the pH or hydrophobicity or when the intramolecular interactions are disrupted due to the binding of salts or metals. Different reagents/solutions like TCA,

chloroform/methanol, methanol/ammonium acetate, ammonium sulfate, acetone, among others, can be used to purify proteins from complex mixtures [122]. The TCA in conjunction with acetone is one of the most common methods to precipitate proteins in proteomics procedures. Being an acid relatively weak to hydrolyze the peptide bonds of proteins, TCA is used to acidify the aqueous medium thus promoting the disruption of the hydrogen-bonded proteins to water molecules. This way, the protein-to-protein interactions are increased and thus proteins tend to form aggregates, becoming no longer soluble [123]. Then, the samples are centrifuged, and the protein precipitates are washed with acetone to remove the TCA.

I.2.3.2 Protein fractionation

The global characterization of the whole proteome of any sample is one of the major challenges in proteomics workflows. The extreme diversity and heterogeneity of the different protein sources allied to the vast dynamic range in their expression levels as well as the absolute detection limits of analytical technologies play a critical role in sample complexity [124].

In complex samples, such as plasma and serum, due to the large dynamic range of protein expression, which in blood-derivate samples can be greater than 10 orders of magnitude, the detection of low-abundant proteins is compromised [125]. To reduce sample complexity, a step to compress the dynamic range of protein concentrations is mandatory. The depletion of the most abundant proteins (MAPs) is the most popular step and embodies a widely implemented approach to reduce sample complexity and improve the detection of proteins with lower concentrations. Depletion strategies include chromatographic affinity (immune or chemical) columns and chemical assisted methods.

I.2.3.2.1 Chromatographic techniques

One of the most common techniques to separate proteins is liquid chromatography (LC). Relying on the differential partitioning of the proteins between the mobile and the stationary phase, a variety of chromatographic modes, such as reverse-phase liquid chromatography (HPLC), hydrophilic interaction liquid chromatography (HILIC), and ion-exchange chromatography (IEX) have been developed [104]. Additionally, the compound immobilization on the stationary phase has also been used for selective interaction [126]. In this mode, selectivity can be useful either to enrich the samples of a specific protein or protein types, such as glycoproteins or phosphoproteins, or alternatively, to deplete the sample of the MAPs. Immunodepleting technologies to remove MAPs have emerged through the combination of specific antibodies. The development of depletion columns started with the two most abundant proteins in the blood, the albumin, and the IgG. Nevertheless, these technologies have evolved increasing their capacity and several depletion columns are currently commercially available for the aforementioned and also other proteins [127]. Among the disposable columns, the Multiple Affinity Removal from Agilent has been shown as one of the most efficient, reproducible, and binding specific disposable columns for six high-abundant

proteins (albumin, IgG, antitrypsin, IgA, transferring, and haptoglobin) [128]. Although most commercially available kits have demonstrated an efficient depletion of MAPs, a significant pitfall of immunoaffinity technologies is the loss of proteins bound to MAPs [129]. To avoid the co-depletion of low abundant proteins, alternative methods are based on the construction of a combinatorial ligand library designed to capture the low-abundant proteins and decrease the amount of MAPs [130]. Presently, the peptide library methodology is commercialized under the trade name of ProteoMiner.

I.2.3.2.2 Chemical-based depletion

The use of the affinity kits for protein depletion has been widely implemented for sample preparation strategies in proteomics [125], such kits are expensive approaches precluding their extensively used in most laboratories. Alternatively, some common chemicals, such as acetonitrile, ACN, and dithiothreitol, DTT, have been proposed as cheaper, simpler, and faster methodologies to reduce sample complexity [131]. The use of the ACN was firstly reported by Kay *et al.* [132], which demonstrated the efficiency of the ACN-depletion strategy in high molecular weight proteins from a serum sample.

Similarly, Warder *et al.* [133] proposed the use of the reducing agent DTT to deplete high abundant disulfide-rich proteins, such as albumin and transferrin, and also to equalize the protein content.

Interestingly, in a comparative study, the efficacy of protein depletion with ACN and DTT relatively to the commercial kit ProteoMiner was assessed by 1D gel electrophoresis and MS analysis [131]. The results have shown that the depletion of high molecular weight (over 75 kDa) proteins was achieved by the ACN method. The ACN extract was found rich in apolipoproteins and 75% of MAPs were not detected. On the other hand, precipitation disulfide-rich proteins were promoted by the DDT method rendering an extract that was found rich in immunoglobulins. Overall, both ACN and DDT methods were found as a feasible alternative depletion method to expensive commercial tools regularly used. The ACN has demonstrated an efficient depletion of higher molecular proteins while DTT has provided a more efficient compression of the protein concentration dynamic range.

I.2.3.3 Protein digestion

Separation and analysis can be performed on intact proteins, the so-called top-down proteomics, however, significant limitations on fractionation methods remains a major drawback to use these strategies in MS-based proteomics [104]. Therefore, most proteomic analysis follow a bottom-up approach, where proteins are digested into peptides to identified and/or characterize de proteome [134].

I.2.3.3.1 Protein digestion pre-treatment: Reduction and alkylation

Reduction and alkylation are essential steps during sample preparation for protein digestion, to ensure complete protein denaturation and allow the efficient cleavage of the peptide chains. Therefore, before protein digestion, some chemical agents are needed to break the disulfide bonds (reduction) and alkylation the sulfhydryl groups of cysteines, which participate in the process of protein folding. The reducing agent generally used is the sulfur-containing reagent DTT which reduces the disulfide bond through thiol-disulfide exchange [135]. Following protein reduction, cysteine alkylation is necessary in order to prevent the reversible effect of sulfhydryl group reduction. Typically, iodoacetamide (IAA) is used to alkylate the free SH-groups through bimolecular nucleophilic substitution and blocking the restoration of the disulfide bonds [135]. Finally, the remaining free IAA in solution is quenched by the addition of more DTT or by diluting the sample solution.

I.2.3.3.2 Protein cleavage into peptides: Trypsinization

As aforementioned, in bottom-up proteomic strategies proteins are digested into shorter peptides before MS analysis. Peptide cleavage can be performed by the action of proteolytic enzymes or through chemical digestion. Despite the adequate selectivity of both methods, enzymatic digestion is by far the prevalent strategy in proteomics [136]. Usually, the hydrolytic break of the peptide bonds is accomplished by endoproteinases being the trypsin the most widely used hydrolytic enzyme. The high selectivity and specificity, which exclusively cleaves arginine and lysine residues at c-terminus, generating peptides with 5-40 amino acids, and relatively reasonable costs are the major advantages of the use of trypsin in peptide-centric workflows [137]. Typically performed as an overnight reaction, trypsin achieves its best performance under a neutral pH (ammonium bicarbonate buffer) at a temperature of 37 ºC. Low concentrations of some reagents, like ACN, SDS, SDC, or urea may be included to improve the efficacy of the digestion [119]. At the end, the addition of an acidic solution is added to stop the enzymatic activity and prevent its autolysis [138]. To overcome such pitfall, immobilized enzymatic reactors have been developed through the proteomic community [139]. In the case of the trypsin, this enzyme has been covalently bonded to micro/nanoparticles [140–144], or physically absorbed onto polymeric membranes [145,146]. Immobilized trypsin has been also applied to monolithic materials [147–149], microchips [150,151], and capillary columns [152]. Overall, trypsin immobilization has provided a powerful tool for proteolytic workflows.

Although the high proteomic performance retrieved by the trypsin enzyme, the efficacy of this protease is affected by the presence of i) proline residue immediately after the cleavage site (in arginine and lysine residues), ii) repeated basic residues, or iii) post-translational modifications (e.g. methylation, acetylation) resulting in missing cleaves [119]. The occurrence of this missing cleavage sites leads to the formation of too large and hydrophobic peptides impairing their identification by MS techniques, and subsequent incomplete protein sequence coverage. To overcome these issues and increase the coverage of the total proteome, a

combination of multiple proteases has been proposed, especially with Lys-C and trypsin. The addition of the endopeptidase Lys-C, which also cleaves the lysine residue and under the same pH conditions as trypsin, overtakes some of the trypsin pitfalls yielding fewer missing values [153]. However, it is noteworthy to mention that the addition of multiple enzymes represents supplementary costs and work.

I.2.3.3.3 Protein digestion approaches

The digestion of the solubilized proteins within the crude or fractionated samples can be performed through two main methods, the i) in-gel digestion, or ii) in-solution digestion [136]. More recently, a third digestion approach has gain momentum, iii) in filter-assisted digestion [154].

I.2.3.3.3.1 In-gel digestion

In gel-based digestion approaches, protein digestion occurs after sample separation, e.g. by gel-based approaches. Being one the first sample treatment used to digest proteins, the SDS-PAGE firstly employs a protein fractionation stage and the complex mixture of proteins is separated either by their molecular weight in the one dimensional (1D) version or by their isoelectric point along with their molecular weight in the two dimensional (2D) variant [119]. The resulting gel displays the pattern of bands or spots representing the sample proteome. Commonly, the protein pattern is visualized by Coomassie staining which reversible binds to tyrosine residues and detects a protein in a concentration range from 10 to 50 ng [155]. Apart of being a compatible technique for MS, several other methods, including silver staining, fluorescent staining, or radioactivity detection, can be used to visualize in-gel matrix proteins [156]. After protein visualization, the typical set-up using in-gel digestion approaches undergoes to differential analysis of protein expression patterns and the use of enzymes for digestion. Since, proteins need to be fixed into the gel matrix after electrophoretic separation to avoid their diffusion resulting in resolution losses, all the pre-treatment, including protein reduction and alkylation, followed by proteolytic reaction need to be performed inside the gel matrix. Afterward, the resulting peptides are extracted from the polyacrylamide matrix and can be analyzed by MS.

I.2.3.3.3.2 In-solution digestion

The proteome of complex samples can also be digested in solution without any previous protein separation technique. In this approach, the proteins, previously solubilized, present in the mixture are reduced and alkylated and then submitted to enzymatic digestion. Since all proteins present in the mixture are broken into peptides, this approach is also referred to as shotgun proteomics, due to analogy to shotgun genomic sequencing [157]. Finally, the resulting peptides are then analyzed through MS. Some inherent advantages, like the higher number of proteins identified at the same time, the need for less analytical steps and the possibility to scale up to mixtures containing less than 10 µg or greater than 1 mg of protein, have turned out this approach more popular over gel-based methods [119].

I.2.3.3.3.3 In-filter digestion

Alternatively, the protein digestion step has been combined with the FASP system and placed on top of the molecular weight cutoff (MWCO) filter [154,158]. In this strategy, the solubilized protein mixture was added and retained on the top of a spin filter, and then reduced and alkylated to subsequent digestion. Digested peptides are finally eluted from the membrane. The advantages of the in-filter based methods include the simplification of the workflow, with less protein handling, and additional cleaning steps usually needed in the two other approaches, are here avoided since any contaminants can be easily washed away throughout the MWCO filter.

I.2.3.4 Peptide treatment

In shotgun approaches, the ultimate goal is to obtain the whole proteome digested into small peptide pieces to inject them in the MS apparatus and proceed with protein identification and/or quantification. However, whilst in digestion assisted by a filter (FASP) the pool of peptides is recovered free of salts and order contaminants as urea, in in-solution approaches an additional step is often required at the end of the procedure to clean the peptide sample. To avoid complex cleaning procedures, the use of micropipette tips packed inside with silica-based sorbents functionalized with C18 groups after protein digestion has been employed for desalting purposes [159]. These on-a-tip solid-phase extraction systems, commercially available under the name of ZipTip®, adsorb the peptides on the sorbent by a hydrophobic interaction allowing to discard of the undesired contaminants.

Peptide fractionation is frequently used in post digestion proteomics approaches. Several strategies have been developed to enrich the sample fraction of a specific target. For instance, post-translational modified peptides, such as with phosphorylations, glycosylations, sumoylations, and others, are present in very low abundance when compared with the other ones [124]. Although the detention capabilities of the current MS instruments have been improved over the last decades, the vast quantification of a very wide dynamic range of complex mixtures is still a challenge.

I.2.3.4.1.1 Phosphopeptide enrichment

In phosphoproteomics, a variety of techniques and strategies, mainly coupled to MS techniques due to their high sensitivity and accuracy, have been developed for the selective enrichment of the phosphopeptides [160]. In Figure I.5 are summarized the most commonly applied strategies for phosphopeptide enrichment.

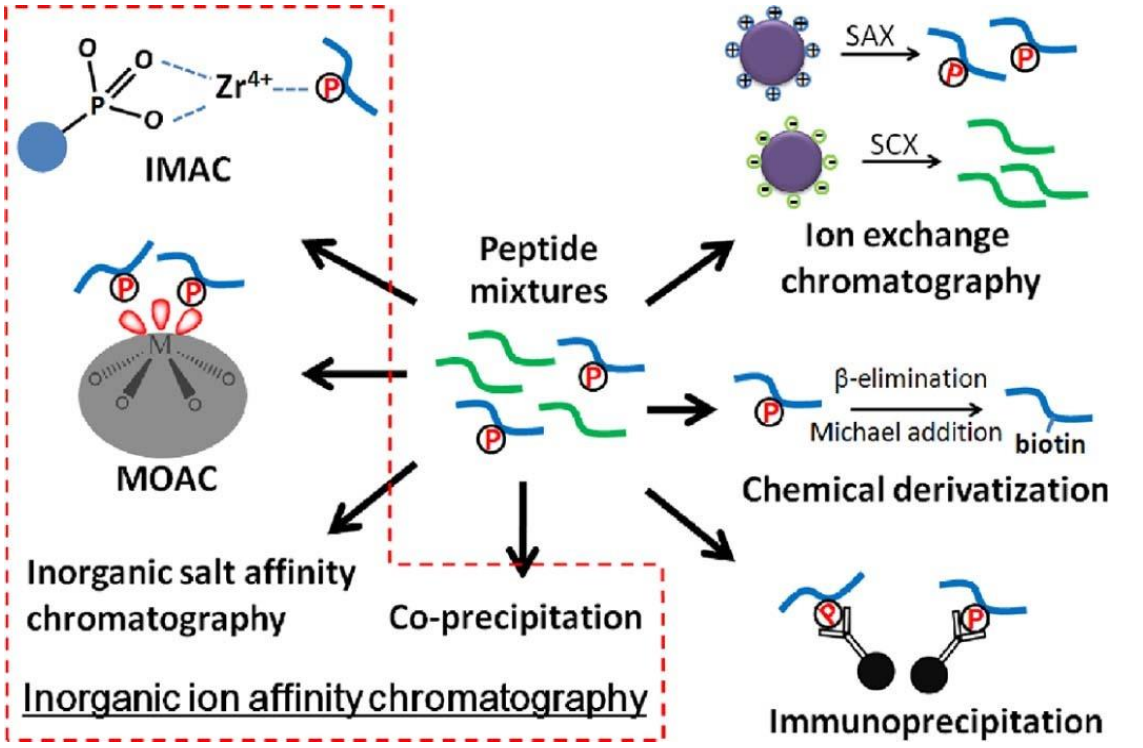

Figure I.5 – The most common phosphopeptide enrichment strategies used in proteomics [160].

The inorganic ion affinity chromatography type approaches, which include the immobilized metal affinity chromatography (IMAC), the metal oxide affinity chromatography (MOAC), the inorganic salt affinity chromatography, and the co-precipitation, are based on the affinity interaction between the phosphate group and the metal ions. Among them, the IMAC consists of the immobilization of metal ions on supporting substrates, whereas the MOAC utilizes metal oxides instead. In the case of the inorganic salt affinity and co-precipitation, the affinity of the metal ion towards the phosphate group is exploited in both [160]. Including the strong anion exchange chromatography (SAX) and strong cation exchange chromatography (SCX) is the ion exchange chromatography type. The principle of this enrichment technique is the reversible interaction of charged species with the ionic exchange matrix, which is composed of cationic or anionic resin in SAX and SCX, respectively. As result, the positively charged groups will have a higher affinity for negatively charged molecules, as phosphopeptides, while the negatively charged resins for positively charged molecules [160]. The other two types are chemical derivatization and immunoprecipitation. In chemical derivatization, the phosphate group is replaced by another affinity group, usually via a β-elimination followed by a Michael addition of a biotin moiety for biotin-labeled peptides enrichment [161]. Immunoprecipitation is based on an antibody separation. Although high specificity and efficiency can be achieved to selectively capture tyrosine-phosphorylated residues, the lack of phosphoserine- and phosphothreonine-specific antibodies limit the utility of this strategy [162].

I.2.4 ADJUVANT METHODOLOGIES FOR SAMPLE PREPARATION

Proteomic approaches are very long workflows with multiple sample treatment steps, making these approaches tedious and time-consuming. Over the last decades, growing demand for faster approaches has been verified and many efforts have been addressed to accelerate diverse steps of the proteomics pipeline [121]. For instance, higher temperatures have been used to accelerate denaturation/solubilization steps as well as the reaction rates of digestion processes. Microwave radiation, infrared energy, and high-pressure approaches have also been proven to accelerate the proteolytic cleavage of proteins [121]. Similarly, the ultrasound energy (UE) has become a powerful tool for speed up proteomics workflows [163], which will be commented bellow.

I.2.4.1 Ultrasound (US) approaches

Defined by the American National Standards Institutes, ultrasound is the sound waves at frequencies higher than 20 kHz, which is out of the upper audible limit of human hearing [164]. Depending on the effects of the ultrasonic wave when passing a liquid medium, the ultrasonic frequency can be divided into two main zones: i) the high frequency comprising a range between 2 MHz and 10 MHZ, and ii) the low frequency which varies from 20 kHz to 10 kHz. In the former, the physical and chemical properties of the liquid media where it is used do not change and are widely used for medical applications while in the latter the liquid media where it is applied undergo some physical and chemical changes [163]. The physical phenomenon caused by the low frequency is known as cavitation and consists of the production of microbubbles in a liquid medium that violently collapse, leading to the formation of hot spots (e.g. extreme local temperatures and pressures). The energy released by the collapse of these microbubbles propagates through the media, in the form of shockwaves, provoking the disruption of solid surfaces [165] and the increasing of chemical kinetics. Taking advantage of these effects, the UE has been applied in diverse analytical processes [166] including in many steps of proteomic workflows [163] to shorten sample treatment time, to diminish the number of steps and to increase sample troughput.

I.2.4.1.1.1 Protein extraction and solubilization

As mention above, in section I.2.3.1.1.1 Protein extraction, the application of sonication can be used to enhance protein extraction and solubilization from solid matrixes. In this case, the use of an ultrasonic field leverages the release of the total protein content through a better disruption of the cell wall and overall tissue homogenization, giving this way higher yields of protein recovery. In fact, the effects on protein extraction when ultrasonication and heating are applied are similar, and so the joint use of both can significantly improve the efficiency of the extraction process [167].

In proteomic workflows, the step of protein precipitation is often required to clean and suit the sample for downstream analysis. However, the difficulty of protein re-solubilization after

this process represents a major issue causing variability in protein recovery yields. Here, the adjuvant effects of UE during protein solubilization have also been proven, by increasing the amounts of protein recovery and overall reproducibility [168].

I.2.4.1.1.2 Protein reduction and alkylation

The ultimate goal of protein denaturation, reduction, and alkylation is to make the protein more accessible for the proteolytic process during digestion. Like the denaturation and solubilization steps, the ultrasound energy can also be used to help the chemical reactions of reduction and alkylation. By accelerating the kinetics of the reactions involved in these processes the time consumed at this stage has been not only considerably reduced, but also sample handling has been simplified [169,170].

I.2.4.1.1.3 Protein digestion

Protein digestion represents one of the most important steps for bottom-up strategies since protein identification relies on the cleavages of proteins into peptides. Although all the preceding steps contribute to the efficiency of the proteomics procedures, effective protein digestion is crucial for the formation of peptide species. Nevertheless, the typical protein digestion step is usually the slowest of the entire workflow. To overcome this shortcoming, the beneficial advantages of the US are being ever more frequently used in protein digestion steps as well, which greatly decreases the overall analysis time consumed in a proteomic pipeline [163]. The synergetic effect created by ultrasonication and the enzyme activity is being pointed out as an enhancement of the mass transfer processes between the enzyme-substrate system [171]. It noteworthy to mention that the enzymatic acceleration by an ultrasonic field has been proven to be efficient in the different proteomics strategies namely in i) liquid-liquid phases like in-solution digestion; ii) solid-liquid phases such as the in-gel approaches, in which the proteins are trapped in the polyacrylamide matrix and the enzyme is in solution or when the enzyme is immobilized in solid support and the proteins are in a liquid extract [144]; and iii) filter-aided approaches, in which the digestion is performed on top of a cutoff membrane where the proteins are retained [172]. In the case of the in-gel digestion, where the UE helps the penetration of the enzyme into the gel matrix, acting as a micro-syringe, the selection of a proper amplitude is of utmost importance. If too low, then the ultrasound would fail in an efficient delivery of the enzyme into the gel, an if too high, the ultrasound will destroy the gel surface and interfere with the downstream analysis [163]. Likewise, in procedures involving solid support, either in immobilized enzymes or in filter-aided approaches, the use of the UE amplitude needs to be carefully chosen for the degradation of the material. Another important aspect to be considered in UE applications is the time of the ultrasonic field since a negative effect on enzyme activity can be provoked in long periods of ultrasonication.

I.2.4.1.1.4 Other applications

Ultrasonication has also been useful in other processes of sample handling. For instance, in a gel-based approach, after electrophoresis and gel staining, for protein

visualization, spots/bands of interest must be excised, and staining reagents eliminated. The dye cleaning steps, which frequently consist of successive washes to remove the staining product, can be accelerated through the aid of UE [173,174]. Protein quantification through the isotopic labeling of the peptides with ^{18}O embodies another proteomics point for UE application. In direct approaches, the ^{18}O labeling is coupled with protein digestion, and through the mean of ultrasonication, this step can be shortened from overnight to a couple of minutes [175,176].

I.2.4.1.1.5 Current ultrasonic tools

Nowadays, many ultrasonic devices have emerged to empower several analytical workflows. In proteomics, the UE can be delivered as a direct or indirect ultrasonication. The former utilizes sonotrodes to apply the ultrasound wave directly dipped into the sample, while in the latter the ultrasonic field reaches the sample through its container's wall, and usually utilizes sonoreactor systems. Concerning the main purpose of the analytical process, the ultrasonic systems applied must be carefully chosen.

Sonotrodes, which are directly immersed in the solution, perform best in extraction and solubilization procedures, where a powerful tool must be used to aid in cell disruption and protein solubilization [167,177]. Nowadays, different sonotrodes typologies are commercially available including silica glass probes, spiral probes are even multiple probes that incorporate two or more probes accomplishing a higher sample treatment throughput [178]. However, a few critical factors must be considered for the correct application of the sonotrode-based system. First is the selection of the probe tip diameter along with the shape of vessel containers, since each tip size will perform better in a certain range of volume, and a conical-shaped will allow a deeper insertion of the probe avoiding aerosol and foaming formation. This way, the formation of zones with no cavitation (dead zone) is minimized. Another remarkable point deals with the constant, though at a slow rate, increasing of bulk temperature caused by high intensity sonotrodes. For this reason, samples usually are placed on the ice during the ultrasonic application, and for long ultrasonication times, a pulsed mode is recommended [163].

One of the most common devices in scientific laboratories are the ultrasonic bath. Although these systems have been applied in certain proteomics tasks [169], their low energy intensity delivery make them inappropriate for diverse proteomic steps, specially protein digestion [163].

By contrast, the sonoreactors and cup-horns are the most promising ultrasonic devices to use in proteomics, because they deliver indirect ultrasonication, where a sealed sample container can be used, and the formation of aerosols and cross-contaminations are avoided [163]. With these devices, ultrasonication has been widely used in either in gel-based or off-gel methods and even in filter assisted systems [172], covering all the different strategies of protein digestion. Like the sonotrodes, diverse sonoreactors with different horn sizes can be found [178]. Nowadays, the microplate ultrasonic horn assembly device represents the most outstanding tool by being able to evenly apply the UE to a 96-well plate, making the proteomics sample treatment to an unprecedented level [174].

Overall, the implementation of the UE in the proteomics pipeline has drastically simplified sample handling by reducing the time spent in each step from hours to minutes and by diminishing the number of steps.

I.2.4.2 Extraction technologies

The main goal of a sample treatment workflow is to prepare the sample for downstream analysis. In proteomics the sample preparation workflows for MS-based analysis are very complex and time-consuming. While the application of the UE has been useful to speed up the whole pipeline, other strategies have also been implemented to clean or reduce the complexity of the biological samples. The separation of the solute of interest from its matrix based on their different solubilities is commonly called as extraction. Two different classes, the liquid-liquid and the solid-phase extraction can be distinguished in analytical extraction processes.

I.2.4.2.1.1 Liquid-liquid extraction (LLE)

In LLE, the analytes of interest are extracted from one liquid to another [122]. Typically, an aqueous sample is added to an organic solvent resulting in the formation of the two immiscible phases, with the polar compounds dissolved in the aqueous phase and the apolar ones in the organic phase [179]. This analytical process can be used for sample cleanup and/or concentration, being widely used in drug extraction from aqueous matrices through the use of volatile organic solvents [179].

I.2.4.2.1.2 Solid-phase extraction (SPE)

Over the years, SPE technology has become one of the most widely implement and powerful tool to improve MS analysis. Based on the use of solvents and a stationary (solid) phase, which can appear as the format of cartridges, pipette tips, discs, magnetic beads, among others, the SPE chemistry is mainly categorized as reverse-phase chromatography (RPC), HILIC, IEX, affinity chromatography or mixed-mode [179].

The RPC is the most widely SPE technique in analytical processes for protein or peptide separations. Based on the reversible hydrophobic interaction between the sample and the stationary phase, RPC uses a polar sample and an apolar stationary phase. Relying on that hydrophobicity principle, RPC is an adsorptive process that uses an organic solvent to desorb the analyte from the stationary phase to elute it. The stationary phase is commonly composed of alkyl chains, such as C4, C8, and C18, linked to porous silica particles [104]. The different composition of those columns will determine the retention of the analyte. For instance, while the C8 column type is more frequently used in intact protein, the C18 is more appropriate to peptides [180]. Currently, this technology is widely used in the separation of complex biological samples due to its high performance and resolution. For instance, the use of the RPC allows the sequential elution of the complex pool of digested peptides into the MS instrument. Here the peptide mixture is loaded into the hydrophilic column under aqueous conditions that will strongly adsorb the more hydrophobic peptides and elute firstly the hydrophilic ones. Afterwards,

through an increased organic solvent gradient, will progressively modify the adsorption affinities and then the peptides are sequentially eluted for further analysis [180]. The use of these alkyl-bonded resins is also frequently used as on-a-tip SPE technologies, which are pipette tips similar to the conventional ones but with a sorbent packed inside. In the proteomics field, these technologies are used to clean the samples from salts and detergents, especially the C18-based tip, which is used after the enzymatic digestion, just before the MS analysis. In this scenario, the peptides adsorbed are eluted by the addition of an organic solvent.

On the other hand, the HILIC SPE strategy appears as the counterpart of the RPC. Based on the use of polar stationary, this technique has been applied to fractionate more polar samples before MS analysis. Here, the analyte partition relies on the mobile phase and a water-enriched region immobilized onto the stationary phase, which is mainly made of underivatized bare silica or uncharged modified silica [180].

The IEX strategy is another SPE technique which involves the electrostatic interaction of the charged residues on the surface of the protein and the opposite charge of the stationary phase. After that, protein elution encompasses an increased salt concentration gradient or by shifting the pH to avoid the incompatibilities between the salts and MS applications [181].

Alternatively, to the previous SPE techniques, the selective extraction of a specific analyte can be performed through an affinity chromatography which uses some specific compounds immobilized onto the stationary phase to interact with the analyte of interest. This analytical technique presents a very good selectivity and can be applied in diverse steps throughout the proteomics pipeline. For instance, as mention above in section I.2.3.2.1 Chromatographic techniques, this technique is regularly used coupled to specific antibodies, to deplete the MAPs and fractionate protein samples. Also, as described in section I.2.3.4.1.1 Phosphopeptide enrichment, affinity chromatographic approaches have been applied to selectively capture phosphopeptides.

I.3 MASS SPECTROMETRY FOR PROTEOMICS

Over the last decades, mass spectrometry (MS) has massively grown and become a versatile and indispensable analytical technique in many fields of science such as chemistry, biochemistry, pharmacy, and medicine [182]. Going back to its origins, in the early 1900s, the great ascension of MS has been marked by numerous Nobel Prizes. Starting in 1906 with the original study of cathode rays by Joseph J. Thompson, who earned the physics prize [183], the first full mass spectrometer was built in 1917 by Francis W. Aston [184], who used it to elucidate the existence of isotopes and earned the chemistry prize in 1922 for his research [185]. Next, Hans G. Dehmelt and Wolfgang Paul jointly received in 1989 the physics Nobel prize for the development of the ion trap technique [186]. More recently, in 2002, the Nobel prize in chemistry was given to John B. Fenn and Koichi Tanaka, for the development of ESI and MALDI ionization methods, respectively [187].

The quintessence of the MS is the separation of isolated ions by their masses in a gas phase. Consisting of a (i) sample introduction system, (ii) an ion source, (iii) a mass separator, (iv) a detector, and a (v) data management system, the mass spectrometer is the instrument used to analyze any ionizable molecule. All of these components, and their variations, will be outlined and discussed throughout this chapter.

I.3.1 SAMPLE INTRODUCTION SYSTEM

The first component of a mass spectrometer is the sample inlet. The main function of this compartment is not only the introduction of the sample molecules into the instrument, but also the decompression of the sample, which is initially at atmospheric pressure (760 Torr) into a very low-pressure state. Typically, mass spectrometers operate in negative pressure (e.g. high vacuum, ≈ 10^{-6} Torr) to prevent the collision of sample ions with neutral molecules, such as air, which would interfere with the performance of the mass spectrometer [188].

There is a variety of samples that can be analyzed through MS technology. Varying from simple gases to large protein complexes, samples can be composed of single species to highly complex mixtures. If direct sample introduction systems, as solid probes or plates, deliver an acceptable efficiency in simple mixtures, the same is not accomplished when analyzing more complex ones. To overcome these issues, chromatographic separation techniques have been coupled to the MS system. Directly interfaced with the mass spectrometry, gas chromatography (GC) and LC are usually used to improve the efficiency of MS analysis. As described previously, in section *I.2.3.2.1 - Chromatographic techniques*, LC technique is frequently used in proteomics to reduce the complexity of the sample mixtures. Similarly, the GC moved forwards in the same direction. Accordingly, while LC is used in liquid format samples, the CG is applied in gaseous samples. Likewise, the combination of these separation techniques together with the MS is described as hyphenated methodologies, like GC-MS and LC-MS, as they add an additional dimension to the analytical measurement [189].

I.3.1.1.1.1 Liquid chromatography (LC)

In LC the analytical separation is based on a stationary and a mobile phase. Typically, the mobile phase, containing the sample in which is in a liquid format, permeates through the stationary phase which will retain the analytes present in solutions according to their chemical affinity. In proteomics, where the sample solution is digested peptides in aqueous solutions, the most common stationary phase used is a C18 packed column [104]. Conventionally, the liquid flows across the column under the application of high pressure (400 – 600 bar). In this case, this technique is usually termed as high-performance LC (HPLC, also known as its former name high-pressure LC), and works with columns packed with particles sized between 3 to 5 µm. However, over the years several efforts have been made to improve the efficiency of chromatographic approaches. The choice of the particle size packed inside chromatographic approaches has been verified to influence the efficiency of analyte separation. The separation efficiency is improved with the reduction of the size of the particles, however, to move forward

the solvent inside the column higher pressures are necessary. As consequence, the use of smaller particle sizes has demanded the development of the ultra-high-performance LC (UHPLC) instrument, which is able to resist higher backpressures (up to 1200 bar). In this case, the use of UHPLC has provided an improvement of resolution and sensitivity, by using packed columns with downsized particles of 1.7 - 2 µm, without compromising the flow rate and gradient length [190].

After reducing the size of stationary phase particles to improve chromatographic resolution, additional approaches to miniaturize chromatographic columns at nanoscale has continuously emerged. The smaller inner diameter of columns (10 to 100 µm) has been associated with a reduction of the flow rate necessary to perform the analytical analysis and is described as nano-LC techniques [191]. In this sense, by reducing the flow rate in nano-LC approaches, less biological samples' amounts will be needed since the analytes will be concentrated inside the columns and eluted in fewer volumes, consequently resulting in higher sensitivity as well. Also, a drastic reduction of solvent consumption is accomplished with the application of these nanocolumns.

A summary with the main characteristics and operating conditions of conventional HPLC and UHPLC along with miniaturized column systems is present in Table I.8.

.

Table I.8 - Principal characteristics and operating conditions of conventional HPLC and UHPLC at micro and nano scale [192,193].

	HPLC	**UHPLC**	
		microscale	nanoscale
Pressure range	400 – 600 bar	up to 1200 bar	
Particle size (stationary phase)	3 – 5 µm	1.7 – 2 µm	
Internal column diameter	3 – 4.6 mm	1 – 2 mm	10 - 100 µm
Flow rates	1 – 2 $mL.min^{-1}$	0.1 – 0.5 $mL.min^{-1}$	≤ 1 $\mu L.min^{-1}$

Overall, the miniaturization achieved by the downsizing the size of the particles in the stationary phase and the reduction of the internal diameter of chromatographic columns to nanoscale has leveraged the LC technique by (i) allowing faster analyte separations with higher resolution; (ii) reducing the sample volume for the analysis; and (iii) reducing the consumption of solvents and consequently their inherent costs [191].

I.3.2 IONIZATION METHODS

In mass spectrometry, the separation and detection of the analytes require charged species, and therefore sample must be submitted to the ionization process, e.g., conversion of their analytes into ions [194]. Ions are atoms or molecules that carry one or more electrical

charges, either positive or negative. Ionization occurs when a neutral analyte gains protons or lose electrons, resulting in an impaired balance of charges of the analyte. Albeit most MS instruments can handle both ion modes, cationic or anionic mode, in proteomics they traditionally dealt with positively charged ions [195]. Generically, the ionization methods are classified as hard or soft procedures, as they cause molecular fragmentation or not, respectively. A more detailed description of the different methods to make ions is present below.

I.3.2.1 Electron ionization (EI)

In EI, the sample in the ionization chamber is bombarded with a beam of high-energy electrons (70 eV). The beam, which consists of a heated filament located inside the vacuum compartment, imparts an excess of energy to the analytes resulting in loss of the electrons from sample molecules and formation of positively charged molecular ions [194]. In this method, the energy delivered to the sample is in considerable excess compared to the required to ionize organic species (≈ 10 eV, typically), resulting in the fragmentation of the analyte into molecular ions. For this reason, EI is generally considered as a hard ionization procedure [195].

I.3.2.2 Chemical ionization (CI)

With the purpose to reduce the amount of energy delivered to the analytes and, consequently, limit the fragmentation of the molecules during the ionization process, led to the development of the first soft ionization method, the chemical ionization (CI). In CI, molecules are charged only with sufficient energy (≈10 eV) to generate adducted ions but insufficient to cause fragmentation. This fact is accomplished by the introduction of reagent gas in excess into the ion source chamber, which will be ionized by an electron beam, similarly to EI. Next, the resultant ions formed from the collision of the electrons and the molecules of the gas, will, in turn, collide with the neutral analyte molecules and transfer a proton, resulting in protonated species. In this way, the ions formed, such as $[M + H]^+$, are directly related to the molecular mass [196].

I.3.2.3 Atmospheric pressure chemical ionization (APCI)

In the two ionization methods described above, EI and CI, the ionization process is performed in (high) vacuum conditions. However, the need to interface the LC with MS to analyze polar compounds has led to the development of ionization techniques at atmospheric pressure. In this way, liquid solutions are firstly evaporated at atmospheric pressures and then ionized, allowing the use of LC separation techniques coupled to ionization chambers. In APCI, liquid samples are injected into the chamber and both, analyte and LC effluent, are evaporated by heating. Then the resulting vapor is swept through the corona discharge needle where ionization starts by creating an ionizing plasma. The plasma, composed of both protonated solvent ions and electrons, is in turn, responsible for the ionization of the analyte molecules.

Basically, APCI is a variation of classical CI performed at atmospheric pressures instead of high vacuum conditions. By replacing the filament used to produce the electron beam, that would burn out the sample at such high pressures, for a corona discharge the ionization can be accomplished at atmospheric pressure, and then the analyte-forming protonated molecules, $[M + H]^+$, follows to a vacuum chamber and proceed to the mass analyzer [196].

I.3.2.4 Atmospheric pressure photoionization (APPI)

At the beginning of the new millennium, the ultraviolet (UV) light was introduced in the ionizing systems as an alternative to the corona-discharged needle, resulting in the APPI method. With the use of the UV light, two competing ionizing processes can be verified. Similarly to CI and APCI which use a reagent gas, in APPI a dopant, like toluene or acetone, is introduced within the system. By passing through the UV source, dopant ions are formed, which will in turn transfer the proton to the analyte molecules, acting this way as reactive proton donors. The second mechanism of ionization in APPI is the direct photoionization of the analytes. As in APPI, the energy provided by the UV light can be high enough, ≈ 10 eV, to expel an electron from some analyte, such as polynuclear aromatic compounds, this technique is commonly used to ionize less polar species. Additionally, the background signal of LC solvents is reduced in this method since the energy of the UV radiation used is sufficient to ionize the analyte molecules but below the energy required to ionize the LC solvents as well as atmospheric gases [196].

I.3.2.5 Electrospray ionization (ESI)

ESI is another technique classified as soft ionization, as the energies involved are barely above those necessary to generate ions. Along with APCI and APPI, ESI is an atmospheric pressure compatible method [197]. Since its discovery in the late 1980s, the ESI has revolutionized MS, and it is now the most prominent ionization technique for polar molecules. Remarkably, this technique provides extremely good results in the analysis of large, non-volatile, chargeable molecules such as proteins, being, for this reason, the method of choice for proteomics studies based on LC-MS approaches [198]. In the ESI process, an electrically charged aerosol is formed through the application of a high voltage on the analyte solution dissolved in a conductive solvent. Usually, the aerosol formation is supported by a sheath gas which nebulizes the elute into a spray of charged droplets. The droplets formed are then continuously evaporated (desolvated), with a secondary flow of heated gas, leading to their disintegration and release of the ionized analyte into the gas phase. The disintegration of the ions occurs due to the increase of the electric field, caused by the desolvation, which distorts and develops a sharper curvature on the droplets surface, called a Taylor cone. This process is termed as droplet jet fission since the formation of the Taylor cone results in a stable stream of droplets from which the ions are released. During this process, the resulting charged gas phase is transferred from the atmospheric pressure environment into the vacuum system of the mass

analyzer [199]. As previously mentioned, downscaling procedures represent a valuable improvement of the different techniques. Similarly, the miniaturization of the electrospray to nanoscale has allowing reduce the size of the droplets formed and, moreover, the flow rate used. In this way, the so-called nanoelectrospray (nanoESI) variant produces droplets with less than 200 nm, against the 1 – 2 µm droplets formed by the conventional ESI, and, flow rates of 20 – 50 nl min^{-1} [197]. Additionally, to the reduction of solvent amounts required due to the reduction of the flow rate, an extreme reduction of sample consumptions is achieved in nanoESI approaches.

The specific technical aspects of ESI techniques, such as the liquid-based system and the requirement of a conductive solvent, make this ionization process compatible with reverse-phase LC and consequently suitable for the analysis of a very large class of compounds, like polar molecules. With the introduction of ESI methods, the MS-based analysis has been expanded into biological systems where most compounds are water-soluble and polar. Moreover, this ionization technique has a larger molecular mass range, as it can be applied from small to larger molecules, such as proteins [200].

I.3.2.6 Matrix-assisted laser desorption/ionization (MALDI)

Along with ESI, the matrix-assisted laser desorption/ionization (MALDI) is another ionization method that has revolutionized the use of MS-based approaches in proteomics. Similarly to ESI, MALDI has a wider range of molecular masses, up to 350 kDa, in opposition to 1 kDa from other ionization techniques. On the other hand, this technique is typically performed under high vacuum conditions, below 10^{-6} mbar. The mechanism of this method relies on the absorption of energy from laser light by a solid sample layer. Basically, first, the analyte is co-crystalized with a matrix solution onto a plate surface, which is mainly made of steel [201]. Different matrices can be used according to their properties and the analyte of interest. For instance, while the α-cyano-4-hydroxycinnamic acid matrix works nicely for peptides acid, the 2,5-dihydrobenzoic acid (DHB) matrix is more appropriate for proteins and synthetic polymers, albeit it can still be used for peptides as well. After sample-matrix crystallization, the plate is introduced inside a vacuum chamber, where the laser beam irradiates the sample causing evaporation and the formation of a plasma. There are two types of irradiation in MALDI, UV and infrared (IR) wavelengths, however, the UV lasers are by far the most wavelength radiation used in MALDI applications. As the matrix must be efficiently absorbed by the laser light, the wavelength used in MALDI also influences the choice of the matrix compound. Upon laser irradiation, the energy absorption by the matrix and the evaporation due to the rise of the temperature leads to the formation of gas-phase composed of protonated matrix ions. Therefore, the neutral sample analyte, which is also carried into the vapor phase, is charged by the matrix ions, yielding protonated molecules $[M + H^+]$.

Other ionization methods, such as fast atom bombardment (FAB), field desorption (FD), field ionization (FI), liquid introduction field desorption ionization (LIFDI), or surface ionization

methods, including the desorption electrospray ionization (DESI) and the direct analysis in real-time (DART), but they have been superseded, such the case of FAB by the ESI, or their limited applications has resulted in a minor use in proteomics approaches [202–204].

For protein analysis, soft ionization methods are required to ionize peptides without considerable fragmentation. Therefore, MALDI and ESI are the two techniques widely used to analyze proteins [101].

I.3.3 MASS ANALYZERS

The main role of the mass analyzer is to separate the ions resulting of the ionization process, according to their *m/z* values. A *m/z* value is a dimensionless number and represents the ratio of the mass of the ion (*m*) by the electrical charges (*z*) acquired by the sample during the ionization process. Analyzers can be classified into two main categories based on ion separation over time or in space. The time systems are scanning instruments where the operational parameters of the electric/magnetic field are changed progressively over time to obtain the spectra. While in the space systems the spectrum is obtained in a nonscanning mode as none of the operational parameters is altered during analysis [205]. Except for MALDI, which cannot be coupled to quadrupole analyzers due to their limited mass range, any type of ion source can be coupled with any type of mass analyzer.

I.3.3.1 Quadrupole (Q)

Quadrupole (Q) is a mass analyzer type consisting of a set of four cylindrical rods, set parallel to each other in a square shape. The rods, which can be made of metal or metal-coated ceramic, are arranged around a central axis with opposing pairs connected electrically. Combinations of radio frequency (RF) voltage and direct current (DC) offset voltage are used to generate an electric field. Ions are then separated according to their trajectories, as at a given RF/DC voltage only a particular *m/z* value will have a stable oscillating trajectory through the rods and reach the detector [206]. Contrary, too large or too small ions will end up colliding with the rods due to their unstable motion. In this way, quadrupole-type analyzers are scanning instruments as the operational parameters (voltage) of the electric field need to change over time to addresses the different stability regions of the different ion masses [207].

I.3.3.2 Ion traps (IT)

Similarly to quadrupole technology, ion trap (IT) instruments are based on the stability of the ions, but in opposite directions. While in Q analyzers only stable ions passed down the electric rods and the unstable species are lost, in IT all ions are retained, trapped in a stable trajectory, and the unstable ones are caught by the detector. In IT systems, the voltage is applied to the stable ions stored inside to lead them to an unstable state and eventually eject them from the trap onto the detector. The voltage needed to render unstable ions is dependent

on their *m/z* values, therefore the voltage applied over time is changing in order to scan all the ions [206]. There different types of IT according to their geometries, as described below.

I.3.3.2.1 Quadrupole ion trap (QIT)

To be precise, the quadrupole ion trap (QIT) instruments should be referred to as ion trap with tree-dimensional (3D) quadrupole field, nevertheless the QIT is the most widely used term being commonly accepted through the scientific community. QIT devices consist of three electrodes, two end caps, and a central ring, all machined to have a hyperbolic geometry inside. The ions produced by the external ion source get inside the analyzer compartment via the holes of one end cap electrode, which only allows the entrance of a controlled number of ions. Once inside, the constant RF voltage applied to the central ring electrode causes the ions to undergo trapped with a stable orbit. As the amplitude of the RF voltage is increased, the orbits of ions destabilize, starting with the lowest *m/z*. Those destabilized ions exit the analyzer cavity through the centrally arranged holes in the opposite end cap electrode striking the detector [208].

I.3.3.2.2 Linear ion trap (LIT)

The development of a linear ion trap (LIT) mass analyzer was a consequence of the limited number of ions that can be placed in a QIT analyzer. Consisting of hyperbolic rods adjacently positioned, the ions are confined inside radially by the two-dimensional RF voltage, and axially by DC potentials applied to end electrodes [209]. In this way, LIT mechanisms can trap and store a higher fraction of ions than QIT. Once inside, ions are steadily ejected towards detection by increasing the RF voltage.

I.3.3.3 Fourier transform (FT) analyzers

Another type of mass analyzer is the Fourier transform (FT) type. FT instruments are also based on the ion trap principle, however, in this mass analyzer type the analytical cell is also the detector. The transient signal is then converted into the frequency domain employing Fourier transformation. In this way, the *m/*z of each ion is measured as a function of ion frequency instead of ion stability [210]. Currently, FT instruments include the orbitrap and the ion cyclotron resonance (ICR) and as the spectra are obtained without changing the operational parameters, they are considered nonscanning systems.

I.3.3.3.1 Orbitrap

The Orbitrap mass analyzer type is composed of a central spindle-shaped electrode and a split outer electrode. Additionally, the operation of this type of instruments requires a complementary quadrupole ion trap, called C-trap due to its C-shaped form, and ultra-high vacuum conditions for a proper ion injection. Ions from ionization are collected in the C-trap and then injected into the orbitrap as high-speed pulses. As charged species and due to the

electrostatic field created, once inside the orbitrap, the ions acquire a trajectory composed of circular motion around the central electrode and an oscillatory movement along the same axis. The resulting motion is detected by the split outer electrode via image current (also called transient). Ultimately the transient signal is converted into frequencies after the FT analysis, with different ions being detected since their axial oscillatory movement is proportional to the *m/z* values [211].

I.3.3.3.2 Ion cyclotron resonance (ICR)

The FT-ICR analyzers are based on the cyclotron frequency, *i.e.* circular orbits, of the ions in a fixed magnetic field due to the Lorentz force. In these instruments, the magnetic field not only imposes the cyclotron motion of the ions but also acts as an ion trap, keeping the ions inside the analytical cell. Then, a transverse electric field must be applied to accelerate the ions and increase the radius of their orbits, which allows their detection by the receiver plates. As the cyclotron frequency is proportional to the mass of the ions, this excitation can be mass-selective and is usually called as resonant excitation. Upon the detection of the ions by the plates, the image currents are recorded and transformed into frequencies by the FT, and ultimately converted into *m/z* values [212].

I.3.3.4 Time of flight (TOF) analyzers

Inherently simple systems, TOF analyzers consist of an acceleration grid and a flight tube. Its principle is based on ions with different *m/z* ratios travel from the source to the detector with disparate times. Essentially, TOF instruments measure the time that a particular ion takes to reach the detector, the mass of the ions can be calculated. Since the ionic separation occurs within the space of the flight tube, the electronic parameters do not need to be changed during the analysis, these analyzers are nonscanning systems. As consequence, ions must be introduced into the mass analyzer with discontinuous pulses produced by the ion source [212]. This principle makes this type of mass analyzer more suitable for MALDI ion sources than continuous ion sources like ESI. In MALDI systems, packets of ions are produced upon each laser pulse, which usually are added up several laser shots to increase the ions signal and give a mass spectrum. However, the spatial distribution and kinetic energy produced by laser-based ion sources are widely distributed within the resultant ion packets, which ultimately fallouts in loss of resolution. To overcome these pitfalls, an acceleration grid has been introduced inside the TOF systems. Addressing the spatial issues, this phenomenon is also called delayed extraction and allows the alignment and concentration of the ionized species before their acceleration into the flying tube [213].

Different architectures can be found among the TOF analyzers. The simplest forms are the linear TOF structures where the ions move in a straight line after the ionization process through the field-free drift region until the detector. However, the delayed extraction grid reduces, to some extent, the inhomogeneity of the energy content among the ions of packets,

the differences between the acquired energies of the ions still remain broadly distributed among them. As consequence, different arrival times will be detected for ions with the same m/z but with different energies. To overcome these problems, TOF analyzers have been incorporating reflectors to counteracting the spreading of energy effects. Reflectors, which are ion mirrors, create an electrical field responsible for the reversion of the ion trajectory and equalization of arrival times onto the detector [212].

I.3.3.5 Multianalyzer systems

The development of multianalyzers arises from the need to investigate not only the molecular mass of the ionized species but also their structure. Also called MS/MS instruments, the multianalyzer systems result from the combination of a first analyzer, usually used to select the ions of interest, and a second analyzer which is often of higher resolving power. Between the two mass analyzers, a fragmentation cell can be incorporated. Different formats can be classified into MS/MS systems, including the tandem and the hybrid instruments [214].

I.3.3.5.1 Tandem MS/MS instruments

Multi-systems composed of the same type of mass analyzer are designated as tandem MS/MS instruments and are described below.

I.3.3.5.1.1 Triple quadrupole (QqQ)

Consisting of two quadrupole devices (Q1 and Q3) the triple quadrupole is also combined with a central component, q2. While the Q1 scan across a range of *m/z* values or selectively filter ions with a specific *m/z* ratio, the q2 can be used as an ion guide or as a collision cell to fragment the ions from Q1. Then, the product ions pass to the Q3 to be analyzed and to obtain the mass spectrum. The q component when used in a wide band pass mode acts as an ion guide where all ions are transmitted regardless of their *m/z* values, whereas when operating a collision cell, the ions from the Q1 are fragmented prior to subsequent analysis by the Q3 unit. As result, QqQ systems can be used either in scanning or static (as selected ion monitoring) [206].

I.3.3.5.1.2 TOF/TOF

Two TOF analyzers can be combined and perform tandem MS. In this approach, the selection of the interested ions is made through an electronic gate called a timed ion selector (TIS) in the first TOF compartment. Then, selected ions are fragmented in a collision cell placed between the two TOF analyzers, and the resultant fragmented ions are analyzed by the second TOF analyzer [206].

I.3.3.5.1.3 Tandem QIT and LIT

Ions traps, like QIT and LIT, can also be set in a tandem mode. However, in these cases, the multiple MS analysis is not separated in space as in QqQ and TOF/TOF instruments,

but the multiple scans are sequentially performed using the same analyzer. Therefore, MS/MS data acquisition through these instruments are considered as separated in time. As result, the QIT and LIT tandem systems allow the analysis of multiple MS experiments, MS^n, using a single mass analyzer device [206].

I.3.3.5.2 Hybrid MS/MS instruments

When different types of mass analyzers are combined these instruments are termed as hybrid systems.

I.3.3.5.2.1 QTOF

A very versatile mass analyzer configuration is the QTOF format. In this hybrid system, ions coming from the ionization source are analyzed by the quadrupole unit and then transferred to the TOF analyzer. Also called as QqTOF, these instruments incorporate the q unit similarly to QqQ systems. Operating in the wide band pass format, all the ions deriving from the quadrupole pass through the q section and transferred to the TOF analyzer. Here the q unit acts as an ion guide and a full scan of the MS data is obtained. In opposition, when the quadrupole unit is operating as a filter and is set in a narrow band pass mode, only ions with a specific *m/z* ratio are passed into the q unit. In this scenario, the q section acts as a collision region, and the fragmented ions are then transferred into the TOF instrument to acquire the MS/MS data [215].

I.3.3.5.2.2 FT-based hybrid instruments

The combination of different types of mass analyzers is also verified in FT-based instruments, including the orbitrap and the ICR-FT. The most common combinations are LIT-FT as the LIT unit is itself an MS/MS device, which can be used for collecting all spectra or for selective ion collection. For instance, orbitrap mass analyzers, which are by themselves single analyzers, are commonly referred to as orbitrap instruments interfaced with LIT units, making them a hybrid instrument [212].

