

EMPOWERING PERSONAL HEALTH IN BLADDER CANCER PATIENTS BY DISCLOS- ING MEDICAL INFORMATION HIDDEN ON COMPLEX URINARY PROTEOMES: DIAGNOSTIC, PROGNOSTIC AND RESPONSE TO THERAPY

LUÍS ANDRÉ BOTELHO DE CARVALHO
Master in Biochemistry

DOCTORATE IN BIOCHEMISTRY
NOVA University Lisbon
April, 2024

**EMPOWERING PERSONAL HEALTH IN BLADDER
CANCER PATIENTS BY DISCLOSING MEDICAL INFOR-
MATION HIDDEN ON COMPLEX URINARY PROTEO-
MES:
DIAGNOSTIC, PROGNOSTIC AND RESPONSE TO THERAPY**

LUÍS ANDRÉ BOTELHO DE CARVALHO

Master in Biochemistry

Adviser: José Luis Capelo Martínez
Full Professor, NOVA University Lisbon

Co-advisers: Luís Manuel Viegas de Campos Pinheiro
Head of Urology Department at Hospital de S. José Lisbon
Invited Assistant Professor, NOVA University Lisbon
Jacek Roman Wiśniewski
Associate Professor, Max-Planck-Institute of Biochemistry (Germany)

Examination Committee:

Chair: António Jorge Dias Parola,
Full Professor, NOVA University Lisbon

Rapporteurs: Francisco Manuel Lemos Amado,
Associate Professor with Habilitation, University of Aveiro
Deborah Penque,
Principal Researcher, National Health Institute Dr. Ricardo Jorge, Uni-
versity of Lisbon

Adviser: José Luís Capelo Martínez,
Full Professor, NOVA University Lisbon

Members: Tamara García-Barrera,
Full Professor, University of Huelva

Agapios Agapiou,
Assistant Professor, University of Cyprus

Carlos Alberto Gomes Salgueiro,
Full Professor, NOVA University Lisbon

António Alfredo Coelho Jacinto,
Principal Researcher, NOVA University Lisbon

Susana Maria Pereira Gaudêncio de Matos,
Assistant Researcher, NOVA University Lisbon

DOCTORATE IN BIOCHEMISTRY

NOVA University Lisbon
April, 2024

Empowering personal health in bladder cancer patients by disclosing medical information hidden on complex urinary proteomes: diagnostic, prognostic and response to therapy

Copyright © Luís André Botelho de Carvalho, NOVA School of Science and Technology, NOVA University Lisbon.

The NOVA School of Science and Technology and the NOVA University Lisbon have the right, perpetual and without geographical boundaries, to file and publish this dissertation through printed copies reproduced on paper or on digital form, or by any other means known or that may be invented, and to disseminate through scientific repositories and admit its copying and distribution for non-commercial, educational or research purposes, as long as credit is given to the author and editor.

To those who made me believe,
and never ceased to inspire.

ACKNOWLEDGMENTS

Throughout my PhD journey I found myself immersed in a roller coaster of emotions. From fulfilling my long-held dream goal and embarking deeper into the academia "marathon" feels rewarding and only possible due to the people around me ...

... who made me believe, and never ceased to inspire.

First and foremost, I would like to thank my main PhD supervisor, Professor Capelo, for his relentless support and for continuously inspire and pushing me into breaking new goals and acquiring new ambitions. I see him as a mentor such like "*Pai Mei*" to "*Bill*" who I can count on and without his help and teachings this thesis was not possible. I also want to extent my thanks to Dr. Hugo M. Santos, for his patience, advice, and teaching in the laboratory. His help in all the projects was immeasurable, and a key component of the success of this work. To my co-supervisors, Dr. Luis Campos Pinheiro and Dr. Jacek R. Wiśniewski, a special thanks for the help and advice during this thesis and the working projects.

In addition, I am indebted to the BIOSCOPE Research Group. A warm thanks to Professor Lodeiro for his support and motivation provided during this thesis. I extend these words to the exceptional current lab and group colleagues, Dr. Elisabete Oliveira, Dr. Javier Lodeiro, Dr. Adrian Lodeiro, André Figueiredo, Silvia Nuti, Frederico Duarte, Joana Galhano, Inês Domingos and Gonalo Pedro.

A huge thanks to Professor Moses, for the kind welcome her group and laboratory at Harvard Medical school, Boston children's hospital. Marking my first experiences in the USA as a visiting scientist, I will treasure the intellectually stimulating discussions, presentations and formation that this academic training during my PhD has provided me.

I am incredibly grateful to the PROTEOMASS Scientific Society for the trust and continuous financial support; to Fundação para a Ciência e a Tecnologia for the PhD grant (SFRH/BD/144222/2019) and financial support as well.

To my close friends, the *wolf pack*, Filipe Marques, João Campos and Manuel Martins a huge thanks for the unconditional support and for being present throughout my life. To my close friends, Susana Jorge and Gonçalo Martins, who shared their PhD journey with me, thanks for the support, laughs and continuous help during these years. To my close friend, Cachucho also known as "Chuchu", thanks for the inspiration and positive energy during this journey.

Para a minha família, o meu amor e suporte incondicional, eu devo-vos tudo. Ao meu pai e mãe, quem eu considero os meus heróis. Constantemente a fazer sacrifícios, a demonstrar coragem, a motivar e a acreditar em mim, um obrigado não é suficiente. Vocês sempre me proporcionaram tudo sem pedir nada em troca, fazendo-me sempre acreditar que sou capaz e que serei sempre o vosso "campeão". Por isso, vocês fizeram o meu hoje possível e são e serão sempre a minha essência. Ao resto da minha grande família, que incondicionalmente estão presentes e prontos a ajudar um enorme obrigado. Por fim, às pessoas que carrego constantemente comigo no coração e que nunca deixam de me ajudar e de inspirar, sou eternamente grato.

To my beautiful beloved wife, who is my emotional support, foundation and source of motivation, words cannot describe how thankful I am for your continuous presence and backing me up throughout this journey that we both share. To you, who stood by me though late nights, early mornings, every rough and good moments you were present giving me a fresh breath and keeping me align and consistent to my own ambitions. For allowing me to keep dreaming and aiming for new goals, your love is what propels and keeps me going. Finally, to my two boys, Sushi and Wasabi, who are my constant companions you all complete me.

"The first and greatest victory
is to conquer yourself;
to be conquered by yourself is of all things most
shameful and vile"

(Plato)

THESIS CONTRIBUTION SUMMARY

Journal articles

1. Luis B. Carvalho; Susana Jorge; Hugo López-Fernández; Carlos Lodeiro; Rajiv Dhir; Luis Campos Pinheiro; Mariana Medeiros; Hugo M. Santos; Jose Luis Capelo. **"Proteomic analysis of chromophobe renal cell carcinoma and benign renal oncocytoma biopsies reveals shared metabolic dysregulation"**. *Clinical Proteomics* (2023). <https://doi.org/10.1186/s12014-023-09443-8>.
2. Luis B. Carvalho; Pedro A.D. Teigas-Campos; Susana Jorge; Michele Protti; Laura Mercolini; Rajiv Dhir; Jacek R. Wisniewski; et al. **"Normalization methods in mass spectrometry-based analytical proteomics: A case study based on renal cell carcinoma datasets"**. *Talanta* (2023). <https://doi.org/10.1016/j.talanta.2023.124953>.
3. Luís B. Carvalho; José Luis Capelo Martínez; Carlos Lodeiro; Rafael Bento; Rajiv Dhir; Jeremiah J. Morrissey; Luis Campos Pinheiro; Mariana Medeiros; Hugo M. Santos. **"Biochemical network analysis of protein-protein interactions to follow-up T1 bladder cancer patients"**. *Journal of Proteomics* 278 (2023): 104865-104865. <https://doi.org/10.1016/j.jprot.2023.104865>.
4. Luís Botelho Carvalho; José Luis Capelo; Carlos Lodeiro; Rajiv Dhir; Luis Campos Pinheiro; Hugo López-Fernández; Gonçalo Martins; et al. **"Pathway-guided monitoring of the disease course in bladder cancer with longitudinal urine proteomics"**. *Nature Communications Medicine* (2023). <https://doi.org/10.1038/s43856-023-00238-4>.
5. Luís B. Carvalho; José-Luis Capelo-Martínez; Carlos Lodeiro; Jacek R. Wisniewski; Hugo M. Santos. **"Ultrasonic-Based Filter Aided Sample Preparation as the General Method to Sample Preparation in Proteomics"**. *Analytical Chemistry* 92 (2020): 9164–9171. <https://doi.org/10.1021/acs.analchem.0c01470>.
6. Carvalho, L.B.; Capelo-Martínez, J.L.; Lodeiro, C.; Wisniewski, J.R.; Santos, H.M. **"Snap-heated freeze-free preservation and processing of the urine proteome using the**

combination of stabilizer-based technology and filter aided sample preparation". *Analytica Chimica Acta* 1076 (2019): 82-90. <https://doi.org/10.1016/j.aca.2019.05.051>.

Oral Communications

1. Oral Communication (10min) at 13^o Encontro Nacional Cromatografia., Faculty of Pharmacy, University of Lisbon, Lisbon (2023) **"Empowering bladder cancer patient management through a tailored non-invasive LC-MS/MS approach"**.
2. Oral Communication (15min) at XXVI Encontro Galego-Portugués de Química. University of Santiago de Compostela. (2022) **"Following-up bladder cancer patient's response to therapy"**.
3. Oral Communication (5min) at 8th International Caparica Conference on Analytical Proteomics (2022). **"Differential proteomic pathway index (dPPI) - A tailored approach for non-invasive bladder cancer patient monitoring"**.
Awarded as an excellent Shotgun Communication Prize.
4. Oral Communication (5min) at 4th International Caparica Christmas Conference on Translational Chemistry (2021). **"Biochemical network analysis of protein-protein interactions to monitor bladder cancer patients"**.
Awarded as an excellent Shotgun Communication Prize.
5. Oral Communication (5min) at 7th International Caparica Conference on Analytical Proteomics (2021). **"Mass spectrometry-based personalized medicine reveals neutrophilic inflammation dysregulation as a predictor of poor clinical outcome in T1 stage bladder cancer patients"**.
6. Oral Communication (5min) at 5th International Caparica Conference on Ultrasonic - based applications from analysis – synthesis (2021). **"High-throughput ultrasound-based workflow for clinical proteomics"**.
7. Oral Communication (5min) at 4th International Caparica Conference on Ultrasonic-based Applications: from analysis – synthesis (2020). **"Ultrafast high-throughput ultrasonic-FASP method for large-scale proteomics"**.
8. Oral Communication (5min) at 6th International Caparica Conference on Analytical Proteomics (2019). **"A new approach to storage and preserve the urine proteome"**.
Awarded as an excellent Shotgun Communication Prize.
9. Oral Communication (5min) at 3rd International Caparica Christmas Conference on Translational Chemistry (2019). **"Urine storage using the snap-heated freeze-free method allows the preservation of protein biomarkers at room temperature"**.

10. Oral Communication (5min) at 4th International Caparica Conference on Urine Omics and Translational Nephrology (2019). **"Urine proteome storage at room temperature in a ready FASP format for biomarker discovery"**.
Awarded as an excellent Shotgun Communication Prize.