I.3.3.6 Mass analyzers comparison

Nowadays, it is very common to combine two or more analyzers to improve and extend the analytical capabilities of MS instruments. In daily practice, the choice of an analyzer, or most likely a multianalyzer, should have into consideration several aspects, including not only the mass resolution and accuracy but also the upper mass limit and the speed of the scan, as it affects the data. The resolution, or mass resolution, is the separation observed in a mass spectrum, while the mass resolving power is the ability of an instrument to resolve neighboring peaks, is commonly expressed in terms of parts per million (ppm). The accuracy of mass measurement is also an important aspect. Expressed in terms of millidaltons (mDa), or ppm, represents the numerical measure of the difference between the mass calculated and experimentally determined of an ion. The maximum measurable *m/z* values of a given instrument are designed as the upper mass limit and influence the suitable mass range The

speed rate that each equipment can achieve to acquire a mass spectrum, measured in masses/unit time, is the other common parameter taken in consideration when evaluating a mass spectrometer the mass. In Table I.9 are outlined the most common aspects assessed in the MS instruments frequently used in proteomics.

Table I.9 - Comparison of performance characteristics of the most commonly used MS for proteomics. Adapted from [103].

Instrument	Mass resolution	Mass accuracy (ppm)	Sensitivity	*m/z* range	Scan rate
QIT	1 000	100 – 1 000	Picomole	200 – 4 000	Moderate
LTQ	2 000	100 – 500	Femtomole	200 – 4 000	Fast
QqQ	1 000	100 – 1 000	Attomole	10 – 4 000	Moderate
QqLIT	2 000	100 – 500	Femtomole	5 – 2 800	Fast
TOF	10 000 – 20 000	< 5	Femtomole	No upper limit	Fast
TOF/TOF	10 000 – 20 000	< 5	Femtomole	No upper limit	Fast
QqTOF	10 000 – 20 000	< 5	Femtomole	No upper limit	Moderate to fast
FT-ICR	50 000 – 750 000	< 2	Femtomole	200 – 4 000	Slow
LTQ-Orbitrap	30 000 – 100 000	< 5	Femtomole	200 – 4 000	Moderate to fast

I.3.4 DETECTORS

The last compartment of a mass spectrometer is the detector. The principle of operation of this section is to detect and determine the abundancies of the emerging ions, resulting in amplified and stored signals. As the movement of the ions traveling across the mass spectrometer, from the source through the analyzer to the detector, constitutes an ion current, this section is more properly called as an ion current detector. The first step of a detector is the conversion of the ion current into an electric current, through the acceleration of the arriving ions onto a conversion dynode, resulting in a release of electrons. Along with ion detection, the detectors are also responsible to multiply (amplify) the signal produced by the ions. The small current of each ion at a particular *m/z* value is amplified when reaches the detector surface, increasing the signal attributed to that mass ratio [216].

Although the principle of operation is fundamentally the same, detectors can be designed in different formats. The first and simplest electronic ion detector is the Faraday cup [217]. Consisting of a metal container, the ions are discharged upon their arrival and a current is generated. However, this format cannot amplify the signal derived from each arriving ion. To improve sensitivity, detectors start to incorporate a combination of amplification steps, electron multipliers, which come in a variety of forms, including the discrete dynode electron multiplier, the continuous dynode electron multiplier, and the multichannel plate detector (MCP). While the discrete and continuous dynode detectors are often used in scanning instruments, the MCP

detector is more suitable for TOF analyzers due to their high demand for rapid detection and amplification. Another format is the Daly detector, which uses a photomultiplier instead of an electron multiplier. In this type of detector, the ions coming from the analyzer sections firstly collide on the conversion dynode to eject the resultant electrons onto the surface of the coated scintillant plate to produce photons. Generated photons are, ultimately, detected and the signal amplified by the photomultiplier. Lastly, in FT-based instruments, the ion detection is not based on ion-counting approaches, as the ones previously described. In these instruments, which employ an image current detection, ions by passing close to a metal surface induce an electric current on that surface, which in turn results in an alternating (sinusoidal) electrical current. Different sinusoidal currents are produced by each ion, resulting in an image current where each frequency is proportional to its *m/z* value. To finish, the resultant image current must be deconvoluted mathematically by FT techniques, to obtain a mass spectrum [212].

I.3.5 FRAGMENTATION STRATEGIES

Over the past years, several strategies have been developed to avoid excessive fragmentation during the ionization processes. Particularly in hard ionization methods, such as EI, the energy imposed on the analytes is higher than the required for the formation of the molecular ion. As result, the molecular bonds of the analyte are dissociated, and non-desired fragments are obtained. This has been overcoming with the development of the soft ionization techniques, including the MALDI and ESI. Nevertheless, fragmentation is a desired process in MS/MS analysis, where a particular precursor is decomposed into smaller product ions. Occurring within the collision cell chamber, this process is found on the principle that once inside the peptide ions interact with the collision gas (usually nitrogen, helium, or argon) and undergo fragmentation [218]. Peptide cleavage can occur in one or more chemical bonds in the peptide, resulting in different fragmentation patterns as illustrated in Figure I.6A [219]. Two ion-series are then formed, the containing the N-terminus of the ion, denoted as a-, b-, and c-ions, and their corresponding containing the C-terminus, denoted as x-, y-, and z-ions. As the most commonly applied fragmentation strategies use low collision energies (< 100 eV), fragmentation mainly occurs along the peptide backbone, originating preferentially b- and y-ions (Figure I.6B) [220]. Thus, in an MS/MS spectrum, the mass of the individual amino acid residue com be inferred as the difference between adjacent y- or b- ions [218].

Although the vast majority of the instruments use low-energy collision-induced dissociation (CID), other types of tandem mass fragmentation include infrared multiphoton dissociation (IRMPD), electron transfer dissociation (ETD), and electron capture dissociation (ECD) [220]. In CID type, ions are accelerated into the collision cell with a provided kinetic energy. Inside, through the multiple interactions with the collision gas, the ions acquire, cumulatively, additional vibrational energy ending up in their fragmentation into the product ions. Contrary to CID, which uses the q intermediary section between the two mass analyzers as collision cells to perform the fragmentation, the IRMPD takes place within an ICR cell and

therefore is the fragmentation type used in FT-ICR instruments. However, its principle is very similar to the CID, since, upon an infrared incidence, the ions interact progressively with the photons until the point that the energy absorbed is high enough to induce fragmentation. Lately, other methods, as ETD and ECD, have been gaining momentum due to their ability to provide structural information of the precursor ion [220].

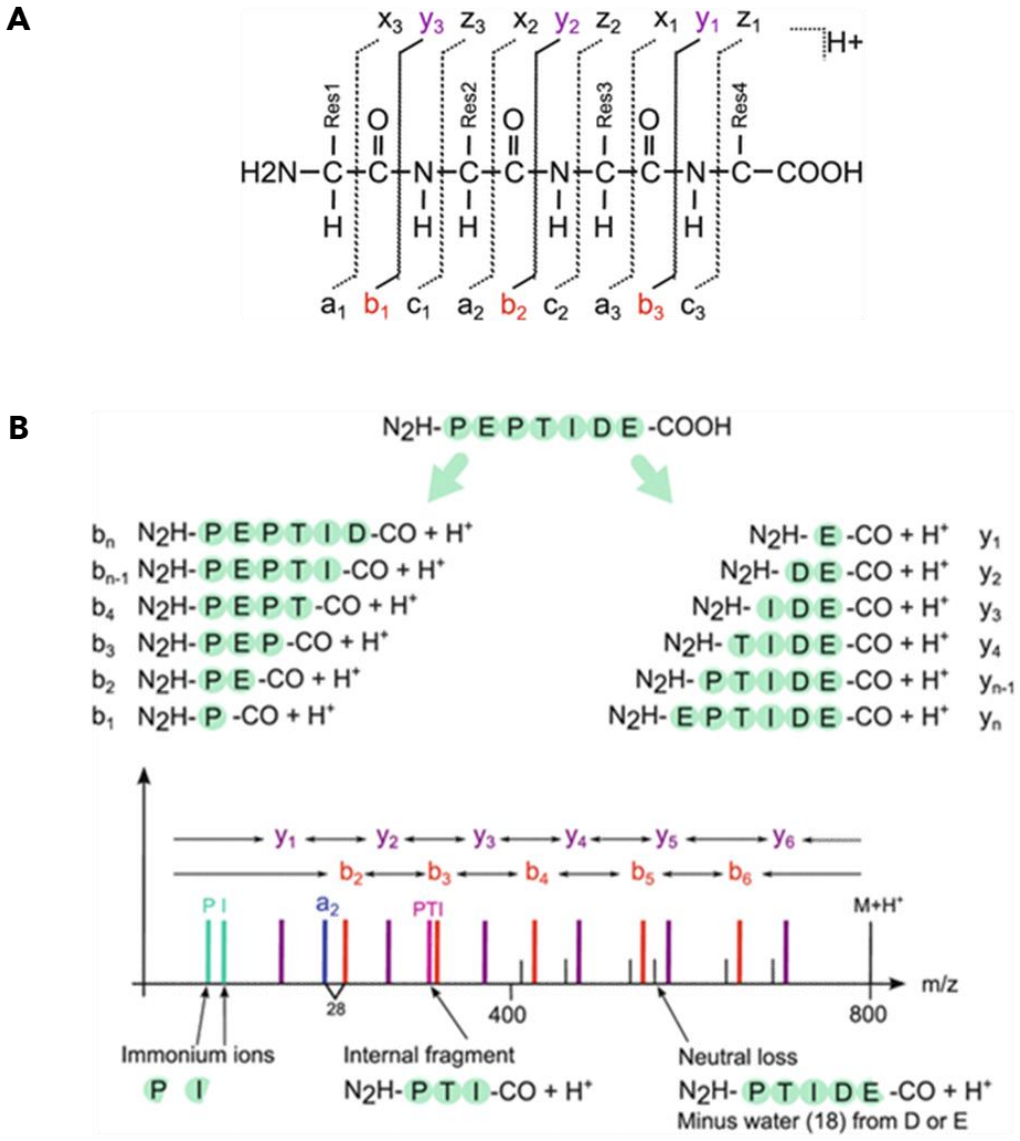

Figure I.6 – Peptide fragmentation A) Two ion-series based on the terminal part of the peptide can occur resulting in different patterns. B) Typical ions observed in low energy fragmentation mass spectrum. Adapted from [219].

I.4 DATA SYSTEMS AND ANALYSIS

The technological advances in MS instruments have greatly improved the ability to explore hundreds of proteins simultaneously. However, the analysis of a large amount of data is *per se* one of the major challenges of modern MS-based proteomics. Consequently, MS-based techniques have shifted from hand-adjusted potentiometers and photographic paper to combinatory systems of computer hardware together with sophisticated software engines. Concluding the MS analysis, the data systems enable not only the acquisition and processing of all data as well as the control of all operational processes either in each MS instrumental section and in integrated peripheral instruments, such as LC systems.

I.4.1 DATA ACQUISITION AND SIGNAL PROCESSING

The main role of data systems is to deliver a mass spectrum as output resultant from the electrical signals detected in the ion detection section. Consisting of a two-dimensional representation, a mass spectrum plots all the ionized molecules detected according to their *m/z* values and their respective intensity. The starting point of an integrated data system is data acquisition, followed by signal processing. During data acquisition, all the electrical data from the detector is captured and stored as raw data. This native data is then processed to detect effective peaks, *i.e.* specific *m/z* with an intensity corresponding to a given ion, present with the analyzed sample. In signal processing, several critical parameters, like noise reduction, peak detection, and monoisotopic peak determination, are the basis to deliver reliable data [221]. The noise present in the systems impacts the detection of the ions, therefore one of the processing steps is to evaluate the signal-to-noise ratio and set the threshold above the noise level to improve the quality of the spectra acquired [222]. After noise filtration, the detection of the peak is achieved through algorithms and tools related to isotope and charge state deconvolution. Most elements are naturally composed of different mass numbers resulting from different numbers of neutrons albeit having nuclei of the same atomic number. Those variants are called isotopes and contribute to the presence of a collection of peaks belonging to the same ionized molecule but incorporating different isotopes. As result, in a mass spectrum, each ion can present an isotopic distribution or pattern, counting the different abundances of the isotopes in nature. In complex samples, where is frequent the superposition of isotopic envelopes, isotopic distribution of a particular ion, deconvolution methods are applied to resolve these peak overlapping issues [223]. Focusing on grouping spectral peaks into isotopic envelopes, the deconvolution allows the efficient determination of the charge state and the monoisotopic mass necessary for accurate calculation of the mass of the corresponding ion [224]. Mass calibration and alignment are also technical aspects often processed in the raw data. Although calibration is rarely needed for mass dimension and the alignment of different LC-MS scans can be conveniently overcome by aligning the retention time dimension [225].

I.4.2 PROTEIN IDENTIFICATION

In mass spectrometry, the resultant peptide signals with a specific *m/z* ratio, intensity, and retention time are called features. In proteomics, the features obtained in the MS and MS/MS spectra are used to identify the proteins present in the sample, which can be accomplished through mass or sequence information [226].

I.4.2.1 Peptide mass fingerprinting (PMF)

PMF consists in the identification of the proteins based on the m/z values of the peptide ions detected in the MS1 spectra, which is uses as a fingermark of the protein. This identification technique is mostly used in instruments with one MS analyzer, such as in the MALDI-TOF mass spectrometer. In PMF, all the masses calculated from the peaks detected in the MS spectra, and a list of peptides masses is generated. Then, taking into consideration that the peptides originate from a specific enzymatic cut, such as trypsin, the resultant list can be compared to a theoretical mass database, composed of predictable tryptic peptides, to determine the identification of the protein [227].

I.4.2.2 Peptide sequencing

A different method to identify proteins is through peptide sequencing. In peptide sequencing the protein identification is made by the interpretation of the tandem MS (MS/MS) spectrum, being more appropriate for large-scale high-throughput analysis. Firstly, the peptide ions are detected in the first mass analyzing step, then precursor ions are selected to the second stage where they are submitted to fragmentation and subsequent analysis [227]. Currently, different methods for selecting the precursor ions are being employed for collecting proteomic data.

I.4.2.2.1 Data-dependent acquisition (DDA)

One of the most widely used strategies for collecting data in tandem MS experiments is the data-dependent acquisition (DDA). In DDA, the selection of the precursors (peptide ions) for fragmentation is based on their occurring intensities, resulting in the top-N most abundant ions selected for MS/MS characterization. This way, precursor ions are chosen and collected when their MS intensities exceed a pre-defined threshold, usually in small isolation windows of ≤ 1 Da wide throughout the entire MS scan [228]. Attempting to minimize the selection of redundant peptides precursors, this acquisition mode is generally coupled with a dynamic exclusion method that prevents the reselection of precursors with the same *m/z* value in a specific time range [229]. However, DDA-based methods reveal some pitfalls in terms of reproducibility, as the overlap between technical replicates is < 75% [230], even though they represent powerful strategies with great proteome coverage and extended dynamic range of detection [229].

I.4.2.2.2 Target data acquisition

A different acquisition mode relies on the selective selection of precursors for further fragmentation in MS2. As the acquisition of the precursors' drives from a pre-selection of specific *m/z* values at an elution time and subsequent exclusive fragmentation, this acquisition mode is defined as target data acquisition. Two main strategies can be used in target acquisition mode, the selected reaction monitoring (SRM), and the parallel reaction monitoring (PRM) [231]. In the former, the set of peptides to analyze are filtered, with a narrow isolation window, in the MS1 scan, submitted to fragmentation, and the resultant fragment ions are once again filtered, according to a specific *m/z* ratio, in the second MS scan. This double selection makes the method highly specific since the probability of two peptides having the same mass at both MS scans along with the exact same retention time is too low [231]. Since SRM can sequentially measure several fragment ions, this technique is often also referred to as multiple reaction monitoring (MRM) [232,233].

Contrasting with SRM/MRM, in PRM acquisition mode all the fragment ions resultant of the pre-selected precursor are simultaneously analyzed at MS2 level. Thus, in PRM the selection in advance of the fragment ion is exempted and the best fragment ions for peptide identification and quantification can be chosen *posteriori*, revealing higher flexibility [231].

However, the knowledge of a peptide's elution time and corresponding MS1 *m/z* value prior to the acquisition along with selective fragmentation, limit these approaches to pre-existent spectral libraries and only to the selected targets [234].

I.4.2.2.3 Data-independent acquisition (DIA)

An alternative method, which acquires all both MS1 and MS2 data in an unbiased manner, is the so-called data-independent acquisition mode. In DIA, all detectable ions, either with sequential isolated *m/*z windows or at a given time point, are selected for fragmentation. The selection of the precursor ions is this way independent of their abundance or *m/z* value [235].

Since the beginning of this century, various DIA acquisition schemes have been implemented. Including but not limited to, the sequential window acquisition of all theoretical mass spectra (SWATH-MS) [236]. The principle of this method falls in the acquisition and fragmentation of all ionized compounds within sequential *m/*z windows. Other methods include diaPASEF from Bruker, PAcIFIC from Thermo Scientific, and SONAR from Waters [235,237].

These acquisition modes have gained their impetus in tandem mass spectrometry as they provide better sensitivity and accuracy, resulting in improved reproducibility and proteome coverage [237]. Nevertheless, the consecutive survey of MS scans and the fragmentation of all ions, generate not only highly complex data sets but also the loss of the link between the precursor and the fragmented ions. As consequence, more elaborate processing algorithms are required in these experiments to fully analyze all the data.

I.4.3 PROTEIN QUANTIFICATION

In proteomics research, the quantitative information about proteins can be evaluated as the total of protein content as well as for individual protein quantification. A wide range of methods comprising the measurement of UV at 280 nm, Bradford, bicinchoninic acid (BCA), and Lowry assays have been used to quantify the total protein content [238]. On the other hand, the individual protein quantification has been performed through the (i) enzyme-linked immunosorbent assay (ELISA), (ii) the western blot analysis, or more recently, (iii) via by mass spectrometry [239].

I.4.3.1 MS-based quantitative methods

Currently, protein quantification using the MS can be classified as i) relative or absolute quantification; or ii) label-based and label-free quantitative proteomics [240].

I.4.3.1.1 Relative versus absolute quantification

Protein quantification methods relying on the determination of the exact amount of a protein, for instance in units of ng/mL or mol/cell, are considered as absolute quantitative methods, whereas relative quantification methods rely on the comparison of the specific protein level in different samples, resulting in quantitative measurements expressed as relative fold change of protein abundance [240].

I.4.3.1.2 Label-based and label-free methodologies

Although the intensity of the ionized peptides is correlated to their abundancies, due to different ionization efficiencies among the different molecules, the accurate measure of protein amounts cannot be inferred directly to their intensities in a mass spectrum [241]. To overcome such pitfalls, various quantitative has been implemented in MS-based proteomics studies. Typically, two main techniques, label-based and label-free, have been widely employed (Figure I.7).

I.4.3.1.2.1 Label-based quantification

In label-based strategies, stable isotope or mass tags are incorporated into proteins or peptides to originate a light and a heavy form with a mass shift, which should be at least 3- or 4-Da between samples to minimize isotopic overlapping [242]. Besides the incorporation of the tags, the workflow entails the mixture of the samples before MS analysis and the comparison between the sample is made in the same MS scan. The different isotope labels can be introduced in samples metabolically, chemically, or enzymatically [243].

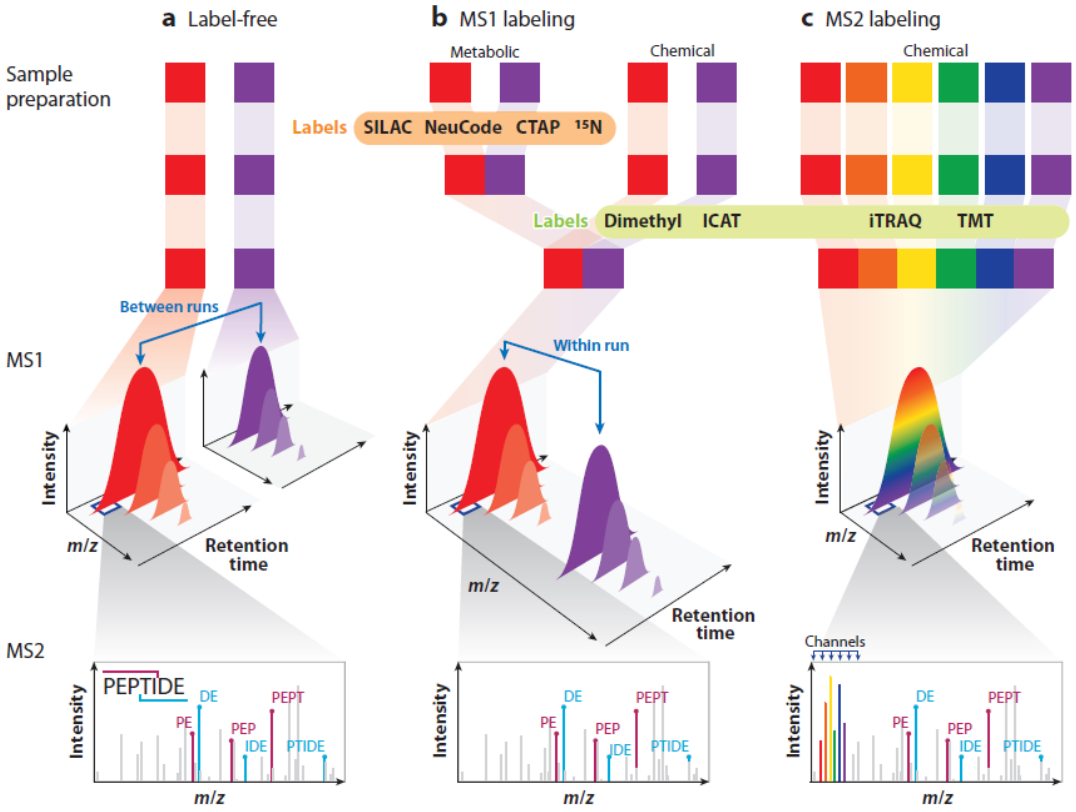

Figure I.7 – Overview of (a) label-free and (b,c) label-based quantification methods. In label-based methods, distinction between (b) MS1 and (c) MS2 (isobaric) labeling is also represented [244].

Metabolic incorporation of amino acids with substituted stable isotopes is the main principle of the stable isotope labeling by amino acids in cell culture (SILAC) method. Also called as amino acid-coded mass tagging (AACT), this approach relies on the incorporation of the light and the heavy form of amino acid, such as lysine labeled with ^{12}C or ^{13}C, respectively, into two cell cultures. In SILAC, the labeling is performed on the proteins, which are then harvested from both cell cultures are mixed and submitted to digestion and MS analysis together within the same sample. Being this way, a technique very straightforward since does not require any chemical reactions, with minimal separate handling between samples. Even though, metabolic labeling has restricted multiplex capacity and also are limited to cell culture and cannot be applied to primary tissues or body fluids, such as clinical specimens [241]. Chemical tagging approaches can circumvent those limitations. In chemically labeling strategies, the incorporation of isotopic labels is accomplished by chemical reagents and can be performed at the protein or peptide levels [245]. At the peptide level, the most commonly used methods are the isobaric tags for relative and absolute quantification (iTRAQ) and the tandem mass tags (TMT) [241]. In these quantitative approaches, different samples are labeled with reporter groups harboring different mass, which are detected in the second scan (MS2) [245].

With a multiplexing capability of 8- and 10-plex, for iTRAQ and TMT, respectively, these quantitative methods overcome the pitfalls of the SILAC strategies [241]. Another chemically labeling approach is the isotope-coded affinity tag (ICAT), which introduces a biotin tag coupled to a linker, via a thiol-reactive group, onto reduced cysteine residues. In this technique, the linker group, which is the stable isotope-labeled linker, can be employed in different samples as light or heavy tag, if incorporate ^{1}H or ^{2}H, respectively [241]. However, similar to SILAC, this strategy also has a lack of multi-sample analysis. Lastly, enzymatic labeling using $^{18}O/^{16}O$ has also been regularly applied in proteomic studies. Though this technique can be performed during proteolytic digestion, the incorporation of ^{18}O into the C-terminus is more commonly performed in a second incubation after proteolysis and results in an introduction of a 4-Da mass shift between the two samples in the MS scan [241,243].

The absolute quantification can be incorporated in these relative approaches by the addition of heavy-labeled peptides to the samples. With the introduction of a known concentration, these peptides act as internal standards on the peptide level [246]. Several absolute quantitative approaches have been developed, including the use of synthetic peptides (AQUA) [247], artificial concatemer of standard peptides (QCAT) [248], isotope-labeled full-length target proteins (PSAQ) [249], among others.

Even though, the restricted applicability of such labeling methods in high throughput cohorts of samples remains a major limitation in proteomics fields.

I.4.3.1.2.2 Label-free quantification

Alternatively, label-free quantification (LFQ) is another approach for protein quantification. LFQ can be divided in two main strategies, the spectral counting (SC) and intensity-based workflows (Figure I.8).

In the first strategy, the number of MS/MS spectra identifying peptides of the same protein are counted and compared across multiple LC-MS/MS runs. Based on an exponentially modified protein abundance index (emPAI), consisting of the exponential form of the number of observed peptides divided by the number of theoretical peptides of a given protein minus one, this strategy can estimate absolute protein content in complex mixtures [250]. Along with the emPAI strategy, the absolute protein expression (APEX) is another methodology used for absolute protein quantification based on spectral counting [251].

On the other hand, in the second approach, quantification is based on the mass spectrometric signal intensity belonging to a precursor ion as determined by the extracted ion chromatogram (XIC) [243]. Conceptually, XIC-based quantification is determined by the area under the curve or peak height, during MS1, of each eluted peptide [241].

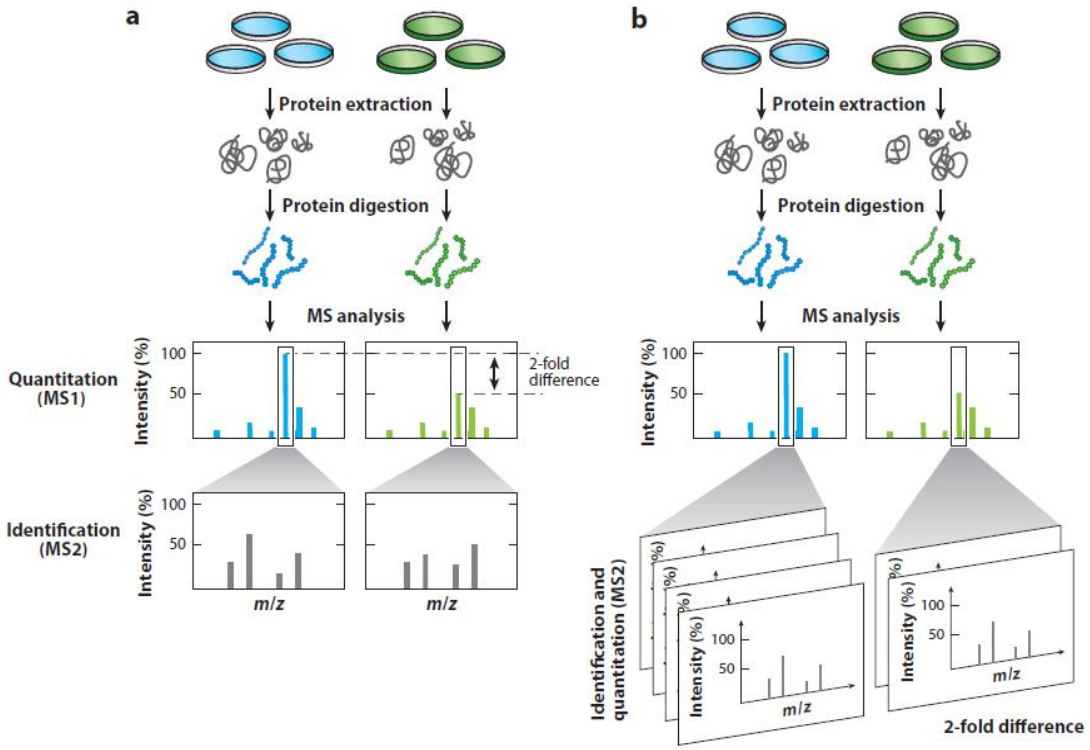

Figure I.8 – Comparison of (a) intensity-based and (b) spectral counting-based strategies for label-free quantification (LFQ) [241].

Likewise to spectral counting strategies, in spectral intensity-based methods, approximate measures of absolute protein abundances can be obtained through computational tools, including the intensity-based absolute quantification (iBAQ) [252], the Top3 [253], and more recently the total protein approach (TPA) [254].

I.4.3.1.2.3 Total protein approach (TPA)

In most LFQ methods, either based on spectral counting or intensity, to infer absolute protein abundance, is necessary some biochemical inputs like the determination of the total amount of protein analyzed or the use of protein standards [255]. Contrary to these methods, in the total protein approach (TPA) the absolute protein quantification is achieved without any specific knowledge of the sample and is standard free [254]. This method is based on two assumptions (i) the total MS signal in a large-scale proteomic analysis reflects the total protein content and (ii) the partial abundance of a given protein in the whole sample corresponds to its MS signal. Thus, the amount of a given protein (p_i) is calculated by the following equation (equation (1)):

$$Total\ protein\ (p_i) = \frac{MS_{signal\ (p_i)}}{Total\ MS_{signal}} \quad (1)$$

Which can be further convertible into molar concentration, expressed in mol per g of total protein, if the molecular weight of the given protein (p_i) is taking into consideration, as present in the following equation (equation (3)):

$$Protein\ concentration\ (p_i) = \frac{MS_{signal\ (p_i)}}{Total\ MS_{signal}\ x\ MW\ (p_i)} \quad (2)$$

In Figure I.9 – Comparison of absolute protein quantification methods used in LFQ [255].is presented a schematic summarization of the absolute protein quantification methodologies used in LFQ.

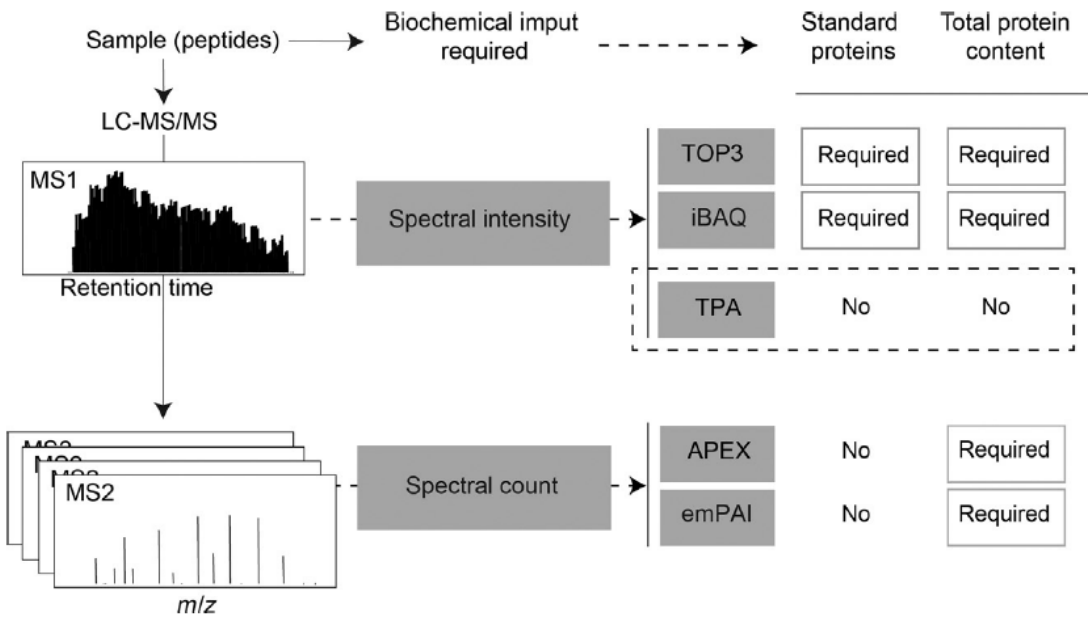

Figure I.9 – Comparison of absolute protein quantification methods used in LFQ [255].

I.4.4 BIOINFORMATIC TOOLS

In bottom-up proteomics, where the original protein is digested into peptides before MS analysis, the association of those peptides with their precursor protein is fundamental. Therefore, several computational tools have been developed to allow the analysis of such high-throughput data.

I.4.4.1 Protein identification and quantification

Several bioinformatic solutions have been developed for large-scale and high-throughput protein analysis. From a computational point of view, several tools have been developed for processing, analyzing, and managing the MS data [226]. Different software tools with different algorithms can be used as platforms to analyze MS-based proteomics data. Nevertheless, a common workflow for the bioinformatic analysis of LC-MS/MS data, present in Figure I.10, consists of (i) binary format data conversion; (ii) pre-processing of the mass spectra including denoising, baseline correction, normalization, and peak detection; (iii) collection of all individual peaks through feature finding; (iv) protein inference based on database searching; (v) protein quantification; (vi) statistical assessment and lastly; (vii) biological analysis [226].

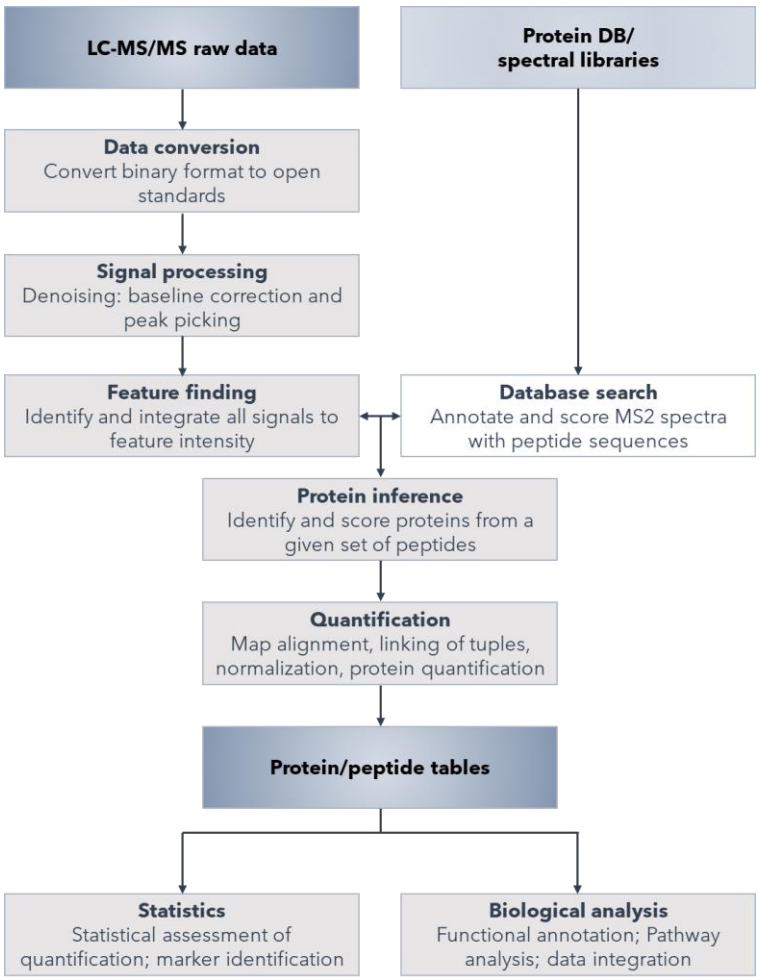

Figure I.10 – Typical bioinformatic pipeline for shotgun proteomics data. Adapted from [226].

While the three first steps are mainly processing parameters to improve the quality of the results, the fourth step is one of the crucial points for accurate protein identification. For protein inference, firstly the peptides need to be identified. Classically, in tandem MS, peptides

are identified by comparing the fragmented spectra, MS2 scan, against the *in-silico* spectra. This task is generally achieved through protein database search engines. As shown in Figure I.11, the experimental spectrum resultant from the MS analysis is compared against a theoretical one, and a match is called a peptide-to-spectrum match (PSM). Then, a list of candidate PSMs for every spectrum, with respective scores is presented as the output of the search engine [256].

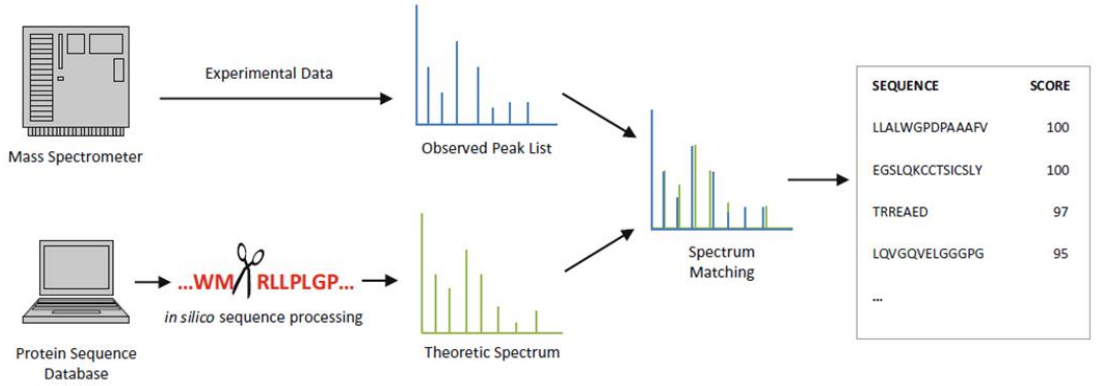

Figure I.11 – Standard workflow for sequence database searching [256].

The resultant PSMs scores are then used to access the best match by the means of statistical assessments. The most common assessment used is the target-decoy strategy, which applied a decoy database composed of reversed, shuffled, or randomized protein sequences as negative controls to estimate the overall false discovery rate (FDR) [257]. Peptide sequences are then assigned usually with a 1% FDR threshold. A wealth of sequence database searching engines are widely used in proteomics, including Mascot, SEQUEST, X!Tandem, Andromeda, VEMS, among others [237].

After peptide identification, the protein inference is accomplished by mapping the peptide sequences identified against available protein sequence resources, like the Universal Protein Resource (UniProt, www.uniprot.org) [258]. At the protein level, a second level of FDR, usually also with an FDR threshold of 1%. For protein families, which generally have identical peptide sequences, it is difficult to distinguish the presence of different isoforms. In this case, those proteins will be grouped together as a protein group.

I.4.5 STATISTICAL ASSESSMENT AND BIOLOGICAL NETWORKS

The last steps of the data analysis pipeline encompass the statistical evaluation based on the proteomics strategy implemented and the rationale of the data. Thus, the ultimate goal of data analysis is to translate large amounts of proteomic data into biological knowledge that can lead to medical decisions [259].

I.4.5.1 Statistical tools

Downstream analysis of proteomic data integrates several aspects including statistical evaluation. As the most popular statistical methods, the *t*-test, analysis of variance (ANOVA), and logistic regression provide a level of significance of the observed result, the so-called *p*-value. The *p*-value measurement, often set as 0.05, 0.01 or 0.01, means that probability of getting by chance a false positive is of 5%, 1% or 0.1% respectively [260]. With that being said, two types of errors can occur in statistical approaches, the type I errors, and type II errors. The former represents the false positives, those situations in which the null hypothesis is erroneously rejected. By contrast, the latter occurs when the null hypothesis is erroneously accepted, originating false negatives [260]. Therefore, diverse multiple hypothesis testing correction methods, including the widely used permutation-based FDR and Benjamini-Hochberg's correction methods have been routinely used to minimize such occurences [259,260]. In proteomics experiments, a standard goal is to compare protein expressions between to samples groups, usually a disease or treatment condition *versus* controls [261]. Resulting in a differential analysis approach, the data retrieved, which is centered on statistical methods, as mentioned above, can be further visualized through several methods, including principal component analysis (PCA), clustering methods, or with a receiver operating characteristics (ROC) curve [260]. It is noteworthy to mention, in discovery approaches those procedures should be performed in an unsupervised manner, where all the input data are unlabelled and analysis is based on the similar attribute profiles instead of samples' labeling (condition) [262].

I.4.5.1.1 Software platforms

In the context of big data, as MS-based data, elaborated computational tools are necessary to perform its analysis. Several programming languages, such as R or python, has been used to fulfill such high demands. However, a massive barrier often arises between those informatic skills and biological researchers. Thus, to translate such complex outputs into valuable biological data with life expectancy significance, some platforms have emerged to overcome those shortcomings. One of those platforms, developed by the Cox group [263] is the Perseus software. Constructed to have an intuitive and user-friendly interface, this computational platform is fitted with an arsenal of algorithms allowing the statistical and downstream analysis [259,263].

I.4.5.2 Biological interpretation

Following statistical analysis, the biological interpretation of such high-throughput proteomics experiments is the next challenge. Functional annotation or enrichment analyses have been implemented to provide biological insights into the underlying mechanisms of different conditions [262].

I.4.5.2.1 Network analysis and visualization

Functional network generation and their visualization is a necessary tool to interpret the data resulting from the statistical approach. Different databases have been useful in systematically collect and organize biological information. For instance, while the main purpose of Gene Ontology (GO) [264] is to assign biological/cellular/molecular terms to the genes, in a hierarchically structured way, the Kyoto encyclopedia of genes and genomes (KEGG) [265] annotates functional pathways to the genes and the Reactome [266] is a curated database of human pathways and reactions. For visualizing complex networks and integrated all the biological information, an open-source platform, the Cytoscape software, has been successfully in workflow-based network analysis. As a general network tool, Cytoscape only allows the visualization of the data, and therefore, several plug-ins, able to import information of the diverse existing databases, have been developed to use the functionalities of the Cytoscape software. This is the case of the ClueGO plug-in which is able to integrate not only GO terms as well as KEGG and Reactome pathways [267]. Another example is the recent StringApp which incorporates the predicted protein-protein interactions [268].

I.5 TRANSLATIONAL AND PROTEOMICS REASEARCH

Basic research results in general knowledge. It is performed without thinking in practical ends. In medical context, the bridge that overcomes the gap between the basic science and the patient care, has been referred as translational research [269].

In modern biological science, more holistic approaches capable of qualitatively and quantitatively characterize complex biological systems are overtaken the traditional hypothesis-driven studies. This novel approaches, referred to as “omics” approaches, have been increasingly adopted towards knowledge discovery which is less hypothesis-driven, where a formulated hypothesis is assumed, and more data-driven, where the whole biological systems are evaluated in a non-targeted and non-biased manner to define a premise that can be further tested [270]. Aiming to unveil the complexity of different biological molecules within a living organism, these ‘omics’-based screening technologies are commonly applied to genes, mRNA, proteins, and metabolites. As one of the most developed ‘omics’ technology, proteomics was initially coined in the mid-1990s by Marc Wilkins [271], and the firsts proteomic studies began with the mapping of proteins from *Escherichia coli*, mouse, and guinea pig by O'Farrell [272], Klose [273], and Scheele [274], respectively. Since then, the potential of proteomic strategies has growth exponentially by the wide range of applications [218]. Over the decades, these applications have allowed not only a deeper understanding of normal physiological processes but also a better knowledge of diseases, such as cancer. Focusing on the investigation of multiple molecules simultaneously, this approach has been also applied to biomarker discovery playing itself a role in screening, diagnosis, and prognosis throughout the scientific and medical community [275].

I.5.1 PROTEOMICS APPLICATIONS

Depicting snapshots of proteins compositions of specific cells or tissues at a particular time point, the study of the proteomes has been fundamental to address several biological questions. Thus, the aim of proteomics goes beyond the identification of all proteins in a cell or a tissue to provide a more holistic assessment of the biochemistry in which such proteins are involved. (Figure I.12) [218].

I.5.1.1 Structural proteomics

Regarding the types of proteomics, the proteomics studies whose main goal attempts to identify all proteins within a protein complex, organelle, or a specific cellular portion are known as structural proteomics. These types of studies generate the three-dimensional structure, which along with the localization of those proteins, map out the overall architecture of the cells [276].

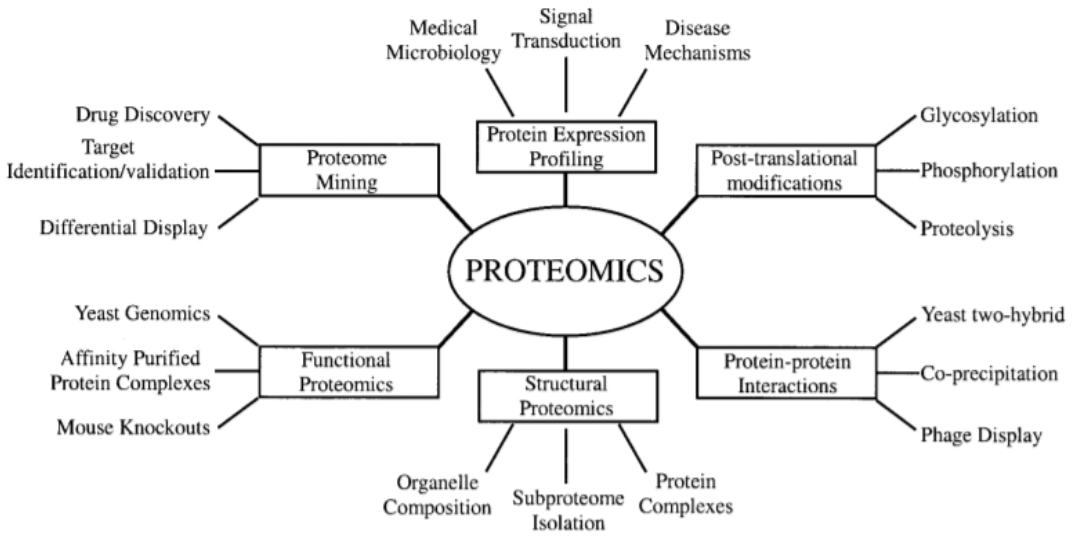

Figure I.12 – Proteomics fields, types and applications [218].

I.5.1.2 Functional proteomics

Transiting to the characterization of those protein activities, multiprotein complexes, and signaling pathways, the goals entering in the domain of the functional proteomics [277]. Addressing towards the elucidation of the biological function of proteins and the definition of cellular mechanisms, the functional proteomics approaches rely on affinity strategies for protein or protein complexes isolation [278], knockout technologies [279], and yeast genomics studies [280].

I.5.1.2.1 Proteome mining

One of the major applications of functional proteomics is the proteome mining, which is focused on the identification and validation of therapeutic targets that subsequently serve as a starting point for drug discovery [281].

I.5.1.2.2 Post-translational modifications (PTMs)

By definition, post-translational modifications (PTMs) of proteins are linked with the structural level, as they are denoted as any chemical changes that occur after a protein has been produced. However, these structural modifications are closely associated with protein activity and turnover being a frequent target of functional proteomics studies [282]. These dynamic modulations are known to be essential for signaling networks and playing a fundamental role in cellular physiology [283]. Catalyzed by specific enzymes, several types of PTMs are currently known as summarized in Figure I.13. They embrace the reversible addition of i) chemical groups, such as phosphate or acetate, ii) more complex molecules, such as carbohydrates or lipids, or ii) the covalent linkage of small proteins, like ubiquitin and ubiquitin-like proteins (UBLs). They also include irreversible reactions like i) the modification of side chain residues of specific amino acids, like deamidation in asparagine and glutamine residues or eliminylation in phosphorylated threonine residue, and ii) cleavage of the peptide bond between adjacent amino acid residues of a protein [283].

I.5.1.2.2.1 Phosphorylation

As one of the most common PTMs, several functional proteomics studies focus on protein phosphorylation [218]. Catalyzed by kinases, this modification type consists of the transfer of a phosphate group from ATP and its reversible attachment on serine, threonine, and tyrosine residues of the targeted protein [283]. The phosphorylation process frequently alters the function of the proteins, resulting in a broad spectrum of cellular processes and states regulated by this mechanism. For instance, phosphorylation is involved in several signaling pathways where many kinases and phosphatases act as enzymes and protein subtracts, and thus forming mutually dependent and hierarchically regulated signaling loops and cascades [284].

I.5.1.2.3 Protein-protein interaction

Conceptually, protein-protein interactions (PPIs) are a vast and complex network, which plays a crucial role in cellular functions. Likewise, in phosphorylation processes, which are themselves a specific type of PPIs, the regulation and execution of the most biological processes in all organisms are modulated by PPIs. In PPI proteomics studies focus on the identification of those interactions and the mechanisms involved in those pathways [285].

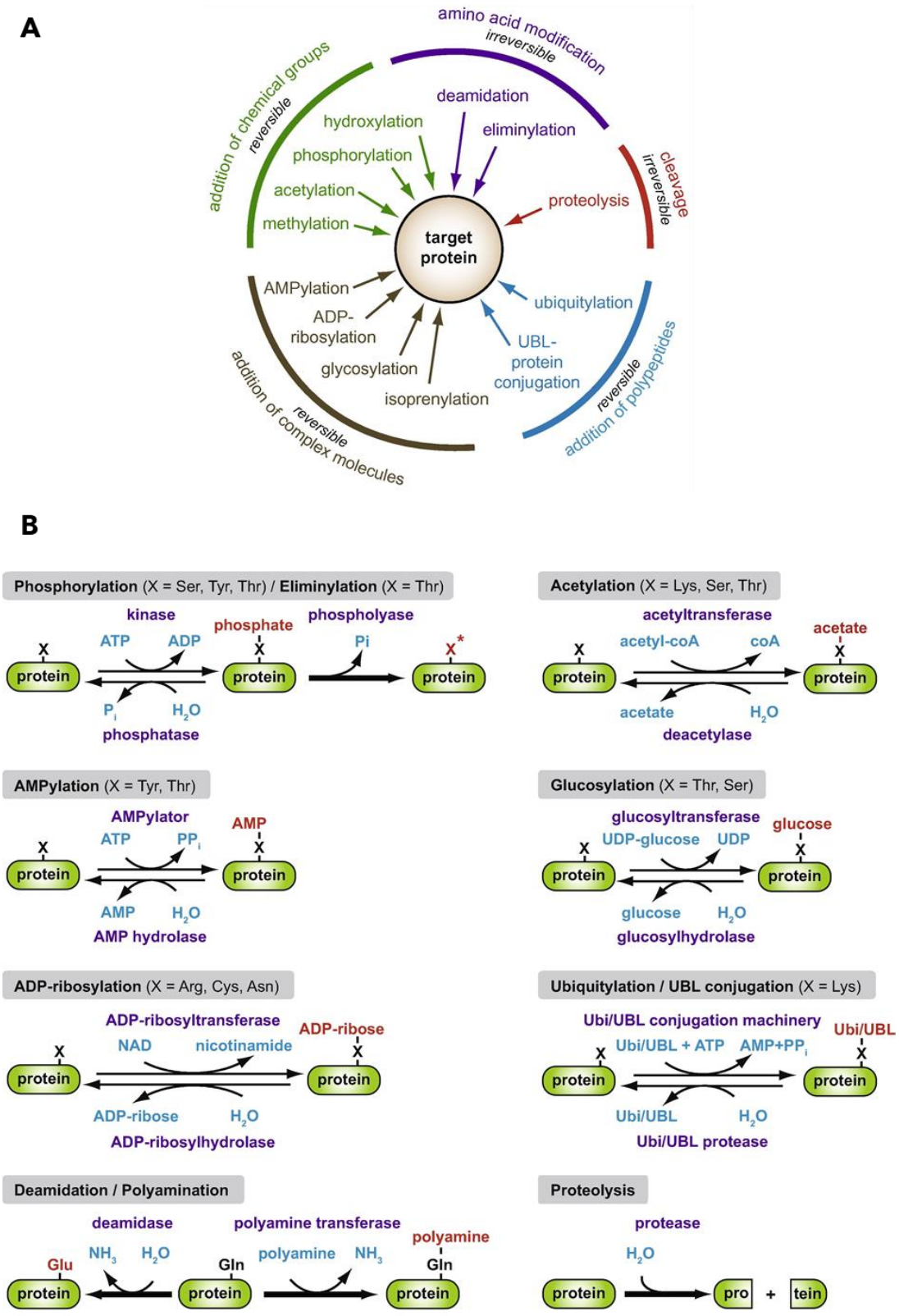

Figure I.13 – Post-translational modifications (PTMs) A) Diversity and B) reactions. For each PTM reaction, the nature of the amino acid most frequently modified is denotated as X and is indicated in the title. In purple are indicated the enzyme catalyzing the modification, in blu the cofactors potentially involved and in red the group added to the target protein, X, unphosphorylated amino acid; P_i, inorganic phosphate, PP_i, inorganic pyrophosphate; NAD, nicotinamide adenine dinucleotide; coA, coenzyme A; Ubi, ibiquitin* [283].

I.5.1.3 Protein expression profiling

Although the study of the structure of the proteome and the functional approach of its proteins have significantly contributed to unveil the complexity of the biology of a living organism, the quantitative study of protein expression continues to be the largest application of proteomics [218]. As the functionality and significance of the expressed proteins cannot be simply inferred by their presence, different MS-based quantitative proteomics are frequently used to address several biological questions including global protein abundance measurements, PTMs and also PPIs in one or more conditions [286]. Thus, the information retrieved from this approach is frequently used in the medical microbiology context, signal transduction and disease mechanisms [218]. The identification of disease-specific proteins is one of the major fundamental of clinical proteomics.

I.5.2 CLINICAL PROTEOMICS

The essence of the clinical proteomics is to address clinically relevant questions by the means of a proteomics analysis. Within a context of clinical studies, the goals of this approach are i) earlier and accurate diagnosis, ii) better evaluation of prognosis and/or prevention of disease, and iii) improvement of therapeutic strategies [287].

Currently, proteomics-based clinical studies can be categorized in main strategies and also correlated with the acquisition methods applied (Figure I.14). For instance, in a discovery branch either to develop a novel diagnostic marker, prognosis factors or therapeutically targets the goal is to identify as many proteins as possible. In these cases, DDA and DIA as more encompassing technologies are applied. In contrast, in targeted approaches the goal is to monitor a selected panel of proteins. Consequently, higher sensitive, reproducible and quantitative accurate techniques, such as SRM/MRM or PRM, are preferential chosen over the previous one [286,288].

I.5.2.1 Proteomic biomarker pipeline

In clinical discovery strategies, the main goal is to identify putative biomarkers to address the different clinical requests. According to the World Health Organization, a biomarker is any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease. Thus, by definition, a biomarker is an objective and quantifiable characteristic of biological processes [289]. Proteins as the most universally affected molecules by disease state are considered valuables sources of biological information and the advances of the MS technologies, has improved their power and utility in biomarker discovery. With the increased throughput and improved precision that has been accomplished with MS-approaches the hypothesis-driven has been superseded by hypothesis-generating in the discovery field, which have broadened the discovery experiments to an unprecedented and unbiased manner [290]. Therefore, a coherent pipeline for the development

of novel biomarkers has been proposed consisting in four phases, i) discovery; ii) qualification; iii) verification, and iv) validation (Figure I.15).

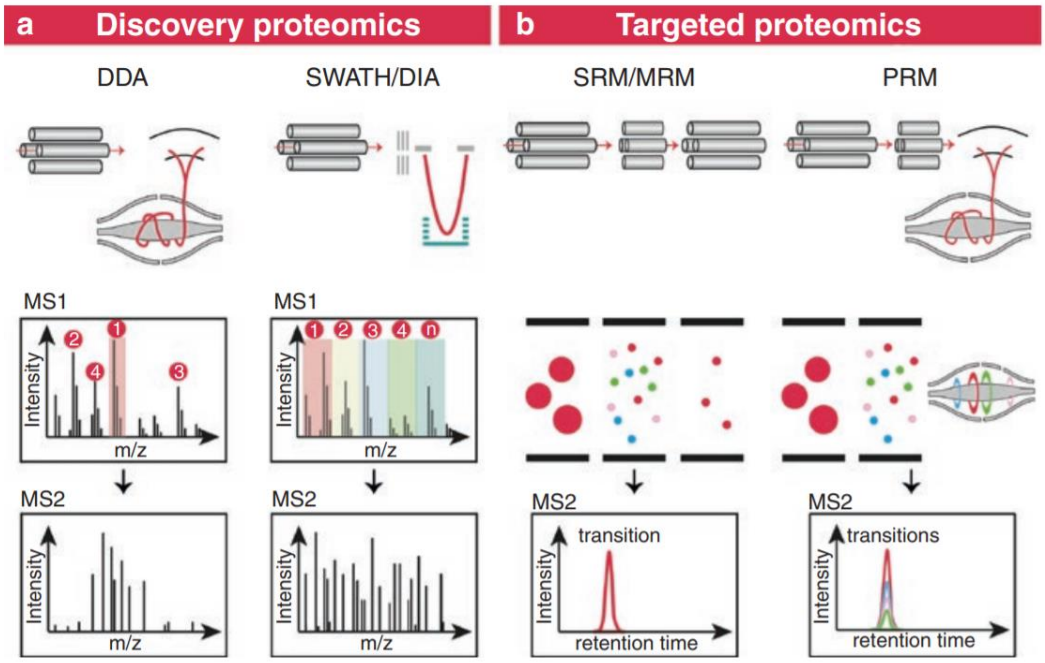

Figure I.14 – Proteomics-based clinical approaches. A) Discovery proteomics based on data-dependent acquisition (DDA) and data independent acquisition (DIA) methodologies, and B) Targeted proteomics approaches including the single/multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM) [286].

Following the pipeline, the candidates determined in the discovery phase are assessed for their clinical utility, though a targeted approach. Undergoing firstly for a qualification and verification phase, the list of candidates decreases ending in the last validation phase with a smaller panel of potential biomarker that are evaluated in hundreds to thousands of samples [291].

Another set of biomarkers types with very similar definition but clear distinct uses are the susceptibility/risk, prognostic and predictive biomarkers. While in the first one indicates the potential for developing a medical condition, prognostic biomarkers evaluate the differential disease outcomes, like progression or recurrences, whereas the predictive biomarkers discriminate those patients who will respond or not to a medical treatment or therapy [292]. In Table I.10 are present a representative list of biomarker discovery studies for clinical applications though MS-based proteomics approaches [293].

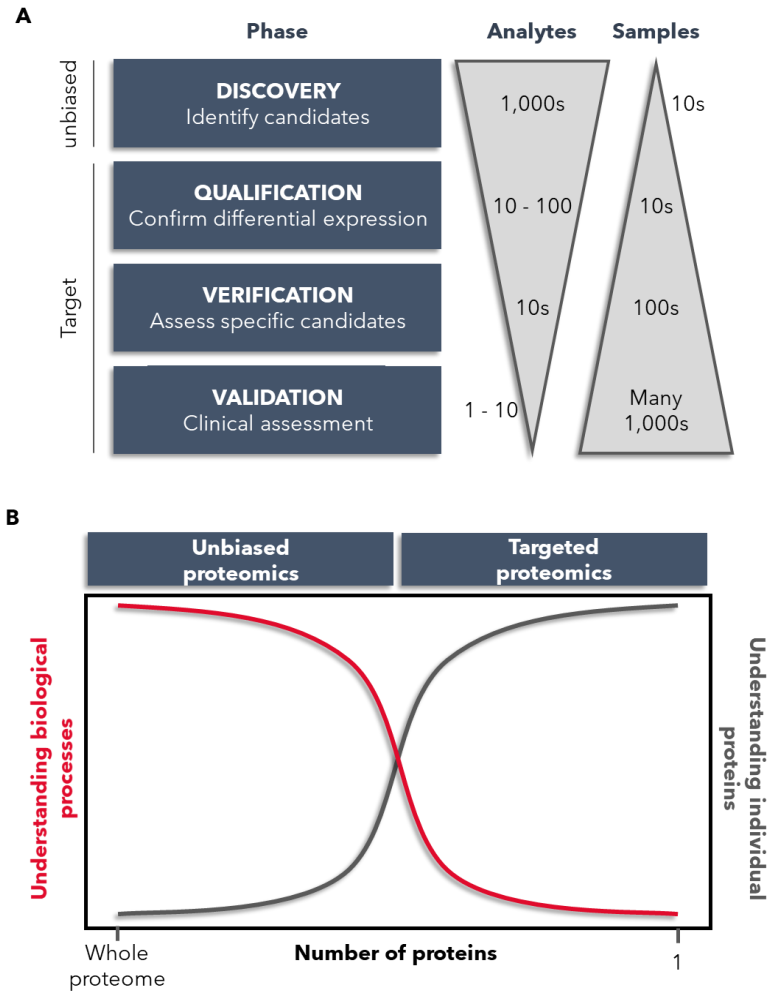

Figure I.15 – Proteomics-based biomarker discovery pipeline A) Biomarker development workflow divided into four main phases, i) discovery, ii) qualification, iii) verification and iv) validation. For each phase is represented at the right panel de number of different protein targets and samples, referred as "Analytes" and "Samples", respectively. Adapted from [291]*. B) The spectrum of protein analysis, ranging from unbiased to targeted proteomics. Adapted from* [294].

Table I.10 – Representative list of MS-based assays for clinical applications [293].

Analyte	MS technique	Clinical purpose
C-peptide	LC-ESI-QqQ	Diabetes millitus [295]
Thyroglobulin	LC-ESI-QqQ	Thyroid carcinoma [296,297]
Alpha-methylacyl-CoA racemase	LC-QqQTOF	Papillary renal cell carcinoma [298]
Vimentin	LC-QqQTOF	Clear cell renal cell carcinoma [298]
Lysosomal-associated membrane protein 1	LC-ESI-QOrbitrap	Chromophobe renal cell carcinoma [299]
Angiotensin-1 (renin activity)	LC-ESI-QqQ	Hypertension [300]
Insulin-like growth factor-1	LC-ESI-QTOF	Growth disorders [301,302]
ADAMTS13 activity	SELDI-TOF	Thrombotic thrombocytopenic purpura [303]
Lipoprotein-associated phospholipase A2	LC-ESI-QqQ	Cardiovascular disease risk [304]
Parathyroid hormone-related peptide	LC-ESI-QqQ	Osteoporosis, osteomalacia, and unexplained hypercalcemia [305]
Immunoglobulin light and heavy chain	MALDI-TOF	Plasma cell disorders [306]
IgG subclasses	LC-ESI-QqQ	IgG4-related disease [307,308]
Vitamin D binding globulin	LC-ESI-QqQ	Vitamin D deficiency [309]
Serum apolipoprotein panel	LC-ESI-QqQ	Cardiovascular disease [310]
Galectin-3-binding protein and scavenger receptor cysteine-rich type 1 protein M130	LC-ESI-QqQ	Malignant lung nodules [311]
High-density lipoprotein particle panel	LC-ESI-Orbitrap	Coronary artery disease [312]
β-Amyloid	LC-ESI-QqQ	Alzheimer disease [313]

REFERENCES

[1] Torday, J., Homeostasis as the Mechanism of Evolution. *Biology (Basel).*, **2015**, *4*, 573–590.

[2] Cooper, GM., The Development and Causes of Cancer. In *The Cell: A Molecular Approach*; Sinauer Associates, **2000**.

[3] Chen, L *et al*., Proteomics for Biomarker Identification and Clinical Application in Kidney Disease. *Adv. Clin. Chem.*, **2018**, *85*, 91–113.

[4] Bello-Reuss, E *et al*., Homeostatic and Excretory Functions of the Kidney. In *The Kidney and Body Fluids in Health and Disease*; Springer US: Boston, MA, **1983**; pp. 35–63.

[5] Ferlay, J *et al*., Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer*, **2019**, *144*, 1941–1953.

[6] Srigley, JR *et al*., The International Society of Urological Pathology (ISUP) Vancouver Classification of Renal Neoplasia. *Am. J. Surg. Pathol.*, **2013**, *37*, 1469–1489.

[7] Hes, O *et al*., The 2012 ISUP Vancouver and 2016 WHO classification of adult renal tumors: changes for common renal tumors. *Diagnostic Histopathol.*, **2016**, *22*, 41–46.

[8] Lopez-Beltran, A *et al*., 2004 WHO Classification of the Renal Tumors of the Adults. *Eur. Urol.*, **2006**, *49*, 798–805.

[9] Ljungberg, B *et al*., European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update. *Eur. Urol.*, **2019**, *75*, 799–810.

[10] https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf (accessed Jul 27, 2020).

[11] Lindgren, D *et al*., Tracing Renal Cell Carcinomas back to the Nephron. *Trends in Cancer*, **2018**, *4*, 472–484.

[12] van Oostenbrugge, TJ *et al*., Diagnostic Imaging for Solid Renal Tumors: A Pictorial Review. *Kidney Cancer*, **2018**, *2*, 79–93.

[13] Muglia, VF *et al*., Renal cell carcinoma: histological classification and correlation with imaging findings. *Radiol. Bras.*, **2015**, *48*, 166–174.

[14] Moch, H *et al*., The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *Eur. Urol.*, **2016**, *70*, 93–105.

[15] Hsieh, JJ *et al*., Renal cell carcinoma. *Nat. Rev. Dis. Prim.*, **2017**, *3*, 17009.

[16] Linehan, WM *et al*., Comprehensive Molecular Characterization of Papillary Renal-Cell

Carcinoma. *N. Engl. J. Med.*, **2016**, *374*, 135–145.

[17] Manley, BJ *et al*., Molecular profiling of renal cell carcinoma:: building a bridge toward clinical impact. *Curr. Opin. Urol.*, **2016**, *26*, 383–387.

[18] Delahunt, B *et al*., Papillary renal cell carcinoma: a clinicopathologic and immunohistochemical study of 105 tumors. *Mod. Pathol.*, **1997**, *10*, 537–544.

[19] Grande, JP *et al*., Renal Cell Cancer. In *Reference Module in Biomedical Sciences*; Elsevier, **2015**.

[20] Campbell, SC *et al*., Guideline for management of the clinical T1 renal mass. *J. Urol.*, **2009**, *182*, 1271–1279.

[21] Israel, GM *et al*., How I Do It: Evaluating Renal Masses. *Radiology*, **2005**, *236*, 441–450.

[22] Cuenod, CA *et al*., Tumor angiogenesis: pathophysiology and implications for contrast-enhanced MRI and CT assessment. *Abdom. Imaging*, **2006**, *31*, 188–193.

[23] Warren, KS *et al*., The Bosniak classification of renal cystic masses. *BJU Int.*, **2005**, *95*, 939–942.

[24] Schoots, IG *et al*., Bosniak Classification for Complex Renal Cysts Reevaluated: A Systematic Review. *J. Urol.*, **2017**, *198*, 12–21.

[25] Rini, BI *et al*., 60 - Kidney. In *AJCC Cancer Staging Manual*; Edge, S.; Byrd, D.R.; Compton, C.C.; Fritz, A.G.; Greene, F.L.; Trotti, A., Eds.; **2017**; pp. 739–749.

[26] Lebret, T *et al*., Percutaneous Core Biopsy for Renal Masses: Indications, Accuracy and Results. *J. Urol.*, **2007**, *178*, 1184–1188.

[27] Zhang, L *et al*., Renal Tumor Biopsy Technique. *Chin. Med. J. (Engl).*, **2016**, *129*, 1236–1240.

[28] Dagher, J *et al*., Clear cell renal cell carcinoma: validation of World Health Organization/International Society of Urological Pathology grading. *Histopathology*, **2017**, *71*, 918–925.

[29] Fuhrman, SA *et al*., Prognostic significance of morphologic parameters in renal cell carcinoma. *Am. J. Surg. Pathol.*, **1982**, *6*, 655–664.

[30] Delahunt, B., Advances and controversies in grading and staging of renal cell carcinoma. *Mod. Pathol.*, **2009**, *22*, S24–S36.

[31] Delahunt, B *et al*., The International Society of Urological Pathology (ISUP) Grading System for Renal Cell Carcinoma and Other Prognostic Parameters. *Am. J. Surg. Pathol.*, **2013**, *37*, 1490–1504.

[32] Samaratunga, H *et al*., The ISUP system of staging, grading and classification of renal cell neoplasia. *J. Kidney Cancer VHL*, **2014**, *1*, 26–39.

[33] Kwak, C *et al*., Sarcomatoid differentiation as a prognostic factor for immunotherapy in

metastatic renal cell carcinoma. *J. Surg. Oncol.*, **2007**, *95*, 317–323.

[34] Zhang, BY *et al*., A novel prognostic model for patients with sarcomatoid renal cell carcinoma. *BJU Int.*, **2015**, *115*, 405–411.

[35] Lam, JS *et al*., Clinicopathologic and molecular correlations of necrosis in the primary tumor of patients with renal cell carcinoma. *Cancer*, **2005**, *103*, 2517–2525.

[36] Renshaw, AA *et al*., Quantitative tumour necrosis is an independent predictor of overall survival in clear cell renal cell carcinoma. *Pathology*, **2015**, *47*, 34–37.

[37] Dekel, Y *et al*., Significance of angiogenesis and microvascular invasion in renal cell carcinoma. *Pathol. Oncol. Res.*, **2002**, *8*, 129–132.

[38] Huang, H *et al*., Microvascular invasion as a prognostic indicator in renal cell carcinoma: a systematic review and meta-analysis. *Int. J. Clin. Exp. Med.*, **2015**, *8*, 10779–10792.

[39] Farber, NJ *et al*., Renal cell carcinoma: the search for a reliable biomarker. *Transl. Cancer Res.*, **2017**, *6*, 620–632.

[40] Nakaigawa, N *et al*., FDG PET/CT as a prognostic biomarker in the era of molecular-targeting therapies: max SUVmax predicts survival of patients with advanced renal cell carcinoma. *BMC Cancer*, **2016**, *16*, 67.

[41] Gofrit, ON *et al*., Diagnostic Challenges of Kidney Cancer: A Systematic Review of the Role of Positron Emission Tomography-Computerized Tomography. *J. Urol.*, **2016**, *196*, 648–657.

[42] Divgi, CR *et al*., Positron Emission Tomography/Computed Tomography Identification of Clear Cell Renal Cell Carcinoma: Results From the REDECT Trial. *J. Clin. Oncol.*, **2013**, *31*, 187–194.

[43] Turkbey, B *et al*., PET/CT imaging of renal cell carcinoma with 18F-VM4-037: a phase II pilot study. *Abdom. Radiol.*, **2016**, *41*, 109–118.

[44] Stillebroer, AB *et al*., Carbonic Anhydrase IX in Renal Cell Carcinoma: Implications for Prognosis, Diagnosis, and Therapy. *Eur. Urol.*, **2010**, *58*, 75–83.

[45] Farber, NJ *et al*., Challenges in RCC Imaging: Renal Insufficiency, Post-Operative Surveillance, and the Role of Radiomics. *Kidney cancer J.*, **2015**, *13*, 84–90.

[46] Rowe, SP *et al*., Imaging of metastatic clear cell renal cell carcinoma with PSMA-targeted 18F-DCFPyL PET/CT. *Ann. Nucl. Med.*, **2015**, *29*, 877–882.

[47] Wu, Y *et al*., Magnetic Resonance Imaging as a Biomarker for Renal Cell Carcinoma. *Dis. Markers*, **2015**, *2015*, 1–9.

[48] Bratulic, S *et al*., The Translational Status of Cancer Liquid Biopsies. *Regen. Eng. Transl. Med.*, **2019**, 1–41.

[49] Chinello, C *et al*., The proteomic landscape of renal tumors. *Expert Rev. Proteomics*,

2016, *13*, 1103–1120.

[50] Pastore, AL *et al*., Serum and Urine Biomarkers for Human Renal Cell Carcinoma. *Dis. Markers*, **2015**, *2015*, 1–9.

[51] Morrissey, JJ *et al*., Urinary Biomarkers for the Early Diagnosis of Kidney Cancer. *Mayo Clin. Proc.*, **2010**, *85*, 413–421.

[52] Vaidya, VS *et al*., Biomarkers of Acute Kidney Injury. *Annu. Rev. Pharmacol. Toxicol.*, **2008**, *48*, 463–493.

[53] de Martino, M *et al*., Serum 20S proteasome is elevated in patients with renal cell carcinoma and associated with poor prognosis. *Br. J. Cancer*, **2012**, *106*, 904–908.

[54] Ahmad, S *et al*., Role of MMP-2, MMP-9 and VEGF as serum biomarker in early prognosis of renal cell carcinoma. *African J. Urol.*, **2018**, *24*, 255–263.

[55] Baldin, A V. *et al*., Autoantibody against arrestin-1 as a potential biomarker of renal cell carcinoma. *Biochimie*, **2019**, *157*, 26–37.

[56] Chen, D *et al*., Detection of survivin expression in bladder cancer and renal cell carcinoma using specific monoclonal antibodies. *Oncol. Rep.*, **2018**, *39*, 2817–2828.

[57] Masuda, A *et al*., Clinical Significance of Serum Soluble T Cell Regulatory Molecules in Clear Cell Renal Cell Carcinoma. *Biomed Res. Int.*, **2014**, *2014*, 1–6.

[58] Zhang, L *et al*., ITRAQ-based quantitative proteomic analysis reveals potential early diagnostic markers of clear-cell Renal cell carcinoma. *Biosci. Trends*, **2016**, *10*, 210–219.

[59] Zheng, Z *et al*., CXCL13/CXCR5 Axis Predicts Poor Prognosis and Promotes Progression Through PI3K/AKT/mTOR Pathway in Clear Cell Renal Cell Carcinoma. *Front. Oncol.*, **2019**, *8*, 682.

[60] Li, C *et al*., Reduced cytosolic carboxypeptidase 6 (CCP6) level leads to accumulation of serum polyglutamylated DNAJC7 protein: A potential biomarker for renal cell carcinoma early detection. *Oncotarget*, **2016**, *7*, 22385–22396.

[61] Kim, KH *et al*., Clinical validation of serum endocan (ESM-1) as a potential biomarker in patients with renal cell carcinoma. *Oncotarget*, **2018**, *9*, 662–667.

[62] Zhang, Y *et al*., ITRAQ-Based Quantitative Proteomic Analysis Identified HSC71 as a Novel Serum Biomarker for Renal Cell Carcinoma. *Biomed Res. Int.*, **2015**, *2015*, 1–6.

[63] White, NMA *et al*., Quantitative proteomic analysis reveals potential diagnostic markers and pathways involved in pathogenesis of renal cell carcinoma. *Oncotarget*, **2014**, *5*, 506–518.

[64] Kargı, A *et al*., Serum levels of HMGB1 have a diagnostic role in metastatic renal cell cancer. *Cancer Biomarkers*, **2016**, *17*, 17–20.

[65] Tanaka, T *et al*., Autoantibody against hypoxia-inducible factor prolyl hydroxylase-3 is a potential serological marker for renal cell carcinoma. *J. Cancer Res. Clin. Oncol.*, **2011**, *137*, 789–794.

[66] Kim, KH *et al*., Prolyl hydroxylase-3 is a novel renal cell carcinoma biomarker. *Investig. Clin. Urol.*, **2019**, *60*, 425–431.

[67] Yildiz, I *et al*., Serum M65 as a Biomarker for Metastatic Renal Cell Carcinoma. *Clin. Genitourin. Cancer*, **2013**, *11*, 290–296.

[68] Papworth, K *et al*., Osteopontin but not parathyroid hormone-related protein predicts prognosis in human renal cell carcinoma. *Acta Oncol. (Madr).*, **2013**, *52*, 159–165.

[69] Rajandram, R *et al*., Tumour necrosis factor receptor-associated factor-1 (TRAF-1) expression is increased in renal cell carcinoma patient serum but decreased in cancer tissue compared with normal: potential biomarker significance. *Pathology*, **2014**, *46*, 518–522.

[70] Lucarini, L *et al*., Plasmatic carbonic anhydrase IX as a diagnostic marker for clear cell renal cell carcinoma. *J. Enzyme Inhib. Med. Chem.*, **2018**, *33*, 234–240.

[71] Yokomizo, A *et al*., Use of quantitative shotgun proteomics to identify fibronectin 1 as a potential plasma biomarker for clear cell carcinoma of the kidney. *Cancer Biomarkers*, **2012**, *10*, 175–183.

[72] Kushlinskii, NE *et al*., Kidney Injury Molecule-1 (KIM-1) in Blood Plasma of Patients with Clear-Cell Carcinoma. *Bull. Exp. Biol. Med.*, **2019**, *167*, 388–392.

[73] Scelo, G *et al*., KIM-1 as a Blood-Based Marker for Early Detection of Kidney Cancer: A Prospective Nested Case–Control Study. *Clin. Cancer Res.*, **2018**, *24*, 5594–5601.

[74] Morrissey, JJ *et al*., Urine Aquaporin 1 and Perilipin 2 Differentiate Renal Carcinomas From Other Imaged Renal Masses and Bladder and Prostate Cancer. *Mayo Clin. Proc.*, **2015**, *90*, 35–42.

[75] Di Meo, A *et al*., Searching for prognostic biomarkers for small renal masses in the urinary proteome. *Int. J. Cancer*, **2020**, *146*, 2315–2325.

[76] Zhang, KJ *et al*., Diagnostic role of kidney injury molecule-1 in renal cell carcinoma. *Int. Urol. Nephrol.*, **2019**, *51*, 1893–1902.

[77] Morrissey, JJ *et al*., Sensitivity and Specificity of Urinary Neutrophil Gelatinase-Associated Lipocalin and Kidney Injury Molecule-1 for the Diagnosis of Renal Cell Carcinoma. *Am. J. Nephrol.*, **2011**, *34*, 391–398.

[78] Kaya, K *et al*., Urinary nuclear matrix protein 22 for diagnosis of renal cell carcinoma. *Scand. J. Urol. Nephrol.*, **2005**, *39*, 25–29.

[79] Ozer, G *et al*., Value of urinary NMP-22 in patients with renal cell carcinoma. *Urology*, **2002**, *60*, 593–597.

[80] MacLennan, GT *et al*., 2 - Neoplasms of the Kidney. In *Urologic Surgical Pathology*; Elsevier, **2020**; pp. 83-163.e23.

[81] Truong, LD *et al*., Immunohistochemical diagnosis of renal neoplasms. *Arch. Pathol. Lab. Med.*, **2011**, *135*, 92–109.

[82] Kim, M *et al*., Comprehensive Immunoprofiles of Renal Cell Carcinoma Subtypes. *Cancers (Basel).*, **2020**, *12*, 602.

[83] Cairns, P., Renal cell carcinoma. *Cancer Biomarkers*, **2011**, *9*, 461–473.

[84] Giribaldi, G *et al*., Proteomic identification of Reticulocalbin 1 as potential tumor marker in renal cell carcinoma. *J. Proteomics*, **2013**, *91*, 385–392.

[85] Masui, O *et al*., Quantitative Proteomic Analysis in Metastatic Renal Cell Carcinoma Reveals a Unique Set of Proteins with Potential Prognostic Significance. *Mol. Cell. Proteomics*, **2013**, *12*, 132–144.

[86] Zhao, Z *et al*., Label-free quantitative proteomic analysis reveals potential biomarkers and pathways in renal cell carcinoma. *Tumor Biol.*, **2015**, *36*, 939–951.

[87] Sun, CY *et al*., Proteomic analysis of clear cell renal cell carcinoma. Identification of potential tumor markers. *Saudi Med. J.*, **2010**, *31*, 525–532.

[88] Raimondo, F *et al*., Proteomic analysis in clear cell renal cell carcinoma: identification of differentially expressed protein by 2-D DIGE. *Mol. Biosyst.*, **2012**, *8*, 1040.

[89] Valera, VA *et al*., Protein Expression Profiling in the Spectrum of Renal Cell Carcinomas. *J. Cancer*, **2010**, *1*, 184–196.

[90] Lichtenfels, R *et al*., Systematic Comparative Protein Expression Profiling of Clear Cell Renal Cell Carcinoma. *Mol. Cell. Proteomics*, **2009**, *8*, 2827–2842.

[91] Wang, Y *et al*., C1QBP Negatively Regulates the Activation of Oncoprotein YBX1 in the Renal Cell Carcinoma As Revealed by Interactomics Analysis. *J. Proteome Res.*, **2015**, *14*, 804–813.

[92] Atrih, A *et al*., Quantitative proteomics in resected renal cancer tissue for biomarker discovery and profiling. *Br. J. Cancer*, **2014**, *110*, 1622–1633.

[93] Kim, DS *et al*., Panel of Candidate Biomarkers for Renal Cell Carcinoma. *J. Proteome Res.*, **2010**, *9*, 3710–3719.

[94] Na, CH *et al*., Identification of Protein Markers Specific for Papillary Renal Cell Carcinoma Using Imaging Mass Spectrometry. *Mol. Cells*, **2015**, *38*, 624–629.

[95] Kowalewski, A *et al*., Caspase 3 as a Novel Marker to Distinguish Chromophobe Renal Cell Carcinoma from Oncocytoma. *Pathol. Oncol. Res.*, **2019**, *25*, 1519–1524.

[96] Conner, JR *et al*., HNF1β and S100A1 are useful biomarkers for distinguishing renal oncocytoma and chromophobe renal cell carcinoma in FNA and core needle biopsies.

Cancer Cytopathol., **2015**, *123*, 298–305.

[97] Ng, KL *et al*., Leptin and its receptor: can they help to differentiate chromophobe renal cell carcinoma from renal oncocytoma? *Pathology*, **2018**, *50*, 504–510.

[98] Molnar, A *et al*., FOXI1 Immunohistochemistry Differentiates Benign Renal Oncocytoma from Malignant Chromophobe Renal Cell Carcinoma. *Anticancer Res.*, **2019**, *39*, 2785–2790.

[99] Clancy, S *et al*., Translation: DNA to mRNA to Protein. *Nat. Educ.*, **2008**, *1*, 101.

[100] Venter, JC *et al*., The Sequence of the Human Genome. *Science (80-.).*, **2001**, *291*, 1304–1351.

[101] Virág, D *et al*., Current Trends in the Analysis of Post-translational Modifications. *Chromatographia*, **2020**, *83*, 1–10.

[102] Resing, KA *et al*., Proteomics strategies for protein identification. *FEBS Lett.*, **2005**, *579*, 885–889.

[103] Han, X *et al*., Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.*, **2008**, *12*, 483–490.

[104] Catherman, AD *et al*., Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.*, **2014**, *445*, 683–693.

[105] Liumbruno, G *et al*., Blood-related proteomics. *J. Proteomics*, **2010**, *73*, 483–507.

[106] Millioni, R *et al*., High Abundance Proteins Depletion vs Low Abundance Proteins Enrichment: Comparison of Methods to Reduce the Plasma Proteome Complexity. *PLoS One*, **2011**, *6*, e19603.

[107] Lygirou, V *et al*., Biological Sample Collection for Clinical Proteomics: Existing SOPs. In *Methods in Molecular Biology*; Humana Press Inc., **2015**; Vol. 1243, pp. 3–27.

[108] Hood, BL *et al*., Proteomic Analysis of Formalin-fixed Prostate Cancer Tissue. *Mol. Cell. Proteomics*, **2005**, *4*, 1741–1753.

[109] Magdeldin, S *et al*., Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics*, **2012**, *12*, 1045–1058.

[110] Loken, SD *et al*., A novel method for freezing and storing research tissue bank specimens. *Hum. Pathol.*, **2005**, *36*, 977–980.

[111] Zhang, W *et al*., Comprehensive proteome analysis of fresh frozen and optimal cutting temperature (OCT) embedded primary non-small cell lung carcinoma by LC-MS/MS. *Methods*, **2015**, *81*, 50–55.

[112] Weston, LA *et al*., Comparative LC-MS/MS analysis of optimal cutting temperature (OCT) compound removal for the study of mammalian proteomes. *Analyst*, **2013**, *138*, 6380–6384.

[113] Vrana, M *et al*., An optimized method for protein extraction from OCT-embedded human kidney tissue for protein quantification by LC-MS/MS proteomics. *Drug Metab. Dispos.*, **2016**, *44*, 1692–1696.

[114] Zhao, X *et al*., Quantitative proteomic analysis of optimal cutting temperature (OCT) embedded core-needle biopsy of lung cancer. *J. Am. Soc. Mass Spectrom.*, **2017**, *28*, 2078–2089.

[115] Shah, P *et al*., Tissue proteomics using chemical immobilization and mass spectrometry. *Anal. Biochem.*, **2015**, *468*, 27–33.

[116] Tian, Y *et al*., Quantitative glycoproteomic analysis of optimal cutting temperature-embedded frozen tissues identifying glycoproteins associated with aggressive prostate cancer. *Anal. Chem.*, **2011**, *83*, 7013–7019.

[117] Chaurand, P *et al*., Monitoring mouse prostate development by profiling and imaging mass spectrometry. *Mol. Cell. Proteomics*, **2008**, *7*, 411–423.

[118] Holfeld, A *et al*., Parallel proteomic workflow for mass spectrometric analysis of tissue samples preserved by different methods. *Anal. Chem.*, **2018**, *90*, 5841–5849.

[119] Rogers, JC *et al*., Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, **2016**; Vol. 919, pp. 43–62.

[120] Shehadul Islam, M *et al*., A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*, **2017**, *8*, 83.

[121] Capelo, JL *et al*., Overview on modern approaches to speed up protein identification workflows relying on enzymatic cleavage and mass spectrometry-based techniques. *Anal. Chim. Acta*, **2009**, *650*, 151–159.

[122] Stone, J., Sample preparation techniques for mass spectrometry in the clinical laboratory. In *Mass Spectrometry for the Clinical Laboratory*; Elsevier, **2017**; pp. 37–62.

[123] Koontz, L., TCA Precipitation. In *Methods in Enzymology*; Academic Press Inc., **2014**; Vol. 541, pp. 3–10.

[124] Ly, L *et al*., Protein and peptide fractionation, enrichment and depletion: Tools for the complex proteome. *Proteomics*, **2011**, *11*, 513–534.

[125] Polaskova, V *et al*., High-abundance protein depletion: Comparison of methods for human plasma biomarker discovery. *Electrophoresis*, **2010**, *31*, 471–482.

[126] Alpert, AJ., Protein Fractionation and Enrichment Prior to Proteomics Sample Preparation. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, **2016**; Vol. 919, pp. 23–41.

[127] Wiederin, J *et al*., Immunoaffinity Depletion of Highly Abundant Proteins for Proteomic Sample Preparation. In *Proteomic Profiling and Analytical Chemistry*; Elsevier, **2016**; pp.

101–114.

[128] Björhall, K *et al*., Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics*, **2005**, *5*, 307–317.

[129] Shen, Y *et al*., Characterization of the human blood plasma proteome. *Proteomics*, **2005**, *5*, 4034–4045.

[130] Boschetti, E *et al*., Hexapeptide combinatorial ligand libraries: the march for the detection of the low-abundance proteome continues. *Biotechniques*, **2008**, *44*, 663–665.

[131] Fernández, C *et al*., A comparison of depletion versus equalization for reducing high-abundance proteins in human serum. *Electrophoresis*, **2011**, *32*, 2966–2974.

[132] Kay, R *et al*., Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. *Rapid Commun. Mass Spectrom.*, **2008**, *22*, 3255–3260.

[133] Warder, SE *et al*., Reducing agent-mediated precipitation of high-abundance plasma proteins. *Anal. Biochem.*, **2009**, *387*, 184–193.

[134] Tholey, A *et al*., Top-down proteomics for the analysis of proteolytic events - Methods, applications and perspectives. *Biochim. Biophys. Acta - Mol. Cell Res.*, **2017**, *1864*, 2191–2199.

[135] Suttapitugsakul, S *et al*., Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. *Mol. Biosyst.*, **2017**, *13*, 2574–2582.

[136] Switzar, L *et al*., Protein Digestion: An Overview of the Available Techniques and Recent Developments. *J. Proteome Res.*, **2013**, *12*, 1067–1077.

[137] Olsen, J V. *et al*., Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol. Cell. Proteomics*, **2004**, *3*, 608–614.

[138] Gundry, RL *et al*., Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. In *Current Protocols in Molecular Biology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, **2009**; Vol. CHAPTER, p. Unit10.25.

[139] Ma, J *et al*., Recent advances in immobilized enzymatic reactors and their applications in proteome analysis. *Anal. Chim. Acta*, **2009**, *632*, 1–8.

[140] Martins, G *et al*., Label-free protein quantification after ultrafast digestion of complex proteomes using ultrasonic energy and immobilized-trypsin magnetic nanoparticles. *Talanta*, **2019**, *196*, 262–270.

[141] Zhang, L *et al*., Recyclable trypsin immobilized magnetic nanoparticles based on hydrophilic polyethylenimine modification and their proteolytic characteristics. *Anal. Methods*, **2018**, *10*, 459–466.

[142] Atacan, K *et al*., Efficient protein digestion using immobilized trypsin onto tannin modified Fe 3 O 4 magnetic nanoparticles. *Colloids Surfaces B Biointerfaces*, **2017**, *156*, 9–18.

[143] Sun, J *et al*., Stability and activity of immobilized trypsin on carboxymethyl chitosan-functionalized magnetic nanoparticles cross-linked with carbodiimide and glutaraldehyde. *J. Chromatogr. B*, **2017**, *1054*, 57–63.

[144] Vale, G *et al*., An assessment of the ultrasonic probe-based enhancement of protein cleavage with immobilized trypsin. *Proteomics*, **2011**, *11*, 3866–3876.

[145] Dong, J *et al*., Limited proteolysis in porous membrane reactors containing immobilized trypsin. *Analyst*, **2017**, *142*, 2578–2586.

[146] Xu, F *et al*., Facile Trypsin Immobilization in Polymeric Membranes for Rapid, Efficient Protein Digestion. *Anal. Chem.*, **2010**, *82*, 10045–10051.

[147] Sproß, J *et al*., A Capillary Monolithic Trypsin Reactor for Efficient Protein Digestion in Online and Offline Coupling to ESI and MALDI Mass Spectrometry. *Anal. Chem.*, **2010**, *82*, 1434–1443.

[148] Naldi, M *et al*., Towards automation in protein digestion: Development of a monolithic trypsin immobilized reactor for highly efficient on-line digestion and analysis. *Talanta*, **2017**, *167*, 143–157.

[149] Jiang, S *et al*., A one-step preparation method of monolithic enzyme reactor for highly efficient sample preparation coupled to mass spectrometry-based proteomics studies. *J. Chromatogr. A*, **2015**, *1412*, 75–81.

[150] Bataille, J *et al*., On-a-chip tryptic digestion of transthyretin: a step toward an integrated microfluidic system for the follow-up of familial transthyretin amyloidosis. *Analyst*, **2018**, *143*, 1077–1086.

[151] Kecskemeti, A *et al*., Preparation and characterization of a packed bead immobilized trypsin reactor integrated into a PDMS microfluidic chip for rapid protein digestion. *Talanta*, **2017**, *166*, 275–283.

[152] Sun, L *et al*., Integrated Capillary Zone Electrophoresis–Electrospray Ionization Tandem Mass Spectrometry System with an Immobilized Trypsin Microreactor for Online Digestion and Analysis of Picogram Amounts of RAW 264.7 Cell Lysate. *Anal. Chem.*, **2013**, *85*, 4187–4194.

[153] Glatter, T *et al*., Large-Scale Quantitative Assessment of Different In-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/Trypsin Proteolysis over Trypsin Digestion. *J. Proteome Res.*, **2012**, *11*, 5145–5156.

[154] Wiśniewski, JR *et al*., Universal sample preparation method for proteome analysis. *Nat. Methods*, **2009**, *6*, 359–362.

[155] Brunelle, JL *et al*., Coomassie Blue Staining. In *Methods in Enzymology*; Academic

Press Inc., **2014**; Vol. 541, pp. 161–167.

[156] Weiss, W *et al*., Protein Detection and Quantitation Technologies for Gel-Based Proteome Analysis. In *Methods in molecular biology (Clifton, N.J.)*; Methods Mol Biol, **2009**; Vol. 564, pp. 59–82.

[157] Yates, JR., Mass Spectrometry and the Age of the Proteome. *J. Mass Spectrom.*, **1998**, *33*, 1–19.

[158] Manza, LL *et al*., Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics*, **2005**, *5*, 1742–1745.

[159] Tubaon, RM *et al*., Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016. *Proteomics*, **2017**, *17*, 1700011.

[160] Li, X-S *et al*., Recent advances in phosphopeptide enrichment: Strategies and techniques. *TrAC Trends Anal. Chem.*, **2016**, *78*, 70–83.

[161] McLachlin, DT *et al*., Improved β-Elimination-Based Affinity Purification Strategy for Enrichment of Phosphopeptides. *Anal. Chem.*, **2003**, *75*, 6826–6836.

[162] Grønborg, M *et al*., A Mass Spectrometry-based Proteomic Approach for Identification of Serine/Threonine-phosphorylated Proteins by Enrichment with Phospho-specific Antibodies. *Mol. Cell. Proteomics*, **2002**, *1*, 517–527.

[163] Araújo, JE *et al*., A journey through PROTEOSONICS. *Talanta*, **2014**, *121*, 71–80.

[164] ANSI/ASA S1.1-2013 - Acoustical Terminology https://webstore.ansi.org/standards/asa/ansiasas12013 (accessed Jul 17, 2020).

[165] Didenko, YT *et al*., Hot Spot Conditions during Cavitation in Water. *J. Am. Chem. Soc.*, **1999**, *121*, 5817–5818.

[166] Luque de Castro, MD *et al*., The role of ultrasound in analytical derivatizations. *J. Chromatogr. B*, **2011**, *879*, 1189–1195.

[167] Araújo, JE *et al*., A comprehensive factorial design study of variables affecting protein extraction from formalin-fixed kidney tissue samples. *Talanta*, **2014**, *119*, 90–97.

[168] Manadas, BJ *et al*., Sample sonication after trichloroacetic acid precipitation increases protein recovery from cultured hippocampal neurons, and improves resolution and reproducibility in two-dimensional gel electrophoresis. *Electrophoresis*, **2006**, *27*, 1825–1831.

[169] Cordeiro, FM *et al*., Simplifying sample handling for protein identification by peptide mass fingerprint using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*, **2007**, *21*, 3269–3278.

[170] Santos, H *et al*., An improved clean sonoreactor-based method for protein identification by mass spectrometry-based techniques. *Talanta*, **2008**, *77*, 870–875.

[171] Galesio, M *et al*., Unravelling the role of ultrasonic energy in the enhancement of enzymatic kinetics. *J. Mol. Catal. B Enzym.*, **2012**, *74*, 9–15.

[172] Carvalho, LB *et al*., Ultrasonic-Based Filter Aided Sample Preparation as the General Method to Sample Preparation in Proteomics. *Anal. Chem.*, **2020**, *92*, 9164–9171.

[173] Galesio, M *et al*., Influence of the Protein Staining in the Fast Ultrasonic Sample Treatment for Protein Identification through Peptide Mass Fingerprint and Matrix-Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry. *J. Proteome Res.*, **2008**, *7*, 2097–2106.

[174] Jorge, S *et al*., Unparalleled sample treatment throughput for proteomics workflows relying on ultrasonic energy. *Talanta*, **2018**, *178*, 1067–1076.

[175] Carreira, R *et al*., Ultrasonic energy as a new tool for fast isotopic 18O labeling of proteins for mass spectrometry-based techniques: Preliminary results. *Talanta*, **2008**, *76*, 400–406.

[176] Carreira, RJ *et al*., Can ultrasonic energy efficiently speed 18 O-labeling of proteins? *Proteomics*, **2009**, *9*, 4974–4977.

[177] Jorge, S *et al*., Development of a Robust Ultrasonic-Based Sample Treatment To Unravel the Proteome of OCT-Embedded Solid Tumor Biopsies. *J. Proteome Res.*, **2019**, *18*, 2979–2986.

[178] Santos, H *et al*., Trends in ultrasonic-based equipment for analytical sample treatment. *Talanta*, **2007**, *73*, 795–802.

[179] Vas, G *et al*., Biomedical sampling. In *Medical Applications of Mass Spectrometry*; Elsevier, **2008**; pp. 37–59.

[180] Kota, U *et al*., Improving Proteome Coverage by Reducing Sample Complexity via Chromatography. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, **2016**; Vol. 919, pp. 83–143.

[181] Staub, A *et al*., Intact protein analysis in the biopharmaceutical field. *J. Pharm. Biomed. Anal.*, **2011**, *55*, 810–822.

[182] Gross, JH., Introduction. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 1–28.

[183] The Nobel Prize in Physics 1906 https://www.nobelprize.org/prizes/physics/1906/summary/ (accessed Sep 15, 2020).

[184] Aston, FW., LXXIV. A positive ray spectrograph. *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, **1919**, *38*, 707–714.

[185] The Nobel Prize in Chemistry 1922 https://www.nobelprize.org/prizes/chemistry/1922/summary/ (accessed Sep 15, 2020).

[186] The Nobel Prize in Physics 1989 https://www.nobelprize.org/prizes/physics/1989/summary/ (accessed Sep 15, 2020).

[187] The Nobel Prize in Chemistry 2002 https://www.nobelprize.org/prizes/chemistry/2002/summary/ (accessed Sep 15, 2020).

[188] Siuzdak, G., An introduction to mass spectrometry ionization: An excerpt from The Expanding Role of Mass Spectrometry in Biotechnology. *J. Assoc. Lab. Autom.*, **2004**, *9*, 50–63.

[189] Gross, JH., Hyphenated Methods. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 831–887.

[190] Nguyen, DT-T *et al*., Fast analysis in liquid chromatography using small particle size and high pressure. *J. Sep. Sci.*, **2006**, *29*, 1836–1848.

[191] Gama, MR *et al*., Nano-Liquid Chromatography in Pharmaceutical and Biomedical Research. *J. Chromatogr. Sci.*, **2013**, *51*, 694–703.

[192] de la Guardia, M *et al*., Downsizing the Methods. In *Comprehensive Analytical Chemistry*; **2011**; Vol. 57, pp. 157–184.

[193] Šesták, J *et al*., Instrument platforms for nano liquid chromatography. *J. Chromatogr. A*, **2015**, *1421*, 2–17.

[194] Rockwood, AL *et al*., Mass Spectrometry. In *Principles and Applications of Clinical Mass Spectrometry*; Elsevier, **2018**; pp. 33–65.

[195] Gross, JH., Practical Aspects of Electron Ionization. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 293–324.

[196] Gross, JH., Chemical Ionization. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 439–496.

[197] Gross, JH., Electrospray Ionization. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 721–778.

[198] Banerjee, S *et al*., Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.*, **2012**, *2012*, 1–40.

[199] Wilm, M., Principles of Electrospray Ionization. *Mol. Cell. Proteomics*, **2011**, *10*, M111.009407.

[200] Bianchi, F *et al*., MS-Based Analytical Techniques: Advances in Spray-Based Methods and EI-LC-MS Applications. *J. Anal. Methods Chem.*, **2018**, *2018*, 1–24.

[201] Gross, JH., Matrix-Assisted Laser Desorption/Ionization. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 651–720.

[202] Gross, JH., Fast Atom Bombardment. In *Mass Spectrometry*; Springer International

Publishing: Cham, **2017**; pp. 613–649.

[203] Gross, JH., Field Ionization and Field Desorption. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 497–537.

[204] Gross, JH., Ambient Desorption/Ionization. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 779–829.

[205] Rubakhin, SS *et al*., A Mass Spectrometry Primer for Mass Spectrometry Imaging. In *Methods in Molecular Biology*; NIH Public Access, **2010**; Vol. 656, pp. 21–49.

[206] Haag, AM., Mass Analyzers and Mass Spectrometers. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, **2016**; Vol. 919, pp. 157–169.

[207] Schermann, J-P., Experimental Methods. In *Spectroscopy and Modeling of Biomolecular Building Blocks*; Elsevier, **2008**; pp. 129–207.

[208] March, RE., Ion Trap Mass Spectrometers. In *Encyclopedia of Spectroscopy and Spectrometry*; Elsevier, **2010**; pp. 1165–1173.

[209] Douglas, DJ *et al*., Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.*, **2005**, *24*, 1–29.

[210] Savaryn, JP *et al*., A researcher's guide to mass spectrometry-based proteomics. *Proteomics*, **2016**, *16*, 2435–2443.

[211] Zubarev, RA *et al*., Orbitrap Mass Spectrometry. *Anal. Chem.*, **2013**, *85*, 5288–5296.

[212] Gross, JH., Instrumentation. In *Mass Spectrometry*; Springer International Publishing: Cham, **2017**; pp. 151–292.

[213] O'Connor, PB *et al*., MALDI Mass Spectrometry Instrumentation. In *MALDI MS*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, **2007**; pp. 29–82.

[214] Yost, RA *et al*., Tandem mass spectrometry: Quadrupole and hybrid instruments. In *Methods in Enzymology*; Methods Enzymol, **1990**; Vol. 193, pp. 154–200.

[215] Chernushevich, I V. *et al*., An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.*, **2001**, *36*, 849–865.

[216] Burinsky, DJ., Mass spectrometry. In *Comprehensive Analytical Chemistry*; Elsevier, **2006**; Vol. 47, pp. 319–396.

[217] Brown, KL *et al*., Faraday-Cup Monitors for High-Energy Electron Beams. *Rev. Sci. Instrum.*, **1956**, *27*, 696–702.

[218] Graves, PR *et al*., Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.*, **2002**, *66*, 39–63.

[219] Hjernø, K *et al*., Interpretation of Tandem Mass Spectrometry (MSMS) Spectra for Peptide Analysis. In *Methods in Molecular Biology*; Humana Press Inc., **2015**; Vol. 1348,

pp. 83–102.

[220] Sadygov, RG *et al*., Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. *Nat. Methods*, **2004**, *1*, 195–202.

[221] Domon, B *et al*., Challenges and Opportunities in Proteomics Data Analysis. *Mol. Cell. Proteomics*, **2006**, *5*, 1921–1926.

[222] Wong, CCL *et al*., Comparison of different signal thresholds on data dependent sampling in orbitrap and LTQ mass spectrometry for the identification of peptides and proteins in complex mixtures. *J. Am. Soc. Mass Spectrom.*, **2009**, *20*, 1405–1414.

[223] Xu, G *et al*., Deconvolution in mass spectrometry based proteomics. *Rapid Commun. Mass Spectrom.*, **2018**, *32*, 763–774.

[224] Liu, X *et al*., Deconvolution and Database Search of Complex Tandem Mass Spectra of Intact Proteins. *Mol. Cell. Proteomics*, **2010**, *9*, 2772–2782.

[225] Matthiesen, R *et al*., Discussion on common data analysis strategies used in MS-based proteomics. *Proteomics*, **2011**, *11*, 604–619.

[226] Codrea, MC *et al*., Platforms and Pipelines for Proteomics Data Analysis and Management. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, **2016**; Vol. 919, pp. 203–215.

[227] Pereira-Medrano, AG *et al*., Proteomics, Protein Engineering. In *Comprehensive Biotechnology*; Elsevier, **2011**; Vol. 2, pp. 421–439.

[228] Barbier Saint Hilaire, P *et al*., Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics. *Metabolites*, **2020**, *10*, 158.

[229] Ting, YS *et al*., Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell. Proteomics*, **2015**, *14*, 2301–2307.

[230] Bateman, NW *et al*., Maximizing Peptide Identification Events in Proteomic Workflows Using Data-Dependent Acquisition (DDA). *Mol. Cell. Proteomics*, **2014**, *13*, 329–338.

[231] Borràs, E *et al*., What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics*, **2017**, *17*, 1700180.

[232] Lange, V *et al*., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.*, **2008**, *4*, 222.

[233] Picotti, P *et al*., Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods*, **2012**, *9*, 555–566.

[234] Aebersold, R *et al*., Applications and Developments in Targeted Proteomics: From SRM to DIA/SWATH. *Proteomics*, **2016**, *16*, 2065–2067.

[235] Zhang, F *et al*., Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics*, **2020**, *20*, 1900276.

[236] Ludwig, C *et al*., Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.*, **2018**, *14*, e8126.

[237] Bilbao, A *et al*., Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics*, **2015**, *15*, 964–980.

[238] Goldring, JPD., Measuring Protein Concentration with Absorbance, Lowry, Bradford Coomassie Blue, or the Smith Bicinchoninic Acid Assay Before Electrophoresis. In *Methods in Molecular Biology*; Humana Press Inc., **2019**; Vol. 1855, pp. 31–39.

[239] Vidova, V *et al*., A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal. Chim. Acta*, **2017**, *964*, 7–23.

[240] Chen, Y *et al*., Mass Spectrometry-Based Protein Quantification. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, **2016**; Vol. 919, pp. 255–279.

[241] Ankney, JA *et al*., Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics. *Annu. Rev. Anal. Chem.*, **2018**, *11*, 49–77.

[242] Ong, S-E *et al*., Mass spectrometry–based proteomics turns quantitative. *Nat. Chem. Biol.*, **2005**, *1*, 252–262.

[243] Bantscheff, M *et al*., Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.*, **2007**, *389*, 1017–1031.

[244] Sinitcyn, P *et al*., Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. *Annu. Rev. Biomed. Data Sci.*, **2018**, *1*, 207–234.

[245] Pontes, AH *et al*., Mass Spectrometry-Based Approaches to Understand the Molecular Basis of Memory. *Front. Chem.*, **2016**, *4*, 40.

[246] Lindemann, C *et al*., Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biol. Chem.*, **2017**, *398*, 687–699.

[247] Gerber, SA *et al*., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci.*, **2003**, *100*, 6940–6945.

[248] Beynon, RJ *et al*., Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods*, **2005**, *2*, 587–589.