Poster communications

1. Poster presentation at 12° Encontro Nacional Cromatografia., Aveiro (2022). **"Universal sample handling for liquid biopsies"**.
Awarded as a honorable mention for best poster presentation.
2. 16. Poster presentation at 8th International Caparica Conference on Analytical Proteomics (2022). **"Differential proteomic pathway index (dPPI) - A tailored approach for non-invasive bladder cancer patient monitoring"**.
3. Poster presentation at 5th International Caparica Christmas Conference on Sample Treatment (2021). **"Room temperature urine proteome storage in a ready-to-use, high-throughput ultrasonic FASP system"**.
4. Poster presentation at 4th International Caparica Christmas Conference on Translational Chemistry (2021). **"Biochemical network analysis of protein-protein interactions to monitor bladder cancer patients"**
5. Poster presentation at 7th International Caparica Conference on Analytical Proteomics (2021). **"Mass spectrometry-based personalized medicine reveals neutrophilic inflammation dysregulation as a predictor of poor clinical outcome in T1 stage bladder cancer patients"**.
6. Poster presentation at 5th International Caparica Conference on Ultrasonic - based applications from analysis – synthesis (2021). **"High-throughput ultrasound-based workflow for clinical proteomics"**.
7. Poster presentation at 4th International Caparica Conference on Ultrasonic-based Applications: from analysis – synthesis (2020). **"Ultrafast high-throughput ultrasonic-FASP method for large-scale proteomics"**.
8. Poster presentation at 3rd International Caparica Christmas Conference on Translational Chemistry (2019). **"Urine storage using the snap-heated freeze-free method allows the preservation of protein biomarkers at room temperature"**.
Awarded as an excellent shotgun poster communication prize.
9. Poster presentation at 4th International Caparica Conference on Urine Omics and Translational Nephrology (2019). **"A proteomic biomarker panel for bladder cancer stage differentiation by label-free quantitative high-resolution mass spectrometry"**.

10. Poster presentation at 6th International Caparica Conference on Analytical Proteomics (2019). **"A new approach to storage and preserve the urine proteome"**.
11. Poster presentation at 4th International Caparica Conference on Urine Omics and Translational Nephrology (2019). **"Urine proteome storage at room temperature in a ready FASP format for biomarker discovery"**.

ABSTRACT

Bladder cancer (BCa), as the most prevalent malignancy within the urinary system, poses a significant global health challenge, with nearly 600,000 new cases emerging annually worldwide. This number is alarmingly projected to double by 2040, according to forecasts by the World Health Organization, spotlighting the urgent need for minimally invasive and cost-effective diagnostic options to improve patient outcomes.

This doctoral thesis addresses the urgent call for innovative approaches in managing BCa and potentially other urological cancers. It details the development and application of cutting-edge proteomic technologies to refine the accuracy of BCa diagnosis and improve prognostic evaluations. By extending this research to include renal cancers, the thesis underscores the proposed methods' versatility and potential broad applicability.

Central to this work is the novel differential personal pathway index (dPPi), a personalized medicine approach derived from extensive work in urine sample processing, proteome analysis, and mass spectrometry data normalization. This thesis presents several key advancements, including (i) The development of a novel urine protein preservation method using filter-aided sample preparation (FASP), ensuring effective concentration and stabilization for long-term storage without compromising sample integrity; (ii) The introduction of a high-throughput ultrafast proteomics method known as US-FASP, which speeds sample processing while maintaining robustness and scalability; (iii) The evaluation of normalization techniques for mass spectrometry data. (iv) Identifying urinary proteomic signatures associated with BCa recurrence, highlighting potential prognostic indicators; (v) Implementing longitudinal urine proteomic analysis using the dPPi approach, to provide insights into BCa progression over time and guide clinical interventions; (vi) Exploring proteomic landscapes in chromophobe renal cell carcinoma (chRCC) and renal oncocytomas (RO), offering insights into diagnostic and therapeutic avenues

Overall, this thesis contributes significantly to understanding BCa and opens up new avenues for developing personalized medical interventions. The findings enhance our grasp of BCa's biological underpinnings and offer a promising framework for extending these innovations to a wider array of urological and other malignancies.

Keywords: Bladder cancer, Proteomics, Mass Spectrometry, Urine sample handling, Longitudinal Analysis, Precision Medicine.

RESUMO

O cancro da bexiga (BCa), sendo o tipo de cancro mais prevalente no sistema urinário, representa um grande desafio para a saúde global, com aproximadamente 600.000 novos casos por ano no mundo. De acordo com a Organização Mundial de Saúde, este número está previsto duplicar até 2040, realçando a necessidade urgente de diferentes opções de diagnóstico não invasivas e economicamente mais viáveis para melhorar o tratamento e diagnóstico dos pacientes.

Esta tese de doutoramento aborda o importante apelo por abordagens inovadoras no tratamento do BCa e de outros cancros urológicos. Aprofunda o desenvolvimento e a aplicação de tecnologias proteómicas de ponta para refinar a precisão do diagnóstico e melhorar as avaliações prognósticas do BCa. Este trabalho ao incluir cancros renais, sublinha também a versatilidade e potencial aplicabilidade ampla dos métodos propostos.

Como trabalho central, está o differential personal pathway index (dPPI), uma abordagem de medicina personalizada derivada de um extenso trabalho no processamento de amostras de urina, análise do proteoma e normalização de dados de espectrometria de massa. Esta tese apresenta vários avanços importantes, incluindo (i) O desenvolvimento de um novo método de preservação de proteínas na urina usando o método filter-aided sample preparation (FASP), garantindo concentração e estabilização eficazes para armazenamento a longo prazo sem comprometer a integridade da amostra; (ii) A introdução de um método de proteómica ultra-rápido em alta produção conhecido como US-FASP, que acelera o processamento de amostras mantendo a robustez e escalabilidade; (iii) A avaliação de técnicas de normalização para dados de espectrometria de massa. (iv) Identificação de assinaturas proteómicas urinárias associadas à recorrência do BCa, destacando indicadores com prognóstico potencial; (v) Implementação de análise proteómica urinária longitudinal usando a abordagem dPPI, para fornecer informações sobre a progressão do BCa ao longo do tempo

e orientar intervenções clínicas; (vi) Análise proteómica no carcinoma renal cromóforo e nos oncocitomas renais, oferecendo perspetivas sobre opções de diagnóstico e terapêuticas.

No geral, esta tese contribui significativamente para a compreensão do BCa e abre novas direções para o desenvolvimento de intervenções médicas personalizadas. Os resultados melhoram a compreensão das bases biológicas do BCa e oferecem um quadro promissor para alargar essas inovações a uma ampla gama de cânceros urológicos e outros.

Palavras chave: Cancro da bexiga, Proteómica, Espectrometria de massa, Manuseamento de amostras de urina, Análise longitudinal, Medicina personalizada.

CONTENTS

1	INTRODUCTION.....	1
1.1	Bladder Cancer.....	1
1.1.1	Bladder Cancer Demographics and Patterns.....	1
1.1.2	The hidden burden of bladder cancer.....	3
1.1.3	Landscape of bladder cancer.....	3
1.1.4	Bladder Cancer Diagnosis and Treatment.....	8
1.1.5	Other oncoproteomics case studies.....	10
1.2	The Promise of Proteomics.....	11
1.2.1	Extracting information from complex proteomes.....	11
1.2.2	Mass spectrometry (MS)-based proteomics.....	12
1.2.3	MS-Proteomics methodologies.....	13
1.2.4	Liquid chromatography in mass spectrometry.....	14
1.2.5	Ionization & Mass Analysers.....	15
1.2.6	Acquisition Methods.....	17
1.2.7	Protein identification & quantification.....	19
1.2.8	Recent Technological Advances of Mass Spectrometry-Based Proteomics.....	22
1.3	Liquid Biopsies in Proteomics.....	23
1.3.1	Urine a source of body information.....	23
1.3.2	Urine proteome landscape.....	24
1.3.3	Sample Treatment for MS-Based Urine proteomics.....	24

1.4	Personalised medicine-based proteomics.....	27
1.4.1	The current extent of clinical protein-based diagnostics	27
1.4.2	Clinical proteomics and towards personalised medicine	28
1.4.3	Preciptomics.....	30
1.5	References.....	30
2	OBJECTIVES	55
3	SNAP-HEATED FREEZE-FREE PRESERVATION AND PROCESSING OF THE URINE PROTEOME USING THE COMBINATION OF STABILIZOR-BASED TECHNOLOGY AND FILTER AIDED SAMPLE PREPARATION	57
3.1	Abstract.....	58
3.2	Introduction.....	58
3.3	Experimental Section	60
3.3.1	Reagents.....	60
3.3.2	Apparatus.....	60
3.3.3	Urine Specimens and Collection	61
3.3.4	Urinary Proteome Extraction.....	62
3.3.5	Heat Shock Preservation	63
3.3.6	Urine Storage.....	63
3.3.7	Protein Digestion Using Filter Aided Sample Preparation	63
3.3.8	Nano-LC-MS/MS.....	64
3.3.9	Protein Identification and Quantification	64
3.4	Results & Discussion.....	65
3.4.1	Quantitative comparison between different storage conditions at T0 (48h)	65
3.4.2	Quantitative comparison between different storage conditions at T1 (6 months) 65	
3.4.3	Quantitative T0 versus T1 comparison for each storage condition	66
3.4.4	Hierarchical cluster analysis of each storage methods and time.....	68

3.4.5	Proof-of-concept: preservation of (i) acute kidney injury urinary biomarkers; (ii) kidney tubular injury; and (iii) FDA -cleared or -approved protein assayed in serum or plasma	70
3.5	Conclusions	72
3.6	Supplementary Material	73
3.6.1	Supplementary data.....	73
3.6.2	Supplementary figures & tables.....	73
3.7	Contributions & Acknowledgements	76
3.7.1	Author Contributions.....	76
3.7.2	Acknowledgments	76
3.8	References.....	76
4	ULTRASONIC-BASED FILTER AIDED SAMPLE PREPARATION AS THE GENERAL METHOD TO SAMPLE PREPARATION IN PROTEOMICS.....	81
4.1	Abstract.....	82
4.2	Introduction.....	82
4.3	Experimental Section	83
4.3.1	Reagents.....	83
4.3.2	Apparatus.....	84
4.3.3	E. coli proteome.....	84
4.3.4	Mouse liver and brain proteomes	85
4.3.5	Total protein and peptide quantification.....	85
4.3.6	Filter aided sample preparation (FASP)	85
4.3.7	Ultrasonic-based filter aided sample preparation (US-FASP)	86
4.3.8	Nano-LC-MS/MS.....	86
4.3.9	Label-free protein quantification and statistical analysis.....	87
4.4	Results & Discussion.....	89
4.4.1	Initial considerations about variables affecting ultrasonic performance.....	89
4.4.2	Effect of Ultrasonic frequency over the FASP membrane.....	89

4.4.3	Effect of time of ultrasonication in the FASP-based digestion protocol	90
4.4.4	Effect of ultrasonication in reduction and alkylation steps	91
4.4.5	Mass spectrometry label free-based protein quantification: ultrasonic-based FASP versus classic overnight FASP	91
4.4.6	Performance of Ultrasonic-based FASP	94
4.4.7	Ultrasonic-based FASP Label-free quantification of a set of spiked proteins in <i>E. coli</i> samples	96
4.4.8	Ultrasonic-based FASP versus classic FASP Label-free quantification of a set of spiked proteins in <i>E. coli</i> samples	97
4.4.9	Application to brain and liver tissue samples	97
4.5	Conclusions	99
4.6	Supplementary material	99
4.6.1	Supplementary data	99
4.6.2	Supplementary figures	100
4.7	Contributions & Acknowledgements	102
4.7.1	Author Contributions	102
4.7.2	Acknowledgments	102
4.8	References	102
5	NORMALIZATION METHODS IN MASS SPECTROMETRY-BASED ANALYTICAL PROTEOMICS: A CASE STUDY BASED ON RENAL CELL CARCINOMA DATASETS	107
5.1	Abstract	108
5.2	Introduction	108
5.3	Experimental Section	110
5.3.1	Patients and sample collection	110
5.3.2	Mass spectrometry raw data	111
5.3.3	Identification and quantification of proteins	111
5.3.4	Bioinformatics data processing	111
5.3.5	Bioinformatic data analysis and normalization comparisons	112