[249] Dupuis, A *et al*., Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics*, **2008**, *8*, 4633–4636.

[250] Ishihama, Y *et al*., Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced

Peptides per Protein. *Mol. Cell. Proteomics*, **2005**, *4*, 1265–1272.

[251] Braisted, JC *et al*., The APEX Quantitative Proteomics Tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics*, **2008**, *9*, 529.

[252] Schwanhäusser, B *et al*., Global quantification of mammalian gene expression control. *Nature*, **2011**, *473*, 337–342.

[253] Silva, JC *et al*., Absolute Quantification of Proteins by LCMS E. *Mol. Cell. Proteomics*, **2006**, *5*, 144–156.

[254] Wiśniewski, JR *et al*., Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.*, **2012**, *8*, 611.

[255] Wiśniewski, JR., Label-Free and Standard-Free Absolute Quantitative Proteomics Using the “Total Protein” and “Proteomic Ruler” Approaches. In *Methods in Enzymology*; Academic Press Inc., **2017**; Vol. 585, pp. 49–60.

[256] Verheggen, K *et al*., Database Search Engines: Paradigms, Challenges and Solutions. In *Advances in Experimental Medicine and Biology*; Springer New York LLC, **2016**; Vol. 919, pp. 147–156.

[257] Elias, JE *et al*., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods*, **2007**, *4*, 207–214.

[258] Bateman, A *et al*., UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **2017**, *45*, D158–D169.

[259] Tyanova, S *et al*., Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. In *Methods in Molecular Biology*; **2018**; Vol. 1711, pp. 133–148.

[260] Lualdi, M *et al*., Statistical analysis of proteomics data: A review on feature selection. *J. Proteomics*, **2019**, *198*, 18–26.

[261] Urfer, W *et al*., Statistics for Proteomics: A Review of Tools for Analyzing Experimental Data. *Proteomics*, **2006**, *6*, 48–55.

[262] Karimpour-Fard, A *et al*., A survey of computational tools for downstream analysis of proteomic and other omic datasets. *Hum. Genomics*, **2015**, *9*, 28.

[263] Tyanova, S *et al*., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **2016**, *13*, 731–740.

[264] Ashburner, M *et al*., Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **2000**, *25*, 25–29.

[265] Kanehisa, M *et al*., KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **2012**, *40*, D109–D114.

[266] Croft, D *et al*., Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **2011**, *39*, D691–D697.

[267] Bindea, G *et al*., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, **2009**, *25*, 1091–1093.

[268] Doncheva, NT *et al*., Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.*, **2019**, *18*, 623–632.

[269] Rubio, DM *et al*., Defining Translational Research: Implications for Training. *Acad. Med.*, **2010**, *85*, 470–475.

[270] van Helden, P., Data-driven hypotheses. *EMBO Rep.*, **2013**, *14*, 104–104.

[271] Lovric, J.,John Wiley & Sons, Ltd, **2011**.

[272] O'Farrell, PH., High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, **1975**, *250*, 4007–4021.

[273] Klose, J., Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues - A novel approach to testing for induced point mutations in mammals. *Hum. Genet.*, **1975**, *26*, 231–243.

[274] Scheele, GA., Chapter 22 Analysis of the Secretory Process in the Exocrine Pancreas by Two-Dimensional Isoelectric Focusing/Sodium Dodecyl Sulfate Gel Electrophoresis. In *Methods in Cell Biology*; **1981**; Vol. 23, pp. 345–358.

[275] Horgan, RP *et al*., 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet. Gynaecol.*, **2011**, *13*, 189–195.

[276] Norin, M *et al*., Structural proteomics: developments in structure-to-function predictions. *Trends Biotechnol.*, **2002**, *20*, 79–84.

[277] Monti, M *et al*., Functional proteomics. *Clin. Chim. Acta*, **2005**, *357*, 140–150.

[278] Cristea, IM *et al*., Affinity Purification of Protein Complexes. *Cold Spring Harb. Protoc.*, **2011**, *2011*, pdb.prot5611.

[279] Zhang, J *et al*., Exploring the functional complexity of cellular proteins by protein knockout. *Proc. Natl. Acad. Sci.*, **2003**, *100*, 14127–14132.

[280] Suter, B *et al*., Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond. *Biotechniques*, **2006**, *40*, 625–644.

[281] Chapman, T., Mining the proteome. *Nature*, **2004**, *430*, 109–109.

[282] Larsen, MR *et al*., Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques*, **2006**, *40*, 790–798.

[283] Ribet, D *et al*., Post-translational modifications in host cells during bacterial infection. *FEBS Lett.*, **2010**, *584*, 2748–2758.

[284] Wang, Y-C *et al*., Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.*, **2014**, *24*, 143–160.

[285] Kuzmanov, U *et al*., Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Med.*, **2013**, *5*, 37.

[286] Niu, L *et al*., Proteomics in the Study of Liver Diseases. In *The Human Gut-Liver-Axis in Health and Disease*; Springer International Publishing: Cham, **2019**; pp. 165–193.

[287] Mischak, H *et al*., Clinical proteomics: A need to define the field and to begin to set adequate standards. *PROTEOMICS – Clin. Appl.*, **2007**, *1*, 148–156.

[288] Doerr, A., Mass spectrometry–based targeted proteomics. *Nat. Methods*, **2013**, *10*, 23.

[289] Strimbu, K *et al*., What are biomarkers? *Curr. Opin. HIV AIDS*, **2010**, *5*, 463–466.

[290] Hoofnagle, AN *et al*., Proteomics. In *Principles and Applications of Molecular Diagnostics*; Elsevier, **2018**; pp. 381–401.

[291] del Campo, M *et al*., Facilitating the Validation of Novel Protein Biomarkers for Dementia: An Optimal Workflow for the Development of Sandwich Immunoassays. *Front. Neurol.*, **2015**, *6*, 1.

[292] Califf, RM., Biomarker definitions and their applications. *Exp. Biol. Med.*, **2018**, *243*, 213–221.

[293] Neubert, H *et al*., Protein Biomarker Quantification by Immunoaffinity Liquid Chromatography–Tandem Mass Spectrometry: Current State and Future Vision. *Clin. Chem.*, **2020**, *66*, 282–301.

[294] MacBeath, G., Protein microarrays and proteomics. *Nat. Genet.*, **2002**, *32*, 526–532.

[295] Taylor, SW *et al*., A high-throughput mass spectrometry assay to simultaneously measure intact insulin and C-peptide. *Clin. Chim. Acta*, **2016**, *455*, 202–208.

[296] Hoofnagle, AN *et al*., Quantification of Thyroglobulin, a Low-Abundance Serum Protein, by Immunoaffinity Peptide Enrichment and Tandem Mass Spectrometry. *Clin. Chem.*, **2008**, *54*, 1796–1804.

[297] Kushnir, MM *et al*., Measurement of Thyroglobulin by Liquid Chromatography–Tandem Mass Spectrometry in Serum and Plasma in the Presence of Antithyroglobulin Autoantibodies. *Clin. Chem.*, **2013**, *59*, 982–990.

[298] Guo, T *et al*., Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.*, **2015**, *21*, 407–413.

[299] Drendel, V *et al*., Proteomic distinction of renal oncocytomas and chromophobe renal cell carcinomas. *Clin. Proteomics*, **2018**, *15*, 25.

[300] Bystrom, CE *et al*., Plasma Renin Activity by LC-MS/MS: Development of a Prototypical Clinical Assay Reveals a Subpopulation of Human Plasma Samples with Substantial

Peptidase Activity. *Clin. Chem.*, **2010**, *56*, 1561–1569.

[301] Bystrom, CE *et al*., Narrow Mass Extraction of Time-of-Flight Data for Quantitative Analysis of Proteins: Determination of Insulin-Like Growth Factor-1. *Anal. Chem.*, **2011**, *83*, 9005–9010.

[302] Bystrom, C *et al*., Clinical Utility of Insulin-Like Growth Factor 1 and 2; Determination by High Resolution Mass Spectrometry. *PLoS One*, **2012**, *7*, e43457.

[303] JIN, M *et al*., A rapid test for the diagnosis of thrombotic thrombocytopenic purpura using surface enhanced laser desorption/ionization time-of-flight (SELDI-TOF)-mass spectrometry. *J. Thromb. Haemost.*, **2006**, *4*, 333–338.

[304] Topbas, C *et al*., Measurement of Lipoprotein-Associated Phospholipase A2 by Use of 3 Different Methods: Exploration of Discordance between ELISA and Activity Assays. *Clin. Chem.*, **2018**, *64*, 697–704.

[305] Kushnir, MM *et al*., LC-MS/MS Measurement of Parathyroid Hormone–Related Peptide. *Clin. Chem.*, **2016**, *62*, 218–226.

[306] Mills, JR *et al*., Comprehensive Assessment of M-Proteins Using Nanobody Enrichment Coupled to MALDI-TOF Mass Spectrometry. *Clin. Chem.*, **2016**, *62*, 1334–1344.

[307] Ladwig, PM *et al*., Quantification of Serum IgG Subclasses by Use of Subclass-Specific Tryptic Peptides and Liquid Chromatography–Tandem Mass Spectrometry. *Clin. Chem.*, **2014**, *60*, 1080–1088.

[308] van der Gugten, G *et al*., Resolution of Spurious Immunonephelometric IgG Subclass Measurement Discrepancies by LC-MS/MS. *Clin. Chem.*, **2018**, *64*, 735–742.

[309] Henderson, CM *et al*., Measurement by a Novel LC-MS/MS Methodology Reveals Similar Serum Concentrations of Vitamin D–Binding Protein in Blacks and Whites. *Clin. Chem.*, **2016**, *62*, 179–187.

[310] van den Broek, I *et al*., Automated Multiplex LC-MS/MS Assay for Quantifying Serum Apolipoproteins A-I, B, C-I, C-II, C-III, and E with Qualitative Apolipoprotein E Phenotyping. *Clin. Chem.*, **2016**, *62*, 188–197.

[311] Silvestri, GA *et al*., Assessment of Plasma Proteomics Biomarker's Ability to Distinguish Benign From Malignant Lung Nodules. *Chest*, **2018**, *154*, 491–500.

[312] Collier, TS *et al*., Rapid Affinity Enrichment of Human Apolipoprotein A-I Associated Lipoproteins for Proteome Analysis. *J. Proteome Res.*, **2018**, *17*, 1183–1193.

[313] Leinenbach, A *et al*., Mass Spectrometry–Based Candidate Reference Measurement Procedure for Quantification of Amyloid-β in Cerebrospinal Fluid. *Clin. Chem.*, **2014**, *60*, 987–994.

CHAPTER II.

Objectives and work plan

II.1 OBJECTIVES

Accurate diagnosis and consequential medical treatment management are of utmost importance in clinical practice, where patient care can be remarkably compromised based on a misdiagnosis. Although in clear cell carcinoma imaging and immunohistochemical techniques have been useful in diagnosis and subtype differentiation, some overlapped features are shared between subtypes, making diagnosis difficult. Although some proteomics studies have been performed to overcome such shortcoming, up to now, an unequivocal classification based on an accurate biomarker or panel of biomarkers for RCC subtype diagnosis remains elusive. Thus, the objectives of the present work are:

- The search of a unique proteomic profile for each one of the different RCC tumors
- Identification of a specific biomarker, or a panel of biomarkers, able to accurately diagnose each tumor type by the application of the Total Protein Approach to solid biopsies.
- To translate the new mass-spectrometry biomarkers into immunohistochemistry biomarkers.
- The development of new phosphoproteomics approach to find new targets for therapy.

II.2 WORK PLAN

To develop the topics projected above renal tissue biopsies were collected from patients diagnosed with different subtypes of RCC. In total, 27 tissue biopsies were included in the present study, as follows, ccRCC = 7, pRCC = 5, chRCC = 5 and RO = 5. For comparative purposes five normal adjacent tissues (NAT) were used as controls. Sample were provided by the University of Pittsburgh Medical Center (UPMC, Pittsburgh, PA, USA).

II.2.1.1.1.1.1 Tissue collection and optimization of protein extraction from OCT-preserved solid biopsies: OCT Cleaning

Tissue biopsy samples will be collected by the University of Pittsburgh Biospecimen Core and preserved in OCT media. In order to enable to use of OCT-embedded samples in MS-based experiments, a first step of cleaning is added to the main experimental workflow.

Deliverables: Implementation of a methodology to improve the OCT removal in sample cleaning steps. Acquisition of tissue samples OCT-free.

II.2.1.1.1.1.2 MS-based proteomics profiles of OTC-clean solid biopsies

The proteome of the biopsies will be solubilized using ultrasonic energy. Then, the extract will be digested and submitted to a sequential extraction using Zip-Tips. The profile of each extract will be obtained using MALDI-TOF-MS and using bioinformatics tools. The most promising profile will be further analyzed using Ultra-High-Resolution LC-MS/MS.

Deliverables: Acquisition of MALDI profiles for chRCC and RO. Identification of the list of proteins present in the best profile.

II.2.1.1.1.1.3 Ultrasonic based Total Protein Approach for the discovery of new biomarkers for RCC pathology

Applying the extraction method developed in WORK PACKAGE 2, the protein content of the biopsies will be quantified by label free-based mass spectrometry. Then the total protein approach will be also applied to find out biomarkers of diagnosis. Biomarkers will be validated using immunohistochemistry.

Deliverables: Disclosing biomarkers of diagnosis for RCC using the Total Protein Approach and validation of a panel of biomarkers by immunohistochemistry.

II.2.1.1.1.1.4 Phosphopeptide enrichment

New nano-IMACs based on LA^{3+} and Ti^{4+} materials for peptide enrichment will be synthetized and characterized. Such nano-IMACs will be used for phosphopeptide enrichment from complex proteomes, including RCC samples.

Deliverables: A new family of nano-IMACs sorbents. Identification of phosphorylated peptides characteristic to each tumor subtype.

CHAPTER III.

Development of a robust ultrasonic-based sample treatment to unravel the proteome of OCT-embedded solid tumor biopsies

Susana Jorge, José L. Capelo, William Laframboise, Rajiv Dhir, Carlos Lodeiro, Hugo M. Santos*

Published in:

Journal of Proteome Research, **2019**, 18, 2979 – 2986

DOI: 10.1021/acs.jproteome.9b00248

ABSTRACT

An effective three step proteomic workflow is proposed to overcome the pitfalls caused by polymers present in OCT-embedded tissue during preparation for mass spectrometry analysis. First, the OCT-embedded tissue biopsies are cleaned using ethanol and water, in a sequential series of ultrasonic washes in an ultrasound bath (35 kHz ultrasonic frequency, 100% ultrasonic amplitude, 2 min ultrasonic duty time). Second, a fast ultrasonic-assisted extraction of proteins is done using an ultrasonic probe (30 kHz ultrasonic frequency, 50% ultrasonic amplitude, 2 min ultrasonic duty time, 1 mm diameter tip). Third, a rapid ultrasonic digestion of complex proteomes is performed using a microplate horn assembly device (20 kHz ultrasonic frequency, 25% ultrasonic amplitude, 4 min ultrasonic duty time). As proof of concept, the new workflow was applied to human normal and tumor kidney biopsies including chromophobe renal cell carcinomas (chRCC) and renal oncocytomas (RO). A successful cluster of proteomic profiles was obtained comprising 511 and 172 unique proteins found in chRCC and RO samples, respectively. The new method provides high sample throughput and comprehensive protein recovery from OCT samples.

Keywords: OCT-embedded tissues; label-free quantification; mass spectrometry; ultrasound energy; chromophobe renal cell carcinoma; renal oncocytoma

III.1 INTRODUCTION

The identification and quantification of proteins in liquid or solid tumor biopsies by mass spectrometry (MS) is essential for optimal medical diagnosis and prognosis and may help to identify novel targets for therapies against disease [1–4]. In the case of solid tumors, tissue preservation involves the use of formaldehyde for protein fixation followed by embedding in paraffin; or flash-freezing, unfixed tissues in polymers such as optimum cutting temperature compound (OCT). Formaldehyde chemically modifies proteins resulting in lower yields after extraction and fewer proteins identified by tandem MS [5–7]. Likewise, polyethylene glycol (PEG) and polyvinyl alcohol polymers (PVA) in OCT interfere with peptide analysis by LC-MS/MS, because of ion signal suppression caused by such polymers [8–10]. Due to the high solubility of PEG and PVA in aqueous solutions, most methods for removal of polymer contaminants from OCT-embedded samples, include washing tissue biopsies with ethanol and water as reported by Loken *et al.* [11], and Zhang *et al.* [12]. This remains the most common method to clean samples embedded in OCT [13–15]. Other approaches include (i) protein precipitation with diethyl ether-methanol [16,17] or TCA [18], (ii) filter-aided sample preparation

(FASP) of tissues [16], (iii) sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) [16] and (iv) solid-phase extraction [19,20].

The advent of ultrasound as a tool in analytical and bioanalytical laboratories has gained momentum in proteomics [21]. Ultrasonic energy can be used (i) in the cleaning of samples [22], (ii) in accelerating digestion of complex proteomes [23], (iii) in reducing the time and handling necessary in tedious proteomics workflows [24] and (iv) can be scaled up for use in high-throughput applications [25,26].

In this work, we present a novel ultrasonic-based pipeline to interrogate the proteome of OCT-embedded tissue biopsies. To this end ultrasonic energy is used first, to speed up the OCT cleaning process, then to increase protein extraction and solubilization and finally to expedite the overall protein digestion workflow. As a proof of concept, the proteome of chRCC and RO tumors was evaluated by MS and compared with normal adjacent kidney tissues (NAT) as biological control samples.

III.2 EXPERIMENTAL SECTION

III.2.1 MICE TISSUE SAMPLES

Mouse kidneys were purchased from Patricell Limited Ltd (UK).

III.2.2 HUMAN BIOPSIES

The human kidney tissue samples were collected by the University of Pittsburgh Biospecimen Core's and the study was approved by the Institutional Review Board at the University of Pittsburgh (IRB # 02-077). All neoplasms contained a minimum of 90% tumor cells and NAT specimens were at least 90% normal cells. Data of patients enrolled in this study are summarized in Table III.1.

III.2.3 OPTIMIZATION OF OCT CLEANING USING MOUSE KIDNEY SAMPLES

Mice kidney were used as a surrogate for optimization purposes. The mice kidney were placed in a mortar filled with liquid nitrogen and reduced to powder using a pestle. The powder was divided in aliquots of approximately 10 mg each. Nine samples were preserved by embedding in OCT compound, ensuring that the tissues were completely covered. Three aliquots were frozen without OCT, and all aliquots were stored in the freezer at -60 ºC. Later, the OCT-embedded tissues were subjected to the OCT cleaning protocol according to the method proposed by Zhang *et al*. [12] with modifications. According to the authors, the polymers in the OCT compound can be removed through a series of ethanol and water washes.

In our approach, the washing steps were done following three different treatments:

(i) Ultrasonic (US) bath, at 35 kHz for 2 min at 100% ultrasonic amplitude;

(ii) US bath at 130 kHz for 2 min at 100% ultrasonic amplitude; and

(iii) Gentle vortex agitation as control.

Washing was performed as follows:

(i) 1 ml 70% (v/v) ethanol at 4 ºC was added to the OCT-embedded samples, washed by one of the three above treatments followed by centrifugation at 4 ºC for 2 min (5,000 g). The supernatant was carefully removed, and the procedure was repeated once.

(ii) The pellet was washed with 1 ml of water at 4 ºC and submitted to the same treatment as before followed by centrifugation at 4 ºC for 2 min (5,000 g). The supernatant was carefully removed, and this cleaning step was performed 5 times.

Table III.1. Description of human kidney biopsies used in the study.

BIOPSY	AGE	GENDER	DIAGNOSIS*	SAMPLE TYPE*
N1	50-59	Male	RCC	NAT
N2	40-49	Female	Papillary	NAT
N3	50-59	Female	RCC	NAT
N4	70-79	Female	RCC	NAT
N5	70-79	Male	RCC	NAT
C6	70-79	Male	RCC	chRCC
C7	60-69	Female	RCC	chRCC
C8	70-79	Male	RCC	chRCC
C9	50-59	Female	RCC	chRCC
C10	80-89	Male	RCC	chRCC
O11	80-89	Male	RCC	RO
O12	60-69	Female	RCC	RO
O13	60-69	Male	RCC	RO

**RCC: renal cell carcinoma; NAT: normal adjacent tissue; chRCC: chromophobe renal cell carcinoma; RO: renal oncocytoma.*

III.2.4 OCT CLEANING OF HUMAN KIDNEY BIOPSIES

Human kidney biopsies embedded in OCT were thawed and excess OCT was removed. The biopsies were cleaned using the optimum methodology obtained with mice kidney samples: US bath at 35 kHz for 2 min at 100% ultrasonic amplitude. At the end of the cleaning steps, the biopsies were frozen with liquid nitrogen and reduced to a powder.

III.2.5 OPTIMIZATION OF PROTEIN EXTRACTION FROM MOUSE KIDNEY SAMPLES

OCT-free tissues were extracted in 8 M urea prepared in 25 mM Ambic buffer (ratio: 100 µL buffer to 10 mg of tissue). Protein extraction was done three times over the same sample, utilizing an ultrasonic processor UP50H (50 w, 30 kHz, 1 mm diameter probe tip) operating at 50% ultrasonic amplitude for 2 min in pulsed mode (10 sec on; 10 sec off). After each ultrasonic extraction, the sample was centrifuged at 10,000 g for 10 min, and then the supernatant was transferred to a new tube. Then the protein content from each supernatant was precipitated using the DOC/TCA and acetone method. Briefly, 1 µL of 2% (w/v) DOC were added to each 100 µL of supernatant and left on ice for 30 min. Subsequently, 25 µL of 100% TCA were added to the mixture and the samples left on ice for 20 min, followed by centrifugation at 4 ºC for 20 min (16,000 g). The supernatant was removed, and the pellets were washed with 200 µL of ice-cold acetone (-20 ºC), followed by centrifugation (16,000 g for 20 min at 4 °C). Then, 20 µL of 0.2 M NaOH were added to the protein pellet followed by a 2 min incubation at room temperature and addition of 80 µL of 6 M urea in 25 mM Ambic. The protein pellet was dissolved using four cycles of 10 sec of ultrasonic energy through an ultrasonic processor UP50H (50 w, 30 kHz, 1 mm diameter probe tip) operating at 50% ultrasonic amplitude. Five second intervals occurred between ultrasonic cycles. Finally, the total protein content of each supernatant (n = 3) was determined by the Bradford protein assay.

III.2.6 PROTEIN EXTRACTION IN HUMAN KIDNEY SAMPLES

Human biopsies with OCT removed were placed in a mortar filled with liquid nitrogen and reduced to powder using a pestle. Proteins where extracted from the resulting power, using 8 M urea/ 25 mM Ambic buffer (ratio: 100 µL buffer to 10 mg of tissue). The solid-liquid extraction process was done under an ultrasonic field provided by an ultrasonic processor UP50H (50 w, 30 kHz, 1 mm diameter probe tip) operating at 50% ultrasonic amplitude for 2 min in pulsed mode (10 sec on; 10 sec off). The ultrasonic-assisted protein extraction procedure was performed twice over each sample, as described in the previous section, and then supernatants were combined. The protein extract was submitted to precipitation and quantification as described for mouse kidney samples.

Reduction and Alkylation. Twenty µL of combined supernatants were reduced with addition of 2 µL of 110 mM DTT. The sample was then vortexed and incubated for 60 min at 37 °C. The resulting cysteines were blocked with 2 µL of 600 mM IAA. The sample was vortexed and incubated for 45 min at room temperature in the dark followed by dilution to a final volume of 100 µL with 25 mM Ambic.

III.2.7 ULTRASONICALLY ASSISTED IN-SOLUTION DIGESTION OF PROTEINS

For protein digestion, 50 µg total protein were mixed with trypsin in a 1:20 (w/w) ratio. Trypsin digestion was performed in the microplate horn assembly device under the following

operating conditions: 25% ultrasonic amplitude, 4 min ultrasonic duty time in pulsed mode (30 sec on, 15 sec off). Finally, formic acid was added to obtain a final concentration of 0.1% (v/v) to stop the enzymatic activity and the digested samples were evaporated to dryness.

III.2.8 MALDI-TOF-MS

To evaluate the efficiency of OCT removal during the cleaning procedure, 10 µL of cleaning solution were withdrawn from the supernatant after each cleaning step for analysis by MALDI-TOF-MS. A solution of DHB (10 mg/mL), resuspended in NaBF4 (10 mg/mL), was used as MALDI matrix. Samples were mixed with matrix solution in a ratio of 1:1, hand-spotted in duplicate on to a MALDI target and allowed to air dry. The MALDI TOF/TOF mass spectrometer was operated in positive ion mode using a reflectron, and spectra were acquired in the m/z range of 600–3500. A total of 500 spectra were acquired for each sample at a laser frequency of 50 Hz. A solution of 0.2 mM PEG 1000 was used as reference.

III.2.9 NANO-LC-ESI-MS/MS ANALYSIS

The LC-MS/MS analysis was carried out using an Ultimate 3000 nLC coupled to an UHR-QqTOF IMPACT HD (Bruker Daltonics) with a CaptiveSpray ion source (Bruker Daltonics). All samples were reconstituted to a final digested protein concentration of 0.25 µg/µL in 3% ACN/0.1% (v/v) aqueous formic acid. Three µL of each sample was loaded into a trap column Acclaim PepMap100, 5 µm, 100 Å, 300 µm i.d. × 5 mm and desalted for 5 min with 3% B (B: 90% ACN/0.1% FA) at a flow rate of 15 µL/min. Chromatographic separation was carried out using an analytical column Acclaim™ PepMap™ 100 C18, 2 µm, 0.075 mm i.d x 150 mm with a linear gradient at 300 nL/min (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) ACN and 0.08% (v/v) FA), 0-5 min with 3% of mobile phase B, 5-95 min from 3% to 35% of mobile phase B, 95-105 min linear gradient from 35% to 95% of mobile phase B, 105-115 with 95% of mobile phase B. The total run time was 130 min. For each sample, two replicate injections were performed. Chromatographic separation was carried out at 35 ºC. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8–32 Hz), cycle time 3.0 seconds, with active exclusion (precursors were excluded from precursor selection for 0.5 min after acquisition of 1 MS/MS spectrum, intensity threshold for fragmentation of 2500 counts). Together with active exclusion set to 1, reconsider precursor if the intensity of a precursor increases by a factor of 3, this mass will be taken from temporary exclusion list and fragmented again, ensuring that fragment spectra were taken near to the peak maximum. All spectra were acquired in the range 150–2200 m/z.

III.2.10 DATA ANALYSIS AND STATISTICS

Raw LC-MS/MS data were processed in DataAnalysis 4.2 and subsequently exported to Protein-Scape 4.0 for automated protein identification. CID-MS2 spectra were first searched

against the M. musculus (16,230 sequences) or H. sapiens (20,266 sequences) subset of the SwissProt database 57.15 (515,203 sequences; 181,334,896 residues), using the Mascot search engine (V. 2.3.02) with the following parameters: (i) two missed cleavage; (ii) fixed modifications: carbamidomethylation (C); (iii) variable modifications: oxidation of methionine, Acetyl (Protein N-term), Glu- > pyro-Glu (N-term E), Gln- > pyro-Glu (N- term Q), (vi) peptide mass tolerance up to 20 ppm, (v) fragment mass tolerance 0.05 Da and (vi) FDR adjusted to 1%. Label-free quantification was carried out using MaxQuant software (version 1.6.0.16). All raw files were processed in a single run with default parameters [27,28]. Database searches are performed using the Andromeda search engine with the SwissProt M. musculus database as reference and a database of common contaminants. Data processing was performed using Perseus (version 1.6.2.3) with default settings [29]. In brief, protein groups LFQ intensities were log_2-transformed to reduce the effect of outliers. Missing LFQ values were imputed using default parameters (with = 0.3 and down shift = 1.8). Log ratios were calculated as the difference in average log_2 LFQ intensity values between the two conditions tested (two-tailed Student's t-test, FDR 0.01 and S0 = 1.5).

III.3 RESULTS AND DISCUSSION

III.3.1 OPTIMIZATION OF OCT CLEANING

To remove OCT embedding material from tissues we modified the method of Zhang *et al*. [12] by performing the cleaning process under an ultrasonic field. Figure III.1 presents the entire workflow that was used to optimize the ultrasonic-based OCT removal from biopsies and the subsequent protein extraction and digestion to peptides. Mouse kidney tissue embedded in OCT was used in the optimization process. The mouse kidneys were powdered using liquid nitrogen, as described in the experimental section, III.2.3 Optimization of OCT cleaning using mouse kidney samples. OCT can be removed from tissues with water [12], but some proteins might be lost during this process. To avoid this problem, proteins were first fixed with ethanol 70% (v/v) under ultrasonication. Next, the OCT was removed by washing the samples with water in an ultrasonic field. Two conditions were tested: (i) ultrasonic (US) bath at 35 kHz, 100% ultrasonic amplitude, and (ii) US bath at 130 kHz, 100% ultrasonic amplitude. For comparative purposes, samples were cleaned for 2 min with water by gentle mixing using a vortexer. After each washing step an aliquot was analyzed by MALDI-TOF to check for the presence of OCT.

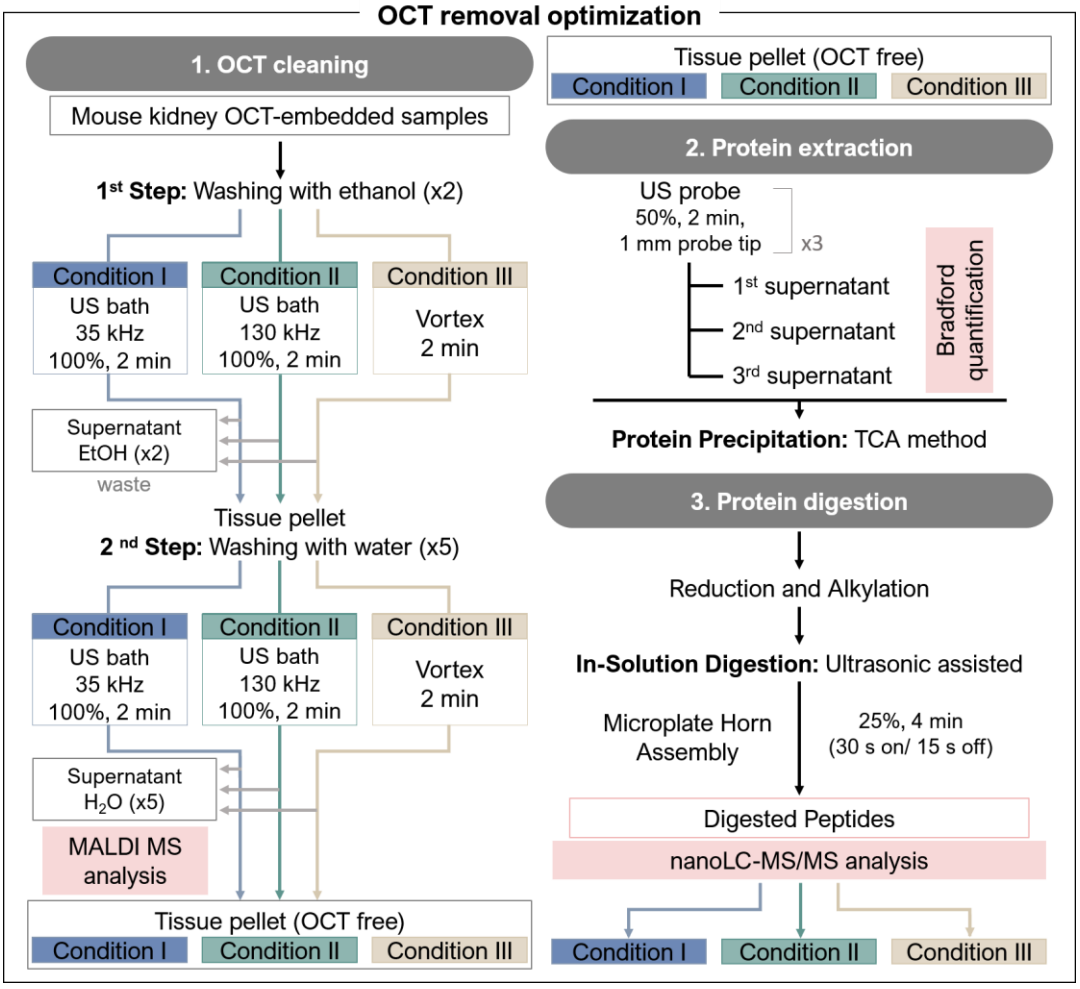

Figure III.1 - Schematic representation of OCT removal from mouse kidney. First the proteins are fixed by washing the tissues with ethanol 70% (v/v). Next, the OCT was removed by washing the samples with water under the effects of an ultrasonic field in an ultrasonic bath. Two conditions were tested: (i) ultrasonic (US) bath at 35 kHz, 100% ultrasonic amplitude and (ii) US bath at 130 kHz, 100% ultrasonic amplitude. A third condition was also performed using vortex shaking for comparison. All procedures were applied for 2 min.

Figure III.2 depicts the spectra obtained after five water cleaning cycles for the three conditions. As it might be seen the spectrum of the supernatant treated with the US bath at 35 kHz (condition I) presents OCT peaks with lower intensity than the US bath at 130 kHz (condition II) and vortex shaking (condition III). In fact, Fernandes *et al*. [30] have already described the use of the US bath at 35 kHz for fast and high throughput sample treatment for polymer characterization.

III.3.2 Optimization of protein extraction

After OCT removal, the proteins were extracted using an ultrasonic processor UP50H following a procedure we previously established [31] with modifications (see optimization of protein extraction from mouse kidney samples from experimental section). In brief, each pellet was extracted three times using an ultrasonic probe and the supernatants were collected for

new tubes, after centrifugation. The total protein content was determined for individual and pooled extracts by Bradford protein assay. As shown in Figure III.3A, the three procedures we tested showed similar protein yields regardless of the cleaning protocol. Figure III.3A shows also that almost all the protein content is recovered in the first cycle of extraction, 1st supernatant, about 87-93%. With a second cycle of extraction the amount of protein is increased up to 10%, while a third cycle only adds up to a 3% of the total protein recovered. For this reason, two cycles of homogenization were the number of extractions chosen for further experiments.

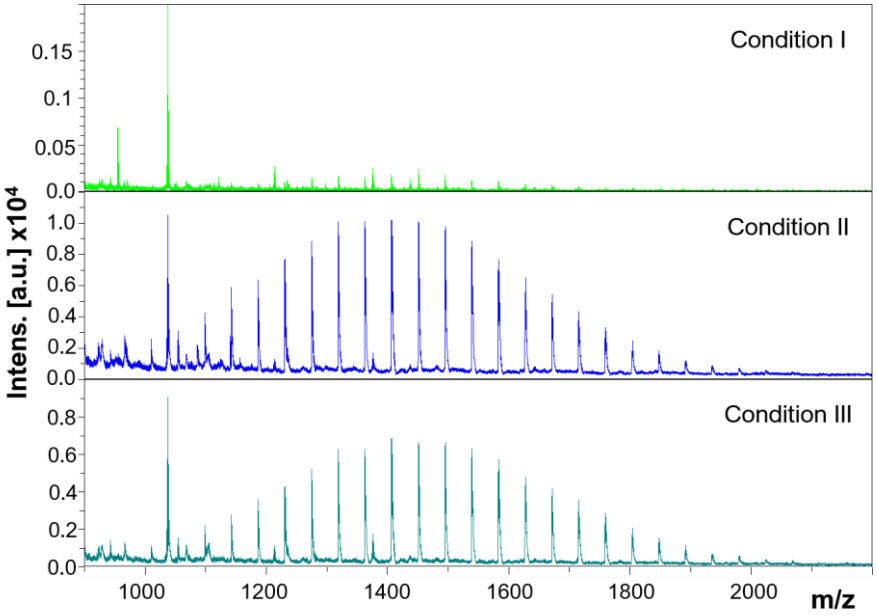

Figure III.2 - MALDI-TOF-MS spectra of supernatants obtained after five water cleaning steps for each condition assessed. The cleaning conditions evaluated were: (i) condition I: US bath at 35 kHz ultrasonic frequency, 2 min ultrasonic duty time and 100 % ultrasonic amplitude, (ii) condition II: US bath at 130 kHz ultrasonic frequency, 2 min ultrasonic duty time and 100 % ultrasonic amplitude, and (iii) condition III: vortexer, 2 min vortexing time.

III.3.3 PROTEIN IDENTIFICATION

The protein extracts were submitted to in-solution protein digestion. Trypsin digestion was performed with the aid of ultrasonication using a microplate horn assembly device as described in Jorge *et al* [26]. Finally, the resultant pools of peptides were analyzed by LC-MS/MS. The analysis resulted in the identification of 1510, 1419 and 1388 unique proteins for samples treated with US bath 35 kHz, US bath 130 kHz and vortexer, respectively (Figure III.3B). This result is consistent with the larger number of unique peptides obtained when the cleaning protocol is done with the ultrasonic bath at 35 kHz. Almost 1000 more unique peptides are obtained when compared with the vortexer method Figure III.3C). This is also consistent with the fact that after five washing cycles, the levels of OCT removed from the samples using

the US bath at 35 kHz are negligible if compared to the other two conditions, thus indicating a more efficient removal of OCT.

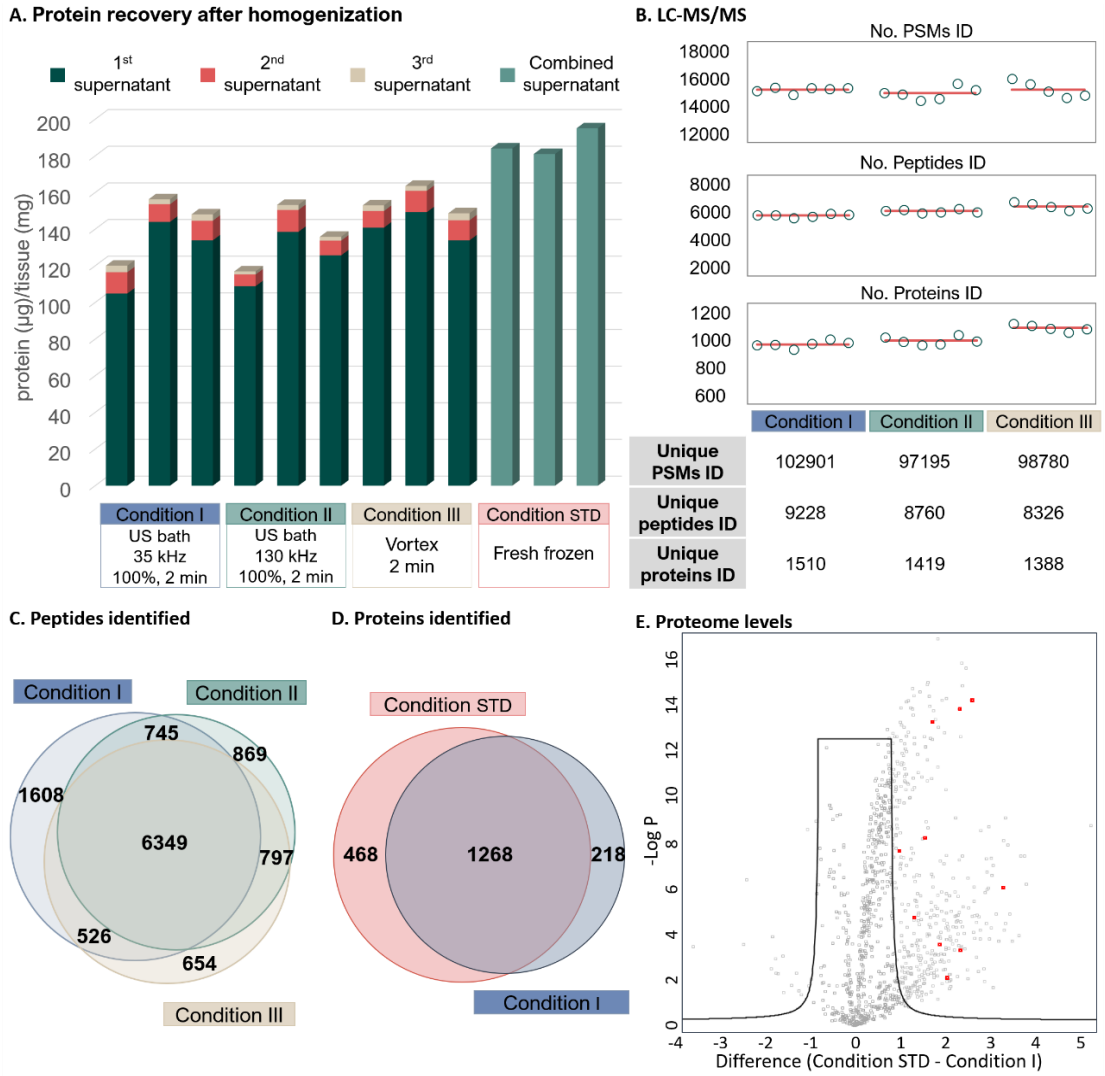

Figure III.3 – Comparison between different cleaning procedures from OCT-embedded mouse kidney samples. A. Amount of protein recovered per mg of tissue homogenized for each one of the three extracting cycles. Green: first extraction supernatant. Red: second extraction supernatant. Grey: third extraction supernatant. Three independent samples for each OCT cleaning procedure were evaluated. B. Number of (i) PSMs (Peptide Spectrum Matches), (ii) peptides, and (iii) proteins identified by LC-MS/MS analysis. C. Venn diagram showing the number of non-redundant peptides, both shared and unique, for each cleaning procedure. D. Venn diagram representing the number of overlapped proteins identified between condition I and condition STD. E. Volcano plot representation of the differentially expressed proteomes between condition I and condition STD. The red squares represent most abundant proteins present in blood (two-tailed, Student's t-test, FDR 0.01, S0 = 1.5). The conditions evaluated were (1) cleaning procedures: (i) condition I: US bath at 35 kHz ultrasonic frequency, 2 min ultrasonic duty time and 100% ultrasonic amplitude, (ii) condition II: US bath at 130 kHz ultrasonic frequency, 2 min ultrasonic duty time and 100 % ultrasonic amplitude, and (iii) condition III: vortexer, 2 min shaking time; (2) protein extraction and protein digestion were the same for each one of the cleaning conditions tested. Condition STD: for comparative purposes, fresh frozen tissues were also assessed as standard treatment. Fresh frozen tissues proteins were extracted and digested with trypsin following the same conditions as the OCT-cleaned samples.

The best performance obtained with the US bath at 35 kHz may be explained by the fact that the cavitation bubbles generated at 35 kHz frequency are larger compared to higher frequencies and their collapse results in shockwaves that promote mechanical shearing and surface disruption, thus leading to improved tissue breakdown. On the other hand, the cavitation bubbles generated at 130 kHz are much smaller, and their collapse induces a higher increase in temperature, but they are less effective for disrupting solid surfaces [32,33]. As a matter of fact, when ultrasound energy is applied on an aluminum foil, the ultrasonic field generated by the 35 kHz frequency promotes the disruption of the foil, whereas when it is treated within the ultrasonic field generated by the 130 kHz frequency it remains intact under visual inspection (Figure SM IX.1, Supplementary information). Therefore, we selected the US bath protocol at 35 kHz ultrasonic frequency, 2 min ultrasonic duty time and 100% ultrasonic amplitude as optimum conditions to clean OCT embedded tissue samples.

III.3.4 COMPARISON OF ULTRASONIC OCT-CLEANING METHOD TO FRESH FROZEN TISSUES

Fresh frozen tissues are considered the gold-standard for clinical mass spectrometry-based research and we included this preparation as an important control in this study.

Both frozen and OCT preparations were submitted to the same protocols for protein extraction and protein digestion. Figure III.3D shows that the number of proteins identified was about 13% higher for the fresh frozen tissue. This is consistent with the results achieved by Zhao *et al*. [18] which have compared OCT-embedded and fresh frozen samples of squamous cell carcinoma. On the other hand, fresh frozen preserved tissues retain some blood, which originates a higher number of identified proteins. Such explanation has been proved by Petris *et al*. [34], who suggest that blood contaminants can be removed from tissue samples through filtration and sequential washings. In our case the blood present in OCT-embedded tissue samples is likely separated during the cleaning steps under the effects of an ultrasonic field, whilst the fresh frozen were not submitted to such cleaning steps. Also, this effect can be seen in Figure III.3A, which shows that the amount of total protein per mg of tissue is higher in the fresh-frozen samples than in the OCT-cleaned ones. Moreover, Figure III.3E shows as red squares most abundant proteins in blood with a fold change higher than 1.5 present in the fresh frozen samples when compared to the OCT-preserved samples. List of quantified proteins are provided in Table ESM IX.1 of Supplementary information.

III.3.5 PROOF-OF-CONCEPT: IMPLEMENTATION OF ULTRASONIC OCT-CLEANING PROCEDURE FOR THE CLASSIFICATION OF HUMAN KIDNEY BIOPSIES

Figure III.4 shows a schematic representation of the workflow for protein extraction from OCT-embedded human kidney biopsies. Ultrasonication was applied three times, once during

the cleaning procedure (ultrasonic bath), another time during protein extraction (ultrasonic probe) and finally during protein digestion (ultrasonic 96-well plate).

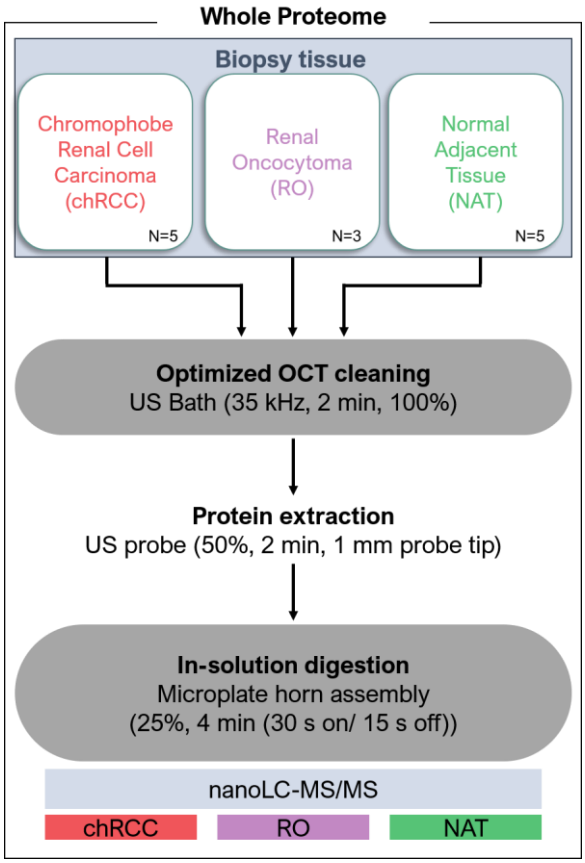

Figure III.4 – Schematic representation of human renal OCT-embedded tissue analysis workflow. The optimized method obtained to remove OCT (condition I: US bath at 35 kHz ultrasonic frequency, 2 min ultrasonic duty time and 100% ultrasonic amplitude) was applied to human biopsies. Proteomic analysis was carried out by nanoLC-MS/MS.

Renal biopsies diagnosed as chromophobe renal cell carcinoma (n = 5) or renal oncocytoma (n = 3) were interrogated by mass spectrometry. Normal adjacent tissue, NAT, (n = 5) were used as control. Intra and inter tumor heterogeneity are presented in Figure SM IX.2 of Supplementary information. Once the mass spectrometry data was obtained (90 min LC-MS/MS runs), the samples were grouped using an unsupervised clustering algorithm. The results shown in Figure 5A depict classification of mass spectrometry replicates for each sample matched together and their clustering. These findings suggest that the proteomics profiles expressed by each tissue type contains a unique fingerprint, that potentially allows discrimination of each sample type.

The number of proteins identified was 1798, 1276 and 1641 (Table ESM IX.2, Table ESM IX.3 and Table ESM IX.4 of Supplementary information), and of unique proteins was 511, 172 and 518 for chRCC, RO and NAT type-biopsies respectively (supporting information Table ESM IX.5, Table ESM IX.6 and Table ESM IX.7 of Supplementary information), as seen in Figure

III.5B. The large number of unique proteins for each group may explain the precise clustering depicted in Figure III.5A.

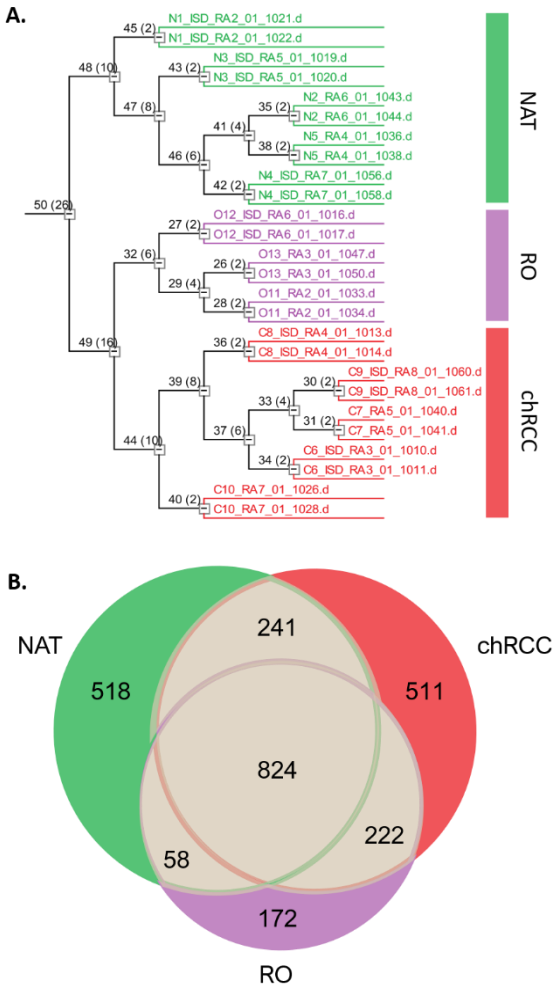

Figure III.5 – A. Unsupervised hierarchical cluster analysis of the proteomic profiles obtained for normal adjacent tissue (NAT), renal oncocytoma (RO) and chromophobe renal cell carcinoma (chRCC). B. Venn diagram showing the number of common and unique proteins achieved for each tissue type.

III.4 CONCLUSIONS

A new ultrasonic-based methodology for proteomic analysis of OCT-embedded tissues and biopsies was developed. The method comprised the OCT cleaning, the protein solid-liquid extraction and the protein cleavage steps using ultrasonication. An ultrasonic bath is used for cleaning purposes, then an ultrasonic probe is used for the solid-liquid extraction of proteins from the solid biopsies. Finally, a microplate horn assembly is used to accelerate the cleavage of the extracted proteome. When compared with traditional cleaning, the ultrasonic based method (US bath 35 kHz at 100% of ultrasonic amplitude for 2 min ultrasonic duty time) delivered a higher number of identifiable proteins. This method can be adapted to a microplate,

making it possible to obtain the proteome of 96 biopsies in one day. This approach will allow high-throughput proteomic interrogation of retrospective biospecimen repositories which contain vast numbers of valuable tissues preserved in OCT. In addition, this proof of concept study suggests that our proteomic protocol may allow us to successfully distinguish chromophobe renal cell carcinoma from renal oncocytoma which is of potential clinical value in kidney cancer diagnosis and therapy.

ACKNOWLEDGMENTS

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura. Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV which is financed by national funds from FCT/MEC (UID/QUI/50006/ 2019). H. M. S. is funded by the FCT 2015 Investigator Program (IF/00007/2015). S. J. thanks FCT/MEC (Portugal) for her research contract as PhD student with the grant SFRH/BD/120537/2016. This project utilized the University of Pittsburgh Hillman Cancer Center shared resource facility (Cancer Genomics Facility) supported in part by award P30CA047904 (Dr. LaFramboise).

REFERENCES

[1] Duarte, TT *et al*., Personalized proteomics: the future of precision medicine. *Proteomes*, **2016**, *4*.

[2] Guo, T *et al*., Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.*, **2015**, *21*, 407–413.

[3] Wang, H *et al*., The clinical impact of recent advances in LC-MS for cancer biomarker discovery and verification. *Expert Rev. Proteomics*, **2016**, *13*, 99–114.

[4] Crutchfield, CA *et al*., Advances in mass spectrometry-based clinical biomarker discovery. *Clin. Proteomics*, **2016**, *13*.

[5] Thavarajah, R *et al*., Chemical and physical basics of routine formaldehyde fixation. *J. Oral Maxillofac. Pathol.*, **2012**, *16*, 400–405.