5.4	Results & Discussion.....	113
5.4.1	Comparing each type of renal carcinoma against NAT.....	113
5.4.2	Comparison of the two, three and four renal carcinomas at a time against NAT 116	
5.4.3	Comparison of the best biomarkers for <i>Immunohistochemistry</i>	119
5.4.4	Exploring the impact of normalization methods on identifying main pathways affected in renal tumours.....	119
5.5	Conclusions	121
5.6	Supplementary material	121
5.6.1	Supplementary data.....	121
5.6.2	Supplementary figures & tables.....	121
5.7	Contributions & Acknowledgements	125
5.7.1	Author contributions	125
5.7.2	Acknowledgements.....	126
5.8	References.....	126
6	BIOCHEMICAL NETWORK ANALYSIS OF PROTEIN-PROTEIN INTERACTIONS TO FOLLOW-UP T1 BLADDER CANCER PATIENTS.....	131
6.1	Abstract.....	132
6.2	Introduction.....	132
6.3	Experimental Section	133
6.3.1	Patients	133
6.3.2	Urine collection and preservation.....	133
6.3.3	Filter-Fided sample preparation.....	134
6.3.4	LC-MS/MS.....	134
6.3.5	Bioinformatics.....	135
6.3.6	MPO in vitro Enzyme-Linked Immunosorbent Assay.....	136
6.4	Results & Discussion.....	136
6.4.1	CUBN to MPO ratio	137

6.4.2	Biochemical networks.....	139
6.5	Conclusions	141
6.6	Supplementary material	142
6.6.1	Supplementary data.....	142
6.6.2	Supplementary figures & tables.....	142
6.7	Contributions & Acknowledgements	145
6.7.1	Author Contributions.....	145
6.7.2	Acknowledgments	145
6.8	References.....	145
7	PATHWAY-GUIDED MONITORING OF THE DISEASE COURSE IN BLADDER CANCER WITH LONGITUDINAL URINE PROTEOMICS.....	149
7.1	Abstract.....	150
7.2	Introduction.....	150
7.3	Experimental Section	151
7.3.1	Patients	151
7.3.2	Sample treatment	152
7.3.3	Filter-aided sample preparation (FASP).....	152
7.3.4	LC-MS/MS	153
7.3.5	Bioinformatics and statistical analyses.....	153
7.3.6	dPPi calculation.....	154
7.4	Results & Discussion.....	156
7.4.1	Selection of pathway hallmarks of cancer	156
7.4.2	Longitudinal follow-up of BCa patients.....	157
7.5	Conclusions	165
7.6	Supplementary material	165
7.6.1	Supplementary data.....	165
7.6.2	Supplementary figures.....	166

7.6.3	Supplementary discussion.....	167
7.7	Contributions & Acknowledgements	173
7.7.1	Author Contributions.....	173
7.7.2	Acknowledgments	173
7.8	References.....	174
8	PROTEOMIC ANALYSIS OF CHROMOPHOBE RENAL CELL CARCINOMA AND BENIGN RENAL ONCOCYTOMA BIOPSIES REVEALS SHARED METABOLIC DYSREGULATION	179
8.1	Abstract.....	180
8.2	Introduction.....	180
8.3	Materials and Methods	181
8.3.1	Study design and sampling.....	181
8.3.2	Ethics approval and consent to participate.....	182
8.3.3	Proteomic analysis.....	182
8.3.4	Data analysis and statistics.....	182
8.4	Results	183
8.4.1	Renal tumor profiling	183
8.4.2	Dysregulated proteins between tumors and NATs.....	185
8.4.3	Main chRCC and RO features	188
8.5	Discussion	190
8.6	Conclusions	192
8.7	Supplementary material	192
8.7.1	Supplementary data.....	192
8.7.2	Supplementary figures & tables.....	192
8.8	Contributions & Acknowledgements	196
8.8.1	Author Contributions.....	196
8.8.2	Acknowledgments	196
8.9	References.....	196

9	GENERAL CONCLUSIONS.....	201
10	FUTURE PERSPECTIVES	203

LIST OF FIGURES

Figure 1-1. Global incidence and mortality of bladder cancer in 2022 in men and women fo all ages.....	2
Figure 1-2. World Health Organization (WHO) 1973 and 2004/2022 bladder cancer clinical clas- sification.....	4
Figure 1-3. Schematic classification and origin of bladder cancer (BCa).	7
Figure 1-4. Urinary makers and tests used to diagnose bladder cancer.....	10
Figure 1-5. Bottom-up proteomic workflow.....	14
Figure 1-6. Filter Aided Sample Preparation (FASP).....	26
Figure 2-1. Rationale and Integration of the doctoral thesis objectives with the different chap- ters of the thesis.	56
Figure 3-1. Graphical abstract.....	57
Figure 3-2. Method Workflow chart.....	62
Figure 3-3. Quantative proteomic analysis for each storage condition.	67
Figure 3-4. Cluster analysis presenting the variation of proteins levels between the different conditions of storage and different time of storage.	69
Figure 3-5. Comparison of LFQ intensities for 24 selected -cleared or -approved protein as- sayed in serum or plasma between T0 (48 h) and T1 (6 months) for the 3 conditions tested in this study.	71
Figure 3-6. System for urinary proteome preservation and storage.	73
Figure 3-7. Volcano plots showing the variation in protein levels with storage condition in dif- ferent times points of storage.....	74
Figure 4-1. Graphical abstract.....	81
Figure 4-2. Ultrasonic-based filter aided sample preparation workflow and optimization.....	88
Figure 4-3. <i>E. coli</i> proteome digestion using the US-FASP and ON-FASP.....	93

Figure 4-4. Ultrasonic space effectiveness in US-FASP.....	95
Figure 4-5. Mouse liver and brain proteome digestion using the US-FASP.....	98
Figure 4-6. Original optical microscopy images (40x) from the FASP membranes using the conditions.....	100
Figure 4-7. Color adjusted optical microscopy images (40x) from the FASP membranes using the conditions.....	100
Figure 4-8. Hierarchal clustering of the E. coli proteome processed with US-FASP and ON-FASP.....	101
Figure 4-9. Custom-built wood plate vivacon support floater.....	101
Figure 5-1. Graphical Abstract.....	107
Figure 5-2. Schematic representation of the data process applied to the proteomic data when comparing Neat Adjacent Tissue, NAT (control group) against multiple renal tumour types.....	113
Figure 5-3. Comparison of the numbers of proteins found statistically differentially expressed using each normalization method when comparing the near adjacent tissue (NAT) against each renal tumour type.....	115
Figure 5-4. Proteins found statistically differentially expressed.....	116
Figure 5-5. Comparison of candidate biomarkers for each renal tumour subtype across all different type of data normalization.....	118
Figure 5-6. Comparison of Top-10 pathways differentially expressed when chRCC is compared against NAT as per the proteins differentially expressed found with each normalization method.....	120
Figure 5-7. Workflow of the data process applied to the proteomics data when comparing by pairs the Neat Adjacent Tissue, NAT (control group) against the other renal tumour types.....	123
Figure 5-8. Comparison of the numbers of proteins overlap using each normalization method.....	124
Figure 5-9. TPA-based concentrations of potential biomarkers for each renal carcinoma tumour subtype.....	125
Figure 6-1. Graphical abstract.....	131
Figure 6-2. Proteomics and bioinformatics workflow.....	136
Figure 6-3. T1 BCa patients MPO/CUBN levels and protein-protein pathway analysis.....	138
Figure 6-4. Data integration for T1 BCa stage patients.....	141
Figure 6-5. Workflow of the experimental design.....	142

Figure 6-6. Volcano plots showing statistically significant changes in protein levels in the group of T1-stage bladder cancer patients that had experienced recurrence, (RG), versus patients that had not, (NRG).....	143
Figure 7-1. Graphical abstract.....	149
Figure 7-2. Proteome analysis of 30 pT1 bladder cancer patients.	155
Figure 7-3. Longitudinal analysis of the urinary proteome of patient I.....	159
Figure 7-4. Longitudinal analysis of the urinary proteome of patient II.....	161
Figure 7-5. Longitudinal analysis of the urinary proteome of patient III.	163
Figure 7-6. Bioinformatics analysis workflow.....	166
Figure 7-7. Proteomics data summary with the number of identified proteins for each patient.	167
Figure 7-8. Longitudinal analysis of the urinary proteome of patient IV.....	168
Figure 7-9. Longitudinal analysis of the urinary proteome of patient V.	170
Figure 7-10. Longitudinal analysis of the urinary proteome of patient VI. A.	172
Figure 8-1. Graphical abstract.....	179
Figure 8-2. Classification of human renal tissue proteomes and protein expression profiles in tumors relative to NATs.	184
Figure 8-3. Metabolic pathway proteomes similarly affected in chRCC and RO tumors.....	186
Figure 8-4. Differences in protein pathways underlying each tumor sub-type.....	190
Figure 8-5. TPA concentration values (pmol/mg of tissue) of respiratory chain complex I (CI), II (CII) and III (CIII) proteins.....	194
Figure 8-6. TPA concentration values (pmol/mg of tissue) of respiratory chain complex IV (CIV) and V (CV) proteins.....	195

LIST OF TABLES

Table 1-1. TNM classification of bladder cancer.	5
Table 1-2. Different biomarkers applied in clinical proteomics for bladder cancer.....	29
Table 3-1. LFQ variation of 24 FDA approved biomarkers in the urine preserved for 6 months. Variation of the 24 FDA approved biomarkers analysed.	75
Table 5-1. Description of human kidney biopsies used in the study.....	121
Table 6-1. List of patients with the respective medical information.....	144
Table 7-1. Pathways and proteins used on the longitudinal follow up of bladder cancer patients.....	157
Table 7-2. The dPPi index column shows the index calculated as explained in the methods section.....	164
Table 8-1. Description of human kidney biopsies used in the study.....	192

GLOSSARY

Andromeda	A search engine integrated into the MaxQuant that is used for peptide identification.
ClueGO	A Cytoscape plug-in that enables pathway visualizing and analysis
Cytoscape	An open-source software platform for visualizing complex networks.
Data-Dependent Acquisition (DDA)	A mode of data acquisition in tandem mass spectrometry (MS/MS) where certain peptides within a restricted mass range are fragmented.
Data-Independent Acquisition (DIA)	A mode of data acquisition in tandem mass spectrometry (MS/MS) mass spectrometry that selects the top N most abundant precursor ions for fragmentation
Filter-Aided Sample Preparation (FASP)	A bottom-up proteomic approach method for preparing protein samples for mass spectrometry analysis using filter devices.
Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)	Technique combining liquid chromatography with tandem mass spectrometry.
MaxQuant	A software platform for mass spectrometry-based proteomics that enables protein identification and quantification.

Parallel Accumulation Serial Fragmentation (PASEF)	A scan mode in mass spectrometry where ions are accumulated in parallel and then released serially for fragmentation allowing for a high peptide sequencing speed with high sensitivity.
Perseus	Software with a user-friendly interface that is used for large data analysis and visualization with a user-friendly interface.
Retention Time (RT)	Time taken for a compound to pass through a chromatography column.
StringApp	A Cytoscape app that focus on protein-protein analysis and enables pathway visualizing.
Tandem Mass Spectrometry (MS/MS)	Technique that uses two mass spectrometers connected that allows the identification and quantification of molecules such as peptides by fragmenting them and then analyzing the fragments
Trapped Ion Mobility Spectrometry (TIMS)	Analytical mass spectrometry technique for separating ions based on their mobility in a gas phase.