[6] Scicchitano, MS *et al*., Protein extraction of formalin-fixed, paraffin-embedded tissue enables robust proteomic profiles by mass spectrometry. *J. Histochem. Cytochem.*, **2009**, *57*, 849–860.

[7] Steiner, C *et al*., Applications of mass spectrometry for quantitative protein analysis in formalin-fixed paraffin-embedded tissues. *Proteomics*, **2014**, *14*, 441–451.

[8] Zhao, C *et al*., Removal of polyethylene glycols from protein samples using titanium dioxide. *Anal. Biochem.*, **2007**, *365*, 283–285.

[9] Nirmalan, NJ *et al*., Initial development and validation of a novel extraction method for quantitative mining of the formalin-fixed, paraffin-embedded tissue proteome for biomarker investigations. *J. Proteome Res.*, **2011**, *10*, 896–906.

[10] Strohalm, M *et al*., Poly[N -(2-hydroxypropyl)methacrylamide]-based tissue-embedding medium compatible with MALDI mass spectrometry imaging experiments. *Anal. Chem.*, **2011**, *83*, 5458–5462.

[11] Loken, SD *et al*., A novel method for freezing and storing research tissue bank specimens. *Hum. Pathol.*, **2005**, *36*, 977–980.

[12] Zhang, W *et al*., Comprehensive proteome analysis of fresh frozen and optimal cutting temperature (OCT) embedded primary non-small cell lung carcinoma by LC-MS/MS. *Methods*, **2015**, *81*, 50–55.

[13] Hood, BL *et al*., Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol. Cell. Proteomics*, **2005**, *4*, 1741–1753.

[14] Chaurand, P *et al*., Monitoring mouse prostate development by profiling and imaging mass spectrometry. *Mol. Cell. Proteomics*, **2008**, *7*, 411–423.

[15] Holfeld, A *et al*., Parallel proteomic workflow for mass spectrometric analysis of tissue

samples preserved by different methods. *Anal. Chem.*, **2018**, *90*, 5841–5849.

[16] Weston, LA *et al*., Comparative LC-MS/MS analysis of optimal cutting temperature (OCT) compound removal for the study of mammalian proteomes. *Analyst*, **2013**, *138*, 6380–6384.

[17] Vrana, M *et al*., An optimized method for protein extraction from OCT-embedded human kidney tissue for protein quantification by LC-MS/MS proteomics. *Drug Metab. Dispos.*, **2016**, *44*, 1692–1696.

[18] Zhao, X *et al*., Quantitative proteomic analysis of optimal cutting temperature (OCT) embedded core-needle biopsy of lung cancer. *J. Am. Soc. Mass Spectrom.*, **2017**, *28*, 2078–2089.

[19] Shah, P *et al*., Tissue proteomics using chemical immobilization and mass spectrometry. *Anal. Biochem.*, **2015**, *468*, 27–33.

[20] Tian, Y *et al*., Quantitative glycoproteomic analysis of optimal cutting temperature-embedded frozen tissues identifying glycoproteins associated with aggressive prostate cancer. *Anal. Chem.*, **2011**, *83*, 7013–7019.

[21] Araújo, JE *et al*., A journey through PROTEOSONICS. *Talanta*, **2014**, *121*, 71–80.

[22] Santos, HM *et al*., An improved clean sonoreactor-based method for protein identification by mass spectrometry-based techniques. *Talanta*, **2008**, *77*, 870–875.

[23] Vale, G *et al*., An assessment of the ultrasonic probe-based enhancement of protein cleavage with immobilized trypsin. *Proteomics*, **2011**, *11*, 3866–3876.

[24] Cordeiro, FM *et al*., Simplifying sample handling for protein identification by peptide mass fingerprint using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. mass Spectrom.*, **2007**, *21*, 3269–3278.

[25] Santos, HM *et al*., On-target ultrasonic digestion of proteins. *Proteomics*, **2013**, *13*, 1423–1427.

[26] Jorge, S *et al*., Unparalleled sample treatment throughput for proteomics workflows relying on ultrasonic energy. *Talanta*, **2018**, *178*, 1067–1076.

[27] Cox, J *et al*., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **2008**, *26*, 1367–1372.

[28] Tyanova, S *et al*., Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics*, **2015**, *15*, 1453–1456.

[29] Tyanova, S *et al*., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **2016**, *13*, 731–740.

[30] Fernandes, L *et al*., Ultrasonic energy as a tool in the sample treatment for polymer

characterization through matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Talanta*, **2008**, *77*, 882–888.

[31] Araújo, JE *et al*., A comprehensive factorial design study of variables affecting protein extraction from formalin-fixed kidney tissue samples. *Talanta*, **2014**, *119*, 90–97.

[32] Capelo-Martínez, J-L., The power of ultrasound. In *Ultrasound in chemistry : analytical applications*; Wiley-VCH, **2009**; pp. 1–6.

[33] Leong, T *et al*., The fundamentals of power ultrasound - a review. *Acoust. Aust.*, **2011**, *39*, 54–63.

[34] De Petris, L *et al*., A novel method for sample preparation of fresh lung cancer tissue for proteomics analysis by tumor cell enrichment and removal of blood contaminants. *Proteome Sci.*, **2010**, *8*..

CHAPTER IV.

Ultrasonic-assisted extraction and digestion of proteins from solid biopsies followed by peptide sequential extraction hyphenated to MALDI-based profiling holds the promise of distinguishing renal oncocytoma from chromophobe renal cell carcinoma

Susana Jorge, Kevin Pereira, Hugo López-Fernández, William Laframboise, Rajiv Dhir, Javier Fernández-Lodeiro, Carlos Lodeiro, Hugo M. Santos, José L. Capelo*

Published in:

Talanta, **2020**, 206, 120180

DOI: 10.1016/j.talanta.2019.120180

ABSTRACT

A novel analytical approach is proposed to discriminate between solid biopsies of chromophobe renal cell carcinoma (chRCC) and renal oncocytoma (RO). The method comprises the following steps: (i) ultrasonic extraction of proteins from solid biopsies, (ii) protein depletion with acetonitrile, (iii) ultrasonic assisted in-solution digestion using magnetic nanoparticle with immobilized trypsin, (iv) C18 tip-based preconcentration of peptides, (v) sequential extraction of the peptides with ACN, (vi) MALDI-snapshot of the extracts and (vii) investigation of the extract containing the most discriminating features using high resolution mass spectrometry. With this approach we have been able to differentially cluster renal oncocytoma and chromophobe renal cell carcinoma and identified 18 proteins specific to chromophobe and seven unique to renal oncocytoma. Chromophobes express proteins associated with ATP function (ATP5I & 5E; VATE1 & G2; ADT2), glycolysis (PGK1) and neuromedin whilst oncocytomas express ATP5H, ATPA, DEPD7 and TRIPB thyroid receptor interacting protein.

Keywords: Renal oncocytoma, chromophobe renal cell carcinoma, renal cancer, MALDI, profiling ESI, sequential extraction

IV.1 INTRODUCTION

Renal cell carcinoma is the 12th most frequently diagnosed cancer worldwide with 5% classified as chromophobe renal cell carcinoma (chRCC) which requires therapeutic radical or partial nephrectomy. Diagnosis of chRCC is complicated by morphological and histological features that overlap with renal oncocytoma (RO), a benign neoplasm that occurs at a similar frequency (5%) but has a positive prognosis and does not require aggressive treatment. For instance, staining for CD117, which is used to distinguish chRCC from other malignant subtypes of RCC, is helpless to discriminate chRCC of RO. As another example, when staining is done with cytokeratin 7 (CK7), a diffuse staining pattern in presented in chRCC with pale cells. Yet, in the eosinophilic variant of chRCC, CK7 staining is limited and very similar to RO [1–3]. Despite the efforts to find biomarkers to distinguish these two tumors, a unique pattern for each one remains to be undisclosed. Therefore, there is an urgent need for new methods that can effectively differentiate between these renal neoplasms.

In biomedical research, sample treatment is considered a bottleneck in analysis due the large number of steps needed to make the samples ready for analysis. To overcome this problem, diverse extraction techniques, including solid-phase extraction (SPE), liquid-liquid

extraction (LLE), or protein precipitation, have been used in sample preparation of biomolecules of interest. Moreover, extraction procedures are currently used in biomedical analysis and research to simplify the matrix, clean-up the sample and/or selectively enrichment of a given analyte [4–10]. Among these extraction procedures, on-a-tip SPE strategies has been widely applied in the proteomic field, specially C18 resin packed on-a-tip. Thus, the sequential extraction of peptides with C18-based tips has been used to reduce proteome complexity of biological samples through sequential extraction of peptides followed by mass spectrometry interrogation with the aim of finding biomarkers of diseases [11]. Taking into consideration the problems and solutions described above, a methodology to elucidate differences between chromophobe renal cell carcinoma and renal oncocytoma has been developed as follows. First, the sample treatment is optimized to reduce proteome complexity using ultrasonic assisted protein extraction from solid biopsies. The substrates then undergo (i) protein depletion with acetonitrile, ACN, (ii) ultrasonic assisted in-solution digestion using magnetic nanoparticle with immobilized trypsin, (iii) peptide preconcentration with C18-tips and (iv) peptide sequential extraction with ACN [12–14]. Finally, MALDI mass spectrometry was employed to provide a snapshot of the tumor proteomes revealing the most discriminative features specific to chRCC versus RO. The extracts were further investigated using nano-HPLC and high-resolution mass spectrometry (nano-LC-HR-MS). This tool may be extended to any disease presenting similar pathological profiles but with different outcomes. The method here presented can be extended to large cohort of samples using the high throughput provided by the 96-well plate-based ultrasonic approach [15], thus holding the promise of being used in routine medical assays.

IV.2 EXPERIMENTAL SECTION

IV.2.1 REAGENTS

All reagents used were HPLC or electrophoresis grade. Coomassie brilliant blue G-250, urea, albumin from bovine serum (BSA), Bradford reagent, iodoacetamide (IAA), ammonium dihydrogen-phosphate ($NH_4H_2PO_4$), N,N,N',N' – tetramethylethylene diamine (TEMED), and trichloroacetic acid (TCA) were purchased from Sigma-Aldrich (Basel, Switzerland). Ammonium bicarbonate (Ambic), α-cyano-4-hydroxy-cinnamic acid (α-CHCA) were purchased from Fluka (Basel, Switzerland). Trifluoroacetic acid (TFA) was purchased from Thermo Fischer Scientific (Waltham, MA, USA). Ammonium persulphate (APS) and tris base were purchased from NZYTech (Lisbon, Portugal). Acetonitrile (ACN) and formic acid (FA) were purchased from Carlo Erba Reagents (Val de Reuil, France). Dithiothreitol (DTT) and 4x Laemmli SDS sample buffer were purchased from Alfa Aesar (Karlsruhe, Germany). 10x tris/glycine/SDS running buffer and precision plus protein™ standards unstained were purchased from Bio-Rad (CA, USA). Pierce™ C18 tips, 100 µL bed were purchased from Thermo Fisher Scientific. Peptide calibration standard II from Bruker (Bremen, Germany) was used as a mass calibration standard for MALDI-TOF-MS measurements.

IV.2.2 MATERIAL

Protein digestion was done in Eppendorf safe-lock tubes of 0.5 mL volume (Hamburg, Germany). A vacuum concentrator centrifuge model UNIVAPO 150 ECH Speed Vac and a vacuum pump model UNIJET II (Munich, Germany) were used for sample drying and sample pre-concentration. A mini incubator from Labnet (New Jersey, USA) was used for protein reduction steps. Vortex models ELMI CM70M-09 SkyLine (Southern California, USA), and Prism™ R refrigerated microcentrifuge, VX-200 Lab vortex mixer, AccuBlock™ digital dry baths from Labnet (New Jersey, USA), were used throughout the sample treatment. CLARIOstar® high performance monochromator multimode microplate reader from BMG LABTECH (Germany) was used for Bradford assays. Mini-PROTEAN tetra cell and PowerPac™ basic power supply from Bio-Rad (CA, USA) was used for SDS-PAGE protein separation. Image gels were obtained using a ProPic II gel imaging (Digilab-Genomic Solutions, USA). An ultrasonic processor UP50H (50 W, 30 kHz, 1 mm diameter probe tip) from Hielscher Ultrasonics (Teltow, Germany) was used for tissue homogenization. An ultrasonic bath, model TI-H-5, from Elma (Singen, Germany) with control of temperature and amplitude was used to sample cleaning and enhance protein depletion, and a sonoreactor model UTR200 from Dr.Hielscher (Teltow, Germany) was used to accelerate enzymatic digestions. Acquisition of mass spectrometry data was done using an Ultraflex II MALDI-TOF/TOF and an UHR-QqTOF IMPACT HD from Bruker Daltonics (Bremen, Germany). Chromatographic separation of peptides was carried out using an Ultimate 3000 nLC nano-system equipped with a trap-column Acclaim PepMap100, 5 µm, 100 Å, 300 µm i.d. × 5 mm (Thermo Fisher Scientific) and an analytical column Acclaim™ PepMap™ 100 C18, 2µm, 0.075mm i.d x 150mm (Thermo Fisher Scientific).

IV.2.3 KIDNEY SAMPLES

The human kidney tissue samples were collected by the University of Pittsburgh Biospecimen Core's and the study was approved by the Institutional Review Board at the University of Pittsburgh (IRB # 02-077). All neoplasms contained a minimum of 90% tumor cells and NAT specimens were at least 90% normal cells. Data of patients enrolled in this study are summarized in Table IV.1.

Table IV.1. Description of human kidney biopsies used in the study.

BIOPSY	AGE	GENDER	DIAGNOSIS*	SAMPLE TYPE*
N1	50-59	Male	RCC	NAT
N2	40-49	Female	Papillary	NAT
N3	50-59	Female	RCC	NAT
N4	70-79	Female	RCC	NAT
N5	70-79	Male	RCC	NAT
C6	70-79	Male	RCC	chRCC
C7	60-69	Female	RCC	chRCC
C8	70-79	Male	RCC	chRCC
C9	50-59	Female	RCC	chRCC
C10	80-89	Male	RCC	chRCC
O11	80-89	Male	RCC	RO
O12	60-69	Female	RCC	RO
O13	60-69	Male	RCC	RO

**RCC: renal cell carcinoma; NAT: normal adjacent tissue; chRCC: chromophobe renal cell carcinoma; RO: renal oncocytoma.*

IV.2.4 EXTRACTING PROTEINS FROM KIDNEY BIOPSIES

The OCT-embedded tissues were treated as described in Jorge *et al.* [16]. Initially washed with 2 ml of 70% (v/v) ethanol at 4 ºC, submitted to ultrasound energy using an ultrasonic bath, 35 kHz, for 2 min, 100% amplitude, and centrifuged at 4 ºC for 2 min (5,000 g). The supernatants were carefully removed, and the procedure was repeated. Cell pellets were then washed with 2 ml of water at 4 ºC, sonicated again using an ultrasonic bath, 35 kHz for 2 min at 100% amplitude, and centrifuged at 4 ºC for 2 min (5,000 g). Then, the supernatant was carefully removed, and the above steps were performed five times for each pellet. After OCT cleaning, each pellet was placed in a mortar and ground to a powder in liquid nitrogen using a pestle. The resulting powder was extracted for protein in 8 M urea/ 25 mM Ambic buffer (ratio: 100 µL buffer to 10 mg of tissue) using an ultrasonic processor UP50H (50 w, 30 kHz, 1 mm diameter probe tip) operating at 50% amplitude for 2 min in a pulsed mode, 10 sec on/10 sec off. The samples were centrifuged at 10,000 g for 10 min and the supernatants were transferred to new tubes. The ultrasonic extraction procedure was repeated for the pellets and the second supernatant was combined with the first to produce the final protein extract. Each protein extract was precipitated using the DOC/TCA and acetone method. Briefly, to each 300 µL of protein extract, 3 µL of 2% DOC were added and left on ice for 20 min, then 75 µL of 100% TCA were added to the mixture and the samples were left on ice for 20 min, followed by centrifugation at 4 ºC for 20 min (16,000 g). The supernatant was removed, and the pellets were washed with 200 mL of ice-cold acetone (-20 ºC), followed by centrifugation (16,000 g for 20 min at 4 °C). Then,

20 µL of 0.2 M NaOH were added to the protein pellet, and after 2 min at room temperature, 80 µL of 6 M urea in 25 mM Ambic were added. The proteins were solubilized using four cycles of 10 sec ultrasound energy through an ultrasonic processor UP50H (50 w, 30 kHz, 1 mm diameter probe tip) operating at 50% amplitude. Finally, total protein content was determined using a Bradford protein assay. This extract was used further as described in the following sections.

IV.2.5 ACN-BASED PROTEIN DEPLETION

Protein depletion was assayed using extracts containing a total of 250 µg of proteins and two different ACN concentrations (v/v): (i) 20% or (ii) 45%. Each extract was prepared in triplicate. Then, samples were sonicated using an ultrasonic bath (35 kHz, 100%, 20 min) and the pellet formation was then observed after centrifugation at 14,000 g for 10 min, thus allowing the separation of the ACN-based precipitated proteins. Thus, two fractions for further study were obtained: the supernatant (SN) and the pellets. The SN fractions were evaporated to dryness.

IV.2.6 SDS-PAGE

To perform the electrophoresis, dried SN samples were resuspended in 55,5 µL of Milli-Q H_2O plus 18.5 µL of 4x Laemmli SDS sample buffer, whilst the pellet was dissolved in 1.9 µL of 4x Laemmli SDS sample buffer plus 5.7 µL of Milli-Q water. Five µL of each sample were loaded on a 4% acrylamide/bis-acrylamide stacking gel and 12% acrylamide/bis-acrylamide running gel at 1mm of thickness. Additionally, 3 µL of molecular weight marker were also loaded. The gels were run at 200 V, and 400 mA during 50 min and then stained overnight with colloidal Coomassie blue. After staining, the gels were washed with Milli-Q water until a clear background was achieved. Gel imaging was carried out with a ProPic II-robot using 14 ms of exposure time and a resolution of 70 µm.

IV.2.7 PROTEIN QUANTIFICATION

Prior to quantification, samples were resuspended in equal amounts of Milli-Q water (184 µL) and then quantified using the Bradford protein assay. Briefly, a BSA standard curve (0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4 µg/µL) was generated in duplicate. In duplicate wells, 5 µL of samples were mixed with 250 µL of Bradford reagent. The unknown samples were diluted with water to an approximate concentration between 0.4 and 1 µg/µL, then 5 µL of each unknown were mixed with 250 µL Bradford reagent. Finally, the samples were incubated at room temperature for 20 min, and the absorbances measured at 590 nm through a high-performance monochromator multimode microplate reader.

IV.2.8 REDUCTION AND ALKYLATION

IV.2.8.1 Supernatant

Supernatant fractions were prepared from 60 µg of extract in a volume of 88 µL so that the SN could later be split into 6 extracts. The pH of the samples was adjusted with 1 µL of 1 M Ambic, to a final concentration of 12.5 mM. Then, proteins were reduced with 9 µL of DTT (110 mM in 12.5 mM Ambic) and incubated at 37 ºC for 1 hour. The resulting cysteines were blocked with 9 µL of IAA (400 mM in 12.5 mM Ambic). Since iodoacetamide is light-sensitive, all tubes were kept in darkness for 45 min at room temperature. Free IAA was inactivated by adding 3 µL of DTT (110 mM in 12.5 mM Ambic) to each sample.

IV.2.8.2 Pellet

The pellet fractions were also reduced and alkylated. Ten µg of total protein were prepared in a final volume 115 µL and 1.5 µL of 1 M Ambic was added to adjust the pH to a final concentration of 12.5 mM. Proteins were reduced with 12 µL of DTT (110 mM in 12.5 mM Ambic) and incubated at 37 ºC for 1 hour. The resulting cysteines were then blocked with 12 µL of IAA (400 mM in 12.5 mM Ambic) in the dark for 45 min at room temperature. Finally, free IAA was inactivated by adding 4 µL of DTT (110 mM in 12.5 mM Ambic) to each sample.

IV.2.9 IN-SOLUTION DIGESTION OF SUPERNATANTS AND PELLETS

In-solution digestion was performed using homemade magnetic nanoparticles with immobilized trypsin [17–19]. 20 µL of 3 mg/mL of immobilized trypsin were added to each sample, and the samples were digested using an ultrasonic sonoreactor device at 50% amplitude for 2 x 2.5 min with temperature constant at 20 ºC. A magnet was subsequently used to immobilize the trypsin magnetic beads. The supernatant was transferred to new microtubes and evaporated to dryness.

IV.2.10 PEPTIDE SEQUENTIAL ELUTION

The samples were resuspended after digestion in 50 µL of 0.1% (v/v) TFA and then loaded onto Pierce™ C18 tips with a 100 µL bed for rapid sample desalting and concentrating peptides. First, C18 tips had to be activated by aspirating and dispensing five cycles of 50 µL of an 80% (v/v) ACN + 0,1% (v/v) TFA solution, and then 50 µL of a 0.1% (v/v) TFA solution (3 cycles) before sample pipetting. The peptides were retained in the C18 tips by aspirating and dispensing the samples 20 cycles. Finally, the C18 tips were washed with 50 µL of 0.1% (v/v) TFA twice to remove the salt content, and then were sequentially eluted with 50 µL of different ACN concentrations (4%, 7%, 10%, 14%, 35% and 60%) by aspirating and dispensing each

concentration 15 cycles, from the lowest to highest concentration. All extracts were then evaporated to dryness.

IV.2.11 MASS SPECTROMETRY

Samples were resuspended in 10 µL of 0.3% formic acid (v/v) for MS analysis and 0.5 µL of the sample was hand-spotted, in quintuplicate, onto a MALDI target plate. 1 µL of a matrix solution of 7 mg α-cyano-4-hydroxycinnamic acid was dissolved in 90 mM $NH_4H_2PO_4$, 50% (v/v) ACN and 0.1% (v/v) TFA was added and allowed to air dry. The MALDI TOF/TOF mass spectrometer was operated in positive ion mode using a reflectron, and spectra were acquired in the m/z range of 600–3500. A total of 500 spectra were acquired for each sample at a laser frequency of 50 Hz. External calibration was performed with the $[M + H]^+$ monoisotopic peaks of bradykinin 1–7 (m/z 757.3992), angiotensin II (m/z 1046.5418), angiotensin I (m/z 1296.6848), substance P (m/z 1758.9326), ACTH clip 1–17 (m/z 2093.0862), ACTH18– 39 (m/z 2465.1983) and somatostatin 28 (m/z 3147.4710).

IV.2.12 HIERARCHICAL CLUSTERING ANALYSIS

The corresponding raw data spectrum of each sample generated by the MS analysis was pre-processed with the Mass-Up v1.0.9 open source program (http://sing. ei.uvigo.es/mass-up/) [20] using the following parameters: (i) intensity transformation (squareroot), (ii) smoothing (none), (iii) baseline correction (snip), (iv) standardization (total ion current), (v) peak detection (MALDIquant: SNR (3), half window size (60) and (vi) minimum peak intensity (0.001). Peaks were matched with the following parameters: (i) intra-sample matching (MALDIquant: tolerance (0.002)), selecting the “generate consensus spectrum” box with a percentage of presence of 60%, (ii) inter-sample matching (MALDIquant: tolerance (0.002)). Then, an agglomerative, hierarchical clustering analysis was executed with the following parameters: (i) minimum variance (0.1), (ii) peak list (NULL > for no peak filtering), (iii) cluster reference value (average), (iv) distance function (hamming), (v) conversion values (presence), (vi) intra-sample minimum presence (0), (vii) deep clustering (No).

IV.2.13 NANO-LC-HR-MS/MS ANALYSIS

The nanoLC-HR-MS/MS analysis was carried out using an Ultimate 3000 nLC nano-system coupled to an UHR-QqTOF IMPACT HD (Bruker Daltonics) with a CaptiveSpray ion source (Bruker Daltonics). All samples were reconstituted in 50 µL of 3% ACN/0.1% (v/v) aqueous formic acid. 5 µL of peptides were loaded into a trap column Acclaim PepMap100, 5 µm, 100 Å, 300 µm i.d. × 5 mm and desalted for 5 min with 3% of mobile phase B (B: 90% ACN 0.1% FA) at a flow rate of 15 µL/min. Chromatographic separation was carried out using an analytical column Acclaim™ PepMap™ 100 C18, 2 µm, 0.075 mm i.d x 150 mm with a linear gradient at 300 nL/min (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v)

ACN and 0.08% (v/v) FA), 0-60 min from 3% to 35% of mobile phase B, 60-70 min linear gradient from 35% to 95% of mobile phase B, 70-80 min 95% B. Total run time was 100 min. For each sample, two replicate injections were performed. Chromatographic separation was carried out at 35 ºC. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8–32 Hz), cycle time 3.0 seconds, with active exclusion (precursors were excluded from precursor selection for 0.5 min after acquisition of 1 MS/MS spectrum, intensity threshold for fragmentation of 2500 counts). Together with active exclusion set to 1, reconsider precursor if the intensity of a precursor increases by a factor of 3, this mass will be taken from temporarily exclusion list and fragmented again, ensuring that fragment spectra were taken near to the peak maximum. All spectra were acquired in the range 150–2200 m/z. Raw data were processed in DataAnalysis 4.2 and subsequently exported to Protein-Scape 4.0 for automated protein identification. For protein identification, CID-MS2 spectra were first searched against the *H. sapiens* (20,266 sequences) subset of the SwissProt database 57.15 (515,203 sequences; 181,334,896 residues), using the Mascot search engine (V. 2.3.02) with the following parameters: (i) two missed cleavage; (ii) fixed modifications: carbamidomethylation (C); (iii) variable modifications: oxidation of methionine, Acetyl (Protein N-term), Glu- > pyro-Glu (N-term E), Gln- > pyro-Glu (N- term Q), (vi) peptide mass tolerance up to 20 ppm, (v) fragment mass tolerance 0.05 Da (vi) Adjust FDR 1%.

IV.3 RESULTS AND DISCUSSION

Figure IV.1 shows the comprehensive scheme used for the work presented herein.

IV.3.1 OPTIMIZATION OF ACN CONCENTRATION TO SIMPLIFY THE PROTEOME

We have previously shown that plasma samples depleted using ACN a supernatant reach in apolipoproteins is obtained [12]. The concentration of proteins of high molecular weight remaining in the supernatant decreases dramatically as the ACN concentration is increased. Conversely, the amount of high molecular weight proteins in the pellet increases. Because the samples used in this case were not plasma, but extracts obtained from tissue biopsies, we first performed a set of experiments to assess the effects of ACN on protein depletion. The ACN concentrations selected were 20% (v/v) and 45% (v/v) based on our experience with this sample treatment [12,21]. As can be seen in Figure IV.2A, protein depletion with 45% (v/v) ACN concentration renders a pellet with higher concentration of proteins reflecting their depletion from the supernatant. We have previously established that clustering from the supernatant using MALDI-based mass spectrometry protein analysis is optimized after depletion of high molecular weight proteins plasma samples using ACN [12]. Therefore, we chose the ACN concentration of 45% (v/v) for further experiments.

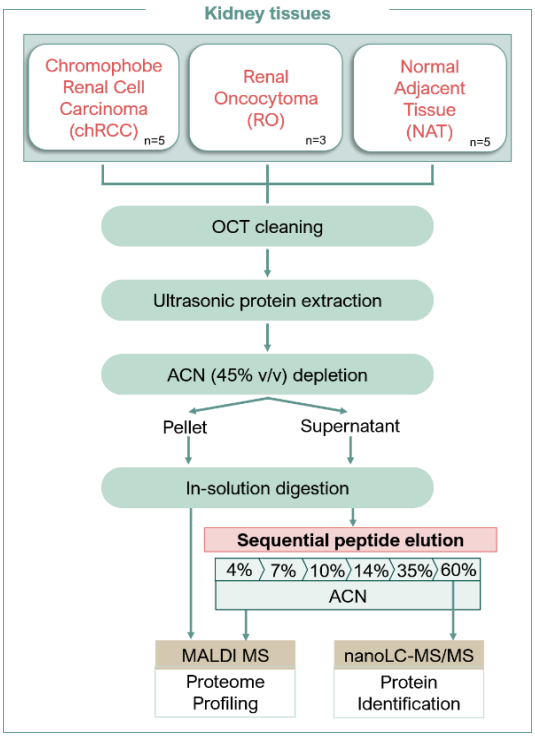

Figure IV.1 – The solid biopsies are cleaned with the aid of an ultrasonic bath. Then, the proteins are solid-liquid extracted with the aid of an ultrasonic probe. The proteins are depleted with ACN 45% v/v, and thus one supernatant and one pellet are obtained. The supernatant is withdrawn, then evaporated to dryness and then resuspended in 12.5 mM Ambic. The pellets were dissolved 12.5 mM Ambic. Proteins contained in supernatant and pellet were further reduced, alkylated and digested with the aid of ultrasonic energy. The peptides from the supernatant were then sequentially extracted using C18 tips and profiled using MALDI-based mass spectrometry. The best ACN fraction, 60% v/v was further interrogated using ESI-based mass spectrometry.

IV.3.2 SEQUENTIAL ELUTION OF PEPTIDES

The sequential extraction of peptides using C18-based tips in conjunction with mass spectrometry has been previously described as a method to discriminate among large cohorts of samples [11,14]. We employed this approach on the (i) supernatants and (ii) pellets after depletion with 45% (v/v) ACN from (i) chromophobe renal cell carcinoma (chRCC) or (ii) renal oncocytoma (RO) or (iii) normal adjacent tissue (NAT). The proteomes were first digested with trypsin and then the peptides were up-loaded in C18 tips. The peptides were sequentially eluted with ACN solutions (% v/v: 4, 7, 10, 14, 35 and 60) based on our previous work with plasma [11].

Protein quantification of supernatants and pellets obtained after depletion with ACN 45% (v/v) concentration revealed that as much as 73% of the protein content remained in the supernatant. The total protein content of some pellets was too low for the sequential extraction procedure (see Figure IV.2B). For example, sequential extractions of some digested pellets produced poor MALDI spectra, e.g. low intensity and few m/z signals at 35% (v/v) ACN (see Figure IV.3A). However, a substantial peptide spectrum was obtained from the correlative

supernatants after digestion as shown in Figure IV.3B. Therefore, we decided to profile through a MALDI-MS approach the supernatants using the sequential extraction approach while interrogating the complete protein content of the pellets.

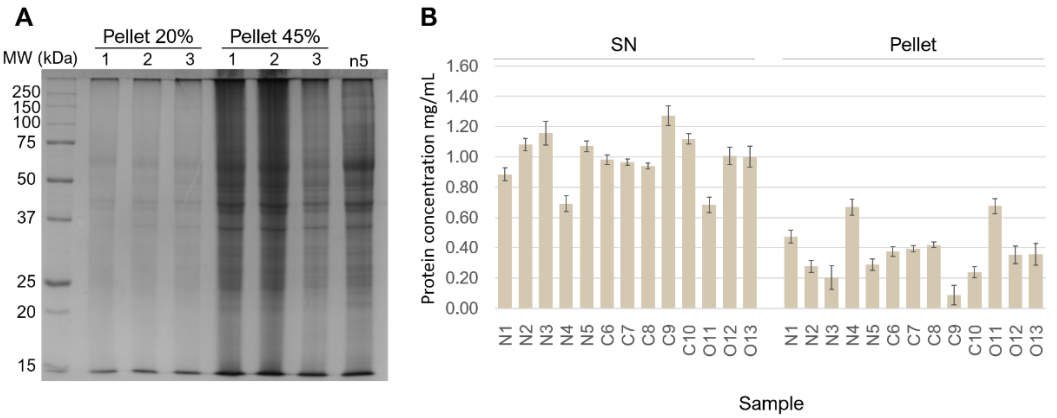

Figure IV.2 – (A): Influence of ACN concentration (v/v) on depletion of protein extracts from solid biopsies. SDS-PAGE of pellet fraction after protein depletion with (i) 20% and (ii) 45% (v/v) of ACN. Depletion was performed in triplicate (3 lanes for each ACN concentration) and N represents a crude protein extract without depletion. (B) Protein content of pellets and supernatants after depletion with 45 % (v/v) ACN.

Quintuple MALDI spectra were obtained for each fraction. The cohorts of m/z signals were used to analyze the samples by unsupervised clustering. The results showed that classification of the chRCC, RO and NAT samples clustered together was not possible at any of the ACN extracting concentrations tested as shown in Figure IV.4A. The best classification was obtained for the oncocytoma samples that were grouped using the cohorts of m/z signals obtained for the ACN concentrations (v/v) of 10% and 60%. It is noteworthy that the clustering using the pellets protein content of the pellets also failed in classifying the samples as may be seen in Figure IV.4A. However, comparing only the oncocytoma and chromophobe biopsies without NAT biopsies showed excellent discrimination using the cohorts of m/z signals obtained for the ACN extraction at a concentration of 60% (v/v) (see Figure IV.4B, panel 7).

IV.3.3 Revealing the protein content of the optimal extraction fraction

These data indicated that the pool of peptides obtained with 60% ACN (v/v) concentration was most informative in distinguishing between the renal tumor classifications. The entire preparation protocol was repeated using the original samples to obtain peptides from the ACN 60% fraction for analysis in the nano-LC-HR-MS. Eighteen proteins specific to chRCC and seven unique to RO were found based on this approach (see Figure IV.5).

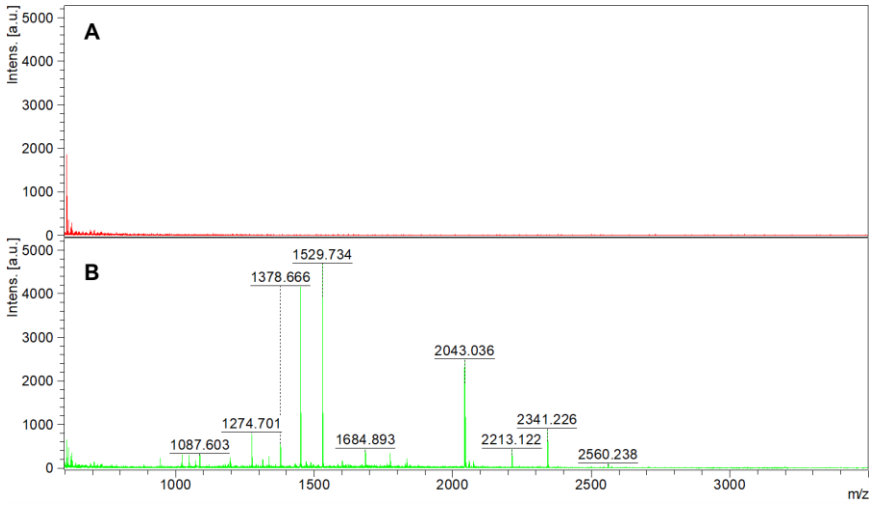

Figure IV.3 – Representative MALDI-MS spectra obtained for the 35% (v/v) ACN sequential fraction from (A) pellet and (B) supernatant. Note that the ACN 35% (v/v) pellet extract does not contain peptides.

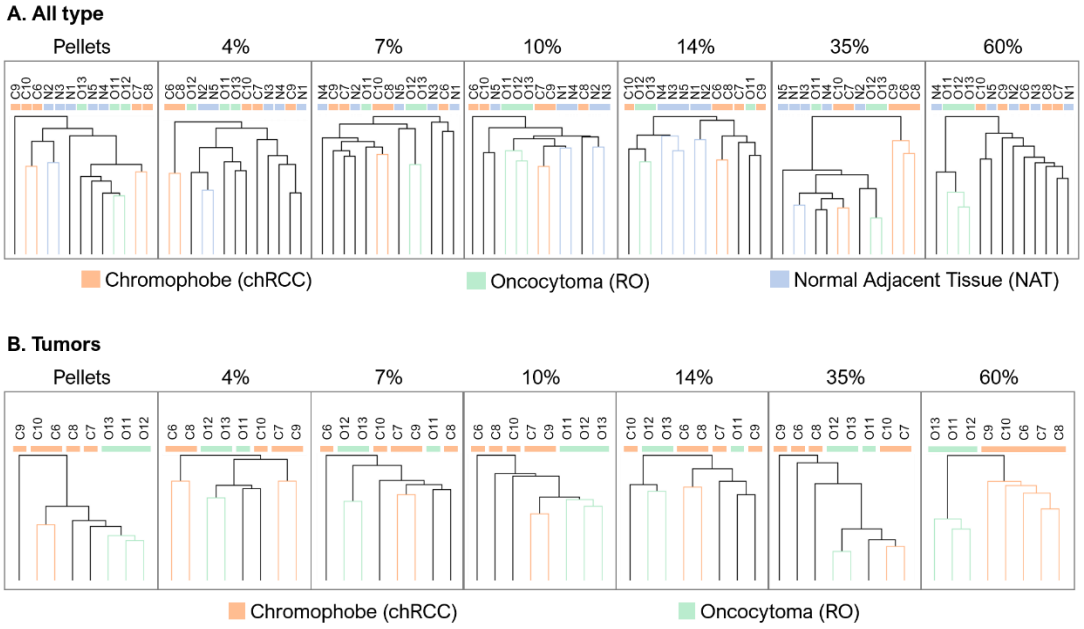

Figure IV.4 – Unsupervised clustering analysis of MALDI-based mass spectrometry data obtained for peptides from (i) the digested pellets and (ii) sequentially eluted fractions from the digested supernatants using C18 tips and 4%, 7%, 10%, 14%, 35% and 60% (v/v) of ACN. Clusters obtained (A) using the three types of solid biopsies, chromophobe, oncocytoma and NAT and (B) using only data from chromophobe and oncocytoma. For the latest case note the 60% ACN (v/v) concentration.

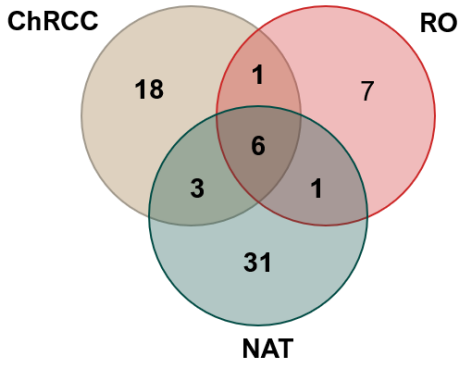

Figure IV.5 – Venn diagram (n_{chRCC}=5, n_{RO}=3, n_{NAT}=5) showing the number of proteins identified for the 60% ACN supernatant fraction by nano-LC-HR-MS/MS. Proteins are listed in Figure IV.6.

The chromophobe tumors expressed 5 proteins associated with canonical ATP function (ATP5I & 5E; VATE1 & G2; ADT2) and one classical glycolytic enzyme (PGK1). In addition, neuromedin was detected specific to the chromophobe tumors. It is interesting to note that changes in the function or expression of these molecules has been previously reported in renal [22–24], pancreatic [25,26] and lung [25] cancers. Oncocytomas expressed ATP5H and ATPA, DEPD7 and TRIPB (thyroid receptor interacting protein). It is noteworthy that both, oncocytoma and chromophobe are characterized by an eosinophilic cytoplasm which contains excessive amounts of mitochondria [27–31]. In the case of the chromophobe pathology, 5 of the proteins we detected were related to mitochondrial function while two ATP-related proteins were found specific to the oncocytomas. The complete list of proteins is presented in Figure IV.6.

Accession	Protein	MW [kDa]	pI	ChRCC	RO	NAT
H2B1K_HUMAN	Histone H2B type 1-K	13.9	10.3			
H4_HUMAN	Histone H4	11.4	11.4			
HBA_HUMAN	Hemoglobin subunit alpha	15.2	8.7			
HBB_HUMAN	Hemoglobin subunit beta	16.0	6.7			
HSPB1_HUMAN	Heat shock protein beta-1	22.8	6.0			
VIME_HUMAN	Vimentin	53.6	5.1			
ACTB_HUMAN	Actin, cytoplasmic 1	41.7	5.3			
COF1_HUMAN	Cofilin-1	18.5	8.2			
ENOA_HUMAN	Alpha-enolase	47.1	7.0			
KCRB_HUMAN	Creatine kinase B-type	42.6	5.3			
VTNC_HUMAN	Vitronectin	54.3	5.6			
ACTG_HUMAN	Actin, cytoplasmic 2	41.8	5.3			
ADT2_HUMAN	ADP/ATP translocase 2	32.9	9.8			
AK1A1_HUMAN	Alcohol dehydrogenase [NADP+]	36.5	6.3			
AL4A1_HUMAN	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	61.7	8.2			
ALBU_HUMAN	Serum albumin	69.3	5.9			
ALDOB_HUMAN	Fructose-bisphosphate aldolase B	39.4	8.0			
ARHGC_HUMAN	Rho guanine nucleotide exchange factor 12	173.1	5.5			
ATP5E_HUMAN	ATP synthase subunit epsilon, mitochondrial	5.8	9.9			
ATP5H_HUMAN	ATP synthase subunit d, mitochondrial	18.5	5.2			
ATP5I_HUMAN	ATP synthase subunit e, mitochondrial	7.9	9.3			
ATPA_HUMAN	ATP synthase subunit alpha, mitochondrial	59.7	9.2			
BTBD3_HUMAN	BTB/POZ domain-containing protein 3	58.4	7.4			
CALD1_HUMAN	Caldesmon	93.2	5.6			
CATB_HUMAN	Cathepsin B	37.8	5.9			
CH60_HUMAN	60 kDa heat shock protein, mitochondrial	61.0	5.7			
CND3_HUMAN	Condensin complex subunit 3	114.3	5.4			
DEN4B_HUMAN	DENN domain-containing protein 4B	163.7	6.9			
DEPD7_HUMAN	DEP domain-containing protein 7	58.3	7.6			
EZRI_HUMAN	Ezrin	69.4	5.9			
FA5_HUMAN	Coagulation factor V	251.5	5.7			
GP123_HUMAN	Probable G-protein coupled receptor 123	137.1	9.2			
H12_HUMAN	Histone H1.2	21.4	10.9			
H14_HUMAN	Histone H1.4	21.9	11.0			
H2A1H_HUMAN	Histone H2A type 1-H	13.9	10.9			
H2AV_HUMAN	Histone H2A.V	13.5	10.6			
HPS5_HUMAN	Hermansky-Pudlak syndrome 5 protein	127.4	5.4			
IF122_HUMAN	Intraflagellar transport protein 122 homolog	141.7	6.1			
IGKC_HUMAN	Ig kappa chain C region	11.6	5.6			
K1C18_HUMAN	Keratin, type I cytoskeletal 18	48.0	5.3			
M2OM_HUMAN	Mitochondrial 2-oxoglutarate/malate carrier protein	34.0	9.9			
MAP4_HUMAN	Microtubule-associated protein 4	120.9	5.3			
MISSL_HUMAN	MAPK-interacting and spindle-stabilizing protein-like	24.3	5.3			
MLL1_HUMAN	Histone-lysine N-methyltransferase MLL	431.5	9.2			
MSH2_HUMAN	DNA mismatch repair protein Msh2	104.7	5.6			
NHRF1_HUMAN	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	38.8	5.6			
NMU_HUMAN	Neuromedin-U	19.7	9.1			
NOL11_HUMAN	Nucleolar protein 11	81.1	5.7			
NOTC2_HUMAN	Neurogenic locus notch homolog protein 2	265.2	5.0			
PEBP1_HUMAN	Phosphatidylethanolamine-binding protein 1	21.0	7.0			
PGK1_HUMAN	Phosphoglycerate kinase 1	44.6	8.3			
PPIA_HUMAN	Peptidyl-prolyl cis-trans isomerase A	18.0	7.7			
PSMD2_HUMAN	26S proteasome non-ATPase regulatory subunit 2	100.1	5.1			
PTMA_HUMAN	Prothymosin alpha	12.2	3.7			
PTMS_HUMAN	Parathymosin	11.5	4.1			
QCR6_HUMAN	Cytochrome b-c1 complex subunit 6, mitochondrial	10.7	4.4			
RADIL_HUMAN	Ras-associating and dilute domain-containing protein	117.4	6.7			
SMBP2_HUMAN	DNA-binding protein SMUBP-2	109.1	9.1			
TARA_HUMAN	TRIO and F-actin-binding protein	261.2	8.9			
TBA1C_HUMAN	Tubulin alpha-1C chain	49.9	5.0			
TPM4_HUMAN	Tropomyosin alpha-4 chain	28.5	4.7			
TRIPB_HUMAN	Thyroid receptor-interacting protein 11	227.5	5.2			
TSTD1_HUMAN	Thiosulfate sulfurtransferase/rhodanese-like domain-containing protein 1	12.5	5.9			
USMG5_HUMAN	Up-regulated during skeletal muscle growth protein 5	6.5	9.8			
VATE1_HUMAN	V-type proton ATPase subunit E 1	26.1	7.7			
VATG2_HUMAN	V-type proton ATPase subunit G 2	13.6	10.3			
VDAC1_HUMAN	Voltage-dependent anion-selective channel protein 1	30.8	8.6			

Figure IV.6 – List of proteins identified for the chRCC, RO and NAT solid biopsies in the fraction eluted from the C18 tips using the ACN 60% (v/v) solution. Grey colored boxes indicate the presence of the protein.

IV.4 CONCLUSION

We have developed a method using ultrasonic energy in conjunction with fast C18 tips sequential extraction and MALDI-snapshotting that is capable of delineating the complex proteomes from solid biopsies of chromophobe renal cell carcinoma and renal oncocytomas. It is also possible to obtain unsupervised clustering of these solid biopsies with optimal discrimination of the chromophobe and oncocytoma specimens using the 60% ACN extraction procedure. This approach yielded 18 proteins unique to chromophobe renal cell carcinoma versus 7 unique to oncocytoma, from a total of 67 proteins found in the fraction of interest. The proposed method is fast, simple and low-cost, and therefore ideal for efficient analysis of solid biopsies. The most informative extracts can be easily identified for further analysis by HR-MS to obtain deep knowledge of a large number of individual features. The method proposed last 14 hours for one sample, from the beginning of the sample treatment to the MALDI analysis. However, with the high throughput provided by the 96-well plate-based ultrasonic approach [15], the total time needed would be reduced from 14 h per sample to 96 samples in 14 h, this is approx. 9 min per sample. This approach holds the promise of discriminating between an indolent renal neoplasms and aggressive kidney tumors that are otherwise difficult to classify using current immunocytochemical methods and could have an impact on the diagnosis and therapy of these patients.

ACKNOWLEDGMENTS

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura. Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV which is financed by national funds from FCT/MEC (UID/QUI/50006/ 2019). H. M. S. is funded by the FCT 2015 Investigator Program (IF/00007/2015). S. J. thanks FCT/MEC (Portugal) for her research contract as PhD student with the grant SFRH/BD/120537/2016. J.F.L. Acknowledges FCT/MEC (Portugal) SFRH/BPD/93982/2013 and FCT-UNL for the DL57/2016 Assistant Researcher Contract. H. López-Fernández is supported by a post-doctoral fellowship from Xunta de Galicia (ED481B 2016/068-0). This project utilized the University of Pittsburgh Hillman Cancer Center shared resource facility (Cancer Genomics Facility) supported in part by award P30CA047904 (Dr. LaFramboise).

REFERENCES

[1] Farber, NJ *et al*., Renal cell carcinoma: the search for a reliable biomarker. *Transl. Cancer Res.*, **2017**, *6*, 620–632.

[2] Jain, S *et al*., Amylase α-1A (AMY1A): A novel immunohistochemical marker to differentiate chromophobe renal cell carcinoma from benign oncocytoma. *Am. J. Surg. Pathol.*, **2013**, *37*, 1824–1830.

[3] Badowska-Kozakiewicz, AM *et al*., Selected tumor markers in the routine diagnosis of chromophobe renal cell carcinoma. *Arch. Med. Sci.*, **2016**, *12*, 856–863.

[4] Huang, S *et al*., Solid-phase microextraction: An appealing alternative for the determination of endogenous substances - A review. *Anal. Chim. Acta*, **2019**.

[5] Venson, R *et al*., A review of the application of hollow-fiber liquid-phase microextraction in bioanalytical methods – A systematic approach with focus on forensic toxicology. *J. Chromatogr. B*, **2019**, *1108*, 32–53.

[6] Tabani, H *et al*., Recent developments in green membrane-based extraction techniques for pharmaceutical and biomedical analysis. *J. Pharm. Biomed. Anal.*, **2018**, *160*, 244–267.

[7] Alexovič, M *et al*., Achievements in robotic automation of solvent extraction and related approaches for bioanalysis of pharmaceuticals. *J. Chromatogr. B*, **2018**, *1092*, 402–421.

[8] Ahmadi, M *et al*., Nanomaterials as sorbents for sample preparation in bioanalysis: A review. *Anal. Chim. Acta*, **2017**, *958*, 1–21.

[9] Nazario, CED *et al*., New materials for sample preparation techniques in bioanalysis. *J. Chromatogr. B*, **2017**, *1043*, 81–95.

[10] Ocaña-González, JA *et al*., New developments in microextraction techniques in bioanalysis. A review. *Anal. Chim. Acta*, **2016**, *905*, 8–23.

[11] Fernández-costa, C *et al*., Sequential depletion coupled to C18 sequential extraction as a rapid tool for human serum multiple profiling. *Talanta*, **2014**, *125*, 189–195.

[12] Fernández, C *et al*., A comparison of depletion versus equalization for reducing high-abundance proteins in human serum. *Electrophoresis*, **2011**, *32*, 2966–2974.

[13] Fernández-Costa, C *et al*., Sequential depletion of human serum for the search of osteoarthritis biomarkers. *Proteome Sci.*, **2012**, *10*, 1–12.

[14] Rappsilber, J *et al*., Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.*, **2003**, *75*, 663–670.

[15] Jorge, S *et al*., Unparalleled sample treatment throughput for proteomics workflows

relying on ultrasonic energy. *Talanta*, **2018**, *178*, 1067–1076.

[16] Jorge, S *et al*., Development of a Robust Ultrasonic-Based Sample Treatment To Unravel the Proteome of OCT-Embedded Solid Tumor Biopsies. *J. Proteome Res.*, **2019**, *18*, 2979–2986.

[17] Martins, G *et al*., Label-free protein quantification after ultrafast digestion of complex proteomes using ultrasonic energy and immobilized-trypsin magnetic nanoparticles. *Talanta*, **2019**, *196*, 262–270.

[18] Atacan, K *et al*., Efficient protein digestion using immobilized trypsin onto tannin modified Fe3O4magnetic nanoparticles. *Colloids Surfaces B Biointerfaces*, **2017**, *156*, 9–18.

[19] Zhang, L *et al*., Recyclable trypsin immobilized magnetic nanoparticles based on hydrophilic polyethylenimine modification and their proteolytic characteristics. *Anal. Methods*, **2018**, *10*, 459–466.

[20] Lopez-Fernandez, H *et al*., Mass-Up: an all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery. *BMC Bioinformatics*, **2015**, *16*.

[21] Prates, J *et al*., Talanta Modulating the protein content of complex proteomes using acetonitrile. *Talanta*, **2018**, *182*, 333–339.

[22] Solaini, G *et al*., Oxidative phosphorylation in cancer cells. *Biochim. Biophys. Acta - Bioenerg.*, **2011**, *1807*, 534–542.

[23] Chevrollier, A *et al*., Adenine nucleotide translocase 2 is a key mitochondrial protein in cancer metabolism. *Biochim. Biophys. Acta - Bioenerg.*, **2011**, *1807*, 562–567.

[24] Harten, SK *et al*., Inactivation of the von Hippel-Lindau tumour suppressor gene induces Neuromedin U expression in renal cancer cells. *Mol. Cancer*, **2011**, *10*, 1–7.

[25] Stransky, L *et al*., The Function of V-ATPases in Cancer. *Physiol. Rev.*, **2016**, *96*, 1071–1091.

[26] Hwang, T-L *et al*., Overexpression and elevated serum levels of phosphoglycerate kinase 1 in pancreatic ductal adenocarcinoma. *Proteomics*, **2006**, *6*, 2259–2272.

[27] Joshi, S *et al*., The Genomic Landscape of Renal Oncocytoma Identifies a Metabolic Barrier to Tumorigenesis. *Cell Rep.*, **2015**, *13*, 1895–1908.

[28] Gasparre, G *et al*., Learning from oncocytic tumors: Why choose inefficient mitochondria? *Biochim. Biophys. Acta - Bioenerg.*, **2011**, *1807*, 633–642.

[29] Rathmell, KW *et al*., Genomics of chromophobe renal cell carcinoma: implications from a rare tumor for pan-cancer studies. *Oncoscience*, **2015**, *2*, 81–90.