ACRONYMS

ACN	Acetonitrile
AKI	Acute Kidney Injury
AmBic	Ammonium Bicarbonate
BCa	Bladder Cancer
BSA	Bovine Serum Albumin
chRCC	Chromophobe Renal Cell Carcinoma
CIS	Cancer Cells in Situ
DDA	Data-Dependent Acquisition
DIA	Data-Independent Acquisition
DTT	1,4-Dithio-DL-threitol
ELISA	Enzyme-Linked Immunosorbent Assay
emPAI	Exponentially Modified Protein Abundance Index
EMT	Epithelial-Mesenchymal Transition
EuroKUP	European Urine and Kidney Proteomics
FA	Formic Acid
FASP	Filter Aided Sample Preparation
FDA	Food and Drug Administration
GO	Gene Ontology

HCl	Hydrochloric Acid
HG	High-Grade
HKUPP	Human Kidney and Urine Proteome Project
HPLC	High-Performance Liquid Chromatography
HRAM	High-Resolution Accurate Mass
IAA	Iodoacetamide
IRB	Institutional Review Board
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
LFQ	Label-Free Quantification
LG	Low-Grade
m/z	Mass-to-Charge Ratio
MIBC	Muscle-Invasive Bladder Cancer
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MWCO	Molecular Weight Cut-Off
nLC	Nano Liquid Chromatography
NMIBC	Non-muscle Invasive Bladder Cancer
OCT	Optimal Cutting Temperature
PASEF	Parallel Accumulation Serial Fragmentation
PRIDE	Proteomics Identification Database
PRM	Parallel Reaction Monitoring
PUNLMP	Papillary Urothelial Neoplasm of Low Malignant Potential
Q-TOF	Quadrupole-Time of Flight
RCC	Renal Cell Carcinoma

RT	Room Temperature
SPE	Solid-Phase Extraction
SRM	Selected Reaction Monitoring
TFA	Trifluoroacetic Acid
TIMS	Trapped Ion Mobility Spectrometry
TNM	Tumour, Node, and Metastasis
TOF	Time of Flight
TPA	Total Protein Approach
TURBT	Transurethral Resection of Bladder Tumour
UA	Ultrasonic Amplitude
UF	Ultrasonic Frequency
UHR-QqTOF	Ultra-High-Resolution Quadrupole-Time of Flight
UICC	Union International Contre le Cancer
UT	Ultrasonic Time
WHO	World Health Organization

INTRODUCTION

1.1 Bladder Cancer

Bladder cancer (BCa) typically develops because of the outgrowth of dysfunctional cells in the bladder lining [1]. Also known as urothelial or transitional cell carcinoma, the condition that constitutes most of the global bladder cancer cases, being about 90%. Starting from the urothelial cells, which line the inner surface of the bladder, it can spread to other sites of the urinary tract, including the renal pelvis, ureters, and urethra [2,3]. Squamous cell carcinomas are also a type of BCa, which is 5% of the cases worldwide. They are typically linked to protozoal infections that mediate inflammation [4]. The remaining cases of BCa constitute uncommon types such as adenocarcinomas, sarcomas, and metastases to the bladder [1–3].

1.1.1 Bladder Cancer Demographics and Patterns

Bladder cancer represents a major world health problem, and the incidence of this neoplasm contributes significantly to cancer-related morbidity and mortality [5,6]. GLOBOCAN data from 2022 indicates that about 600,000 cases of bladder cancer are diagnosed, and it remains the most common malignancy within the urinary tract, being the 11th among all types of cancers worldwide [7]. Even though it predominantly affects the male population, with a four times higher incidence than in women across the world, being ranked 6th in men among all types of cancer.

By Fig.1-1A, it is evident that the incidence of BCa is the highest in the higher income regions such as Europe, North America, and Western Asia and also increased in endemic regions affected by *Schistosoma* parasites such as northern Africa. Such variations in incidence

have been associated with risk factors such as sedentarism, rate of tobacco use, and exposure to various chemicals. From the mortality level in Fig.1-1B, bladder cancer ranks 13th in the overall mortality of all sexes worldwide, being more severe in Europe and Northern Africa. The difference between the incidence and mortality indicates that while there are continued efforts to diminish risk factors, improve diagnosis, and follow up BCa patients, it seems to be declining in understanding sex differences and new avenues of patient monitoring to broaden therapy to decrease global rates of diagnosis and death [8,9].

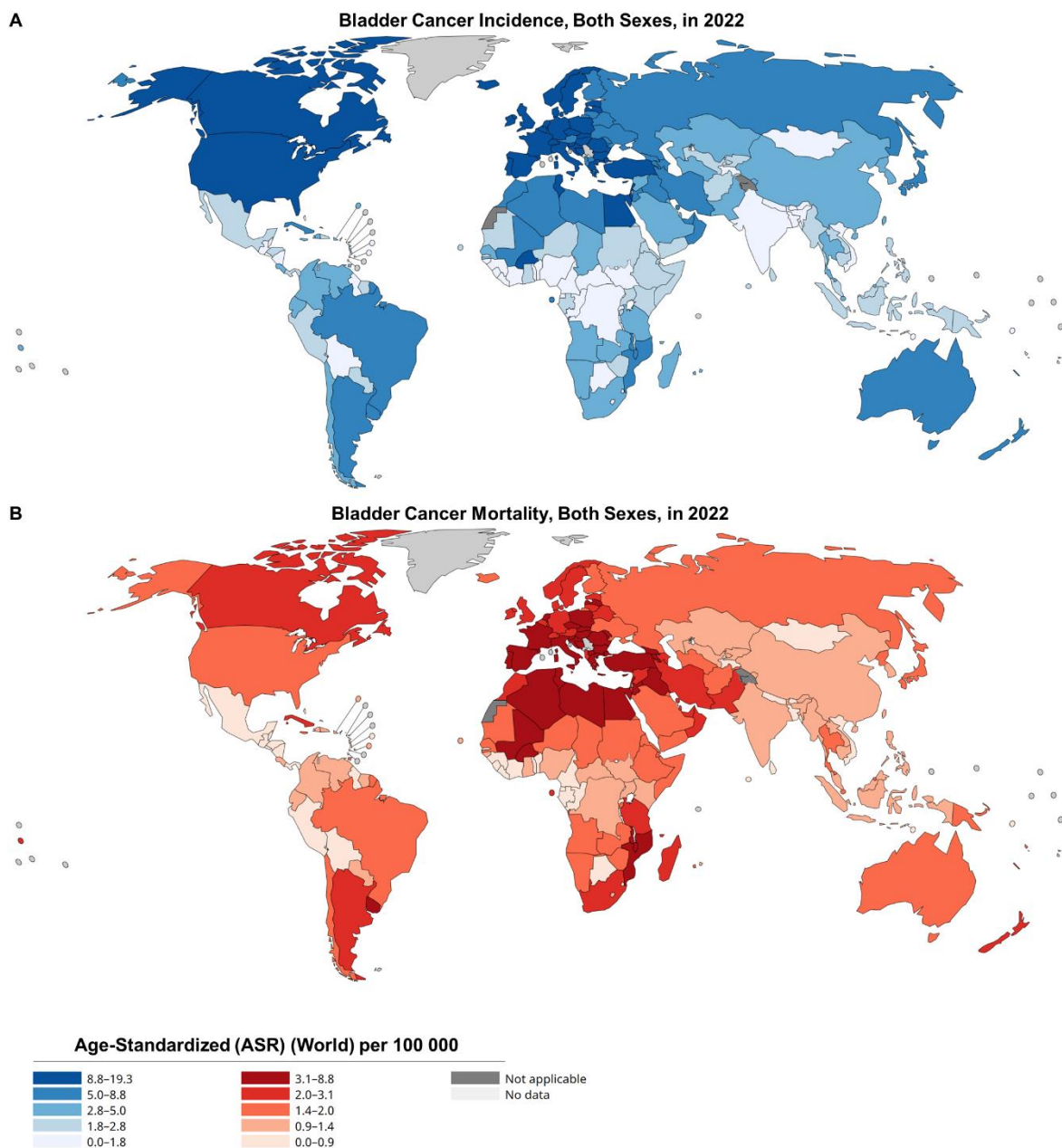


Figure 1-1. Global incidence and mortality of bladder cancer in 2022 in men and women fo all ages. A. Global estimated incidence of bladder cancer in 2020 in men and women of all ages. B. Global mortality of bladder cancer. Data are expressed as age-standardized rates (ASRs; adjusted to World Standard Population) to account for differing

age profiles among regions. Data were obtained from GLOBOCAN 2022, and maps were generated by the World Health Organization (WHO) (<https://gco.iarc.fr/today>) [7]).

1.1.2 The hidden burden of bladder cancer

Recent advancements have led to modest improvements in survival rates for BCa compared to other cancers [10]. However, despite these gains, the five and ten-year survival rates for BCa patients remain disappointingly low [11,12]. Adding to this challenge, BCa has the highest recurrence rate among all cancers, with figures ranging from 30% to 70% [13–15]. For those with high-risk tumours, the likelihood of progression is notably high, between 10% and 30% [15]. Consequently, BCa patients face a demanding and prolonged treatment journey, necessitating regular and invasive monitoring throughout their lives [16]. These realities underscore the critical need for improved early detection methods for BCa, enhanced patient care strategies, and the assurance of consistent, in-depth follow-up care.

1.1.2.1 The Economic Toll of Bladder Cancer

Despite progress in detection and treatment methods, BCa continues to contribute significantly to the global cancer burden, with its effects felt most in developed countries [17,18]. A notable aspect of BCa care is the substantial cost of treatment per patient, driven by the need for ongoing follow-up and meticulous monitoring of the disease's progression. This is particularly true for high-grade T1-stage bladder cancer tumours, which are known for their high rates of recurrence and progression [19,20]. BCa patients represent the highest lifetime treatment costs among all cancers, with expenses per patient ranging between \$140,000 and \$220,000 [21]. Annually, the United States allocates over \$4 billion to bladder cancer medical expenses, whereas the European Union's expenditure reaches €4.9 million [22,23]. This substantial economic impact underscores the importance of developing non-invasive and cost-effective diagnostic methods to mitigate its escalating financial strain.

1.1.3 Landscape of bladder cancer

1.1.3.1 Classification and staging

Bladder cancer (BCa) classification and staging are essential in diagnosing and identifying the proper treatment. Several grading and classification systems for bladder tumours have been suggested of which the two most used and endorsed in clinical practice are the 1973 and 2004/2022 World Health Organization (WHO) grading systems. Such classification systems aim to classify BCa tumours in grades defined by the morphologic properties of cancer cells and

their similarity to normal bladder tissue. The WHO 2002/2022 classification system is the latest classification system recommended for clinical application, yet the WHO 1973 is still in use [24,25].

The 2004/2022 WHO Grading System (Fig.1-2) grading system classifies tumours into Papillary Urothelial Neoplasm of Low Malignant Potential (PUNLMP), Low-Grade (LG) and High-Grade (HG) according to the degree of malignant cytologic features and cell architectural abnormalities. On the other hand, the still used in 1973 WHO grading system differentiates between Grade 1 (G1) corresponding to LG, Grade 3 (G3) to HG, and Grade 2 (G2) as an intermediate grade (Fig.1-2) [26]. Both classifications are efficient and recommended by the European Association of Urology (EAU), but it was shown that the 1973 WHO classification is more suitable for the prediction of tumour recurrence [25,27].

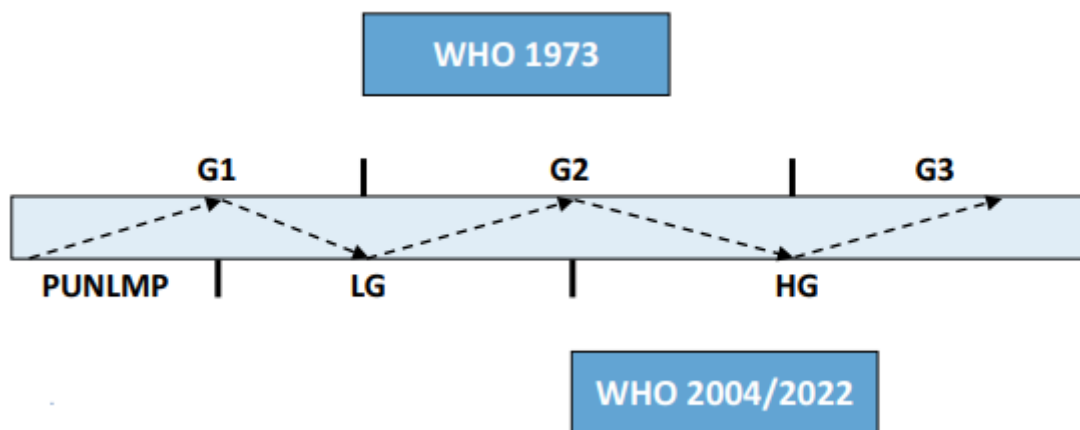


Figure 1-2. World Health Organization (WHO) 1973 and 2004/2022 bladder cancer clinical classification. Adapted from the non-muscle Invasive bladder cancer European Association of Urology guideline [26,27]

Regarding the pathological and depth of invasion staging, the tumour, node, and metastasis staging (TNM) system is the most prevalent cancer spread severity classification method authorized by the Union International Contre le Cancer (UICC). As shown in Table1-1, T stage of bladder tumours refers to the depth of tumour penetration into the layers of the bladder tissue, N describes lymph node involvement, and M is the indicator of distant metastasis [28].

Stage Tis is the classification of cancer cells in situ (CIS), confined to the urothelium, consisting of poorly differentiated cancer cells. However, PUNLMP is characterized by papillary lesions with increased cellular proliferation but without cellular atypia, rendering a very low risk of progression, but recurrence is frequent. Pathologically, non-invasive papillary carcinomas, graded as Ta, present in a papillary architecture within the bladder wall and show diversity in nuclear features and chromatin patterns. T1 has metastasized into the connective tissue but

did not invade the bladder muscle layers. The prognostic significance of depth and extent of invasion into the lamina propria (T1 sub-staging) has been shown in some retrospective cohort studies. Stages T2, T3, and T4 triplets are defined respectively for muscle layer invasion, perivesical tissue invasion, and other organ metastases [28].

Table 1-1. TNM classification of bladder cancer. Adapted from tumour TNM classification [28].

TNM classification of urinary bladder cancer

T - Primary tumour	
TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Ta	Non-invasive papillary carcinoma
Tis	Carcinoma <i>in situ</i> . 'flat tumour'
T1	Tumour invades subepithelial connective tissue
T2	Tumour invades muscle
T2a	Tumour invades superficial muscle (inner half)
T2b	Tumour invades deep muscle (outer half)
T3	Tumour invades perivesical tissue
T3a	Microscopically
T3b	Macroscopically (extravesical mass)
T4	Tumour invades any of the following: prostate stroma, seminal vesicles, uterus, vagina pelvic wall, abdominal wall
T4	Tumour invades prostate stroma, seminal vesicles, uterus or vagina
T4b	Tumour invades pelvic wall or abdominal wall
N – Regional lymph nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in a single lymph node in the true pelvis (hypogastric, obturator, external iliac, or presacral)
N2	Metastasis in multiple regional lymph nodes in the true pelvis (hypogastric, obturator, external iliac, or presacral)
N3	Metastasis in common iliac lymph node(s)
M - Distant metastasis	
M0	No distant metastasis
M1	Non-regional lymph nodes
M1b	Other distant metastases

1.1.3.2 The Origin and evolution of bladder cancer

Bladder stem cells can produce various subtypes of urothelial cells, a process normally well-regulated to prevent tumour growth [29]. On the other hand, stem cells or more differentiated progenitor cells can become tumorigenic due to genetic or epigenetic alterations, often caused by prolonged exposure to risk factors and lifestyle habits. Additionally, specific changes are already known to initiate bladder cancer, correlating with specific clinical pathologies [30].

BCa falls into two main categories (Fig.1-3A), non-muscle invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC), each with distinct pathogenesis pathways and molecular features. Understanding such different features corresponding also to the different tumour stages, such as the risk of recurrence and progression can improve the patient treatment (Fig.1-3B). Despite mutational signatures being consistent across tumour grades and stages, reflecting various underlying mechanisms, the clinical and therapeutic burdens are significantly higher for MIBC than for NMIBC (Fig.1-3C) [1].

In both NMIBC and MIBC, chromosome 9 deletions, which include deletions of the CDKN2A gene locus, are frequent [31]. Such deletions interfere with critical regulatory pathways in tumour suppression, highlighting their importance in bladder cancer progression. In NMIBC, other common mutations occur in genes such as FGFR3 and PIK3CA, thus activating signalling pathways involved in cell growth and proliferation [31,32]. However, in MIBC, there is an exceptional intratumor genetic heterogeneity with mutations in genes such as TP53, RB1, and ATM, which are crucial for cell cycle checkpoint regulation and DNA repair [33,34]. Moreover, chromatin modifiers are altered in NMIBC and MIBC, with differences only in the distribution of mutations. Indeed, in MIBC, the RAS-MAPK and PI3K pathways, associated with cell growth and division, are often activated [35]. Such activation can be produced by mutations of genes such as ERBB2 and ERBB3 or loss of genes such as PTEN and TSC1, which inhibit these pathways in normal situations [35,36]. Additionally, the epithelial-mesenchymal transition (EMT) and the Notch signalling pathway represent additional mechanisms involved in the progression of MIBC, illustrating the complexity of BCa development and progression [37,38].

The origin of bladder cancer presents multiple risk factors most of which are lifestyle-based and can significantly contribute to its prevalence [39]. Tobacco smoking stands as the leading risk factor, accounting for about 50–65% of new cases per year, with three to four times the risk compared with non-smokers [40]. Furthermore, environmental, and occupational exposures such as aromatic amines and polycyclic aromatic hydrocarbons that can be found in different industrial settings contributes to 18% of BCa cases [41,42]. In addition, the effect of alcohol consumption is not statistically significant but may cause a mild increase of risk, and

high consumption of red and processed meats increases the risk by 17% to 22%, which is due to nitrosamines present in processed meats [43]. Obesity also contributes, increasing the risk by around 10% regardless of other lifestyle factors [44]. Last, pathogen exposure particularly from schistosomiasis in endemic regions increases bladder cancer risk through processes involving inflammation and DNA damage [45].

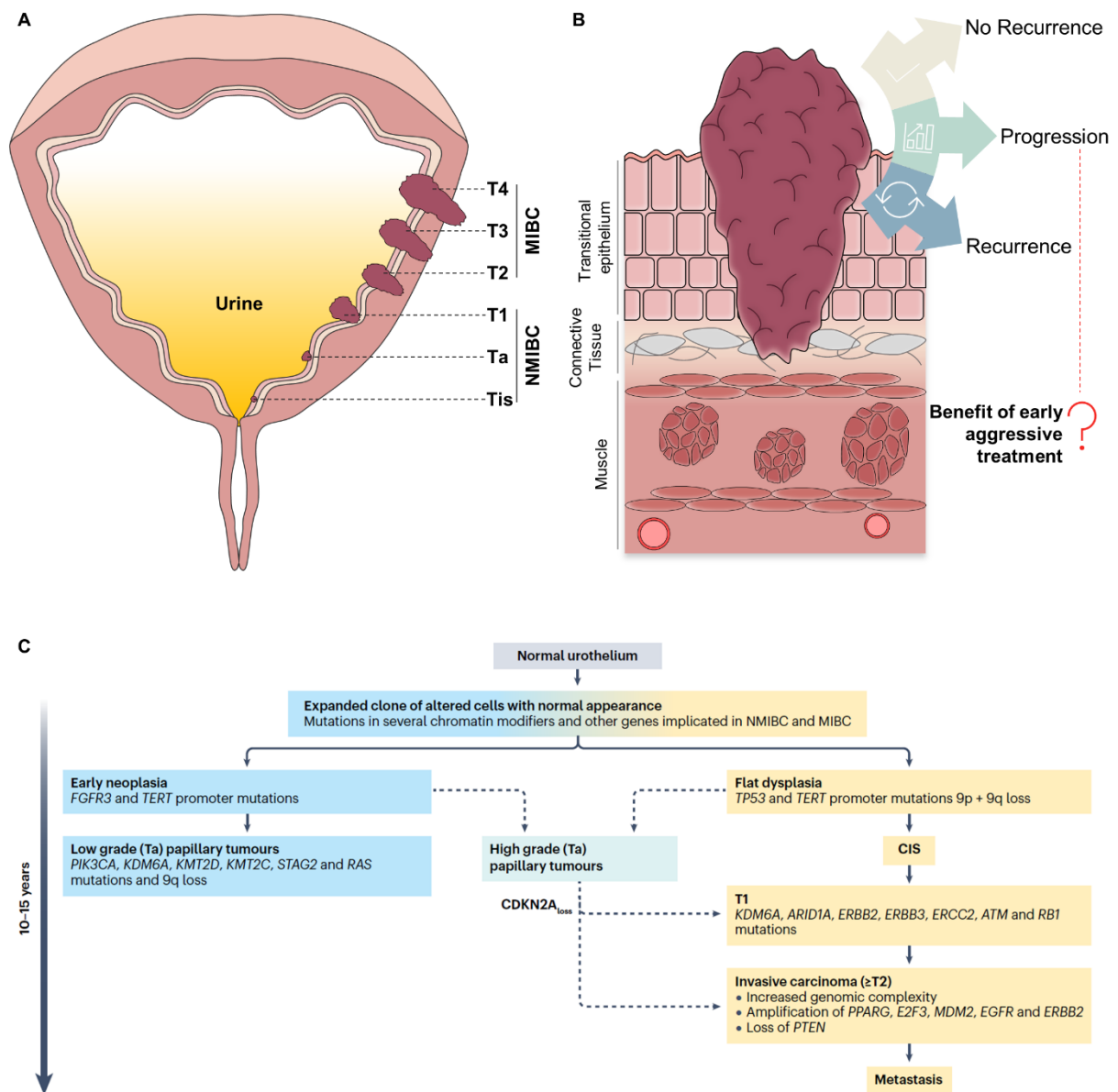


Figure 1-3. Schematic classification and origin of bladder cancer (BCa). **A.** Classification and grouping of BCa tumours. **B.** Importance of the type of tumour into understanding the best course for medical treatment according to the possibility of no recurrence, recurrence, and/or progression. **C.** Different associated pathways and gene mutations to the development of BCa. Adapted from Dyrskjøl *et al.* [1].

1.1.4 Bladder Cancer Diagnosis and Treatment

1.1.4.1 Clinical Presentation and Diagnostic Approach

Usually, BCa begins with visual blood in the urine, which typically indicates the need for cystoscopy and upper tract imaging to locate and evaluate the stage of the disease [46]. Furthermore, a detailed patient history is essential, with gross haematuria being a common finding in BCa and often associated with higher-stage disease [47]. Also, symptoms of the lower urinary tract may suggest BCa. However, these symptoms are not always concomitant with cancer, and as such, a meticulous diagnostic assessment is necessary [48].

Among the diagnostic tools available, urine cytology stands out as the most cost-effective urine-based method for diagnosing bladder cancer. Despite its less-than-optimal sensitivity, urine cytology is notably precise in detecting high-grade urothelial carcinoma. However, invasive cystoscopy, employing various imaging technologies like white light, remains the definitive method for accurately diagnosing and subsequently removing bladder tumours [49]. Thus, despite the advent of various marker-based urine tests, the combination of urine cytology and cystoscopy retains its status as the cornerstone in bladder cancer diagnosis.