[30] Yun, SH *et al*., Antibiotic treatment modulates protein components of cytotoxic outer membrane vesicles of multidrug-resistant clinical strain, Acinetobacter baumannii DU202. *Clinical Proteomics*, **2018**, *15*.

[31] Lee, W., Imprint cytology of the chromophobe renal cell carcinoma: Correlation with the histological and ultrastructural features. *J. Cytol.*, **2011**, *28*, 77–80.

CHAPTER V.

The proteome of tumor biopsies as a tool to distinguish Chromophobe Renal Cell Carcinoma and Renal Oncocytoma

Susana Jorge, José L. Capelo, William LaFramboise, Rajiv Dhir, Carlos Lodeiro, Hugo M. Santos*

In preparation

ABSTRACT

Background: The anatomical and histological similarities of malignant chromophobe renal cell carcinoma (chRCC) and benign renal oncocytoma (RO) compromise the accuracy of their diagnosis. To understand functional and molecular consequences of biological aberrations and establish a potential discriminative protein panel of these two tumour types, we present a label-free quantitative proteomic approach to discriminate clinical tissue biopsies diagnosed as chRCC and RO versus normal adjacent tissue (NAT).

Methods: The label-free quantitative proteomic analysis was performed by comparing protein abundances of frozen OCT-embedded human renal tissue biopsies diagnosed with chRCC (n = 5) and RO (n = 5). NAT specimens (n = 5) were used as control.

Results: Proteomic analysis of these two tumour types revealed a common dysregulation of biochemical pathways involving energy metabolism and mitochondrial activity. Mitochondrial pathways were dominant in RO, while phagosome maturation displayed greater representation in chRCC. A panel of 109 proteins was used to discriminate between chRCC and RO and both from NAT.

Conclusions: Two different approaches are proposed to discriminate between chRCC, and RO based on protein expression profiles. One approach utilizes clustering of the entire proteome, and the other employs a selected panel of proteins with significant power of differentiation.

Keywords: OCT-embedded tissues; label-free quantification; mass spectrometry; chromophobe renal cell carcinoma; renal oncocytoma

V.1 INTRODUCTION

Renal cell carcinoma (RCC) is a disease typically involving abnormal cell growth in the epithelial cells of the proximal convoluted tubules of the kidney (clear cell renal cell carcinoma) which accounts for 400,000 new cases of adult kidney cancer worldwide each year [1]. However, the RCC classification comprises a variety of tumour types with distinct histological and cytological phenotypes [2]. Chromophobe renal cell carcinoma (chRCC) and renal oncocytomas (RO) are two RCC classifications that constitute approximately 10% of all renal tumors [3]. They share similar anatomical origins and histological characteristics, making it difficult to determine their diagnostic classification [4]. Nonetheless, it is critical to discriminate

between these tumors as they result in markedly different clinical outcomes and require distinct therapeutic protocols [4]. Specifically, oncocytoma is a clinically benign neoplasm which often originates from the kidney collecting ducts and can be treated conservatively while chromophobe renal carcinoma is a malignant tumour also arising in distal kidney nephrons but that must be aggressively treated to prevent dire consequences for the patient [3]. Therefore, new diagnostic methods and protocols designed to rapidly and definitively differentiate between these tumour types are needed by the nephrology, pathology and oncology communities [5].

Recent information obtained from transcriptomic studies, including differential expression profiling, in particular, has been used to develop multi-RNA biomarker panels for the identification of a variety of challenging tumors [6,7]. Also, the information retrieved from such profiling studies can be used as large-scale orthogonal validation of proteins identified by mass spectrometry.

Recently, we developed a fast, inexpensive, high sample throughput, ultrasonic-based methodology to extract, identify and quantify the proteome of solid renal biopsies prepared in optimum cutting temperature (OCT) substrate, including chromophobe, oncocytoma and adjacent normal tissue biopsies [8]. In the present work, we perform a deep bioinformatics analysis to identify a cohort of proteins that can distinguish RO and chRCC from NAT (normal adjacent tissue) and at the same time RO from chRCC. Furthermore, we use RNA transcripts to validate the proteins identified by mass spectrometry and immunohistochemical analysis on tissue micro-array (TMA) to validate novel biomarkers candidates.

V.2 EXPERIMENTAL SECTION

V.2.1 STUDY DESIGN AND SAMPLING

The present work utilized high resolution mass spectrometry to analyse the proteomes of 15 flash-frozen, OCT-embedded, human renal tissue biopsies from chromophobe renal cell carcinoma (chRCC, n = 5) and renal oncocytoma (RO, n = 5) including normal adjacent renal tissue (NAT, n = 5). The human kidney tissue samples were collected by the University of Pittsburgh Biospecimen Core and the study was approved by the Institutional Review Board at the University of Pittsburgh (IRB # 02-077). All neoplasms contained a minimum of 85% tumour cells. Data of patients enrolled in this study are summarized in Table SM IX.1 of Supplementary information.

V.2.2 PROTEOMIC ANALYSIS

Biopsies were handled as described in Jorge *et al.* [8]. Briefly, tissues were first cleaned of OCT and then proteins extracted with the aid of an ultrasonic bath (model TI-H-5 from Elma, Singen, Germany) and an ultrasonic probe (UP50H from Hielscher Ultrasonics, Teltow, Germany), respectively. Next, protein digestion was carried out over four minutes using an

ultrasonic microplate horn assembly device (QSonica, Newtown, CT, USA). The extracts containing the digested proteomes were subsequently analysed by a label-free nanoLC-HR-MS/MS approach (UHR-QqTOF IMPACT HD from Bruker Daltonics, Bremen, Germany).

V.2.3 DATA ANALYSIS AND STATISTICS

Relative label-free quantification was carried out using MaxQuant software V1.6.0.16. All raw files were processed in a single run using defaults settings [9,10]. Database searches were performed using Andromeda search engine with the UniProt-SwissProt Human database as a reference and a contaminants database of common contaminants. Data processing was performed using Perseus (version 1.6.5.0) with default settings [11,12]. In brief, reverse hits, and proteins only identified by site were removed from the protein list, and LFQ intensities were log_2-transformed to reduce the effect of outliers. Protein groups were filtered based on a minimum presence of 70% in at least one group. Pearson correlation was performed on filtered LFQ values. Missing LFQ values were imputed through generation of random numbers that were drawn from a normal distribution (width = 0.5 and down shift = 1.8). PCA was performed on the filtered and imputed LFQ intensity data. Log ratios were calculated as the difference in average log_2 LFQ intensity values between the two conditions tested in volcano plots (two-tailed Student's t-test, FDR = 0.01 and S0 = 0.1). Differential expression analysis was performed on z-scored log_2 LFQ intensities through a multiple-sample test (ANOVA test with an 1% of permutation-based FDR filter and preserving randomization for technical replicates). Unsupervised hierarchical clustering was performed based on Euclidean distance.

V.2.4 TRANSCRIPTOMIC ANALYSIS

Transcriptome analysis was performed on these specimens using established methods for frozen tissues [13,14]. Briefly, excess OCT was dissected away on a block of dry ice, the specimens were homogenized in ice cold lysis buffer and then RNA purification was performed using the Qiagen RNeasy Mini- kit (Qiagen, Valencia, CA). RNA underwent extensive QC to confirm quality, yield and diversity and then was subjected to in vitro transcription producing biotin-labelled cRNA followed by hybridization on Human Exon 1.0 ST arrays according to the manufacturers protocol (Affymetrix, Santa Clara, CA). Following washing, staining and scanning (Affymetrix Fluidics 450, Scanner 3000 7G), signal intensities were calculated using the Affymetrix Expression Console MAS 5.0 software.

V.2.5 IMMUNOHISTOCHEMISTRY (IHC) ANALYSIS

A TMA was constructed using 4 cores of RO (n =30), chRCC (n = 12) and 1 core of NAT (n = 20). IHC was performed using Abcam, Rabbit monoclonal antibodies for proteins: Hexokinase 1/HK-1 [clone EPR10134 (B)] and lysosome associated membrane protein-1/LAMP-1 (clone: EPR4204). Granular cytoplasmic staining of any intensity was considered as

positive. Granules were scored as (0 = negative, 1 = focal, 2 = moderate, 3 = abundant) for HK-1 and (0 = negative; D = diffuse, A = apical, F = focal) for LAMP-1. Percent positivity of the cells was also recorded.

V.2.6 DATA AVAILABILITY

Detailed data presented in results are provided in supplementary material as Table SM IX.1 –. Description of human kidney biopsies used in the study.; Table ESM IX.8 – List of pathways enriched in the chRCC and RO tumors when compared to NAT specimens; Table ESM IX.9 – List of proteins with significant differential abundance between chRCC and RO tissues and Table ESM IX.10 – List proteins with significant differential abundance between each tumour and NAT. The mass spectrometry proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE [15] partner repository with the dataset identifier PXD022018.

V.3 RESULTS AND DISCUSSION

V.3.1 TUMOUR CLASSIFICATION BY PCA AND CLUSTERING

The identification of differentially expressed proteins across the three renal phenotype classifications (RO, chRCC, NAT) resulted in a high-resolution principal component analysis (PCA, Figure V.1A) with each specimen class uniquely segregated in distinct clusters. It is interesting to note that the chRCC and RO tumour groups displayed overlapping characteristics of the component 1 variable but both tumour clusters remained distinct from each other. Furthermore, the normal adjacent tissue cluster based on other principal components. Figure V.1B displays the hierarchical clustering obtained from the chromatogram data (90 min. chromatographic runs) which reinforced the PCA results and also validated the hierarchical clustering reported by us previously [8]. It should be noted that the present results include two additional biopsies from which the proteome was extracted, digested and analysed approximately two years later than the other 13 samples. Figure V.1C displays the reproducibility of the mass spectrometry biological replicates by Pearson correlation along with a multi-scatter plot. Pearson correlations ranged from 0.82 to 0.99 for chRCC, from 0.72 to 0.98 for RO and from 0.85 to 0.98 for NAT. The Pearson coefficient was higher when the tumors were compared to each other (0.65 - 0.90), in contrast to comparison with NAT (0.48 - 0.73).

V.3.2 A PROTEOMIC INVESTIGATION INTO MECHANISMS TRIGGERING CHRCC AND RO NEOPLASMS COMPARED TO NAT.

An advantage of high-resolution mass spectrometry is that large numbers of proteins can be simultaneously identified, quantified and compared across multiple samples to

interrogate important differences in protein content and expression. The chRCC versus NAT and the RO versus NAT label-free quantification shown in Figure V.2A and Figure V.2B as volcano plots revealed a similar number of differentially expressed proteins in each comparison (chRCC vs NAT = 393; RO vs NAT = 419). In Figure V.2C the most differentially expressed proteins found in each tumour type are provided (top-10 increased and top-10 decreased, chRCC vs NAT and RO vs NAT).

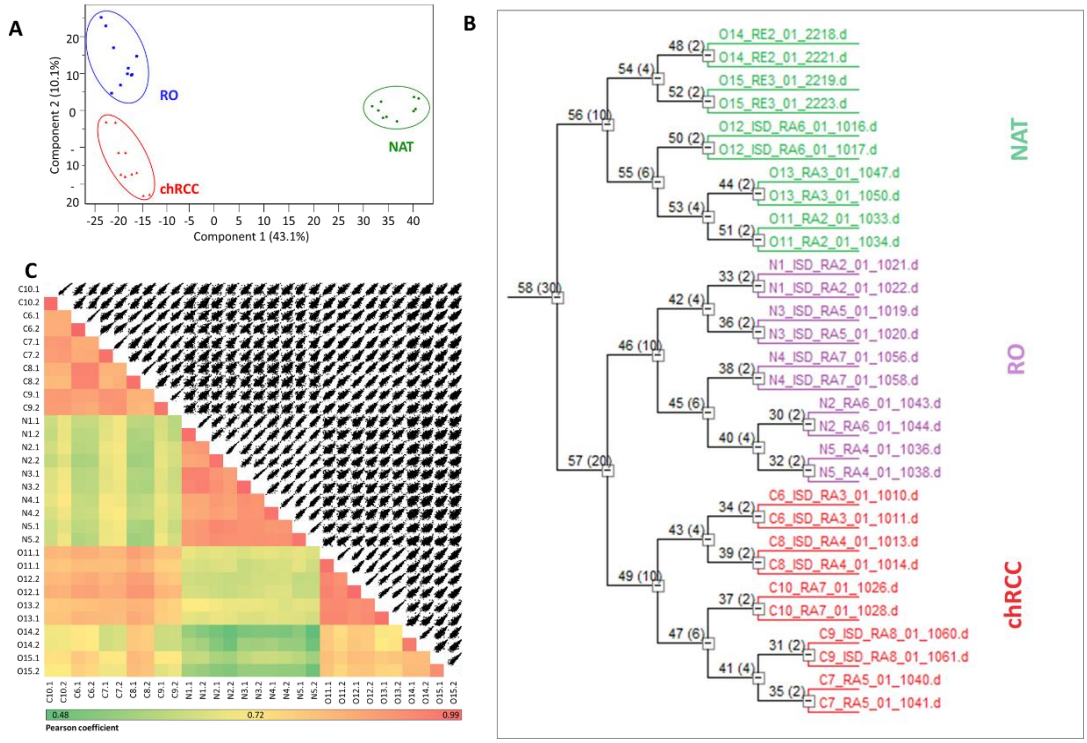

Figure V.1 – Classification of the proteomes. Two instrumental replicates were run for each biopsy resulting in 30 chromatograms that identified a total of 1610 proteins. Of these proteins, a total of 882 were obtained in at least one group (ChRCC, RO, or NAT) with a reproducibility between 70% to 100% of all chromatograms. A) Principal component analysis (PCA) of chRCC, RO, and NAT group samples. B) Unsupervised hierarchical clustering of biopsy tissues' proteome based on all molecular features detected by MS. C) Quality assessment of MS data: the color-coded Pearson correlation of biological (n=5, each condition) and technical (n=2, each sample) replicates along with protein normalized label-free quantification (LFQ) scatterplot matrix represent the reproducibility.

V.3.2.1 Common features between chRCC and RO

The similarity of both chRCC and RO tumors was reflected in the large number of proteins commonly found deregulated (Figure V.2A and Figure V.2B). Our results suggest that multiple proteins directly associated with the creatine phosphokinase high energy transfer pathway were overexpressed in both renal tumour types compared to normal tissues. Among such proteins, cytosolic creatine kinase B (CKB) and mitochondrial creatine kinase S (CKMT2) were found. Another protein involved in metabolism and energy flux that showed increased expression in both classes of tumors in this study was the V-type ATPase 166 kDa subunit a

isoform 4 (ATP6V0A4). Proteins involved in signal transduction and cytoskeletal remodulation were also found dysregulated among tumour subtypes. We found markedly increased protein 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2 (PLCG2) and galectin-3 (LGALS3) in both renal tumour types compared to normal kidney tissue. The similar profile of these two tumors was also verified in down-regulated proteins as the majority of the top ten proteins under expressed were found in both tumors to be the same.

In addition to the analysis of individual protein expression profiles, we carried out a comprehensive ClueGo network analysis [16] of the cumulative MS protein results in order to identify functional pathways enriched in the chRCC and RO tumors when compared to NAT analysis. This is shown in Figure V.2D and the complete list of pathways is provided in Table ESM IX.8 of Supplementary information. The network analysis identified the TCA cycle as one of the most overexpressed pathways in both tumour types. The deregulation of the TCA cycle was revealed by the overexpression of several proteins localized in the mitochondria of both tumour types compared to NAT. For example, isocitrate dehydrogenase [NAD] subunit alpha (IDH3A) and beta (IDH3B) along with dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex (DLAT) proteins were found among the highest overexpressed proteins in both specimens

V.3.2.2 Differential features between chRCC and RO

The functional pathway analysis showed divergent features between chRCC and RO. The lysosome component was found significantly dysregulated in chRCC samples. Mitochondrial networks were significantly dysregulated in RO biopsies with greater representation of mitochondrial biogenesis pathway, mitochondrial protein import pathway among other mitochondrial pathways shown in Figure V.2D and in Table ESM IX.8 of Supplementary information. The oxidative phosphorylation pathway was found also dysregulated on RO (Figure V.2D).

V.3.3 PROTEINS DYSREGULATED BETWEEN CHRCC AND RO

Direct comparison of protein abundance between chRCC and RO tissues revealed a total of 168 differently expressed proteins (Table ESM IX.8 of Supplementary information). The abundances of these proteins were standardized across all specimens based on z-scores and subjected to an unsupervised hierarchical clustering analysis (see Figure V.3A), which clearly stratifies all biopsy types. From these proteins, we found that 109 were deregulated between both neoplasia and at the same time between each neoplasia and NAT (Figure V.3B). The results are presented in Figure V.3C and Figure V.3D for chRCC and RO, respectively. The 109 proteins were subjected to a functional analysis using the ClueGo network. The enriched pathways and their differences in the chRCC and RO are depicted in Figure V.4.

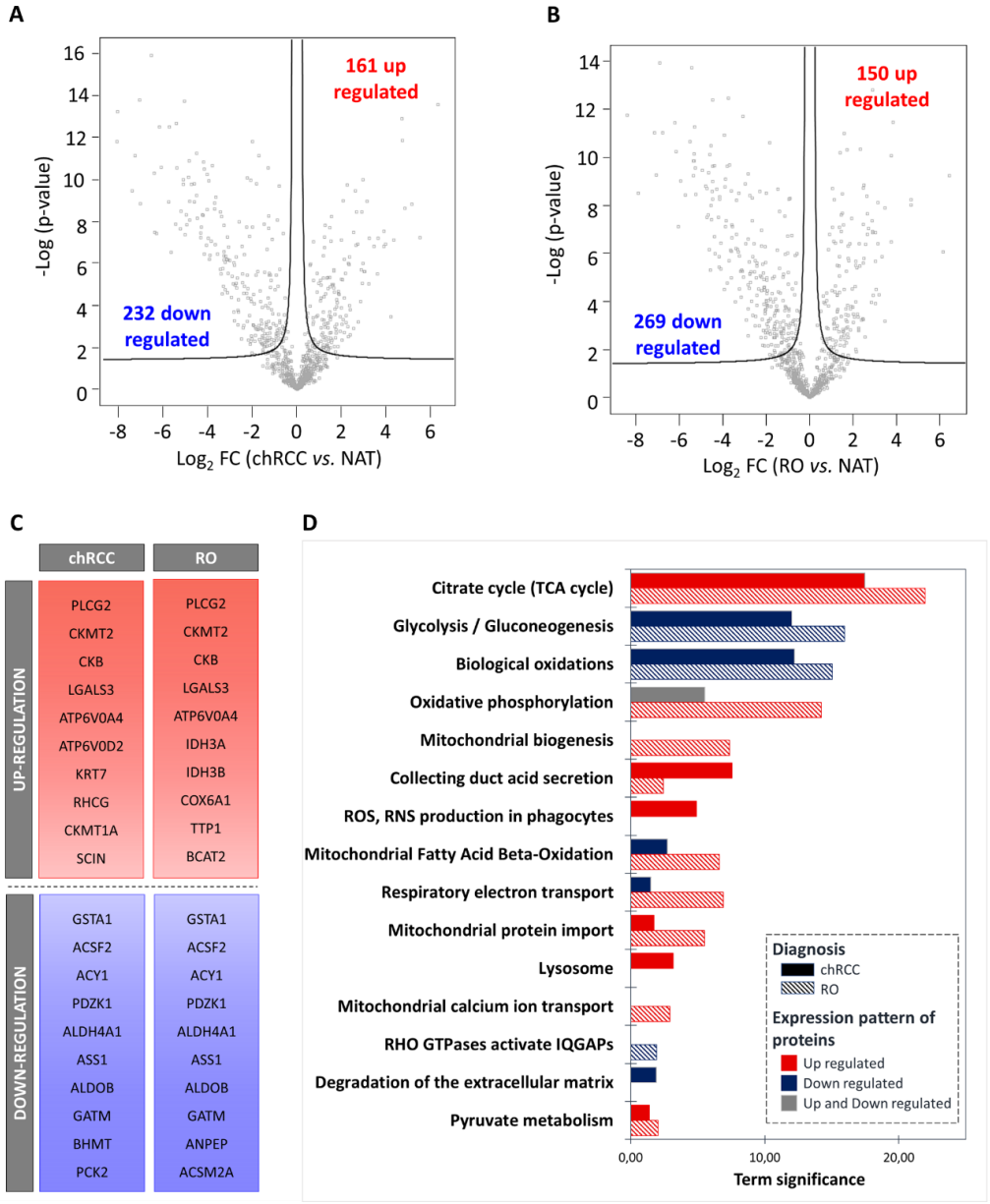

Figure V.2 – Protein expression profiles tumor vs. NAT. Volcano plots illustrating the significant difference (FDR = 0.01, S0 = 0.1) of protein expression levels between A) chromophobe renal cell carcinoma (chRCC) and NAT samples group, and B) renal oncocytoma (RO) and NAT samples group. C) Top 10 upregulated proteins and Top 10 down-regulated proteins for chRCC and RO when compared with NAT levels. D) Representation of a subset of differentially regulated pathways with a p-value < 0.05. Full colored bars represent pathways for chRCC biopsy samples and diagonal stripped bars represent pathways for RO. Blue, red and grey colored denote pathways with at least 50% of downregulated proteins, at least 50% of protein upregulated and equally deregulation of proteins (up and down), respectively. The significance of each term was calculated as -log (term p-value corrected with Bonferroni step-down). The full protein name is given in the abbreviation section.

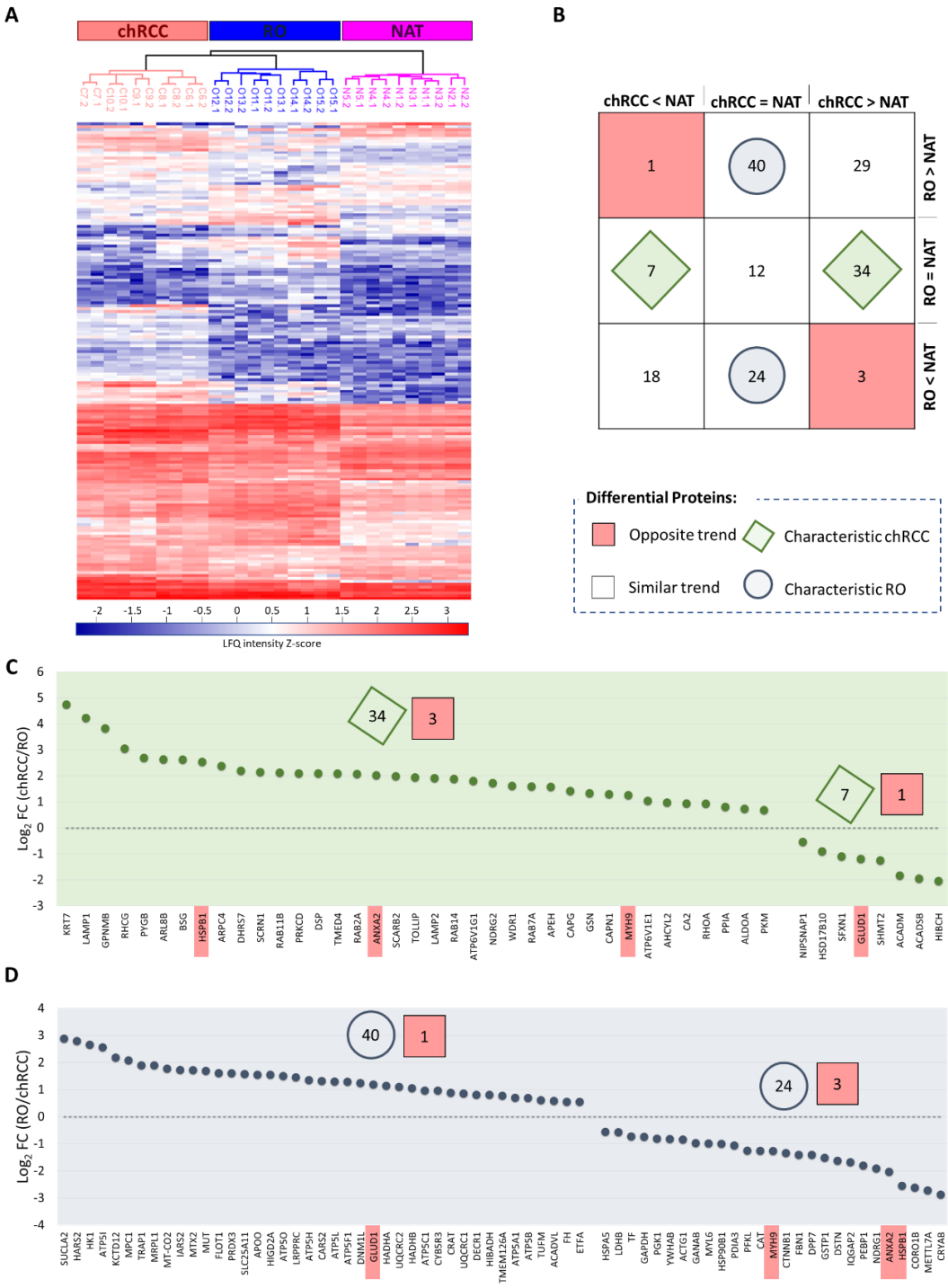

Figure V.3 – Differentiation between chRCC and RO abundances. A) Heatmap representation and unsupervised hierarchical cluster depicting the proteins (n = 168) with significant changed levels (ANOVA, FDR = 1%) between tumor groups, chRCC against RO. Color scale reports the Z-Score of log2 transformed of normalized LFQ intensity values. B) Comparison of dysregulated pattern of the differential proteins. Proteins pattern were firstly achieved by comparison to control levels (NAT) resulting in a dysregulated pattern, then intrinsic pattern was compared between tumors. Red and white colored squares represent the number of proteins with opposite and similar deregulated trend, respectively. Blue diamond shape and circle embody the number of differential proteins characteristic, i.e. only deregulated, to chRCC and RO, respectively. C and D) Plot representation of fold change (FC) profiles for tumor-specific proteins for chRCC and RO, respectively. Pink boxes highlight the four proteins with opposite expression pattern for each tumor subtype. The list of proteins can be found in Table ESM IX.10 of Supplementary information

V.3.4 ORTHOGONAL VALIDATION OF PROTEINS IDENTIFIED BY MASS SPECTROMETRY

V.3.4.1 mRNA validation

To validate the proteins identified by mass spectrometry we use mRNA identification as an orthogonal validation method. Figure V.5A shows that of the list of 882 proteins that are present in 70% of patients of at least one group (chRCC, RO, NAT), 808 proteins matched the corresponding transcript. Thus, remarkably, in terms of validation of proteins, the 92% overlap with transcriptomic data.

V.3.4.2 Immunohistochemistry validation

IHC was applied on a TMA to validate novel candidate biomarkers. Figure V.5B displays the immnunostaining of HK1 and LAMP1 in chRCC, RO and NAT. Figure V.5C and Figure V.5D present the score and pattern evaluation of the IHC results. In HK1 a cut-off of 2 or above score and > 90% cells positivity results in a sensitivity of 96.7% and a specificity of 91.7% for distinguishing RO from chRCC, while diffuse LAMP1 staining has 91.7% sensitivity and 100% specificity for distinguishing chRCC from RO.

V.4 DISCUSSION

Our results delineate multiple proteins and pathways that are similarly affected in both chromophobe and oncocytoma specimens, as shown in Figure V.2. These overlapping findings likely underlie their common neoplastic behaviours despite their different prognoses. On the other hand, differences in protein expression and pathways between these two tumors and normal tissue can provide insight into the potential drivers of malignancy, aggressiveness, and invasiveness related to the chRCC phenotype.

From an initial panel of 168 proteins differentially expressed between both tumour types, a set of 109 proteins have been found to be significantly different between both neoplasia and at the same time significantly different between them and the NAT. Such proteins are presented in Figure V.3C and Figure V.3D and in Table ESM IX.10 of Supplementary information. Interestingly, some have been also reported in literature as chRCC or RO biomarkers, as indicate in Table V.1, what gives further confidence in the data reported in this manuscript. The 109 set of proteins were further analysed to elucidate the intrinsic biological mechanisms underlying each renal tumour subtype, which might be crucial for their distinct outcomes. This is shown in Figure V.4, which reflects the most dysregulated pathways in both neoplasia.

Thus, we have found overexpressed phagosome maturation in the chRCC tumors. This finding has been reported in literature previously and suggests that autophagy provides an adaptative mechanism for chRCC to compensate for high metabolic substrate demand [17]. Also, during phagosome maturation, V-type ATPases are trafficked to the phagosome resulting

in acidification of the internal medium, while RAS related proteins are recruited to fuse the phagosome and lysosome [18]. As a matter of fact, we found in chRCC overexpression of V-type ATPases and RAS related proteins. We hypothesise that the overexpression of RAS-related proteins in chRCC suggests the involvement of MAP kinase signalling in this neoplasia [19].

Table V.1. Proteins proposed as biomarkers found also described in literature.

Gene name	Protein name	Expression profile*		Ref*
		Our study	Lit.	
Deregulated in chRCC				
KRT7	Keratin, type II cytoskeletal 7	up	up	2018 Drendel; 2016 Ng
LAMP1	Lysosome-associated membrane glycoprotein 1	up	up	2018 Drendel
RHCG	Ammonium transporter Rh type C	up	up	2017 Lindgren
PYGB	Glycogen phosphorylase, brain form	up	up	2018 Drendel
HSPB1	Heat shock protein beta-1	up	up	2010 Yusenko
ANXA2	Annexin A2	up	up	2018 Drendel
SCARB2	Lysosome membrane protein 2	up	up	2018 Drendel
LAMP2	Lysosome-associated membrane glycoprotein 2	up	up	2018 Drendel
RAB14	Ras-related protein Rab-14	up	up	2018 Drendel
RAB7A	Ras-related protein Rab-7a	up	up	2018 Drendel
CA2	Carbonic anhydrase 2	up	up	2010 Yusenko
ALDOA	Fructose-bisphosphate aldolase A	up	up	2018 Drendel
Deregulated in RO				
MRPL1	39S ribosomal protein L1, mitochondrial	up	up	2018 Drendel
MUT	Methylmalonyl-CoA mutase, mitochondrial	up	up	2018 Drendel
PRDX3	Thioredoxin-dependent peroxide reductase, mitochondrial	up	up	2018 Drendel; 2010 Yusenko
APOO	Apolipoprotein O	up	up	2017 Kurschner
ATP5H	ATP synthase subunit d, mitochondrial	up	up	2018 Drendel; 2010 Yusenko
HADHA	Trifunctional enzyme subunit alpha, mitochondrial	up	up	2010 Yusenko
UQCRC2	Cytochrome b-c1 complex subunit 2, mitochondrial	up	up	2015 Joshi
HADHB	Trifunctional enzyme subunit beta, mitochondrial	up	up	2018 Drendel
LDHB	L-lactate dehydrogenase B chain	down	down	2010 Yusenko
GSTP1	Glutathione S-transferase P	down	down	2017 Kurschner
HSPB1	Heat shock protein beta-1	down	down	2010 Yusenko
ANXA2	Annexin A2	down	down	2017 Kurschner

Lit. literature

These molecules are critical regulators of intracellular protein trafficking, and disruption of this pathway has been associated with tumour proliferation and metastasis [19]. We have found also elevated levels of lysosomal related proteins in chRCC compared to RO. The process of autophagy is linked with the lysosomal activity, involving cytolytic degradation that results in the production of free amino acids, that can also serve as metabolic substrate to meet high energy demands [20]. Furthermore, our data is consistent with data previously described by Drendel *et al.* [21] and also at the RNA level by Lindgren *et al.* [22]. The potential for invasive and metastatic behaviour in chromophobe tumors may be reflected by the differences in the activation of molecules and pathways that regulate cell motility in chRCC versus RO, as it is shown in Figure V.4. For instance, it is well known that RO uses to be non-invasive, and in line with this fact we have found downregulated in oncocytoma the Rho-GTPase activate IQGAPs pathway. The IQGAP proteins modulate microtubule function associated with cell adhesion and the Rho-GTPase pathway play a role in cancer cell motility [23].

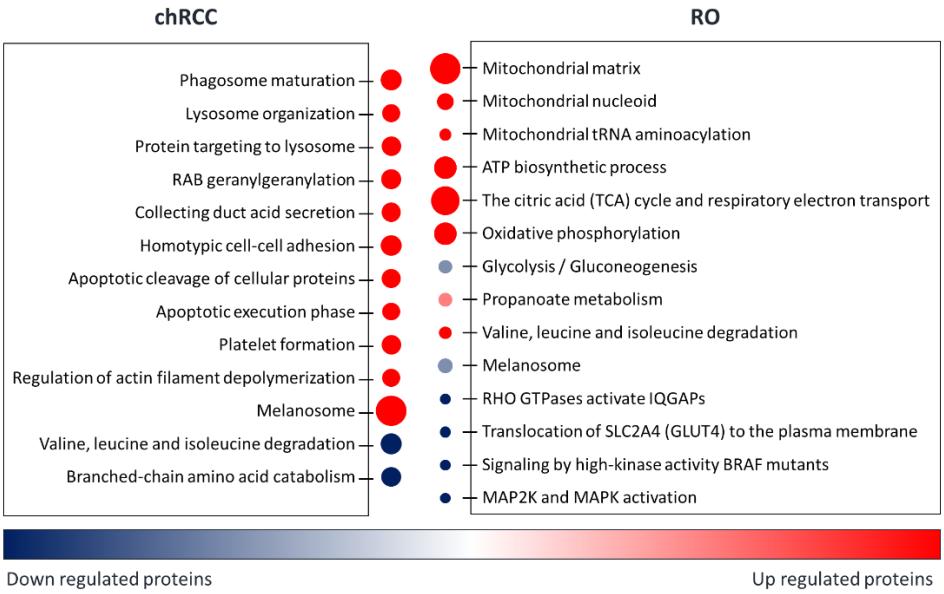

Figure V.4 – Protein enriched functional analysis. The 109 key proteins described in Fig. 3C, Fig. 3D and in Table ESM IX.10 of Supplementary information were searched against GO terms, Reactome and KEGG databases to make the protein enriched functional analysis. Blue to red color represents the percentage of proteins found down or up- regulated, respectively, in each pathway, when each tumor is compared with NAT. Circle size infers the significance of each pathway calculated as -log (term p-value corrected with Bonferroni step-down).

On the other hand, it is worth to mention that 31 proteins belonging to mitochondria are found overexpressed in oncocytoma. This result is in agreement with data reported by Mayr *et al.* [24] who suggest that oncocytoma complex I from oxidative phosphorylation (OXPHOS) is not functional due to several gene mutations but highlights that the proteins related to the other OXPHOS complexes are overexpressed. Also, accumulation of mitochondria is a landmark of oncocytoma [25–27]. In line with these findings, several mitochondrial related pathways were found up-regulated in oncocytomas (Figure V.4).

Further validation as well as the potential utility of the approach here developed for the pathology, is shown via immunohistochemistry with proteins LAMP1 and HK1 (Figure V.5B). Within chRCC tumor cells, LAMP1 protein has shown a diffuse cytosolic staining while within RO a different staining pattern, apical or focal, was present in majority cases. These findings are in agreement with Drendel *et al.* [21] who have seen an enrichment of the LAMP1 protein in both, LC-MS/MS approach and IHC analysis. Additionally, the higher levels of HK1 in RO tissue biopsies were confirmed by IHC results.

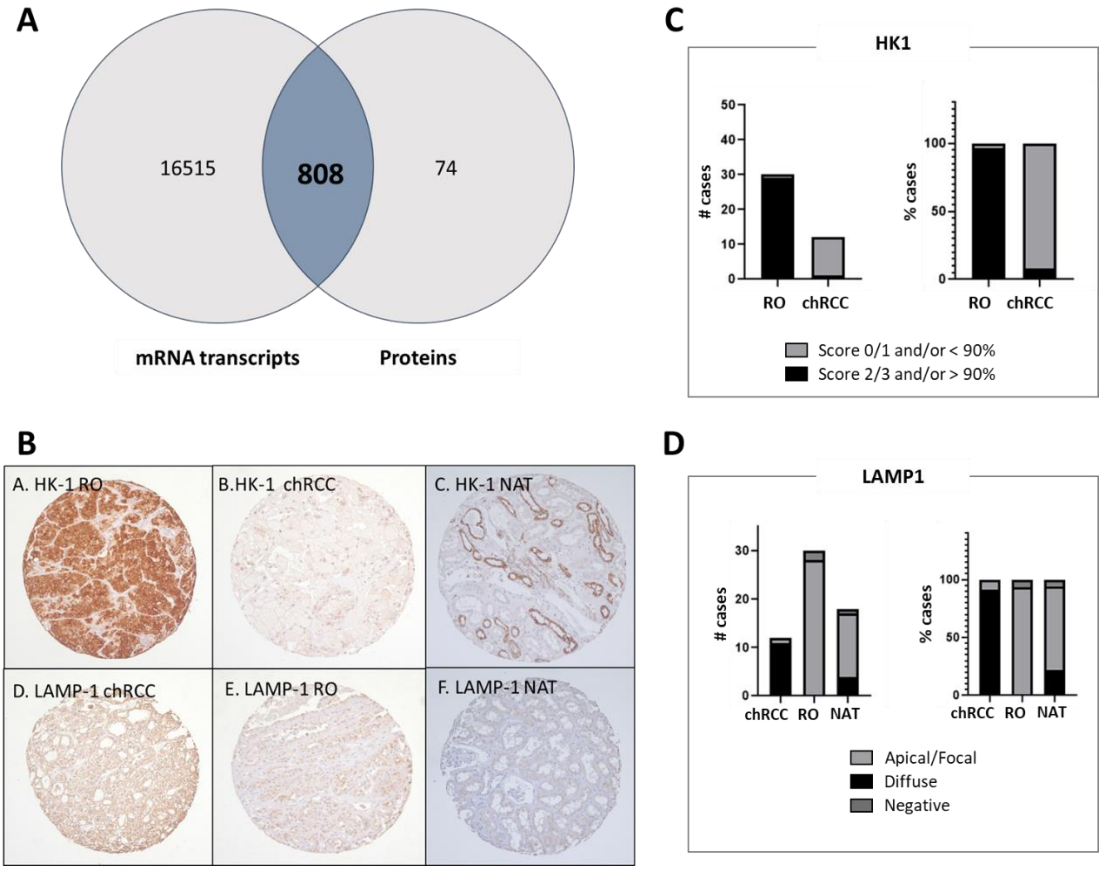

Figure V.5 – Protein validation. A) Proteins identifications were validated by mRNA transcriptomics. Venn diagram present the high confidence the proteomic data, where 808 of identified proteins (92%) were confirmed by mRNA. B) Immunohistochemistry expressions of hexokinase 1 (HK1) and lysosome associated membrane protein-1 (LAMP1) in chromophobe renal cell carcinoma (chRCC), renal oncocytoma (RO) and normal adjacent tissues (NAT). C) and D) represent in bar diagrams the expressions of HK1 and LAMP1, respectively. Granular cytoplasmic staining of any intensity was considered as positive. Granules were scored as (0 = negative, 1 = focal, 2 = moderate, 3 = abundant) for HK1 and (apical/focal, diffuse, negative) for LAMP1.

Thus, these proteins found dysregulated via mass spectrometry were also able to differentiate both neoplasms via IHC.

In conclusion, we have found a panel of 109 proteins that helps to distinguish chRCC from RO and both from NAT. HK1 and LAMP1 have been validated by IHC, being found both as promising biomarkers to differentiate ROs and chRCCs. Some of such proteins have been

already described in literature and some others are here presented for the first time. Also, the proteomics study revealed that mitochondrial networks were dominant in RO, while phagosome maturation displayed greater representation in chRCC, thus paving the way to focus on these pathways as potential sources of biomarkers for diagnosis and prognosis in future works. Moreover, the large number of proteins identified as biomarkers candidates in this study represents a fertile protein library for further stratification of chRCC and RO via IHC. Finally, these findings open new avenues to potentially use the urinary proteome as a source of information to diagnose chRCC and RO.

ACKNOWLEDGMENTS

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura. The authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV, which is financed by national funds from FCT/MEC (UID/QUI/50006/ 2020). H. M. S. is funded by the FCT 2015 Investigator Program (IF/00007/2015). S. J. thanks FCT/MEC (Portugal) for her research contract as PhD student with the grant SFRH/BD/120537/2016. This project utilized the University of Pittsburgh Hillman Cancer Center shared resource facility (Cancer Genomics Facility) supported in part by award P30CA047904 (Dr. LaFramboise).

REFERENCES

[1] Cancer Today https://gco.iarc.fr/today/online-analysis-table?v=2018&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1 (accessed May 7, 2020).

[2] Moch, H *et al*., The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *Eur. Urol.*, **2016**, *70*, 93–105.

[3] Cairns, P., Renal cell carcinoma. *Cancer Biomarkers*, **2011**, *9*, 461–473.

[4] Ljungberg, B *et al*., European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update. *Eur. Urol.*, **2019**, *75*, 799–810.

[5] Wobker, SE *et al*., Modern Pathologic Diagnosis of Renal Oncocytoma. *J. Kidney Cancer VHL*, **2017**, *4*, 1–12.

[6] Batai, K *et al*., Whole-transcriptome sequencing identified gene expression signatures associated with aggressive clear cell renal cell carcinoma. *Genes Cancer*, **2018**, *9*, 247–256.

[7] Casamassimi, A *et al*., Transcriptome Profiling in Human Diseases: New Advances and Perspectives. *Int. J. Mol. Sci.*, **2017**, *18*, 1652.

[8] Jorge, S *et al*., Development of a Robust Ultrasonic-Based Sample Treatment To Unravel the Proteome of OCT-Embedded Solid Tumor Biopsies. *J. Proteome Res.*, **2019**, *18*, 2979–2986.

[9] Cox, J *et al*., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **2008**, *26*, 1367–1372.

[10] Tyanova, S *et al*., The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.*, **2016**, *11*, 2301–2319.

[11] Tyanova, S *et al*., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **2016**, *13*, 731–740.

[12] Tyanova, S *et al*., Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. In *Methods in Molecular Biology*; **2018**; Vol. 1711, pp. 133–148.

[13] Ma, C *et al*., In Vitro Transcription Amplification and Labeling Methods Contribute to the Variability of Gene Expression Profiling with DNA Microarrays. *J. Mol. Diagnostics*, **2006**, *8*, 183–192.

[14] LaFramboise, WA *et al*., Acute molecular response of mouse hindlimb muscles to chronic stimulation. *Am. J. Physiol. Physiol.*, **2009**, *297*, C556–C570.

[15] Perez-Riverol, Y *et al*., The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.*, **2019**, *47*, D442–D450.

[16] Bindea, G *et al*., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, **2009**, *25*, 1091–1093.

[17] Goldsmith, J *et al*., Autophagy and Cancer Metabolism. In *Methods in Enzymology*; Academic Press Inc., **2014**; Vol. 542, pp. 25–57.

[18] Kinchen, JM *et al*., Phagosome maturation: going through the acid test. *Nat. Rev. Mol. Cell Biol.*, **2008**, *9*, 781–795.

[19] Dhillon, AS *et al*., MAP kinase signalling pathways in cancer. *Oncogene*, **2007**, *26*, 3279–3290.

[20] Pavlova, NN *et al*., The Emerging Hallmarks of Cancer Metabolism. *Cell Metab.*, **2016**, *23*, 27–47.

[21] Drendel, V *et al*., Proteomic distinction of renal oncocytomas and chromophobe renal cell carcinomas. *Clin. Proteomics*, **2018**, *15*, 25.

[22] Lindgren, D *et al*., Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Rep.*, **2017**, *20*, 1476–1489.

[23] Warner, H *et al*., Control of adhesion and protrusion in cell migration by Rho GTPases. *Curr. Opin. Cell Biol.*, **2019**, *56*, 64–70.

[24] Mayr, JA *et al*., Loss of Complex I due to Mitochondrial DNA Mutations in Renal Oncocytoma. *Clin. Cancer Res.*, **2008**, *14*, 2270–2275.

[25] Joshi, S *et al*., The Genomic Landscape of Renal Oncocytoma Identifies a Metabolic Barrier to Tumorigenesis. *Cell Rep.*, **2015**, *13*, 1895–1908.

[26] Kürschner, G *et al*., Renal oncocytoma characterized by the defective complex I of the respiratory chain boosts the synthesis of the ROS scavenger glutathione. *Oncotarget*, **2017**, *8*, 105882–105904.

[27] Gasparre, G *et al*., Learning from oncocytic tumors: Why choose inefficient mitochondria? *Biochim. Biophys. Acta - Bioenerg.*, **2011**, *1807*, 633–642.

CHAPTER VI.

Towards TPA-based pathology

Susana Jorge, José L. Capelo, William LaFramboise, Rajiv Dhir, Jacek R. Wiśniewski, Carlos Lodeiro, Hugo M. Santos*

In preparation

ABSTRACT

Renal cell carcinoma is a heterogeneous disease comprising several distinct subtypes including clear cell, papillary, and chromophobe. However, those renal neoplasms demonstrate histo-morphologic overlapped features, including with the benign renal oncocytoma, leading to frequent misdiagnosis. Therefore, novel biomarkers able to subclassified them are needed for clinical practice.

In the present work, we propose to use for the first time the total protein approach (TPA) to identify an effective panel of proteins and their concentration ranges to diagnose and characterize diverse RCC subtypes.

A proteomics tumor profiling was performed on biopsies of seven clear cell, five papillary, and five chromophobes subtypes. Five benign oncocytomas and five normal adjacent tissues were also included. Label-free quantification was performed on proteomes extracted from renal tissue biopsies and the LFQ values were transformed into TPA values for absolute quantitative analysis.

The analysis of the proteome of 27 renal tissue biopsies has revealed a total of 850 differentially expressed proteins between renal neoplasms and normal adjacent tissues. Based on TPA concentrations, a 24 panel, composed of the six proteins with the highest significant differential expression for each tumor subtype, was achieved which can be used to classify the biopsies.

In a near future, we envision the TPA based pathology as the next step in solid biopsy-based cancer diagnosis and prognosis.

VI.1 INTRODUCTION

Renal cell carcinoma (RCC) is the most frequent cancer diagnosed in the adult kidney in the world [1,2]. Arising from the epithelium of renal tubules, RCC is a heterogeneous disease comprising several subtypes, being the clear cell RCC (ccRCC) subtype the most prevalent [3]. This subtype accounts for approximately 75% of all cases and is followed by papillary RCC (pRCC) with approximately 10% and chromophobe RCC (chRCC) with approximately 5% of the cases [4]. Another frequently diagnosed renal neoplasm, with approximately 5% of the renal masses diagnosis, is the benign renal oncocytoma (RO) [4]. Less predominant subtypes include medullary, collecting duct, mucinous tubular and spindle cell carcinoma, and MiTF family translocation renal cell carcinoma [5,6]. Such heterogeneity results in a complex disease in which individual RCC subtypes carry distinct clinical outcomes and, most importantly, therapies. According to the 2012 ISUP classification and the 2016 WHO classification [7], the cytogenetic

profile of ccRCC is characterized by VHL gene mutations, hypermethylation of VHL gene promotor, and loss of heterozygosity (LOH) at 3p25. Mutations in PBRM1, BAP1, and SETD2 have also been linked to this disease [8–10]. To diagnostic ccRCC, the most useful and practical immunohistochemical (IHC) profile comprises the strong nuclear marker PAX-8, the typical membranous staining of carbonic anhydrase IX (CA9), and the traditionally used vimentin protein [7]. Epithelial markers such as EMA, cytokeratins AE1-AE3, and CAM are also used [7]. On the other hand, pRCC is classified into two subtypes, type 1 and type 2, as firstly proposed by Delahunt *et al.* [11]. These two subtypes have been reported with different cytogenetic alterations and gene mutations [12]. While pRCC type 1 typically shows chromosomal gains of 7, 12q, 16p 17 and 20 and losses of 9p with mutations in the MET gene, pRCC type 2 has been associated with losses of 1p and 9p and genetic aberrations in CDKN2A, SETD2, and NRF2 genes [12]. IHC profiles revealed to be positive for cytokeratins (AE1-AE3, CAM 5.2, HMWCK), racemase, vimentin, and CD10 in pRCC, and CK7 is more frequently expressed in type 1 than type 2. The chRCC subtype is characterized by several chromosomal losses such as in chromosomes 1, 2, 6 10, 13, 17, and 21 [12]. The genetic profile has been associated with mutations in TP53 and PTEN [12]. Morphologically, is characterized by huge pale cells with reticulated cytoplasm, diffuse cytoplasmic Hale's iron colloid staining, and prominent cell membrane [13]. The RO subtype is associated with some cytogenetic changes, including losses of chromosomes 1 and Y, rearrangements of CCND1, and mutations in genes of complex I of mitochondrial OXPHOS, such as COX1, COX2, MTND4 MTCYB [12]. IHC panel includes vimentin, CK 7, CD117, and E-cadherin [7]. Although RCCs can be classified on the basis of immunohistological characteristics, some overlapped features are common between subtypes, making diagnosis difficult. For instance, the presence of CK7 is characteristic for pRCC and generally negative for ccRCC and RO. However, cases with focal positivity are also described in these two subtypes [12]. Therefore, further research to find a way for effective RCC subtyping is necessary. As summarized above, immunohistochemical (IHC) assays are widely used for diagnosis of cancers, as tumor specific antigens are expressed de novo or deregulated in tumor tissues [14]. However, IHC is dependent of the availability of the appropriate antibodies and the number of antibodies tested in a single assay is limited.

Mass spectrometry (MS)-based proteomics has become a valuable approach to identify, quantify, and characterized large numbers of proteins in solid and liquid biopsies, as demonstrated by Gilbertson *et al.* [15] who have blindly compared the determination of amyloid fibril protein in amyloidotic tissues by IHC and MS techniques. Overall, a positive concordance was achieved between the two techniques, however, whilst the IHC- based approach has a diagnostic accuracy of 76%, the MS-based approach has it of 94%. In MS-based proteomic approaches, different strategies can be used to evaluate the expression of the proteome, however, the label-free LC-MS approaches are gaining momentum. The advantages of label-free techniques, over labeling approaches, include i) no limitation of the number of experiments that can be compared; ii) higher dynamic range of quantification, and iii) less time-consuming

steps [16]. Thus, significant changes can be measured across an entire proteome and can be compared in a large cohort of samples.

Recently, Wiśniewski *et al.* [17] proposed the total protein approach (TPA) as a label- and standard- free method for absolute protein quantitation of proteins using large-scale proteomic data. The method relies on the assumption that the total MS signal from all identified proteins in the dataset reflects—in a biochemical sense—the total protein and the MS signal from a single protein corresponds to its abundance in the studied sample. Remarkably, this approach has been applied in other proteomic studies, confirming its accuracy and utility [18–20].

In RCC tumors, many MS-based proteomic studies using label-free quantification have been applied to decipher their molecular landscape [21]. One of the early studies to outline differences in protein levels among different subtypes of RCCs was carried out by Valera *et al.* [22]. In this study, several subtypes were addressed using a 2DE/MS approach. Since then, and with the advances of mass spectrometry technology, the conventional gel-based proteomic approaches have been replaced by gel-free strategies, including labeled-based quantitative studies used to evaluate the differential expression between the proteomes [23–25]. Likewise, other studies have applied a label-free quantitative proteomic analysis in RCC samples [26–28]. As the most frequent tumor subtype, the ccRCC is the most common histotype investigated throughout the scientific research community [21]. Nevertheless, the high heterogeneity of the RCC tumors and the shared features among the subtypes, a clear differentiation between them is still needed.

In this work, we successfully used for the first time the TPA quantitative proteomic approach to identify an effective panel of proteins and their concentration ranges to diagnose and characterize diverse RCC subtypes.

VI.2 EXPERIMENTAL SECTION

VI.2.1 STUDY DESIGN AND SAMPLING

The present work utilized high-resolution mass spectrometry to analyze the proteomes of 27 flash-frozen, OCT-embedded, human renal tissue biopsies from clear cell renal cell carcinoma (ccRCC, n = 7), papillary renal cell carcinoma (pRCC, n = 5), chromophobe renal cell carcinoma (chRCC, n = 5) and renal oncocytoma (RO, n = 5) including normal adjacent renal tissue (NAT, n = 5). The human kidney tissue samples were collected by the University of Pittsburgh Biospecimen Core and the study was approved by the Institutional Review Board at the University of Pittsburgh (IRB # 02-077). All neoplasms contained a minimum of 85% tumor cells. Data of patients enrolled in this study are summarized in Table SM IX.1 of Supplementary information.