1.1.4.2 Imaging Modalities

In the clinical stage of BCa, imaging is very important in the diagnostic aspect to screen whether there is no malignancy or metastasis. Computed tomography (CT) scan with contrast substance is the most popular screening examination protocol because it has the highest resolution and the easiest accessibility [47]. However, magnetic resonance imaging (MRI) and magnetic resonance urography are widely preferred in patients with contraindications that cannot undergo CT [50]. Dual-energy spectral CT can potentially solve the problem of blurred tumour characterization over the standard CT scanner [51]. Secondly, Multiparametric magnetic resonance imaging (mpMRI) with T2-weighted (T2W) can provide a higher spatial resolution and reduce staging error that can be used for local staging and assessment of treatment response stages through systems such as the scoring system vesical imaging reporting and data system (VI-RADS) [52,53]. CT is valuable and arguably used more than other technologies for evaluating bone and lymph node involvement and metastasis staging. However, CT has drawbacks in the small lesion's detection. Furthermore, Diffusion-weighted magnetic resonance imaging (DW-MRI) and ^{18}F FDG-PET-CT are preferred imaging methods for lymph nodes and distant metastasis detection as they have more sensitivity than conventional CT [54].

1.1.4.3 Cystoscopy and Transurethral Resection of Bladder Tumour (TURBT)

The diagnostic gold standard in bladder cancer is the invasive cystoscopy examination combined with a histological assessment [55]. Transurethral resection of bladder tumour (TURBT) is frequently carried out during cystoscopy as part of diagnosis and treatment for BCa. As the quality of tumour resection can directly affect the patient outcome, ensuring a good quality TURBT is crucial to prevent future tumour recurrence and/or progression. Hence, re-resection is frequently required, for high grade tumours like T1 or non-intersected muscle layer resections [56].

Nonetheless, the unified standard method of diagnosing and treating BCa, brings numerous drawbacks and high patient treatment morbidity. Such side effects can be attributed to the thermal destruction, absence of detrusor muscle, tumour seeding, incomplete resection and wrong pathological diagnosis [11].

1.1.4.4 Cytology and Urinary Markers in Bladder Cancer Detection

Urine cytology is a valuable tool, but its sensitivity level depends on the grade and stage of the tumour. Although the low sensitivity, urine cytology remains the only marker that is recommended by the treatment guidelines for BCa detection.

Different sources of urinary markers showed in Fig.1-4 can provide more information, especially to diagnose and follow up BCa patients [57,58]. Non-invasive urine-based novel markers and tests are promising in prospective evaluations, improving the negative predictive value, and potentially modifying follow-up strategies [59]. These non-invasive approaches have gained high popularity and as a result a lot of urinary tests and markers have been developed, with focus on urinary proteins, exfoliated tumour cells, RNA, and DNA (Figures 1-4). However, their clinical utility remains as a non-recommended approach to diagnose BCa even though they are market available, and some have got the approval of the food and drug administration (FDA). More exactly, only two protein-based urinary markers, bladder tumour antigen (BTA) and nuclear matrix protein 22 (NMP22), and two cell-based urinary markers, Immunocyte/uCyt + and UroVysion, are approved to the date [60–63]. The major challenge in implementing urine markers in practice is the absence of strong evidence from prospective randomized trials.

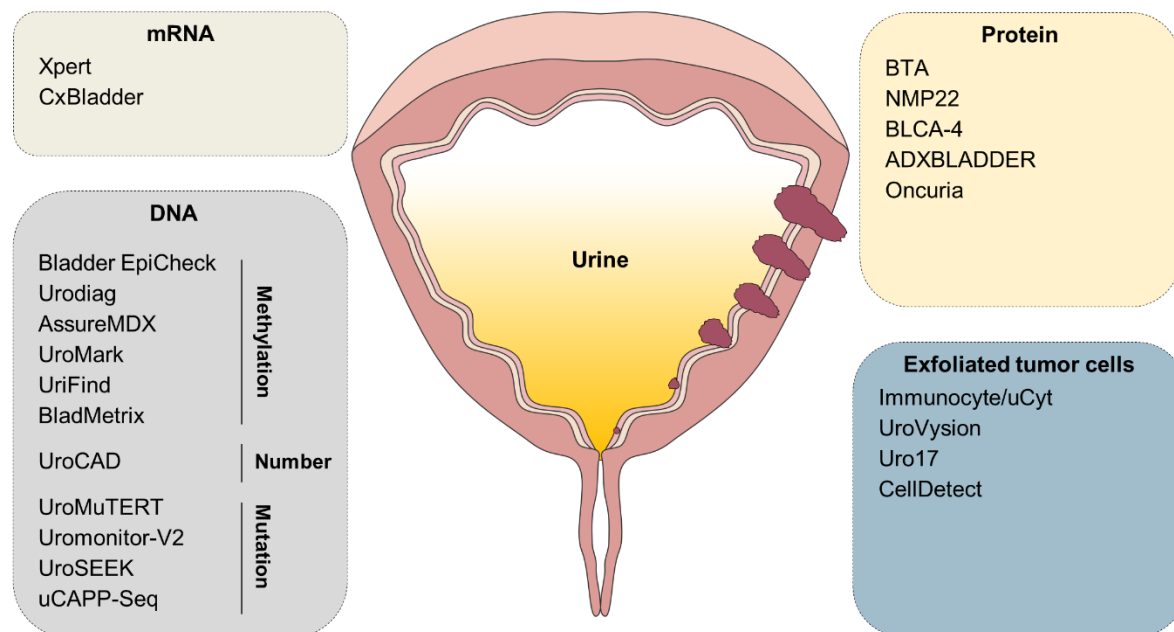


Figure 1-4. Urinary makers and tests used to diagnose bladder cancer. Protein based makers: BTA [60,61], NMP22 [60], BLCA-4 [64], ADXBLADDER [65], Oncuria [66]. mRNA-based markers: Xpert [67]; CxBladder [68]. DNA-based markers: Bladder EpiCheck [69]; Urodiag [70]; AssureMDX [71]; UroMark [72]; UriFind [73]; BladMetrix [74]; UroCad [75]; UroMuTERT [76]; Uromonitor-V2 [77]; UroSEEK [78]; uCAPP-Seq [79]. Exfoliated tumour cells-based markers: Immunocyte/uCyt [60,62], UroVysion [63]; Uro17 [80] and CellDetect [81]. Figure adapted from Tomiyama *et al* [58].

1.1.5 Other oncoproteomics case studies

During my Ph.D., I worked with several human cancers for standalone studies, one of which should be briefly mentioned and introduced here.

1.1.5.1 Renal Cell Carcinoma: Navigating Diversity and Diagnostic Challenges

Kidney neoplasms and renal cell tumours account for approximately 90% of all kidney cancers, predominantly affecting individuals in their early 60s [82]. This type of cancer is diagnosed twice as often in men as in women. Being a heterogeneous disease, kidney tumours vary from benign growths such as renal oncocytoma (RO) to more aggressive renal cell carcinomas such as clear cell (ccRCC), papillary carcinoma (pRCC) and chromophobe (chRCC) [83]. Clear cell renal carcinoma is the most common and aggressive variant that comprises 75% of renal cell cancer cases. Papillary carcinoma accounts for about 10% of cases and has two subtypes with different behaviour [84]. Type 1 papillary presents itself with less aggressive features and basophilic cells encircling a membrane, while type 2 is more heterogeneous with granular cells, prominent nucleoli and necrosis, which makes it more aggressive [85]. Following this, chromophobe renal carcinoma is the third most common subtype, occurring at 5%. Coming with large

cells and reticulated cytoplasm, are more likely to metastasize and develop liver metastases [86]. Eventually, benign renal oncocytoma makes up 5% of the kidney's tumour diagnoses, with round, polygonal cells with dense eosinophilic cytoplasm and uniform round nuclei [85].

Precise tumour typing is critical for individualising treatment approaches and avoiding over- or under-treatment. More specific markers complement inconclusive imaging scans to direct proper clinical decisions [87]. Surgery is the mainstay for localized tumours; however, active surveillance is gaining ground as a viable option in selected cases, offering clinical sparing patients from potential surgical morbidities. Additional attempts to seek molecular markers for early recognition and prediction also persist, aimed at perfecting patient care in the complicated web of renal cell carcinoma management [88].

1.2 The Promise of Proteomics

The field of proteomics, with its focus on the comprehensive analysis of the proteome, offers a powerful approach to understanding the complexity of biological systems. The proteome encompasses all the proteins expressed by an organism at a given time, reflecting not just the genetic blueprint but also how various factors, such as environmental influences, disease states, and developmental stages, can affect protein expression and function [89,90]. The complexity of the proteome is further amplified by post-transcriptional modifications of mRNA and post-translational modifications (PTMs) of proteins themselves, which can significantly alter protein function and interactions. These modifications contribute to the existence of over a million distinct protein species, or proteoforms, highlighting the vast functional diversity within biological systems [91].

Proteomics presents an invaluable resource in biochemical and biomedical research, facilitating a deeper understanding of the molecular phenotypes of living organisms across different levels, from whole organisms to specific parts like cells, tissues, or organs [92]. The field is closely associated with large-scale, high-throughput approaches to protein measurement, which are essential for capturing the full scope of the proteome's complexity. These methodologies contrast single-protein studies, offering a more comprehensive view of protein functions within complex biological networks [93,94].

1.2.1 Extracting information from complex proteomes

Mass spectrometry (MS) has become a crucial tool in biochemistry and systems biology, leveraging its unmatched speed, throughput, and sensitivity to analyse complex biological

systems [95]. MS can identify and quantify vast arrays of proteins and other biomolecules simultaneously, allowing for detailed characterisation of cellular processes and molecular interactions within these complex biological systems [96]. Additionally, MS is invaluable for detecting structural modifications and PTM and providing insights into the dynamic changes and functional states of proteins and other biomolecules [97,98].

The move towards industrial-scale high-throughput proteomic profiling has expanded the toolkit beyond MS-based methods to include techniques like antibodies or aptamer-based affinity agents. Innovative proteomic technologies such as SomaLogic and Olink, along with genomic data, are now providing extensive insights into protein expression profiles of samples [99,100]. While both SomaLogic and Olink are affinity-based platforms, they employ distinct mechanisms; SomaLogic leverages aptamers for specific protein targeting, with each aptamer uniquely binding to its target protein, whereas Olink relies on an immunoassay approach that utilises a pair of antibodies to detect specific proteins [99,101,102].

However, the capacity of SomaLogic and Olink to analyse post-translational modifications (PTMs), such as phosphorylation, is limited by the design of aptamers or antibodies that can specifically recognise these modifications. For instance, differentiating proteins based on phosphorylation sites poses a significant challenge. Phosphorylation often occurs at specific amino acid residues (serine, threonine, or tyrosine), drastically altering a protein's function [103]. However, because this modification only causes a subtle change in the protein's structure, designing aptamers that can distinguish between phosphorylated and non-phosphorylated forms of a protein, particularly with high specificity to the phosphorylation site, would be exceptionally difficult.

While affinity-based techniques, such as those employed by SomaLogic and Olink, provide valuable processes like those achievable with LC-MS, they trade off selectivity in favour of characterization [104,105]. In contrast, MS-based proteomics excels with its untargeted approach, which enables precise protein identification and quantification and overcomes the limitations of affinity-based techniques. This distinct capability positions MS-based proteomics as a method of choice for biomarker discovery, offering unparalleled accuracy and comprehensive proteome coverage.

1.2.2 Mass spectrometry (MS)-based proteomics

Since its first application in 1912, mass spectrometry has been used in numerous scientific and technological industries. It is a very useful technique for determining the mass-to-charge ratio (m/z) of charged ions in the gas phase. This process allows for accurately

determining molecular weights and structural characterisation of molecules. Initially, the application of MS to the analysis of biomolecules was limited because of the harsh conditions of traditional ionization techniques, which often led to the fragmentation of these delicate structures [106,107].