VI.2.2 PROTEOMIC ANALYSIS

Biopsies were handled as described in Jorge *et al*. [29] Briefly, tissues were first cleaned of OCT and then proteins extracted with the aid of an ultrasonic bath (model TI-H-5 from Elma, Singen, Germany) and an ultrasonic probe (UP50H from Hielscher Ultrasonics, Teltow, Germany), respectively. Next, protein digestion was carried out over four minutes using an ultrasonic microplate horn assembly device (QSonica, Newtown, CT, USA). The extracts containing the digested proteomes were subsequently analyzed by a label-free nanoLC-HR-MS/MS approach (UHR-QqTOF IMPACT HD from Bruker Daltonics, Bremen, Germany).

VI.2.3 DATA ANALYSIS AND STATISTICS

Relative label-free quantification was carried out using MaxQuant software V1.6.0.16. All raw files were processed in a single run using default settings [30,31]. Database searches were performed using Andromeda search engine with the UniProt-SwissProt Human database as a reference and a contaminants database of common contaminants. Data processing was performed using Perseus (version 1.6.5.0) with default settings [32,33]. In brief, reverse hits, and proteins only identified by site were removed from the protein list, and LFQ intensities were log_2-transformed to reduce the effect of outliers. Protein groups were filtered based on a minimum presence of 70% in at least one group. Pearson correlation was performed on filtered LFQ values. Missing LFQ values were imputed through the generation of random numbers that were drawn from a normal distribution (width = 0.5 and down shift = 1.8). PCA was performed on the filtered and imputed LFQ intensity data. Log ratios were calculated as the difference in average log_2 LFQ intensity values between the two conditions tested in volcano plots (two-tailed Student's t-test, FDR = 0.01 and S0 = 0.1). Differential expression analysis was performed on z-scored log_2 LFQ intensities through a multiple-sample test (ANOVA test with a 1% of permutation-based FDR filter and preserving randomization for technical replicates). Unsupervised hierarchical clustering was performed based on Euclidean distance.

Absolute protein quantification was calculated using the total protein approach (TPA) as described by Wiśniewski *et al*. [17]. Briefly, protein concentration was calculated as follows (equation (3)):

$$Protein\ concentration\ (p_i) = \frac{MS_{signal\ (p_i)}}{Total\ MS_{signal}\ x\ MW\ (p_i)} \quad (3)$$

VI.2.4 IMMUNOHISTOCHEMISTRY (IHC) ANALYSIS

Tissue Micro Arrays (TMAs) were constructed using 4 cores of ccRCC (n = 40), pRCC (n = 26), RO (n = 30), chRCC (n = 12), unclassified renal cell carcinoma [rhabdoid RCC (n =

29), sarcomatoid RCC (n = 43)] and 1 core of NAT (n = 20). IHC was performed using mouse monoclonal antibody (clone 2C5A3, Abcam) for PLIN2 protein. Membranous or droplet like staining of any intensity was considered as positive. Percent positivity of tumor cells was scored as (0 = Negative; 1:1-10%; 2: > 10-50%; 3: > 50%). All TMA cores were scored, and average was taken to give a final score.

VI.3 RESULTS AND DISCUSSION

VI.3.1 RENAL SAMPLES

The proteomes of 27 human specimens, 22 diagnosed with renal tumors and 5 NAT samples, were interrogated by high-resolution mass spectrometry in duplicate. Data corresponding to patients diagnosed as ccRCC (n = 7), pRCC (n = 5), chRCC (n = 5), RO (n = 5) is provided in Table SM IX.1 of Supplementary information. NAT (n = 5) tissue samples were used as a control to determine protein expression deregulations in each tumor subtype.

VI.3.2 PROTEOMIC PERFORMANCE

MS analysis of tissue biopsies was performed in duplicate resulting in 54 LC-MS/MS runs. Figure VI.1A shows the reproducibility of technical and biological replicates in each sample group. Pearson correlations ranged from 0.65 to 0.97. Worth to note, technical replicates presented a correlation higher than 0.94 for all tumor subtypes studied, being the papillary subtype the less homogeneous in terms of biological variability. These results are in agreement with the well-known characteristics of this cancer subtype. A total of 2547 proteins were identified across all tissue biopsies and Figure VI.1B depicts the number of proteins identified in each tumor subtype. To ensure the robustness of the identified proteins for quantification, only proteins with a reproducibility higher than 70% in at least one tumor subtype were considered for further analysis. A final set of 1234 proteins fulfilled the previous criterion and they were used for quantification.

VI.3.3 PROTEIN QUANTIFICATION

Relative protein quantification of the selected set of proteins (1234) was done by label-free quantification using the MaxQuant software. Correlation by principal component analysis (PCA) of all samples is presented in Figure VI.1C, which shows that tumor proteomes are different enough to clearly differentiate between RCC subtypes and each one of them from NAT.

Absolute quantification of proteins was calculated based on raw intensities using the total protein approach, TPA, method proposed by Wiśniewski *et al.* [17]. Figure VI.1D shows the dynamic range of protein abundances expressed in TPA-based concentrations achieved in this

study, spanning around seven orders of magnitude. This is in line with previous studies reported on the TPA approach to solid biopsies [18].

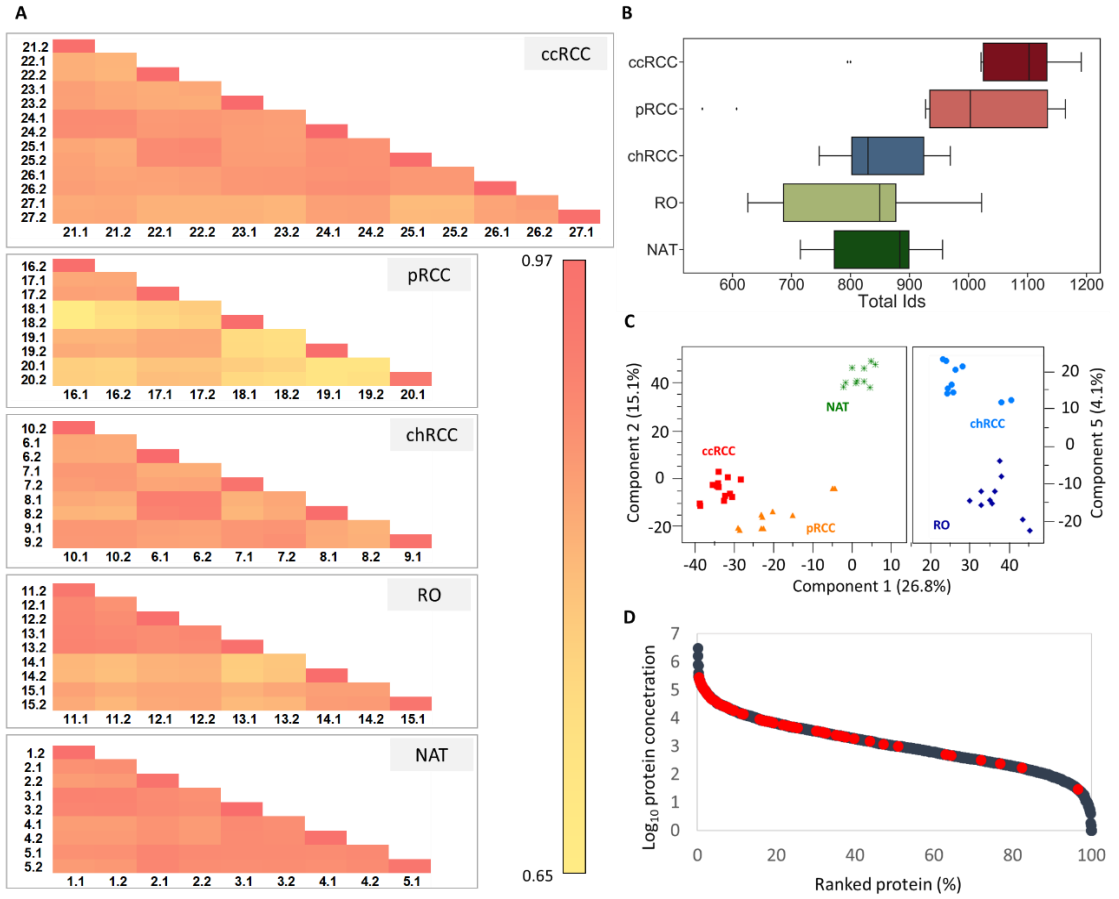

Figure VI.1 – Data statistical analysis. A. Person correlation of biological and technical replicates per tumor type and NAT, e.g.: 1.1 and 1.2 correspond to sample 1 replicate1 and sample 1 replicate 2 respectively. B. Number of proteins identified per tumor type and NAT. C. Principal component analysis, PCA, using the set of proteins as described in the text (1234). D. Distribution of TPA-based protein concentrations (mol/gr total protein) with the common proteins described in literature as potential ccRCC markers (red dots).

VI.3.4 Label-Free Renal Cell Carcinomas Protein-Based Signatures

To assess significantly upregulated or downregulated proteins in each tumor type within the selected set of proteins (1234), a multi-sample test (ANOVA with a permutation-based FDR < 1% filter) was applied. Expression levels were found to be statistically different for 850 proteins between sample groups. As shown in Figure VI.2A, the unsupervised clustering analysis performed on such dysregulated proteins clearly divide the samples into two groups, one comprising ccRCC and pRCC and another one comprising chRCC, RO and NAT.

Taking into consideration the 850 proteins statistically different a search was done to undisclosed biomarkers capable of diagnosis each tumor type. A schematic representation of the candidate biomarker panel selection workflow is shown in Figure VI.2B. The comparison of

protein abundances between sample groups undisclosed a set of proteins that allows to distinguish each tumor subtype (ANOVA, $p < 0.01$). Such panels comprise 81 proteins for ccRCC; 60 proteins for pRCC; 25 proteins for chRCC and 39 proteins for RO. The proteins are summarized in Table ESM IX.11 of Supplementary information.

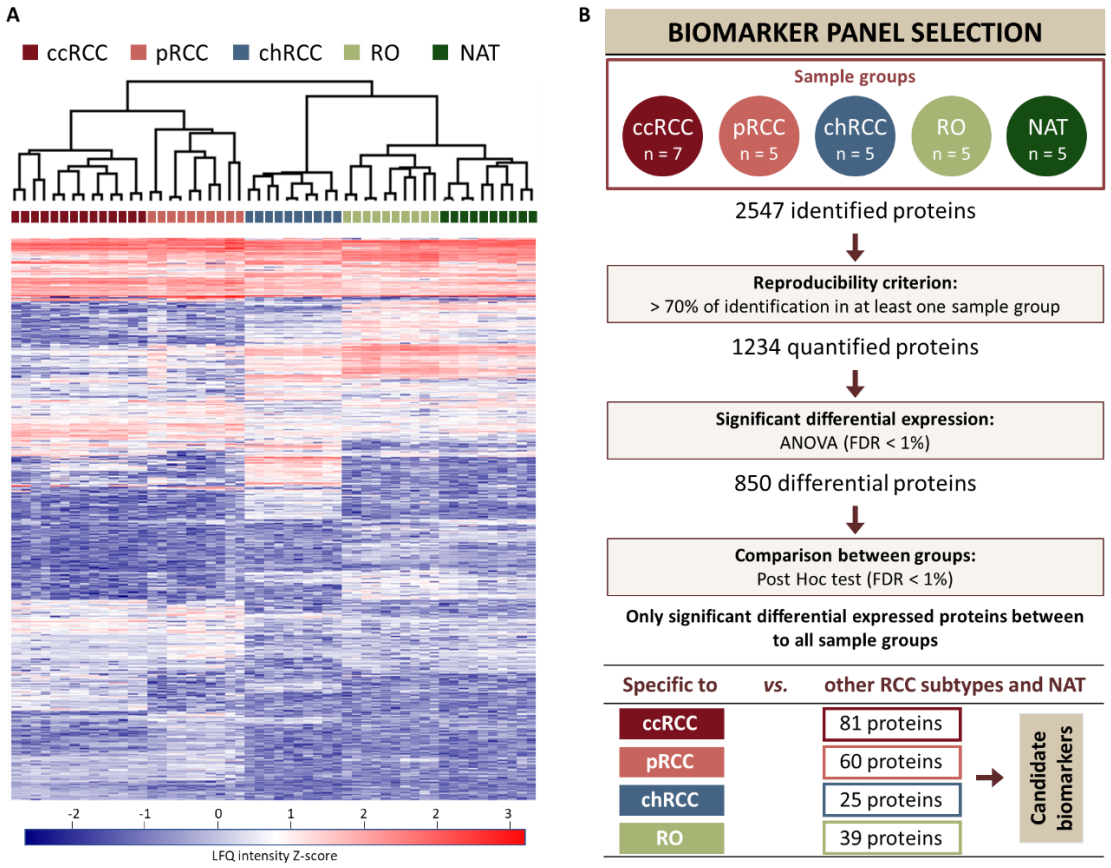

Figure VI.2 – Significant differential expression in RCC. A. Unsupervised hierarchical clustering analysis of 850 differential proteins (ANOVA, FDR < 1%). B. Statistical proteomic workflow applied to select significant protein to discriminate the four RCC subtypes, ccRCC, pRCC, chRCC and RO.

VI.3.5 TPA-BASED CONCENTRATION RANGE FOR DIAGNOSTIC PROTEINS

Next the TPA approach was used to transform the LFQ values of the protein panels selected as depicted in Figure VI.2B and shown in Table ESM IX.12 of Supplementary information in concentration values. Then for each tumor type we selected the group of 6 proteins presenting the highest differential concentrations levels. Such proteins are presented in Figure VI.3.

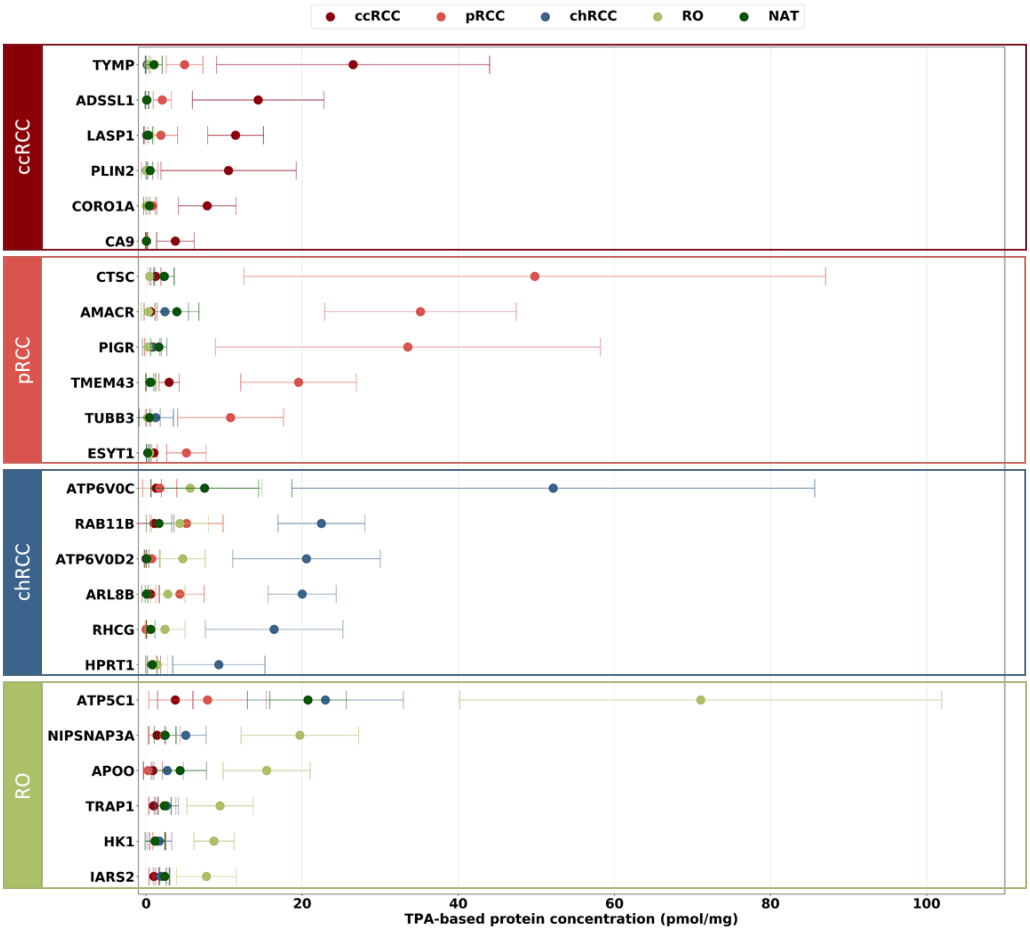

Figure VI.3 – TPA-based concentration of top 24 proteins with the highest significant differential expression between tissue biopsies. Absolute protein concentration expressed in pmol/mg was calculated through the TPA method [34].

VI.3.6 Evaluating the TPA approach with data retrieved from literature.

As proof-of-concept the ccRCC diagnostic proteins found with our TPA-based approach were compared with those ones already described in the literature. Thus, of circa 90 proteins described as putative diagnostic markers, 46 matched our own ones.

Table VI.1 – Protein list of common deregulated proteins achieved in literature and our data.

Protein name	Gene name	ccRCC/NAT expression	Ref.
Acetyl-CoA acetyltransferase, mitochondrial	ACAT1	down	[27]
Alcohol dehydrogenase [NADP(+)]	AKR1A1	down	[35]
Retinal dehydrogenase 1	ALDH1A1	up	[35]
Fructose-bisphosphate aldolase	ALDOA	up	[28]
Aminopeptidase N	ANPEP	down	[36]

Table VI.1 (Cont.)

Annexin A4	ANXA4	up	[37]
Annexin A5	ANXA5	up	[35]
Aquaporin-1	AQP1	down	[38]
Carbonic anhydrase 9	CA9	up	[39]
Calbindin	CALB1	down	[35,40]
Macrophage-capping protein	CAPG	up	[41]
Cofilin-1	CFL1	up	[28]
Coronin-1A	CORO1A	up	[26]
Dipeptidase 1	DPEP1	down	[38]
Alpha-enolase	ENO1	up	[22,25,42]
Gamma-enolase	ENO2	up	[35,41]
Fatty acid-binding protein, brain	FABP7	up	[43,44]
Gelsolin	GSN	up	[40]
Glutathione S-transferase P	GSTP1	up	[35]
Heat shock protein beta-1	HSPB1	up	[22,25,41]
Plastin-2	LCP1	up	[41]
L-lactate dehydrogenase A chain	LDHA	up	[23,25]
Galectin-1	LGALS1	up	[24,43]
Major vault protein	MVP	up	[23]
Nucleoside diphosphate kinase A	NME1	up	[41]
Nicotinamide N-methyltransferase	NNMT	up	[23,28,41,45]
Profilin-1	PFN1	up	[24,28]
Pyruvate kinase PKM	PKM	up	[46,47]
Perilipin-2	PLIN2	up	[23,26,48]
Peptidyl-prolyl cis-trans isomerase A	PPIA	up	[43]
Peroxiredoxin-4	PRDX4	up	[41]
UV excision repair protein RAD23 homolog B	RAD23B	up	[22]
Histone-binding protein RBBP7	RBBP7	up	[41]
Reticulocalbin-1	RCN1	up	[49]
Protein S100-A10	S100A10	up	[50]
Protein S100-A11	S100A11	up	[50,51]
Plasma protease C1 inhibitor	SERPING1	up	[23]
Serpin H1	SERPINH1	up	[52]
ADP/ATP translocase 3	SLC25A6	down	[35]
Protein-glutamine gamma-glutamyltransferase 2	TGM2	up	[35]
Triosephosphate isomerase	TPI1	up	[22]
Tubulin alpha-1B chain	TUBA1B	up	[41]
Thymidine phosphorylase	TYMP	up	[23,41]
Vimentin	VIM	up	[35,42,53];
14-3-3 protein zeta/delta	YWHAZ	up	[24]

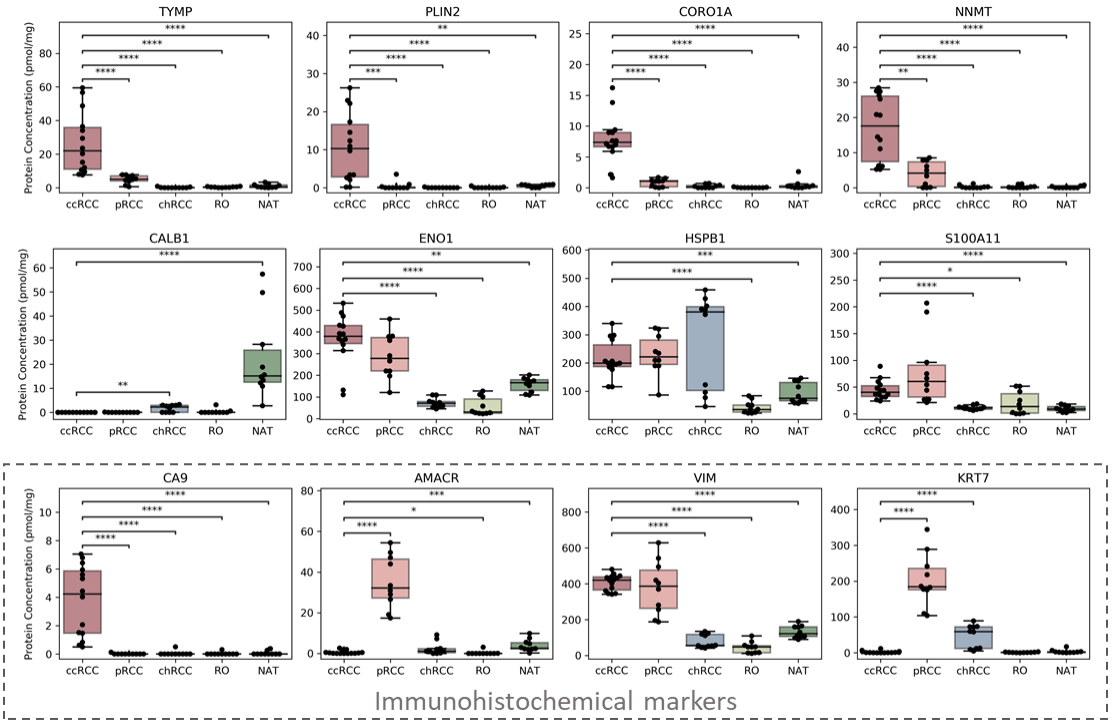

Figure VI.4 – Protein abundances of a set of protein described in literature. In the dashed square are marked proteins widely used in immunohistochemical diagnosis. Statistical analysis was performed using pairwise Mann Whitney test (p <= 0.005; ** p <= 0.001; *** p <= 1.00e-03; p <= 1.00e-04).*

VI.3.7 VALIDATION OF PLIN2 PROTEINS BY IMMUNOHISTOCHEMISTRY

PLIN2 protein was evaluated by IHC (Figure VI.5). The majority of the ccRCC were positive for PLIN-2 whereas pRCC, chRCC and RO were negative. A few rhabdoid and sarcomatoid RCC cases showed focal (< 10%) positivity with a score of 1. PLIN2 staining score of 2 or 3 (> 10% positivity) has a sensitivity of 90% and specificity of 100% for distinguishing ccRCC from other renal neoplasms.

VI.4 DISCUSSION

For many pathological conditions such as cancer, mass spectrometry-based analysis of solid and liquid biopsies has become an indispensable tool to diagnose and to prognose. The advent of high-resolution mass spectrometry in conjunction with powerful software has allowed a step forward from biomarker discovery towards personalized medicine. The diagnosis of RCC still presents some challenges to pathologists because there is no consensus about which biomarkers must be used for diagnosis and prognosis all different variants of this neoplasia. In addition, there is an urgent need to find urine or blood biomarkers of this disease, so the chirurgic intervention to obtain a solid biopsy for cancer diagnostic can be overcome. In the

pursuing of the aforementioned challenges, we interrogate the proteome of 27 RCC samples comprising the main four subtypes using high resolution mass spectrometry.

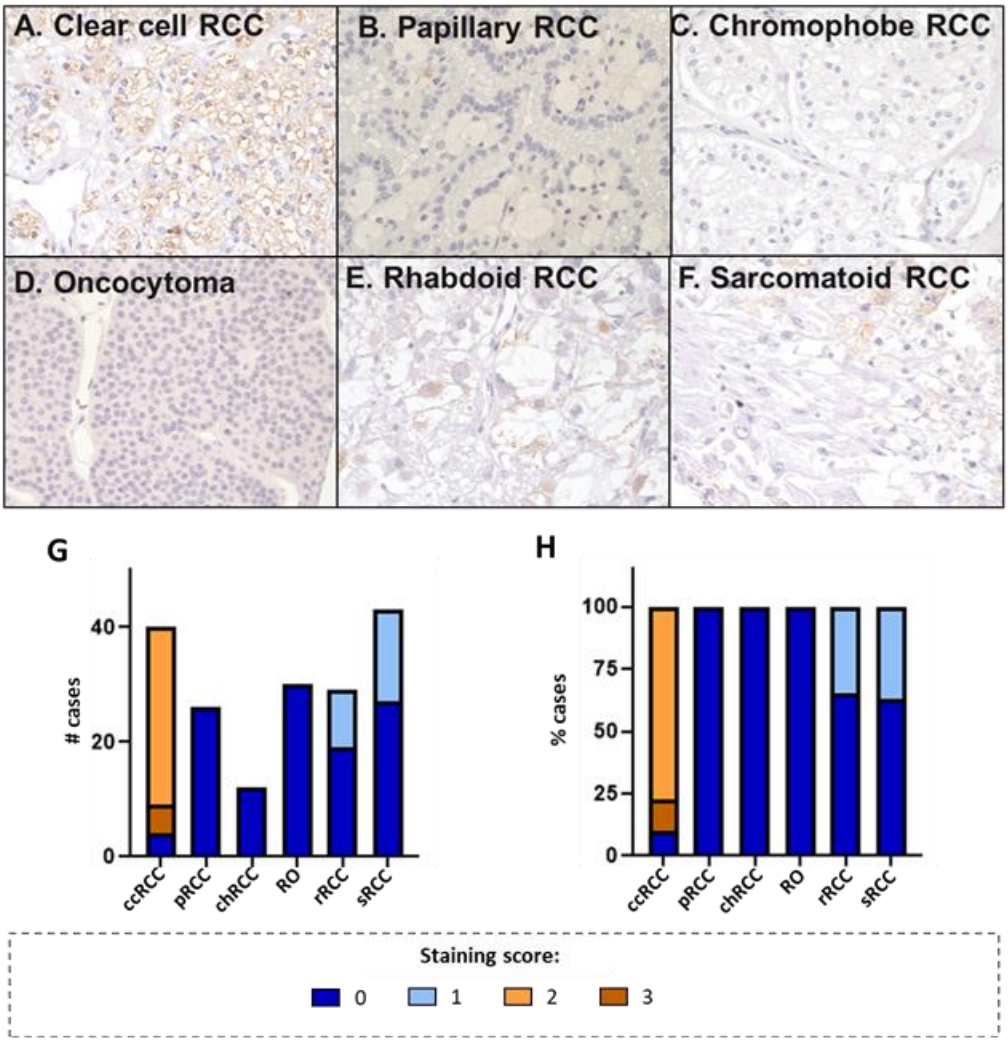

Figure VI.5 – Immunohistochemistry expressions of PLIN2 in A) clear cell RCC, B) papillary RCC, C) chRCC, D) oncocytoma, E) rhabdoid RCC and F) sarcomatoid RCC. G) and F) represent, in bar diagrams, the number of cases and the percentage of cases positively stained for PLIN2, respectively. Percent positivity of tumor cells was scored as (0 = Negative; 1: 1-10%; 2: > 10-50%; 3: > 50%).

Figure VI.1A presents the quality of the mass spectrometry data obtained by comparing the biological and technical replicates expressed as person correlations. It might be seen that technical replicates have a similarity higher than 97%, showing the excellent performance of this technique. In terms of biological replicates, it can be seen that pRCC presents the higher proteome variability between biopsies whilst the NAT presents the higher homogeneity. These results are in agreement with literature, as it is well known that a hallmark of pRCC is the large variability in the morphological expressions of this type of RCC [54]. Once it was verified the quality of the mass spectrometry data, the proteome expressed by each RC tumor subtype was compared using a PCA model, which is shown in Figure VI.1C. The PCA analysis suggests that the proteomes of the ccRCC and pRCC subtypes are more similar between them than with the

proteomes expressed by chRCC and RO. This finding is further confirmed by the clustering presented in Figure VI.2A, which shows two clusters, one constituted by ccRcc and pRCC and the other one by chRCC, RO and NAT. These results are consistent with the different cellular origin of ccRCC and pRCC subtypes on the one hand and with the chRCC and the RO subtypes on the other one. Thus, the first two originate from proximal tubule cells and the second two originate from intercalated cells of the distal nephron and collecting ducts. Interestingly, the most benign, subtype, RO, matches closer to NAT than any other subtype.

The use of mass spectrometry LFQ-based values as depicted in Figure VI.2B allowed us to identify a total of 850 differentially expressed proteins between groups. Further analysis of these 850 proteins address a number of proteins that specifically identify each RC subtype versus all other subtypes and NAT. Next, we transformed the LFQ values of this proteins in TPA-based concentrations [34], so a value in pmol of protein / mg of total protein are obtained. Using the TPA values and statistics, large sets of unique proteins constituting also unique panels to identify each tumor subtype were disclosed (Figure VI.2B). These proteins are present in all tumor subtypes, but the range of concentrations found are unique for each one, and so such proteins can be certainly used to classify the biopsies.

As a proof-of-concept, we have compared the 81 proteins found as unique for ccRCC with those reported in literature as depicted in table 1. Remarkably 46 of our suggested biomarkers have been described in literature as well. For instance, TYMP, PLN2 and CORO1A proteins have been already proposed as putative markers for ccRCC. TYMP protein is associated with pro-angiogenic and anti-apoptotic effects in cancer cells [55,56], and higher levels of TYMP in RCC tissue versus non-neoplastic kidney tissues have been described [57]. Such findings are in agreement with our study, which shows TYMP as one of the most up-regulated proteins in ccRCC, as shown in Figure VI.3. The same finding was obtained for PLIN2, known to be highly expressed in clear cells [23,26,48]. PLIN2 regulates lipids metabolism and storage and positively correlates with HIF-2α which drives cell proliferation and survival [58,59]. Accordingly, we were also able to confirm the highly expression of HIF-2α. As a last example, the levels of COROA1 protein were also found exclusively upregulated in clear cell tumors (Figure VI.3). COROA1 is crucial for cytoskeleton modulation. High levels of this protein have been reported in renal cancer cells with tumor-infiltrating lymphocytes [26].

In traditional immunohistochemistry, the positive stain of CA9 and the negative stain of AMACR and KRT7 are hallmarks of clear cell subtype. Also, when AMACR and KRT7 stains are positive the subtype is identified as pRCC [60]. Remarkably, our results confirm the utility of CA9 in ccRCC diagnosis (Figure VI.4C). Also, high AMACR and KRT7 levels were confirmed for pRCC (Figure VI.4C). As another example, immunoreactivity of VIM is also widely used in RCC diagnosis, as its stain pattern stands positive for the diagnostic of both ccRCC and pRCC and negative for chRCC. Interestingly, our results confirm these findings since VIM protein levels were found over expressed in ccRCC and pRCC versus NAT while in chRCC they were found slightly lower than in NAT (Figure VI.4C).

The advantage of TPA-based pathology stands in the possibility of quantifying hundreds of proteins at the same time, thus allowing to unequivocally diagnostic RCC cancer subtypes. The sample treatment is easy to reproduce, and modern mass spectrometry allows to quantify the proteins of an extract of any solid biopsy in just 15 min. Thus, in this work each RCC cancer subtype is hallmarked with a large set of unique proteins, which we show correlated with data available in literature and with the standard proteins used in current immunohistochemistry practice to diagnostic RC cancer subtypes. The correlation of our data with literature and classic immunohistochemistry corroborates the utility of the TPA based approach to unequivocally establish panels of proteins and their concentrations ranges to diagnostic renal carcinomas. In particular for PLIN2, TYMP, COROA1, and CA9, in diagnosing clear cell carcinoma. Additionally, the AMACR protein has also shown to be useful for the papillary subtype.

IHC analysis was used to validate the promising proteins achieved with the TPA-based methodology. Our preliminary data suggests that PLIN2 may emerge as a sensitive and specific marker for ccRCC strengthen the utility of TPA based methods for biomarker discovery.

We envision TPA based pathology as the next step in solid biopsy-based cancer diagnosis and prognosis and we anticipated this methodology is going to be implemented in all hospitals within the next 10 years.

ACKNOWLEDGMENTS

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura. Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV which is financed by national funds from FCT/MEC (UID/QUI/50006/ 2020). H. M. S. is funded by the FCT 2015 Investigator Program (IF/00007/2015). S. J. thanks FCT/MEC (Portugal) for her research contract as PhD student with the grant SFRH/BD/120537/2016. This project utilized the University of Pittsburgh Hillman Cancer Center shared resource facility (Cancer Genomics Facility) supported in part by award P30CA047904 (Dr. LaFramboise).

REFERENCES

[1] Kidney cancer statistics | World Cancer Research Fund https://www.wcrf.org/dietandcancer/cancer-trends/kidney-cancer-statistics (accessed Jul 21, 2020).

[2] Ferlay, J *et al*., Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer*, **2019**, *144*, 1941–1953.

[3] Cairns, P., Renal cell carcinoma. *Cancer Biomarkers*, **2011**, *9*, 461–473.

[4] van Oostenbrugge, TJ *et al*., Diagnostic Imaging for Solid Renal Tumors: A Pictorial Review. *Kidney Cancer*, **2018**, *2*, 79–93.

[5] Srigley, JR *et al*., Uncommon and recently described renal carcinomas. *Mod. Pathol.*, **2009**, *22*, S2–S23.

[6] Crumley, SM *et al*., Renal cell carcinoma: Evolving and emerging subtypes. *World J. Clin. Cases*, **2013**, *1*, 262–275.

[7] Hes, O *et al*., The 2012 ISUP Vancouver and 2016 WHO classification of adult renal tumors: changes for common renal tumors. *Diagnostic Histopathol.*, **2016**, *22*, 41–46.

[8] Varela, I *et al*., Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature*, **2011**, *469*, 539–542.

[9] Peña-Llopis, S *et al*., BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.*, **2012**, *44*, 751–759.

[10] Dalgliesh, GL *et al*., Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, **2010**, *463*, 360–363.

[11] Delahunt, B *et al*., Papillary renal cell carcinoma: a clinicopathologic and immunohistochemical study of 105 tumors. *Mod. Pathol.*, **1997**, *10*, 537–544.

[12] Hsieh, JJ *et al*., Renal cell carcinoma. *Nat. Rev. Dis. Prim.*, **2017**, *3*, 17009.

[13] Yusenko, M V., Molecular pathology of chromophobe renal cell carcinoma: A review. *Int. J. Urol.*, **2010**, *17*, 592–600.

[14] Kaliyappan, K *et al*., Applications of immunohistochemistry. *J. Pharm. Bioallied Sci.*, **2012**, *4*, 307–309.

[15] Gilbertson, JA *et al*., A comparison of immunohistochemistry and mass spectrometry for determining the amyloid fibril protein from formalin-fixed biopsy tissue. *J. Clin. Pathol.*, **2015**, *68*, 314–317.

[16] Bantscheff, M *et al*., Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.*, **2007**, *389*, 1017–1031.

[17] Wiśniewski, JR *et al*., A “Proteomic Ruler” for Protein Copy Number and Concentration Estimation without Spike-in Standards. *Mol. Cell. Proteomics*, **2014**, *13*, 3497–3506.

[18] Wiśniewski, JR *et al*., Absolute Proteome Analysis of Colorectal Mucosa, Adenoma, and Cancer Reveals Drastic Changes in Fatty Acid Metabolism and Plasma Membrane Transporters. *J. Proteome Res.*, **2015**, *14*, 4005–4018.

[19] Wiśniewski, JR *et al*., In-depth quantitative analysis and comparison of the human hepatocyte and hepatoma cell line HepG2 proteomes. *J. Proteomics*, **2016**, *136*, 234–247.

[20] Wiśniewski, JR *et al*., The Impact of High-Fat Diet on Metabolism and Immune Defense in Small Intestine Mucosa. *J. Proteome Res.*, **2015**, *14*, 353–365.

[21] Chinello, C *et al*., The proteomic landscape of renal tumors. *Expert Rev. Proteomics*, **2016**, *13*, 1103–1120.

[22] Valera, VA *et al*., Protein Expression Profiling in the Spectrum of Renal Cell Carcinomas. *J. Cancer*, **2010**, *1*, 184–196.

[23] Siu, KWM *et al*., Differential Protein Expressions in Renal Cell Carcinoma: New Biomarker Discovery by Mass Spectrometry. *J. Proteome Res.*, **2009**, *8*, 3797–3807.

[24] Masui, O *et al*., Quantitative Proteomic Analysis in Metastatic Renal Cell Carcinoma Reveals a Unique Set of Proteins with Potential Prognostic Significance. *Mol. Cell. Proteomics*, **2013**, *12*, 132–144.

[25] White, NMA *et al*., Quantitative proteomic analysis reveals potential diagnostic markers and pathways involved in pathogenesis of renal cell carcinoma. *Oncotarget*, **2014**, *5*, 506–518.

[26] Atrih, A *et al*., Quantitative proteomics in resected renal cancer tissue for biomarker discovery and profiling. *Br. J. Cancer*, **2014**, *110*, 1622–1633.

[27] Zhao, Z *et al*., Label-free quantitative proteomic analysis reveals potential biomarkers and pathways in renal cell carcinoma. *Tumor Biol.*, **2015**, *36*, 939–951.

[28] Neely, BA *et al*., Proteotranscriptomic Analysis Reveals Stage Specific Changes in the Molecular Landscape of Clear-Cell Renal Cell Carcinoma. *PLoS One*, **2016**, *11*, e0154074.

[29] Jorge, S *et al*., Development of a Robust Ultrasonic-Based Sample Treatment To Unravel the Proteome of OCT-Embedded Solid Tumor Biopsies. *J. Proteome Res.*, **2019**, *18*, 2979–2986.

[30] Cox, J *et al*., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **2008**, *26*, 1367–1372.

[31] Tyanova, S *et al*., The MaxQuant computational platform for mass spectrometry-based

shotgun proteomics. *Nat. Protoc.*, **2016**, *11*, 2301–2319.

[32] Tyanova, S *et al*., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods*, **2016**, *13*, 731–740.

[33] Tyanova, S *et al*., Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. In *Methods in Molecular Biology*; **2018**; Vol. 1711, pp. 133–148.

[34] Wiśniewski, JR., Label-Free and Standard-Free Absolute Quantitative Proteomics Using the “Total Protein” and “Proteomic Ruler” Approaches. In *Methods in Enzymology*; Academic Press Inc., **2017**; Vol. 585, pp. 49–60.

[35] Sun, CY *et al*., Proteomic analysis of clear cell renal cell carcinoma. Identification of potential tumor markers. *Saudi Med. J.*, **2010**, *31*, 525–532.

[36] Raimondo, F *et al*., Comparative membrane proteomics: a technical advancement in the search of renal cell carcinoma biomarkers. *Mol. Biosyst.*, **2015**, *11*, 1708–1716.

[37] Weißer, J *et al*., Quantitative proteomic analysis of formalin–fixed, paraffin–embedded clear cell renal cell carcinoma tissue using stable isotopic dimethylation of primary amines. *BMC Genomics*, **2015**, *16*, 559.

[38] Raimondo, F *et al*., Protein profiling of microdomains purified from renal cell carcinoma and normal kidney tissue samples. *Mol. BioSyst.*, **2012**, *8*, 1007–1016.

[39] Tostain, J *et al*., Carbonic anhydrase 9 in clear cell renal cell carcinoma: A marker for diagnosis, prognosis and treatment. *Eur. J. Cancer*, **2010**, *46*, 3141–3148.

[40] Lichtenfels, R *et al*., Systematic Comparative Protein Expression Profiling of Clear Cell Renal Cell Carcinoma. *Mol. Cell. Proteomics*, **2009**, *8*, 2827–2842.

[41] Kim, DS *et al*., Panel of Candidate Biomarkers for Renal Cell Carcinoma. *J. Proteome Res.*, **2010**, *9*, 3710–3719.

[42] Morgan, TM *et al*., Imaging the Clear Cell Renal Cell Carcinoma Proteome. *J. Urol.*, **2013**, *189*, 1097–1103.

[43] Raimondo, F *et al*., Proteomic analysis in clear cell renal cell carcinoma: identification of differentially expressed protein by 2-D DIGE. *Mol. Biosyst.*, **2012**, *8*, 1040.

[44] Jones, EE *et al*., MALDI imaging mass spectrometry profiling of proteins and lipids in clear cell renal cell carcinoma. *Proteomics*, **2014**, *14*, 924–935.

[45] Lebdai, S *et al*., Identification and validation of TGFBI as a promising prognosis marker of clear cell renal cell carcinoma. *Urol. Oncol. Semin. Orig. Investig.*, **2015**, *33*, 69.e11-69.e18.

[46] Johann, DJ *et al*., Combined Blood/Tissue Analysis for Cancer Biomarker Discovery: Application to Renal Cell Carcinoma. *Anal. Chem.*, **2010**, *82*, 1584–1588.

[47] Yao, Y *et al*., Metabolism-related enzyme alterations identified by proteomic analysis in human renal cell carcinoma. *Onco. Targets. Ther.*, **2016**, *9*, 1327.

[48] Song, Y *et al*., Data-Independent Acquisition-Based Quantitative Proteomic Analysis Reveals Potential Biomarkers of Kidney Cancer. *PROTEOMICS - Clin. Appl.*, **2017**, *11*, 1700066.

[49] Giribaldi, G *et al*., Proteomic identification of Reticulocalbin 1 as potential tumor marker in renal cell carcinoma. *J. Proteomics*, **2013**, *91*, 385–392.

[50] Oppenheimer, SR *et al*., Molecular Analysis of Tumor Margins by MALDI Mass Spectrometry in Renal Carcinoma. *J. Proteome Res.*, **2010**, *9*, 2182–2190.

[51] Gabril, M *et al*., S100A11 is a potential prognostic marker for clear cell renal cell carcinoma. *Clin. Exp. Metastasis*, **2016**, *33*, 63–71.

[52] Qi, Y *et al*., SERPINH1 overexpression in clear cell renal cell carcinoma: association with poor clinical outcome and its potential as a novel prognostic marker. *J. Cell. Mol. Med.*, **2018**, *22*, 1224–1235.

[53] Guo, T *et al*., Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.*, **2015**, *21*, 407–413.

[54] Marsaud, A *et al*., Dismantling papillary renal cell carcinoma classification: The heterogeneity of genetic profiles suggests several independent diseases. *Genes, Chromosom. Cancer*, **2015**, *54*, 369–382.

[55] Mori, S *et al*., Thymidine phosphorylase suppresses Fas-induced apoptotic signal transduction independent of its enzymatic activity. *Biochem. Biophys. Res. Commun.*, **2002**, *295*, 300–305.

[56] Bijnsdorp, I V. *et al*., Thymidine phosphorylase in cancer cells stimulates human endothelial cell migration and invasion by the secretion of angiogenic factors. *Br. J. Cancer*, **2011**, *104*, 1185–1192.

[57] Takayama, T *et al*., High Levels of Thymidine Phosphorylase as an Independent Prognostic Factor in Renal Cell Carcinoma. *Jpn. J. Clin. Oncol.*, **2006**, *36*, 564–569.

[58] Qiu, B *et al*., HIF2α-Dependent Lipid Storage Promotes Endoplasmic Reticulum Homeostasis in Clear-Cell Renal Cell Carcinoma. *Cancer Discov.*, **2015**, *5*, 652–667.

[59] Mylonis, I *et al*., Hypoxia-Inducible Factors and the Regulation of Lipid Metabolism. *Cells*, **2019**, *8*, 214.

[60] Alshenawy, HA., Immunohistochemical Panel for Differentiating Renal Cell Carcinoma with Clear and Papillary Features. *Pathol. Oncol. Res.*, **2015**, *21*, 893–899.

CHAPTER VII.

Phosphopeptide enrichment

Susana Jorge, Gonçalo Martins, José L. Capelo, William LaFramboise, Rajiv Dhir, Carlos Lodeiro, Hugo M. Santos*

Work in progresso

ABSTRACT

Phosphoproteins are estimated to be 30% of the entire proteome, playing a significant role in regulating several biological processes including, signaling networks. To analyze protein phosphorylation, mass spectrometry-based techniques have become the preferred tool. However, the low concentration of phosphopeptides represents a significant limitation. To overcome such obstacles, ion metal affinity chromatography, IMAC, is currently used. In this work, we synthesized new nano-sized IMACs from a polystyrene matrix. These polystyrene matrices were also evaluated using a 2^3 experimental design to unravel the optimum conditions to create the nanomaterial in a specific range of sizes. Metals used for the IMACs were titanium and lanthanum. Phosphopeptide enrichment efficiency of the new proposed material was assessed on a protein model, α-casein. A total of 99 phosphopeptides were identified for α-casein protein. Both IMACs proved to be efficient for phosphopeptide enrichment; however, further optimization is needed to achieve quantitative phosphopeptide recovery from complex samples and different biological conditions.

VII.1 INTRODUCTION

Phosphorylation is a post-translational modification (PTM) that plays a crucial role in many biological functions, and its deregulation is considered a hallmark of many human diseases, including cancer. Therefore, a comprehensive analysis of phosphorylated proteins and the signaling networks they are involved in is fundamental for understanding the related biological processes. Currently, high-throughput proteomics analysis is performed by mass spectrometry-based techniques. However, at a molecular level, the identification of phosphorylated peptides within complex mixtures pose a considerable technical challenge, as only around 30% of the entire proteome is estimated to be affected by this PTM. Currently, several approaches, including immunoprecipitation, affinity purification, and strong cation exchange chromatography, have been developed and used to pre-concentrate the low abundant phosphopeptides [1]. However, these enrichment techniques, in addition to being expensive and time-consuming, their applicability has only been proven in peptides containing phosphotyrosine [2]. Currently, the most commonly used alternative for phosphoproteomics analysis involves chromatography. Immobilized metal affinity chromatography (IMAC) or metal oxide affinity chromatography (MOAC) are two strategies based on the immobilization of with metal ions, such as Fe^{3+}, Ga^{3+}, or Ti^{4+}, or metal oxides, usually TiO_2, respectively, which have coordination capabilities and high preference for phosphate group that successfully enrich sample fractions with phosphopeptides [1]. However, some limitations including i) nonspecific

absorption of nonphosphorylated peptides containing multiple acidic amino acids, and ii) a phosphopeptide bias due to specific targeting amino acid sequences for each method [3,4]. To overcome such shortcomings the use of lanthanide ions such as La^{3+}, Eu^{3+}, Te^{3+}, and Ho^{3+}, has been proposed for selective phosphopeptide enrichment has been proposed [5–8]. Though these new materials have shown an increase in peptide recovery when compared to the conventional IMAC, the amount and ratio of phosphopeptides recovered remain unclear. Here, we propose the synthesis of a new material using monodisperse polystyrene nanoparticles (Ps-NPs) coordinated with lanthanide ions to increase the specificity and ratio of phosphopeptide enrichment. In this way, the high stable chemical properties of the Ps-NPs along with the large ratio of exposed surface area to volume and the hydrophilic surface that minimizes nonspecific adsorption, these nanomaterials also benefit from the strong phosphate binder ability of the lanthanide ions which provide, in addition, more coordination sites for phosphopeptide binding compared to any other transition-metal already used.

VII.2 EXPERIMENTAL SECTION

VII.2.1 REAGENTS

All reagents used, were of maximum purity available without any further purification. Ammonium persulfate (APS), polyvinyl alcohol (PVA), glycidyl methacrylate (GMA), trimethylolpropane trimethacrylate (TMPTMA), 2,2'-Azobis(2-methyl-propionitrile) (AIBN), acetone, titanium chloride and lanthanum chloride heptahydrate were purchased form Sigma-Aldrich. Sodium dodecyl sulphate (SDS), toluene and formaldehyde were purchased from Panreac. Styrene, 1-pentanol, tetrahydrofuran (THF) HCl and methanol were purchased from Carlo Erba Reagents. Ethylenediamine was purchased from Scharlau. Ethanol, acetonitrile (ACN), trifluoroacetic acid (TFA), phosphoric acid were purchased from Alfa Aesar. MilliQ-H2O (18.2 MΩ.cm @ 25 °C) was produced in a Simplicity Water Purification system from Millipore.

VII.2.2 TWO-LEVEL FACTORIAL DESIGN (2^3) OPTIMIZATION

Initially, different factors belonging to the synthesis of the monodisperse polystyrene nanoparticles (Ps-NPs) were optimized. Three reagents, APS, SDS, and styrene, were evaluated as two-levels (minimum or maximum), resulting in a 2^3 experimental design. The variables studied along with their values are shown in Table VII.1.

VII.2.3 SYNTHESIS OF POLYSTYRENE NANOPARTICLES

VII.2.3.1 Synthesis of Ps-NPs

The synthesis of monodisperse polystyrene nanoparticles was adapted from He *et al.* [9] and Zhou *et al.* [1]. A total of 8 experiments resultant of the two-level factorial design ($2^3 = 8$) were carried out in triplicate. For each experiment, the amount of APS and SDS was dissolved in water (48.3 mL for experiments 1 - 4, and 42 mL for experiments 5 – 8). The mixture was sonicated in an ultrasonic bath at 35 kHz for 10 min at 100% amplitude. The solution was then transferred to a round-bottom flask and heated to 80 ºC. After reaching 80 ºC, the styrene and 1-pentanol solution was continuously added by a pump at a flow rate of 250 µL/min. Afterward, the reaction system was maintained at 80 ºC for one h. The concentration of Ps-NPs present was calculated by drying 1 mL of solution and weighing the precipitate. At the end of the reaction, the size of the produced nanoparticles was analyzed by dynamic light scattering (DLS) using a Zetasizer (Malvern) equipment.

Table VII.1 – 2^3 factorial design experimental matrix.

Experiment	Variables					
	APS		SDS		Styrene + 1-pentanol	
1	-	0.004 g	-	0.2 g	-	0.7 + 0.01 mL
2	+	0.04 g	-	0.2 g	-	0.7 + 0.01 mL
3	-	0.004 g	+	2 g	-	0.7 + 0.01 mL
4	+	0.04 g	+	2 g	-	0.7 + 0.01 mL
5	-	0.004 g	-	0.2 g	+	7 + 0.1 mL
6	+	0.04 g	-	0.2 g	+	7 + 0.1 mL
7	-	0.004 g	+	2 g	+	7 + 0.1 mL
8	+	0.04 g	+	2 g	+	7 + 0.1 mL

VII.2.3.2 Synthesis of poly(GMA-*co*-TMPTMA)

Next, a 15 mL solution containing 450 mg of Ps-NPs and 1% (w/w) PVA and 0.25% (w/w) SDS was prepared and sonicated for 1 min using an ultrasonic bath at 100% ultrasonic amplitude and 35 kHz ultrasonic frequency. Afterward, this solution was transferred to a three-necked round-bottom flask. Then, an oil-phase solution containing 6.7 mL of GMA, 6.7 mL of TMPTMA, 140 mg of AIBN, and 16.6 mL of toluene in 150 mL of 1% (w/w) PVA and 0.25% (w/w) SDS solution. The oil-phase solution was emulsified by magnetic agitation at 1200 rpm for 15 min. Afterward, the emulsion was added to the seeds solution and sonicated for 1 min using

an ultrasonic bath at 100% ultrasonic amplitude and 35 kHz ultrasonic frequency to obtain a milk-like emulsion. Finally, the mixture was stirred at 200 rpm for 20 h at 30 ºC. Then the temperature was raised to 70 ºC, and the reaction was kept for another 24 h with stirring. Next, the solution was transferred to centrifugal tubes and washed seven times, with 20 mL of 50% (v/v) THF/ 50% (v/v) acetone. The solution was centrifuged between washes at 8500 g for 10 min, and the washing solution was discarded. Washed NPs were dried, resulting in a white powder.

VII.2.3.3 Synthesis of poly(GMA-*co*-TMPTMA-NH_2)

Seven mg of the white powder resulting from the previous step was mixed with 150 mL of ethylenediamine. The reaction was conducted for 3 h, at 80 ºC with constant stirring. The solution was transferred to centrifugal tubes and washed five times with 30 mL of 50% (v/v) ethanol. The washing solution was centrifuged at 8500 g for 10 min between washes. The resulting particles were allowed to dry.