Soft ionisation methods, notably matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI), were introduced towards the end of the 20th century. The groundbreaking development of these soft ionisation techniques led to the Nobel Prize in Chemistry awarded in 2002 to Koichi Tanaka and John B. Fenn [108–110]. Their contributions enabled the ionisation of large, complex biomolecules without significant degradation. This pivotal advancement greatly expanded the applicability of MS across diverse fields, including biochemistry, biomedical research, and clinical applications [111]. This innovation transformed MS into a versatile tool capable of characterising a broad spectrum of biomolecules, including proteins, metabolites, DNA, and polymers at exceptional sensitivity.

1.2.3 MS-Proteomics methodologies

In modern proteomics, researchers adopt two primary strategies to dissect the complexity of the proteome: the bottom-up and top-down approaches. Each of these methodologies offers unique advantages and is suited to different types of proteomic analyses [112].

Bottom-up proteomics is the most widely used approach in proteomic studies, it involves the enzymatic digestion of proteins into smaller peptides, typically using enzymes like trypsin, before their analysis by MS. This method allows the identification and quantitation of individual peptide sequences within complex mixtures. It is particularly effective for characterising PTMs and for studies where the protein mix is too complex to be resolved by top-down approaches.

The top-down proteomics approach has become popular in the last few years as a replacement for the standard shotgun proteomics methods. Unlike this, which concerns peptide analysis, top-down proteomics deals with direct whole-protein analysis. In this method, proteins are directly ionised and then analysed by MS. Inside the MS system, these intact protein ions are broken, and all parts of the protein are examined [113,114]. Although the top-down approach presents some potential for the MS-based analysis of intact proteins, the proteomics community still needs to rely on bottom-up methods for comprehensive proteome coverage.

Bottom-up proteomics (Fig.1-5) benefits from a wealth of established protocols optimised over the years for a wide range of sample types. Also, it offers high sensitivity [115]. By digesting proteins into peptides, the method increases the likelihood of detecting low-abundance proteins. Peptides are generally more amenable to MS analysis than intact proteins,

partly because they ionize more efficiently and produce more straightforward spectra. This higher sensitivity is crucial for uncovering the full depth of the proteome, including detecting proteins that are present in very small quantities but may play significant roles in biological processes or disease states [116].

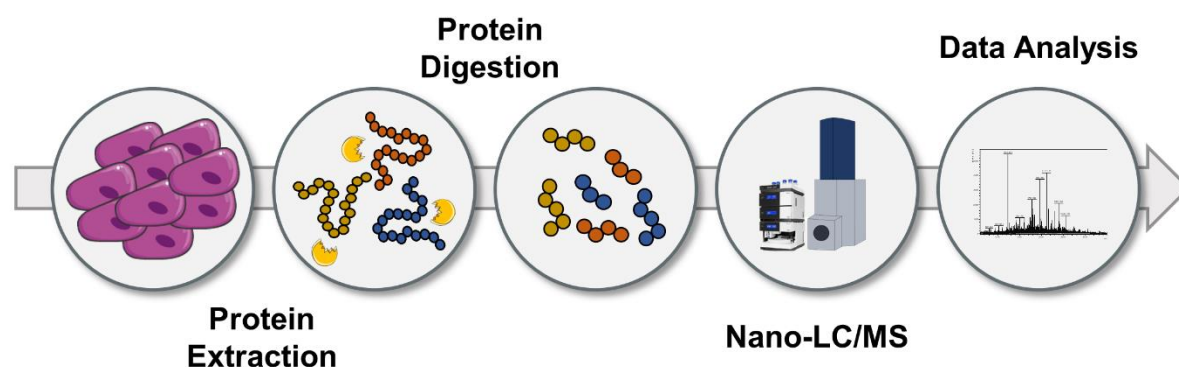


Figure 1-5. Bottom-up proteomic workflow. A typical bottom-up proteomic approach includes several steps: (i) Isolation and extraction of the protein content followed by quantification of the isolated proteins (e.g. Bradford assay); (ii) Protein digestion typically performed by trypsin or a combination of multiple-enzyme digestion to achieve a better proteome coverage [117]; (iii) Mass spectrometry analysis of the resulting peptides and (iv) Data computation for protein identification and quantification.

1.2.4 Liquid chromatography in mass spectrometry

Liquid Chromatography coupled with tandem Mass Spectrometry (LC-MS/MS) represents a pivotal technology in modern bioanalytical chemistry and biochemistry, especially in proteomics, which is indispensable for analysing complex biological samples. This technique combines the separation capabilities of liquid chromatography (LC) with the detection and quantification capabilities of MS, offering unmatched sensitivity and specificity [118,119].

The LC component of LC-MS/MS separates the components of a complex mixture based on their differential interactions with a stationary phase (typically packed in a column) and a mobile phase (a solvent that moves through the column). The separation principle is governed by molecules' affinity for the stationary phase versus the mobile phase, which can be manipulated by choosing different column chemistries (e.g., reverse phase, ion exchange, size exclusion) and by varying the composition of the mobile phase [118,119].

The performance of chromatographic methods is critical to the precision and reliability of LC-MS/MS analysis. In clinical LC-MS/MS analysis, after the sample preparation, the sample extract is loaded onto the chromatography column. High-pressure pumps carefully deliver mobile phases to the column at constant flow rates. Gradient elution is the most common

technique where the mobile phase composition changes throughout the analysis, and it usually takes the required time for analytes to elute. Furthermore, in a tightly controlled environment, the retention time of the analytes eluting from the column is reproducible [120]. The choice of the chromatography column, comprising aspects like length, internal diameter, and packing material, is essential, often determined by the analytes being studied and the pressure limitations of the liquid chromatography system [121].

1.2.5 Ionization & Mass Analysers

As highlighted before, mass spectrometry analysis has become an indispensable tool for modern research, especially in proteomics. The process begins with eluent from the liquid chromatography containing the analyte being fed to the mass spectrometer source, where it undergoes ionization to produce ions in the gas phase. After that, the ions are guided into the mass analyser, where the mass-to-charge ratio of the intact ions is measured. This step precedes the tandem mass spectrometry (MS/MS) experiments, which are essential for proteomic analysis [122]. Upon establishing the m/z values of the intact ions, typically peptides in proteomics, a controlled fragmentation is induced, producing a series of fragment ions. These fragments, each carrying pieces of the original molecule's structural information, are analyzed in the subsequent stage of the tandem mass spectrometer. This analysis provides detailed insights into the peptide sequences and any post-translational modifications.

1.2.5.1 Electrospray Ionization

To travel the trajectory of the electrodynamic flow in the mass spectrometer, molecules must be ionized and brought into the instrument. Ionization before MS analysis can be done in several ways, including electron ionization, chemical ionization, electrospray ionization (ESI), photoionization, inductively coupled plasma, and matrix-assisted laser desorption ionization (MALDI). For protein analysis, it is crucial to utilize soft ionization methods that ionize peptides with minimal fragmentation. Therefore, MALDI and ESI are adopted as the main methods of protein analysis. These methods gently ionize peptides and thus preserve their structural integrity, enabling precise determination of protein composition and structure [123,124].

Electrospray ionization (ESI) operates at atmospheric pressure and is compatible with nearly all analytical instruments, marking its versatility and wide applicability in mass spectrometry. In the ESI process, a solution containing the analyte dissolved in a conductive solvent is subjected to a high voltage. This application of voltage generates an electrically charged aerosol by inducing the formation of a Taylor cone—a sharp, conical shape at the tip of the liquid

flow. The formation of the Taylor cone is crucial, as it is the point from which the charged droplets are emitted into the aerosol. Assisted by a sheath gas, the solution is then nebulized, transforming into a stream of charged droplets [125].

As these droplets progress, they encounter a heated gas stream, leading to solvent evaporation and a decrease in droplet size. This process continues until the droplets reach a critical juncture—the Rayleigh limit—where the surface tension can no longer hold back the Coulomb force of repulsion among the charged molecules within the droplet. At this critical point, a phenomenon known as a Coulomb explosion or Coulomb fission occurs. This critical dispersion mechanism breaks the droplet apart, effectively releasing the ionized analyte into the gas phase. Once in the gas phase, the analyte is ready for analysis by the mass spectrometer [126].

ESI technical aspects, such as being a liquid-based system, and the need for a conductive solvent, make it compatible with reverse-phase LC, which makes it appropriate for the analysis of many compounds and especially of polar molecules. The implementation of ESI has effectively broadened the range of MS- based analyses, especially in biological systems where most compounds are water-soluble and polar. In addition, ESI is adaptable and can be utilized to analyse various molecular masses, from small molecules to large biomolecules such as proteins [127].

1.2.5.2 Mass analysers

The principle of mass spectrometry is that ions are separated according to their mass-to-charge ratio (m/z). Different kinds of mass analyzers are used for mass spectrometry analyses, including quadrupole, magnetic sector, radio-frequency ion trap, time of flight (TOF), orbitrap, and ion cyclotron resonance analyzers [128,129].

These mass analysers can be connected for tandem mass spectrometry (MS/MS), which is the idea of conducting two consecutive mass analyses, usually with a fragmentation step in between them. One of the methods of MS/MS is implemented by applying consecutive identical mass analysers, e.g., triple quadrupole or TOF/TOF analysers. Hybrid instruments can also be developed using various types of mass analysers, such as (i) quadrupole and TOF for Q-TOF analysers, (ii) trapped ion mobility(tims) and Q-TOF for timsTOF instruments, which introduces the parallel accumulation serial fragmentation (PASEF®) [130], or (iii) quadrupole, orbitrap and astral analyzer for Orbitrap Astral MS [131].

1.2.6 Acquisition Methods

In mass spectrometry-based proteomics, data acquisition strategies are typically categorized into discovery and targeted approaches. Discovery strategies, represented by data-dependent acquisition (DDA) and data-independent acquisition (DIA), strive to achieve the most extensive identification and quantification of peptides within the available measurement duration. Nevertheless, both DDA and DIA suffer from their inherent bias, and due to this, they have limitations in the consistent identification of low-abundance peptides. This is problematic, particularly with reproducible analysis of specific peptides expected to be low in abundance, as are those derived from proteins such as kinases, transcription factors or post-translationally modified peptides [132].

On the other hand, targeted methods like selected reaction monitoring (SRM), also known as multiple reaction monitoring (MRM), and parallel reaction monitoring (PRM), provide extreme sensitivity, high reproducibility, and fast detection, as well as accurate quantification of a predefined set of peptides. In a standard SRM or PRM run, the mass spectrometer collects spectra at the predicted m/z ratio and chromatographic retention time (RT) of the desired peptides. During SRM experiments, precursor ions are first filtered by the first quadrupole. Subsequently, they are fragmented in the second quadrupole or in a collision cell, and finally, in the third quadrupole, target-specific fragment ions are selected for detection. On the other hand, PRM experiments employ systems that can record whole fragment spectra and monitor all fragment ions of a chosen precursor simultaneously [133,134].

For SRM and PRM configurations, an inclusion list is supplied to the mass spectrometer with the required target precursor m/z ratio windows and RT window, thus guaranteeing the targeted acquisition of the determined peptides of interest. This specialized procedure provides increased sensitivity and reproducibility, especially useful for quantitative proteomic analyses in different research and clinical settings [135].

The two different approaches are complementary to each other. First, a discovery approach is run, where the aim is to identify and explore the all-proteome landscape, while second, it is possible to track targeted peptides that have a biological mean.

1.2.6.1 Data-dependent acquisition (DDA)

DDA is one of the major approaches used in tandem MS experiments for data collection. DDA is an MS method that selects the top N most abundant precursor ions for fragmentation based on their intensity levels during the MS scan. This selection occurs within narrow isolation

windows, typically around 1 Da wide, which may broaden for higher m/z values to ensure the complete isotopic cluster is isolated for analysis [136].