VII.2.3.4 Coupling of phosphonate groups onto the NP

To functionalize the NP, 7 g of the compound obtained were dissolved in 100 mL of water. Next, 5.1 mL of phosphorous acid, 10 mL of 37% HCl and 8 mL of formaldehyde was added to the solution. The reaction was submitted to 100 ºC at 100 rpm with stirring. After 24 h, the solution was transferred to centrifugal tubes and washed five times with 30 mL of 50% (v/v) ethanol. The washing solution was centrifuged at 8500 g for 10 min between washes, and at the end, the solution was dried to obtain a light-yellow colored powder (pre-IMAC).

VII.2.3.5 Lanthanides and Ti^{4+} immobilization

The pre-IMAC (m = 100 mg) was incubated with 20 mL of the chosen metal (0.09 M in 20% (v/v) HCl) for 8 h at RT under constant stirring (200 rpm). Afterward, the solution was transferred to centrifugal tubes and washed five times with 10 mL of 30% (v/v) of ACN and 0.1% (v/v) of TFA. The washing solution was centrifuged at 8500 g for 10 min between washes, and the washing solution discarded. Finally, 10 mL of 30% (v/v) of ACN and 0.1% (v/v) of TFA was added to 100 mg of IMACs.

Overall, the reaction scheme is represented in Figure VII.1.

VII.2.4 STANDARD PROTEIN DIGESTION

A protein standard, α-casein was initially used to test the efficiency of the nano-IMACs produced. Each standard protein solution was prepared to a final concentration of 1 mg/mL in 25 mM AMBIC containing 0.001% (w/v) SDS.

VII.2.4.1 Protein reduction and alkylation

Disulfide bonds were reduced and alkylated as described in Jorge *et al.* [10]. Briefly, 40 µl of standard protein solution was reduced with 4 µL of 110 mM DTT and incubating for 45 min at 37 ºC. Then, the resulting cysteines were blocked with 4 µL of 400 mM IAA at RT for 45 min. The samples were diluted to a final volume of 200 µL with 25 mM AMBIC / 2% (v/v) ACN.

VII.2.4.2 Protein digestion

Samples were digested in-solution by the ultrasonic method as previously described [10].

VII.2.5 PHOSPHOPEPTIDE ENRICHMENT

Phosphopeptide enrichment was carried out using a in house gel-loading spin-tip assembly. Briefly, C8 membrane disks (1 mm diameter) were place into a gel-loading tip and washed with 20 µL of methanol. Then, 50 µL of slurry IMAC (10 mg/mL) were added onto the gel-loading tip until reaching a packed column of 5 mm height. Between each addition, the tip column was centrifuged at 200 g for 2.5 min. The resulting IMAC spin tips were equilibrated with 50 µL of loading buffer (80% (v/v) ACN / 6% (v/v) TFA) and centrifuged at 200 g for 2.5 min, twice. 100 µL of digested protein sample (20 µg) was added to the IMAC spin tip and centrifuged at 20 g for 10 min. The IMAC column was washed by centrifugation at 100 xg for 3 min firstly with 50% (v/v) ACN / 0.5% (v/v) TFA / 200 mM NaCl buffer and then with 50% (v/v) ACN / 0.5% (v/v) TFA buffer. Phosphopeptides were eluted with 20 µL of 80% (v/v) ACN / 2% (v/v) TFA, and recovered in a centrifugal tube containing 35 µL of 10% (v/v) FA after centrifugation at 100 xg for 3 min. In the end, 3 µL of FA was added to the recovered sample to lower the pH. Before MS analysis, samples were purified by zip-tip.

VII.2.6 NANO-LC-ESI-MS/MS ANALYSIS

The LC-MS/MS analysis was carried out using an Ultimate 3000 nLC coupled to an UHR-QqTOF IMPACT HD (Bruker Daltonics) with a CaptiveSpray ion source (Bruker Daltonics). Phosphopeptide fraction were resuspended in 15 µL of 3% ACN/0.1% (v/v) aqueous formic acid. Seven µL of each sample was loaded into a trap column Acclaim PepMap100, 5 µm, 100 Å, 300 µm i.d. × 5 mm and desalted for 5 min with 3% B (B: 90% ACN/0.1% FA) at a flow rate of 15 µL/min. Chromatographic separation was carried out using an analytical column Acclaim™ PepMap™ 100 C18, 2 µm, 0.075 mm i.d x 150 mm with a linear gradient at 300 nL/min (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) ACN and 0.08% (v/v) FA), 0-5 min with 3% of mobile phase B, 5-90 min from 3% to 35% of mobile phase B, 90-100 min linear gradient from 35% to 95% of mobile phase B, 100-110 with 95% of mobile phase B. The total run time was 130 min. For each sample, two replicate injections were performed. Chromatographic separation was carried out at 35 ºC. MS acquisition was set to cycles of MS (2

Hz), followed by MS/MS (8 – 32 Hz), cycle time 3.0 seconds, with active exclusion (precursors were excluded from precursor selection for 0.5 min after acquisition of 1 MS/MS spectrum, intensity threshold for fragmentation of 2500 counts). Together with active exclusion set to 1, reconsider precursor if the intensity of a precursor increases by a factor of 3, this mass will be taken from temporary exclusion list and fragmented again, ensuring that fragment spectra were taken near to the peak maximum. All spectra were acquired in the range 150–2200 m/z. Raw data were processed in DataAnalysis 4.2 and subsequently exported to Protein-Scape 4.0 for automated protein identification. For protein identification, CID-MS2 spectra were first searched against the *Other Mammalia* (46,036 sequences) subset of the SwissProt database 57.15 (515,203 sequences; 181,334,896 residues), using the Mascot search engine (V. 2.3.02) with the following parameters: (i) two missed cleavage; (ii) fixed modifications: carbamidomethylation (C); (iii) variable modifications: oxidation of methionine, Acetyl (Protein N-term), Phosphorylation (ST), Phosphorylation (Y), (vi) peptide mass tolerance up to 20 ppm, (v) fragment mass tolerance 0.05 Da (vi) Adjust FDR 1%.

VII.3 RESULTS AND DISCUSSION

VII.3.1 POLYSTYRENE SYNTHESIS OPTIMIZATION

The synthesis of polystyrene seeds started with a polymerization step via thermal-initiated free radical formation. A 2^3 experimental design was executed to optimize this first step of the production of the Ps-NPs, resulting in eight different experiences. The factors under optimization were the APS as the initiation of the polymerization reaction, the SDS as the surfactant, and the styrene as a functional monomer. The amount of the three reagents are described in Table VII.1., For each batch, three replicates were assessed in terms of size, population size, and polydispersity. The results are displayed in Table VII.2.

The size, percentage of the population, and polydispersity (PDI) along with PS seeds concentration are some critical attributes to obtain nano-IMACs [11]. To this end, the size of PS seeds was chosen between 20-70 nm, as the swelling step will increase the particles' size. Seeds with sizes representing more than 90% of the population together with PDI inferior to 0.3 (0 for totally monodisperse to 1 for totally polydisperse) were selected. Taking these thresholds into consideration, experiments 1, 3, and 5 were excluded. Regarding the concentration of nanoparticles, presented in Table VII.2, the critical concentration was established above 40 mg/mL, producing material for three independent syntheses.

As might be seen in Table VII.2, only experiments 6, 7, and 8 meet the criteria for further use in IMAC preparation. It is noteworthy to mention that the concentration was slightly lower for the same amount of monomer styrene (7 mL) when the minimal amount of APS was used (experiment 7). Also, in experiments 1, 3, and 5, where a minimal amount of initiator was used, the produced seeds were polydisperse, suggesting the importance of initiator

concentration to fine-tune the size and monodispersity of the particles. For experiments 6 and 8, both had rendered comparable results, 39 ± 1 nm (PDI 0.21 ± 0.01) and 43 ± 5 nm (PDI 0.28 ± 0.03), respectively. However, population size % was discreetly higher in experiment 6 (99.3 ± 0.3%) than experiment 8 (92 ± 8%). Overall, the reproducibility of experiment 6 was higher; therefore, justifying its use for further steps.

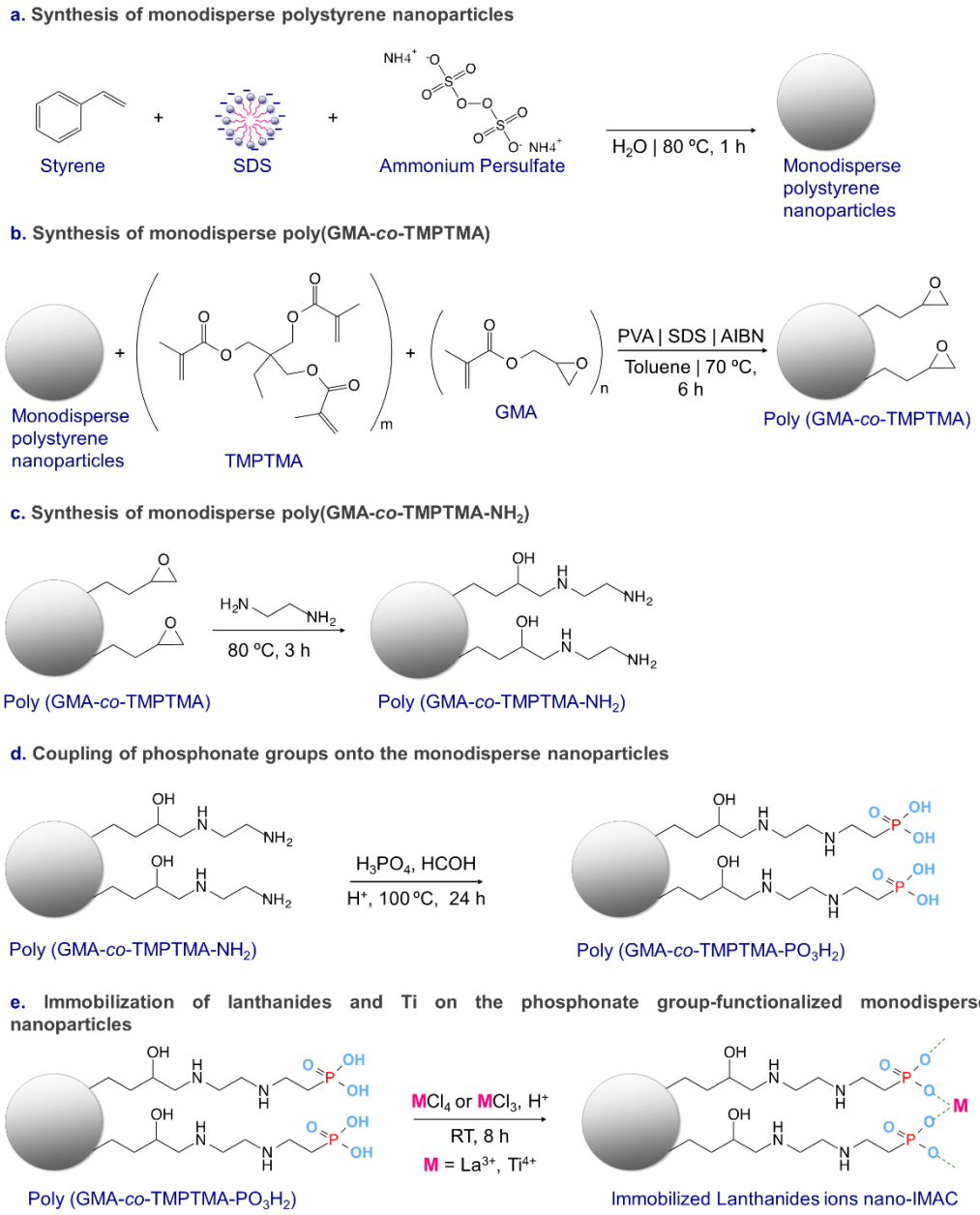

Figure VII.1 – Immobilized lanthanide ions and Ti^{4+} nano-IMACs synthesis. This scheme was adapted from [9] *and* [1].

VII.3.2 SYNTHESIS OF IMAC

The addition of a flexible linker attached to the PS seeds was achieved via a swelling step, which is critical for the final size of the IMACs. After seed swelling, the reactional mixture involves a functionalization step with phosphate groups followed by the metal's immobilization; in this work, lanthanum and titanium were tested. Figure VII.2 shows a representative SEM image of the lanthanum IMAC

Table VII.2 – Size, population, and PDI of PS seeds of each experiment.

Experiment	Size (nm)	Population Size (%)	PDI	Concentration (mg/mL)	Selective criteria
1	132 ± 116	94 ± 5	0.3 ± 0.2	- n.d.	Excluded*
2	18.7 ± 0.4	100 ± 0	0.09 ± 0.00	9 ± 2	Excluded*
3	1.8 ± 0.1	56 ± 6	0.4 ± 0.1	- n.d.	Excluded*
4	13.6 ± 0.9	96 ± 3	0.23 ± 0.01	22 ± 4	Excluded*
5	1473 ± 1729	98 ± 4	0.31 ± 0.07	- n.d.	Excluded*
6	39 ± 1	99.3 ± 0.8	0.21 ± 0.01	105 ± 33	Included
7	49.6 ± 0.9	100 ± 0	0.20 ± 0.00	69 ± 10	Included
8	43 ± 5	92 ± 8	0.28 ± 0.03	110 ± 3	Included

*n.d. = not defined. * = Failed parameters are marked as darker squares. Inclusion parameters: particle size < 70 nm; population size > 90%; PDI < 0.3; critical concentration > 40 mg/mL*

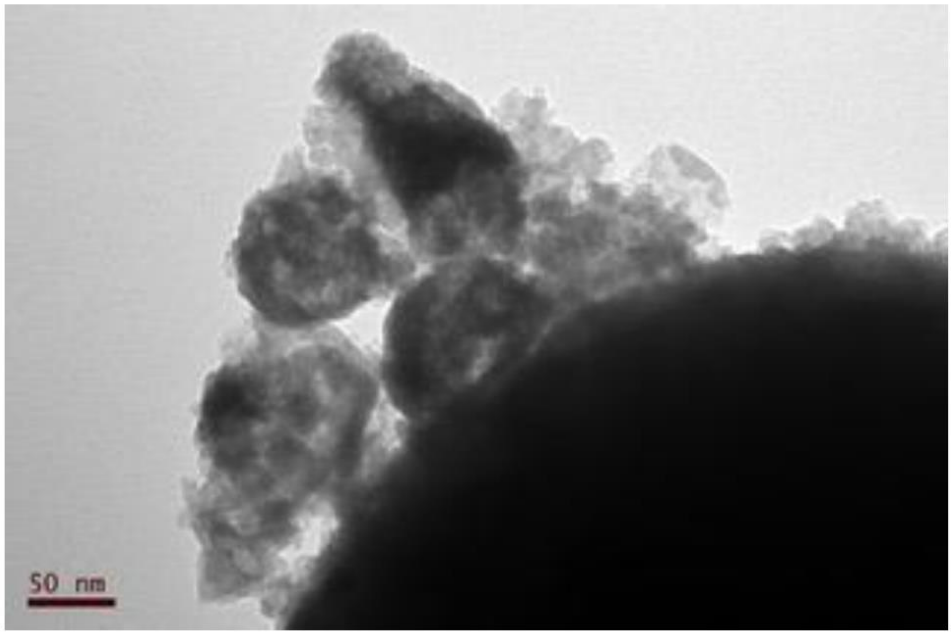

Figure VII.2 – SEM image of lanthanum IMAC produced.

VII.3.3 PHOSPHOPEPTIDE ENRICHMENT

Phosphopeptide enrichment of α-casein digest was caried out using the $Ti^{4'}$ and La^{3+} IMACs. The number of unique peptides recovery by the two IMAC was very similar, 79, and 76 unique phosphopeptides in Ti^{4+}IMAC and La^{3+}IMAC, respectively. This result was noteworthy higher when compared to the 18 or 20 phosphopeptides as expected according to Zhou *et al.* and Yu *et al.*, suggesting a significant improvement of these new materials in phosphopeptide enrichment. With the MS analysis, it was also assessed the presence of two proteoforms of protein used, the α-casein 1 and α-casein 2. In fact, the existence of α-casein 1 (24.5 kDa) and α-casein 2 (26 kDa) has been described elsewhere, and with our technology in combination with the use of high-resolution mass spectrometry we were able to identify both isoforms in our study. In Figure VII.3 is represented the recovery of phosphopeptides using the two IMAC species

With this experiment, the two types of IMAC have recovered similar number of phosphopeptides, Ti^{4+} IMAC = 50 and La^{3+} IMAC = 48 for α-casein 1 and Ti^{4+} IMAC = 28 and La^{3+} IMAC = 27 for α-casein 2. Overall, approximately 70% of the phosphopeptides were commonly identified in both IMACs species. A phosphopeptide preference was observed between the two IMACs for both proteins since only about 20% of the adsorbed peptides were non-phosphorylated.

To evaluate the efficiency of the phosphopeptide binding of both IMACs, the flow-through fraction was also analyzed (Figure VII.4). As might be seen in the Venn diagrams represented in Figure VII.4, some phosphopeptides were not trapped by the IMAC nanoparticle.

Although the majority of the phosphopeptides were retained in the enriched fraction, a significant portion of those peptides were still missed in the flow-through. The phosphopeptide fraction, fraction 1, was analyses by MS (Figure VII.5). Additionally, the flow through was passed through the Ti^{4+}IMAC materials twice, resulting in fraction 2 and 3, Figure VII.5. As might be seen, a considerable amount of peptide signals was still detected in the succeeding fractions. Consequently, the quantification of such phosphopeptides is compromised and additional optimization of the IMAC pre-concentration conditions, including peptide-IMAC ratio, is still required to guarantee quantitative recovery of phosphopeptides.

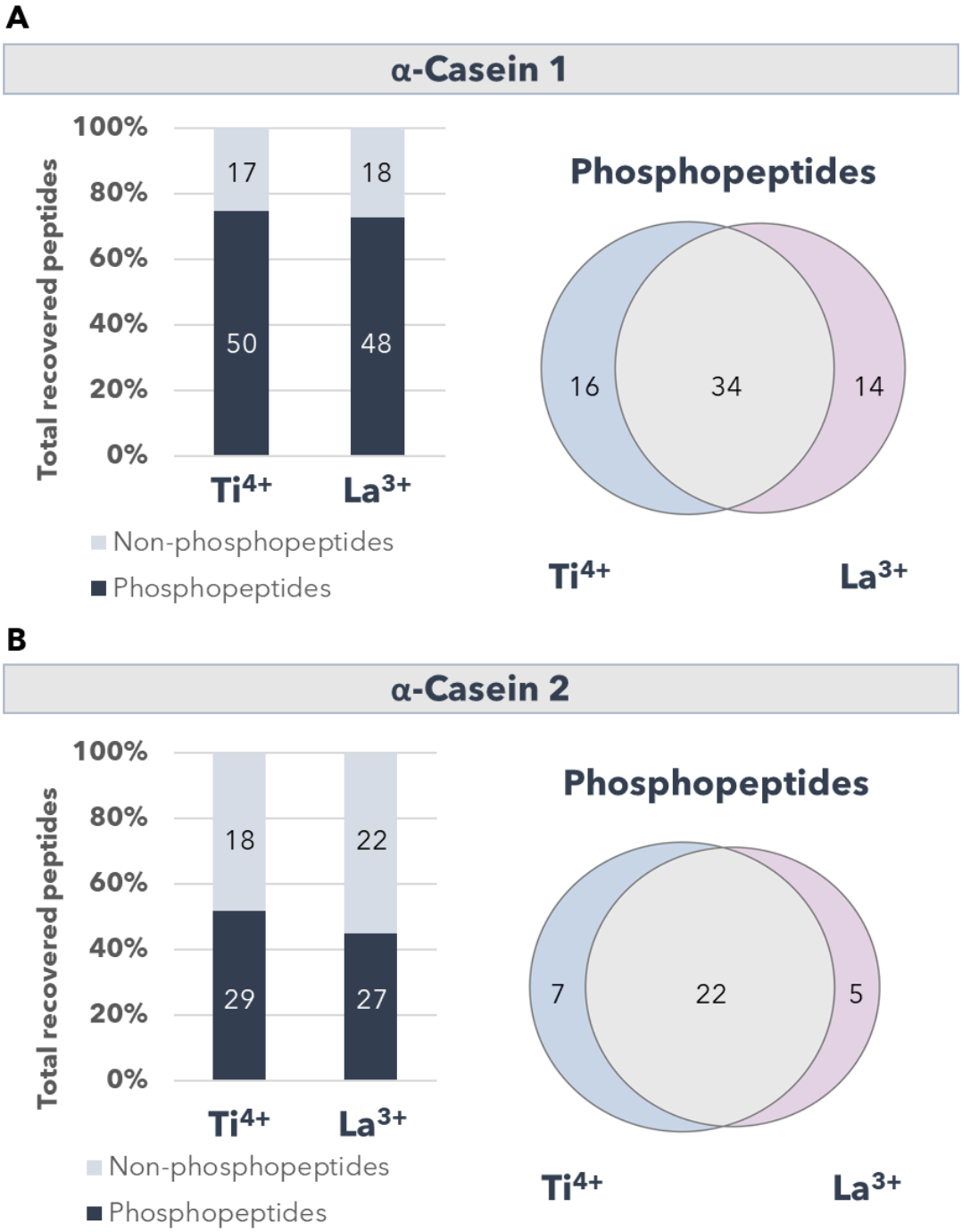

Figure VII.3 – Analysis of the number of phosphopeptides and nonphosphopeptides identified using titanium and lanthanum IMACs nanoparticles for the enrichment of α-casein 1 and 2 proteins.

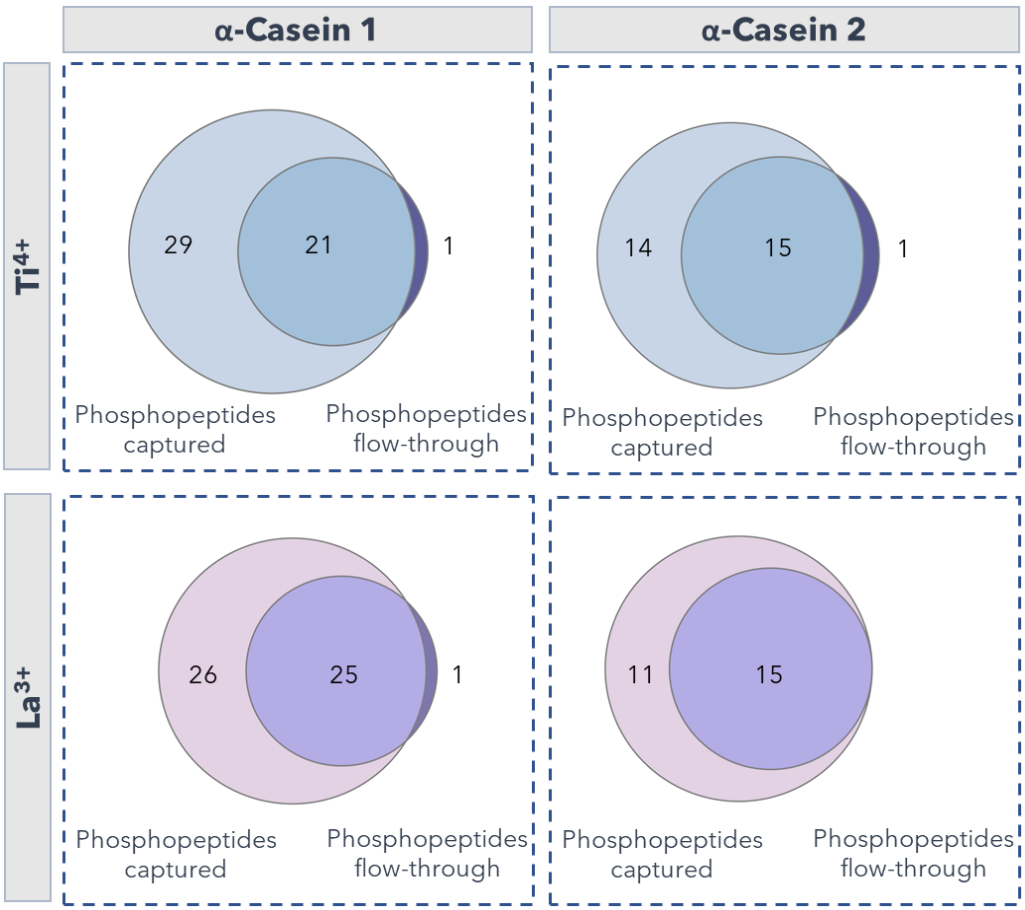

Figure VII.4 – Comparison of the number of phosphopeptides identified captured by the IMAC and eluted in the flow-through fractions for both titanium and lanthanum IMACs nanoparticles and for the enrichment of both α-casein proteins.

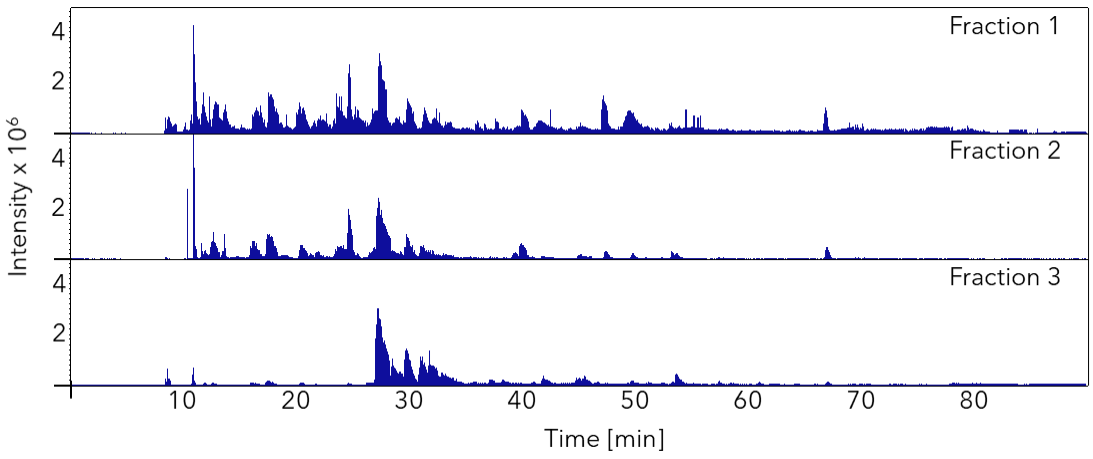

Figure VII.5 – MS/MS Total ion current (TIC) chromatogram of phosphopeptide enrichment fractions using Ti^{4+}nano-IMAC. The peptide pool resultant from the digestion of the α-casein was passed through of the nano-IMAC, fraction 1 resulted from the first enrichment. Then the flow-through was passed once again through the sorbent twice, resulting in fraction 2 and 3.

VII.4 CONCLUSIONS AND FUTURE WORK

Protein phosphorylation is reversible and one of the most important post-translational modifications. MS-based techniques have become the most convenient, high throughput, and efficient approach for proteomics studies. Being also the leading choice for post-translational modification (PTM) proteomics analysis due to their high sensitivity and accuracy. However, the detection and identification of phosphoproteins by MS is still a challenging task, mainly due to their low abundance, lower ionization efficiency compared with non-phosphorylated proteins, and suppression caused by the abundant non-phosphopeptides. Several sample treatments approaches have been proposed to overcome such pitfalls, including phosphopeptide enrichment using IMACs. Here we proposed a new synthesis and family of nano-IMAC materials based on titanium (Ti^{4+}) and lanthanum (La^{3+}) metal ions. The maximal amount of the initiator APS and the monomer styrene combined with the minimal amount of the surfactant SDS was chosen as the ideal condition of monodisperse polystyrene nanoparticles. The two classes of IMAC revealed similar efficiency, resulting in a total of 79 and 76 phosphopeptides identified by Ti^{4+}IMAC and La^{3+}IMAC, respectively. Some phosphopeptides have shown a metal preference for binding; nevertheless, approximately 70% of the peptides identified were shared by both nanomaterials. Although a significant improvement was achieved with both nano-IMAC materials, a quantification gap was verified as a portion of phosphopeptides were still flushed away in the flow-through. As perspectives for the future, optimization of the IMAC material, and enrichment conditions are needed in order to ensure single-shot quantitative recovery of phosphopeptides from complex samples and in different biological conditions for true-quantitative phosphorylation analysis.

ACKNOWLEDGMENTS

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura. Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV which is financed by national funds from FCT/MEC (UID/QUI/50006/ 2020). H. M. S. is funded by the FCT 2015 Investigator Program (IF/00007/2015). S. J. thanks FCT/MEC (Portugal) for her research contract as PhD student with the grant SFRH/BD/120537/2016.

REFERENCES

[1] Zhou, H *et al*., Robust phosphoproteome enrichment using monodisperse microsphere–based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.*, **2013**, *8*, 461–480.

[2] Fíla, J *et al*., Enrichment techniques employed in phosphoproteomics. *Amino Acids*, **2012**, *43*, 1025–1047.

[3] Lai, AC-Y *et al*., Complementary Fe 3+ - and Ti 4+ -immobilized metal ion affinity chromatography for purification of acidic and basic phosphopeptides. *Rapid Commun. Mass Spectrom.*, **2012**, *26*, 2186–2194.

[4] Matheron, L *et al*., Characterization of Biases in Phosphopeptide Enrichment by Ti 4+ -Immobilized Metal Affinity Chromatography and TiO 2 Using a Massive Synthetic Library and Human Cell Digests. *Anal. Chem.*, **2014**, *86*, 8312–8320.

[5] Pink, M *et al*., Precipitation by lanthanum ions: A straightforward approach to isolating phosphoproteins. *J. Proteomics*, **2011**, *75*, 375–383.

[6] Mirza, MR *et al*., A novel strategy for phosphopeptide enrichment using lanthanide phosphate co-precipitation. *Anal. Bioanal. Chem.*, **2012**, *404*, 853–862.

[7] Mirza, MR *et al*., A new type of metal chelate affinity chromatography using trivalent lanthanide ions for phosphopeptide enrichment. *Analyst*, **2013**, *138*, 2995–3004.

[8] Güzel, Y *et al*., Highly selective recovery of phosphopeptides using trypsin-assisted digestion of precipitated lanthanide-phosphoprotein complexes. *Analyst*, **2013**, *138*, 2897–2905.

[9] He, G *et al*., Synthesis of Polystyrene and Polystyrene/Poly(methyl methacrylate) Nanoparticles. *Macromol. Rapid Commun.*, **2004**, *25*, 1545–1548.

[10] Jorge, S *et al*., Unparalleled sample treatment throughput for proteomics workflows relying on ultrasonic energy. *Talanta*, **2018**, *178*, 1067–1076.

[11] Clayton, KN *et al*., Physical characterization of nanoparticle size and surface modification using particle scattering diffusometry. *Biomicrofluidics*, **2016**, *10*, 054107.

CHAPTER VIII.

Conclusion and future perspectives

VIII.1 CONCLUSIONS

- A novel ultrasonic-based pipeline was accomplished to interrogate the proteome of OCT-embedded tissue biopsies. The best condition achieved for the OCT cleaning was the use of US bath 35 kHz at 100% of amplitude for 2 min of ultrasonic duty time. Along with the OCT cleaning step the workflow comprised the solid-liquid protein extraction and the protein cleavage steps under the effects of an ultrasonic field. This methodology was later used as the standard one and then applied to all renal tissue biopsies interrogated in this doctoral thesis.

- A new analytical approach able to discriminate between solid biopsies of chRCC and RO was proposed. Based on a peptide sequential extraction hyphenated to MALDI-based profiling, the method comprised i) ultrasonic extraction of proteins from solid biopsies; ii) protein depletion with acetonitrile; iii) ultrasonic assisted in-solution digestion using magnetic nanoparticles with immobilized trypsin; iv) C18 tip-based preconcentration of peptides; v) sequential extraction of the peptides with acetonitrile, and; vi) MALDI-snapshotting of the extracts. Using the 60% ACN extraction, this methodology was capable of delineating the complex proteomes of each tumor subtype. An unsupervised clustering of these samples rendered a discriminative tool for chromophobe and oncocytoma specimens holding the promise of clinical differentiation between a malignant tumor and benign neoplasm. Also, this approach may be extended to any disease presenting similar pathological profiles but with different outcomes.

- A label-free quantitative proteomic analysis using high resolution mass spectrometry to discriminate tissue biopsies diagnosed with chRCC and RO was also presented. Based on protein expression profiles a successful differentiation of chRCC, RO and NAT was obtained after the application of an unsupervised clustering analysis, suggesting that those neoplasms are different enough to efficiently classify them and distinguish from normal tissue. In addition, a selected panel of 109 proteins was found with significant power of differentiation. Two proteins belonging to the proposed panel of biomarkers candidates, the HK1 and the LAMP1, are proposed as novel immunohistochemistry biomarkers to differentiate chRCC, RO and NAT. The large cohort of differentially expressed proteins opens new avenues for immunohistochemistry, as confirmed by the validation of LAMP1 and HK1 proteins.

- Biochemical pathways involving energy metabolism were revealed dysregulated in chRCC and RO. However, chRCC displayed a specific association with phagosome maturation, while RO had a greater representation of mitochondrial pathways.

- A TPA-based comprehensive study, including three RCC subtypes, the clear cell, the papillary and the chromophobe, along with the benign renal oncocytoma, was developed. The analysis, through a label free quantification approach, of the whole proteome of each tumor type

rendered 850 differentially expressed proteins between samples. After transformation into absolute protein concentrations through the total protein approach (TPA) method, a top 24 protein panel was able to differentiate each tumor subtype, displaying their utility for clinical diagnosis. In this work, The TPA quantitative proteomic approach was used for the first time to identify an effective panel of proteins and their concentration ranges to diagnose and characterize diverse renal neoplasms.

- The use of the TPA approach was found promising in rendering potential biomarkers for immunohistochemistry as demonstrated with PLIN1 via pathological analysis.

- A new family of nano-IMACs sorbents, using two different metals, was designed for phosphopeptide enrichment. The nano-sized IMACs revealed an improvement of efficacy allowing the recovery a total of 99 phosphopeptides. However, the absolute recovery of the phosphopeptides in just one fraction was not achieve, missing some phosphopeptides in the flow-through.

VIII.2 FUTURE PERSPECTIVE

Regarding the developments made in renal neoplasms diagnosis it is expected soon the validations of more protein biomarkers candidates through IHC analysis (a work which is currently being done by the pathology team of Dr. Rajiv Dhir, at the University Pittsburg Medical Center). The validation of such immunobiomarkers will help to develop an effective and proper protein panel to accurately differentiate not only the renal neoplasms studied in this thesis but also other subtypes. Therefore, we expect a revolution in the area of cancer renal pathology on the near future. Also, the implementation of our TPA approach to other type of neoplasia will open a new field or research to pathology in general.

The use of mass spectrometry either in short runs (15 min) with the latest TIMS-TOF technology or the use of the fast SRM or the MRM techniques will slowly, but constantly, be implemented in the pathology departments as a faster way to diagnostic solid biopsies. Furthermore, the potential link between the biomarkers found in solid biopsies with the ones present in urine is expected to help to avoid or to reduce the need for solid biopsies to diagnose and prognose cancer, as we visualize urine as the ultimate gold sample for diagnosis.

The advent of paper-based technology for diagnosis will also allow point-of care medical diagnosis. The new biomarkers found will allow the easy development of these devices.

The deep proteomics analysis of data retrieved from the samples used in this work reveals the promise of disclosing new information of the mechanisms undelaying tumorigenesis of the clear cell and the papillary RCC subtypes. A better understanding may also bring the possibility of finding new therapeutic targets. In this sense, the work in phosphorylation is expected to be key, as we expect to be able to quantitatively recover more than 90% of the phosphoproteome through the new immobilized lanthanide ions nano-IMAC and and Ti^{4+} nano-IMAC.

VIII.3 THESIS OUTPUT

VIII.3.1 PEER-REVIEWED MANUSCRIPTS PUBLISHED IN INTERNATIONAL SCIENTIFIC JOURNAL

2. *Title:* Ultrasonic-assisted extraction hyphenated to MALDI-based profiling holds the promise of distinguish renal oncocytoma from chromophobe renal cell carcinoma

Authors: Susana Jorge, Kevin Pereira, Hugo López-Fernández, William A. LaFramboise, Rajiv Dhir, Javier Fernández-Lodeiro, Carlos Lodeiro, Hugo M. Santos, José Luis Capelo.

Journal: Talanta, 2020, 206, 120180. Impact factor (2019): 5.339. Rank position (2019): 11/86, Q1, 87.791%.

*https://doi.org/*10.1016/j.talanta.2019.120180.

Presented in Chapter IV

1. *Title:* Development of a robust ultrasonic-based sample treatment to unravel proteome of OCT-embedded solid tumor biopsies

Authors: Susana Jorge, José Luis Capelo, William A. LaFramboise, Rajiv Dhir, Carlos Lodeiro, Hugo M. Santos.

Journal: Journal of Proteome Research, 2019, 18 (7), 2979-2986. Impact factor (2019): 4.074. Rank position (2019): 12/77, Q1, 85.065%.

*https://doi.org/*10.1021/acs.jproteome.9b00248

Presented in Chapter III

VIII.3.2 MANUSCRIPTS IN PREPARATION

2. *Title:* Towards TPA-based pathology

Authors: Susana Jorge, José Luis Capelo, William A. LaFramboise, Rajiv Dhir, Jacek R. Wiśniewski, Carlos Lodeiro, Hugo M. Santos.

In preparation

Presented in Chapter VI

1. *Title:* The proteome of tumour biopsies as a tool to distinguish chromophobe renal cell

carcinoma and renal oncocytoma

Authors: Susana Jorge, José Luis Capelo, William A. LaFramboise, Rajiv Dhir, Carlos Lodeiro, Hugo M. Santos.

In preparation

Presented in Chapter V

VIII.3.3 PARTICIPATION IN NATIONAL AND INTERNATIONAL CONFERENCES

Oral communications

1. *Title:* Underlying proteomic signatures to distinguish chromophobe renal cell carcinoma and renal oncocytoma

Authors: Susana Jorge, José Luis Capelo, William A. LaFramboise, Rajiv Dhir, Carlos Lodeiro, Hugo M. Santos.

Congress: 6th International Caparica Conference on Analytical Proteomics (ICAP 2019). July 8th-11th, 2019. Capuchos – Caparica, Portugal

Poster presentations

9. *Title:* Role of Hexokinase-1 and LAMP-1 immunohistochemistry to differentiate oncocytoma and chromophobe renal cell carcinoma from other renal neoplasms

Authors: Swati Satturwar, Dimitrios Korentzelos, Susana Jorge, Gabriela Quiroga-Garza, Sheldon Bastacky, Hugo, M. Santos, Jose L. Capelo, Rajiv Dhir

Will be present at: USCAP 110th Annual Meeting, March 13-18th, 2021, virtual meeting, USA

8. *Title:* Renal cell tumors diagnosis towards TPA-based pathology

Authors: Susana Jorge, J.L. Capelo, William LaFramboise, Rajiv Dhir, Jacek R. Wiśniewski, Carlos Lodeiro, H.M. Santos.

Congress: 3rd International Conference NOVAhealth Chronic Disease and Infection. October 8th, 2020. Virtual meeting, Portugal.

7. *Title:* Translating tumor proteome to accurately diagnose chromophobe renal cell

carcinoma and renal oncocytoma

Authors: Susana Jorge, Rajiv Dhir, William LaFramboise, Carlos Lodeiro, J.L. Capelo, H.M. Santos.

Congress: 3rd International Caparica Christmas Conference on Translational Chemistry (IC3TC 2019). December 2nd-5th, 2019. Costa da Caparica, Portugal.

6. *Title:* Deciphering differences between chromophobe renal cell carcinoma and renal oncocytoma through an integrative proteogenomics analysis

Authors: Susana Jorge, Rajiv Dhir, William LaFramboise, Carlos Lodeiro, J.L. Capelo, H.M. Santos.

Congress: IV NOVAhealth Genetics Workshop, March 21st, 2019. Lisbon, Portugal

5. *Title:* Pinpointing protein differences between chromophobe renal cell carcinoma and renal oncocytoma

Authors: Susana Jorge, Rajiv Dhir, William LaFramboise, Carlos Lodeiro, J.L. Capelo, H.M. Santos.

Congress: XXI Encontro Luso-Galego de Química, November 21st-23rd, 2018. Porto, Portugal

4. *Title:* Unique protein signatures unravel chromophobe renal cell carcinoma and renal cell oncocytoma

Authors: Susana Jorge, Carlos Lodeiro, J.L. Capelo, H.M. Santos, Rajiv Dhir, William LaFramboise.

Congress: 30th Annual UPMC Hillman Cancer Center Scientific Retreat. June 20th – 22th, 2018. University of Pittsburgh, Greensburg, USA

3. *Title:* Overcoming OCT drawbacks for renal tissue proteomics

Authors: Susana Jorge, Rajiv Dhir, William LaFramboise, Carlos Lodeiro, J.L. Capelo, H.M. Santos.

Congress: XII EUPA CONGRESS Translating genomes into biological functions. June 16th – 20th, 2018. Santiago de Compostela, Spain

2. *Title:* The need to distinguish chromophobe renal cell carcinoma and renal oncocytoma: a mass spectrometry-based proteomic approach using OCT-embedded tissues

Authors: Susana Jorge, Rajiv Dhir, William LaFramboise, Carlos Lodeiro, J.L. Capelo, H.M. Santos.

Congress: 2nd Edition of the Instruct Training course: "From protein structure to biological function through interactomics – an integrated view". February 5th – 9th, 2018. UC-Biotech, Cantanhede, Portugal

1. *Title:* Enhancing protein recovery from OCT-embedded tissues for mass spectrometry proteomics analysis

Authors: Susana Jorge, Rajiv Dhir, William LaFramboise, Elisabete Oliveira, Javier Fernandéz-Lodeiro, Carlos Lodeiro, J.L. Capelo, H.M. Santos.

Congress: 2nd International Caparica Christmas Conference on Translational Chemistry (IC3TC 2017). December 4th-7th, 2017. Capuchos – Caparica, Portugal

Awards

1. Excellent Shotgun Presentation Award

Congress: 6th International Caparica Conference on Analytical Proteomics (ICAP 2019). July 8th-11th, 2019. Capuchos – Caparica, Portugal

Title: Underlying proteomic signatures to distinguish chromophobe renal cell carcinoma and renal oncocytoma

CHAPTER IX.

Supplementary information

IX.1 SUPPLEMENTARY MATERIAL CHAPTER III

IX.1.1 SUPPLEMENTARY FIGURES

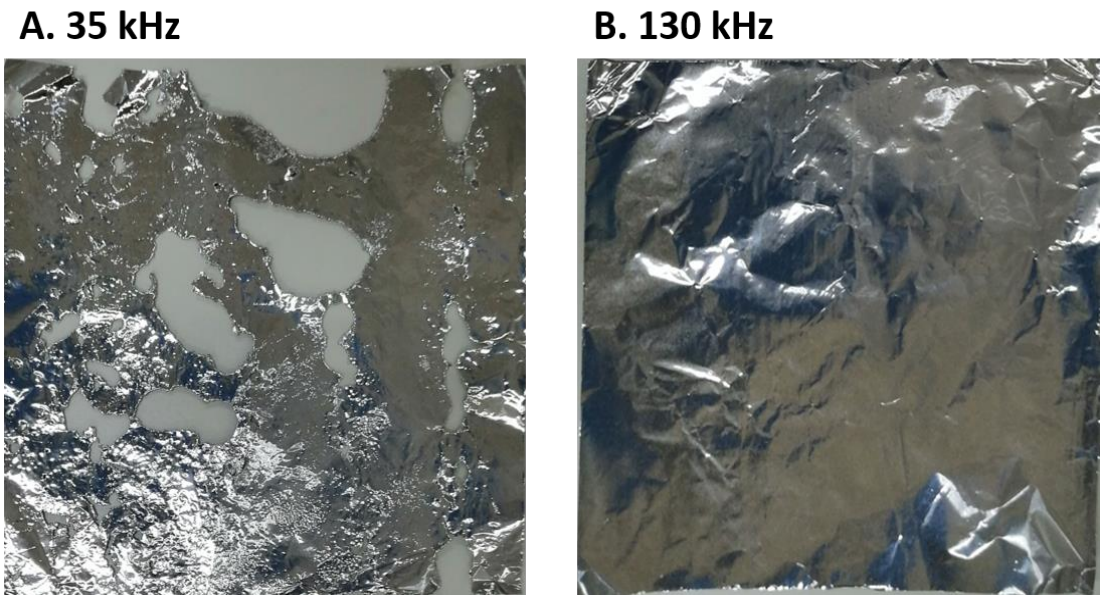

Figure SM IX.1 Effects of ultrasound frequency on an aluminum foil. The two pieces of aluminum foil were submitted to an ultrasonic field generated by an US bath for two min at 100% of ultrasonic amplitude, but with the two different ultrasonic frequencies, A) 35 kHz and B) 130 kHz

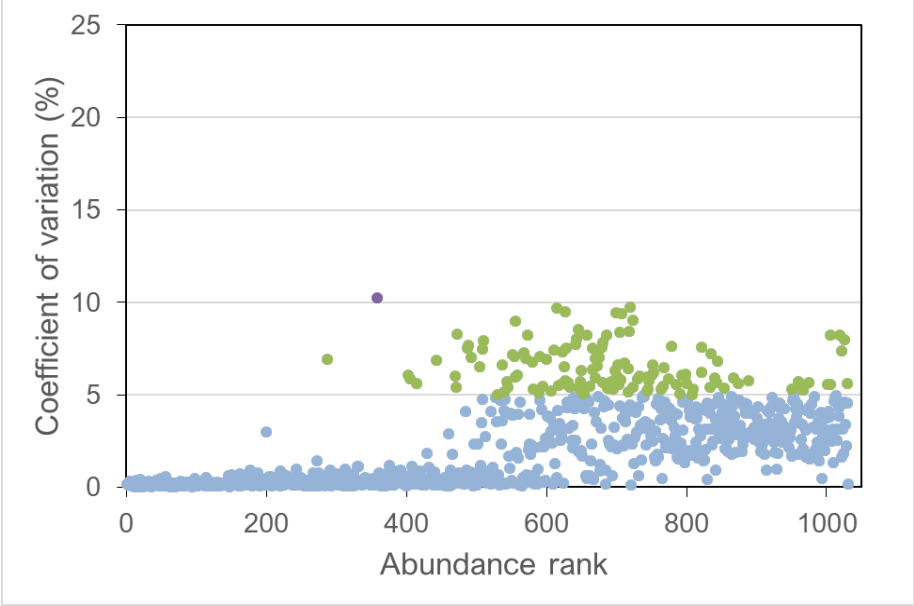

Figure SM IX.2 – Coefficient of variation (CV%) of the technical replicated of mouse kidney samples treated with US bath 35 kHz. The coefficient of variation (CV%) of all quantified proteins were plotted according to their abundance. Proteins with CV´s <5% are colored in blue and those with CV's between 5 to 10% in green. Only one protein was found with a CV > 10%, colored in purple.

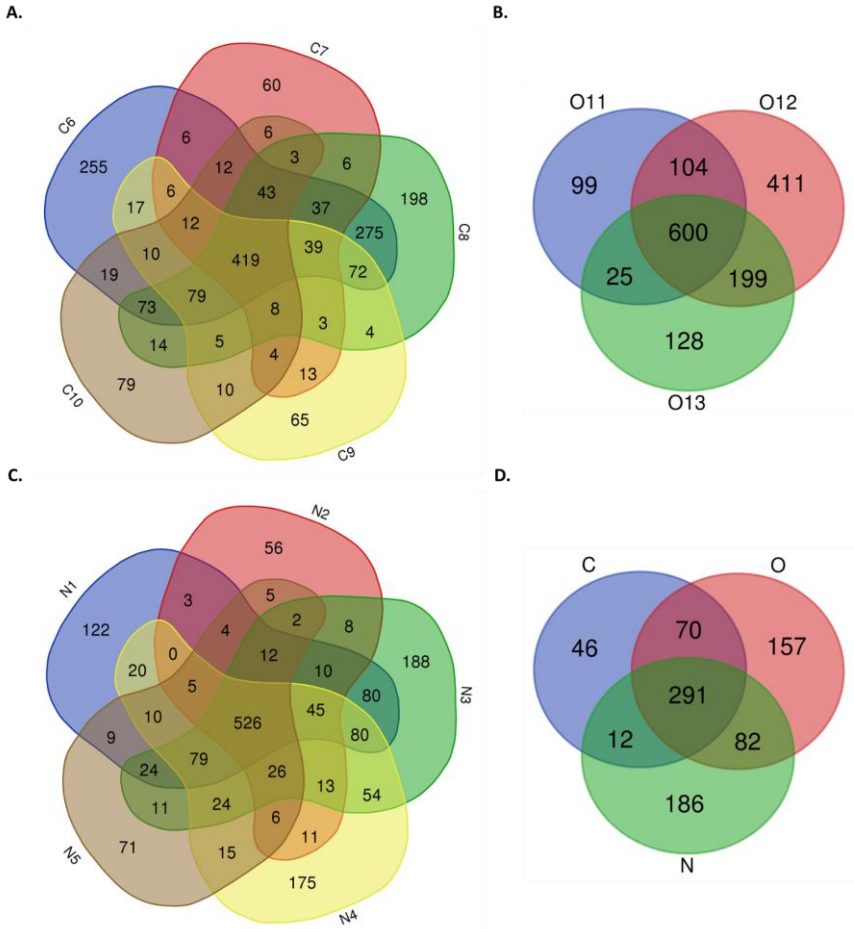

Figure SM IX.3 – Inter and intra-tumor heterogeneity based on the proteins identified in each sample. A) Venn diagram showing the chromophobe intra-tumor heterogeneity of the proteins identified in the five biopsies. B) Venn diagram showing the oncocytoma intra-tumor heterogeneity of the proteins identified in the three biopsies. C) Venn diagram showing the heterogeneity of the five normal adjacent tissue samples. D) Diagram showing the proteome inter-heterogeneity between chRCC, RO and NAT (The proteins consistently detected in 100% of the samples, 419 chRCC, 600 RO and 526 NAT, were used to generate the Venn diagram D).

IX.1.2 ELECTRONIC SUPPLEMENTARY TABLES

Table ESM IX.1 – Protein groups of fresh frozen and OCT mice samples

Table ESM IX.2 – Proteins identified in chRCC tumor samples. The 12 proteins with the highest levels of enriched expression in kidney1 are highlighted in green.

Table ESM IX.3 – Proteins identified in RO tumor samples. The 12 proteins with the highest levels of enriched expression in kidney1 are highlighted in green.

Table ESM IX.4 – Proteins identified in NAT samples. The 12 proteins with the highest levels of enriched expression in kidney1 are highlighted in green.

Table ESM IX.5 – Unique proteins identified in chRCC tumor samples.

Table ESM IX.6 - Unique proteins identified in RO tumor samples.

Table ESM IX.7 – Unique proteins identified in NAT samples.

IX.2 SUPPLEMENTARY MATERIAL CHAPTER V

IX.2.1 SUPPLEMENTARY TABLES

Table SM IX.1 –. Description of human kidney biopsies used in the study.

BIOPSY	AGE	GENDER	DIAGNOSIS*	SAMPLE TYPE*
N1	54	Male	RCC	NAT
N2	49	Female	Papillary	NAT
N3	58	Female	RCC	NAT
N4	72	Female	RCC	NAT
N5	70	Male	RCC	NAT
C6	73	Male	RCC	chRCC
C7	67	Female	RCC	chRCC
C8	71	Male	RCC	chRCC
C9	58	Female	RCC	chRCC
C10	81	Male	RCC	chRCC
O11	80	Male	RCC	RO
O12	69	Female	RCC	RO
O13	63	Male	RCC	RO
O14	62	Female	RCC	RO
O13	55	Female	RCC	RO

**RCC: renal cell carcinoma; NAT: normal adjacent tissue; chRCC: chromophobe renal cell carcinoma; RO: renal oncocytoma.*

IX.2.2 ELECTRONIC SUPPLEMENTARY TABLES

Table ESM IX.8 – List of pathways enriched in the chRCC and RO tumors when compared to NAT specimens;

Table ESM IX.9 – List of proteins with significant differential abundance between chRCC and RO tissues

Table ESM IX.10 – List proteins with significant differential abundance between each tumour and NAT

IX.3 SUPPLEMENTARY MATERIAL CHAPTER VI

IX.3.1 ELECTRONIC SUPPLEMENTARY TABLES

Table ESM IX.11 – LFQ values of proteins' panel

Table ESM IX.12 – TPA values of proteins' panel