To mitigate the repetitive selection of the same peptide precursors, DDA employs a dynamic exclusion strategy. This technique temporarily blocks the re-selection of precursor ions with previously analyzed m/z ratios for a set period, enhancing the diversity of analyzed precursor ions. However, the semi-stochastic nature of m/z selection in DDA, where different precursor ions may be chosen in subsequent analyses, presents a challenge with data completeness, resulting in missing data. An innovative resolution to this challenge lies in enhancing the speed of mass spectrometry instruments. Theoretically, if the instrument possesses the capability to sequence all or nearly all multiply charged MS1 features in every run, it could achieve completeness in data, at least for all precursors detectable in MS1 scans [136]. This concept finds practical application in instruments equipped with advanced hardware capable of rapid ion m/z detection, such as TOF mass analyzers. The Orbitrap mass analyzer, which revolutionized MS-based proteomics in the early 2000s through its high-resolution spectra, encounters a limitation in speed due to the lengthy ion transient times required for high resolution, typically in the tens of milliseconds. In contrast, TOF analyzers perform scans of the entire mass range in a single pulse lasting around 100 μ s, making them exceptionally well-suited for swift detection. This capability is exploited in the PASEF scan mode, where precursors, after being eluted from a mobility device, are further isolated within a narrow m/z window, and then fragmented. The rapid scanning ability of the TOF, alongside the controlled release of ions from the mobility device and the quadrupole's switching time, becomes the focal point in this setup [130,137]. Alongside technological advancements, DDA is lauded for its efficiency. It enables comprehensive proteome coverage across a wide dynamic range, making it a pivotal tool for identifying and characterizing a vast array of peptides and proteins. Additionally, DDA lays the groundwork for creating peptide libraries, which are vital for the execution of DIA analyses.

1.2.6.2 Data-independent acquisition (DIA)

Another data acquisition technique in tandem mass spectrometry is the data-independent acquisition (DIA) mode that acquires both MS1 and MS2 data unbiasedly [138]. Introduced in the early 2010s, DIA performs MS2 scans across a wide range of precursor m/z values using expansive isolation windows, ensuring no precursor is left unfragmented. However, this method generates MS2 spectra that include ion series from multiple peptides that are not identifiable by typical database matching. It was not until recently that the scientific community established precise guidelines for the reliable identification of peptides in DIA studies [139],

and advances in software have started to facilitate the widespread use of DIA methodologies [140,141]. Recent progress indicates that DIA can be effectively utilized across various experimental setups, incorporating additional layers such as ion mobility with TOF scans with parallel accumulation serial fragmentation[®], or the Orbitrap Astral [131,140].

1.2.7 Protein identification & quantification

1.2.7.1 Protein identification

Protein identification by MS is a powerful analytical technique that leverages mass spectrometry principles to elucidate the identity of proteins in a sample. In bottom-up proteomics, peptide ions are analysed in the mass spectrometer to determine their m/z values, which act as distinctive molecular signatures that aid in protein identification. However, protein identification in complex mixtures requires tandem mass spectrometry (MS/MS) and peptide sequencing [142]. Tandem MS takes protein identification a step further by measuring the m/z of peptide ions and also fragmenting these ions into smaller pieces and analysing the resultant fragment ions. This process, known as peptide sequencing, provides sequence information about the peptides, which can be matched against theoretical or experimental databases to identify the parent protein. Bioinformatics tools and protein databases greatly aid the interpretation of MS and MS/MS data [143]. The measured peptide mass and sequence data are compared against known protein sequences stored in databases, facilitating the identification of proteins based on the best match. This computational approach allows for the rapid and accurate identification of thousands of proteins from a single experiment.

1.2.7.2 Protein Quantification

Protein quantification in proteomics is broadly categorised into two main approaches: Stable Isotope Labelling methods and label-free quantitative proteomics [144].

Stable Isotope Labelling is a technique that introduces isotopic tags into proteins or peptides, resulting in the creation of 'light' and 'heavy' versions that correspond to different biological conditions under investigation. The workflow for stable isotope labelling involves tagging samples, followed by their combination before MS analysis. This allows for direct comparison within the same MS scan, as the mass difference introduced by the labels makes it possible to distinguish between the two states of interest [145].

Isotope labels can be introduced into samples through various means, including metabolic incorporation, where cells are grown in media containing labelled amino acids, and

chemical labelling, where reactive groups target specific amino acid termini. Certain labelling methods, such as the TMTpro 16-plex, facilitate the simultaneous analysis of up to 16 conditions within a single mass spectrometry run [146,147]. This high degree of multiplexing significantly enhances throughput and data consistency by analysing multiple samples concurrently. However, integrating 16 samples in a single run introduces several limitations:

- i. Combining many samples increases the complexity of the analyte mixture, potentially leading to ion suppression effects and compression of the dynamic range of detection. The presence of more abundant samples hinders the detection of lower-abundance peptides. This can impact detection sensitivity for specific proteins and may result in missing data for low-abundance species.
- ii. While TMTpro 16-plex offers high multiplexing capability, the ratio compression phenomenon can affect quantification accuracy. This occurs when co-isolated peptides from different samples interfere with each other's quantification, leading to underestimation or overestimation of fold changes.
- iii. In clinical proteomics, the stringent experimental planning requirements associated with labelling methods pose significant challenges, especially for long-term studies or studies where sample availability may be variable.

The previous limitations have led the proteomics community to increasingly adopt Label-free Quantitative (LFQ) methods, which offer greater experimental flexibility. LFQ approaches do not require introducing isotopic or chemical labels into the sample. Instead, they rely on directly analysing the MS signal intensities or using spectral counting techniques to infer protein abundances. This methodological shift simplifies the sample preparation process. It allows for the analysis of virtually unlimited numbers of samples, constrained only by practical considerations such as instrument time and computational resources rather than by the inherent limitations of labelling strategies [148].

1.2.7.2.1 Spectral counting quantification

Spectral counting represents a simple and user-friendly approach to label-free protein quantification in mass spectrometry-based proteomics. This approach is based on the principle that the abundance of a protein in a sample can be estimated by counting the number of times peptides from a specific protein are detected. The frequency with which peptides from a specific protein are detected—termed the spectral count—serves as an indirect measure of the protein's abundance in the sample [149].

Typically, one method used in conjunction with spectral counting is the exponentially modified protein abundance index (emPAI). emPAI is calculated from the observed spectral

count for a given protein and the number of observable peptides for that protein (theoretical peptides that can be detected by the mass spectrometer under experimental conditions). The formula for emPAI is designed to reflect the exponential relationship between the protein abundance and the number of detected peptides, thereby aiming to provide a direct proportionality to the protein concentration. This approach normalises the raw spectral counts by accounting for the total number of peptides that could potentially be detected for each protein, providing a more accurate reflection of protein abundance [150].

1.2.7.2.2 Precursor Intensity-Based Label-Free Quantification

For precursor signal intensity-based quantification, the approach is based on the mass signal intensity of a precursor ion represented by the extracted ion chromatogram. Peptide signals are first observed at the level of MS1 and differentiated from the chemical noise by their typical isotopic distributions. These patterns are followed throughout the retention time dimension and help reconstruct a chromatographic elution profile for the monoisotopic peptide mass. Further, under high-resolution MS instruments, the total ion current of the peptide signal is integrated and used as a quantification measure of the original peptide concentration [151].

Out of the precursor signal intensity techniques, the total protein approach (TPA) developed by Wiśniewski offers a standard-free and label-free absolute quantification method. Essentially, the TPA relies on the assumption that the total MS signal from all identified proteins in the dataset reflects a biochemical sense of total protein, and the MS signal from a single protein corresponds to its abundance in the studied sample. The method offers an easy way to quantify thousands of proteins per sample [152,153].

Two primary constraints are associated with employing the TPA for protein quantification. Initially, this method is best suited for large-scale proteomic studies. Specifically, for human cultured cells, a robust analysis necessitates the identification of c.a. 12,000 peptides per dataset [154]. When analysing body fluids, where a few highly abundant proteins can constitute up to 90% of the total protein mass, the threshold for accurate protein abundance assessment through TPA tends to be lower. For instance, an evaluation of human plasma, identifying 5400 unique peptides, was sufficient for accurately determining protein concentrations .

The second challenge with TPA concerns the inherent characteristics of specific proteins. Due to their primary structure, these proteins exhibit a disproportionately low number of cleavage sites available for enzymatic digestion in bottom-up proteomics. This issue is common to TPA but also impacts other label-free and targeted proteomics methodologies, where

incomplete digestion of proteins is a common obstacle. However, unlike other quantitative techniques that rely on comparisons to standards, TPA adopts a different strategy by aggregating the intensities of all peptides in a dataset. This approach effectively diminishes the impact of partial digestion on the accuracy of protein concentration estimates [152,153].

1.2.8 Recent Technological Advances of Mass Spectrometry-Based Proteomics

Advancements in mass spectrometry have been the driving force behind proteomics, leading to significant progress in understanding complex biological systems [155]. With each technological development, MS has seen substantial enhancements in sensitivity, resolution, and throughput. These advancements enable the comprehensive identification and quantification of a wider array of proteins within complex biological matrices. Consequently, proteomic characterization has become more thorough, offering improved biological and biochemical insights.

Two of the newest breakthroughs in mass spectrometry-based proteomics are the TIMS-TOF Pro and the Orbitrap Astral [131,137]. The TIMS-TOF Pro features Trapped Ion Mobility Spectrometry (TIMS) technology, a breakthrough ion mobility-mass spectrometry technology. It provides the best ion mobility separation and unmatched sensitivity, improving proteomic profiling capacity. In a few words, the ions enter the tunnel of the TIMS system and are separated according to their collision cross-sections in a buffed gas environment. A selective release of ions is facilitated by manipulating the electric field within the TIMS, improving separation efficiency. The TIMS-TOF Pro system is also enhanced with Parallel Accumulation Serial Fragmentation (PASEF®), a redeveloped MS/MS technology dedicated to the speed requirements of omics applications [140]. PASEF® uses the trapped ion mobility spectrometry to separate ions that are then detected in the quadrupole time-of-flight (QTOF) analyser to produce the TIMS MS heat map. This setup allows maximal depth of proteome coverage in single-shot proteomics even with fast LC gradients [156].

On the other hand, the Orbitrap Astral Mass Spectrometer integrates three mass analysers: a quadrupole mass analyser for selectivity and ion transmission, an Orbitrap mass analyser for high dynamic range and resolution, and the novel Astral analyser, a high-resolution accurate mass (HRAM) analyser for rapid and sensitive measurements. Particularly, it is almost lossless ion transmission via the aligned ion injection and precise 3D ion motion manipulation over a long asymmetric track provides high sensitivity and analytical throughput [131]. The front of

the instrument, until the segmented quadrupole mass filter, is designed with great care to provide maximum sensitivity and robustness. Ions are trapped and concentrated in the ion routing multipole, making their delivery to the HRAM analysers possible to make accurate and sensitive measurements [157].

1.3 Liquid Biopsies in Proteomics

Body fluid derived liquid biopsies represent highly sought after samples for all clinical applications and medical research areas. These biopsies, wandering through the organism and are closely linked to different tissues, carry very specific protein profiles that can be used as disease markers [158,159].

The combination of liquid biopsies and proteomics has significant meaning to the diagnosis and monitoring of diseases. Samples such as urine represent a non-invasive way to peek into the molecular intricacies of the patient's disease.

Additionally, with the advances of liquid biopsy approaches and methods further states their power on the shifting the system of healthcare delivery. They provide clinicians with a unique insight into the pathogenesis of diseases and lead to individually designed treatment modalities. Liquid biopsies represent the light at the end of the tunnel in the ever-changing landscape of precision medicine and lead the path toward better and patient-specific health interventions [160,161].

1.3.1 Urine a source of body information

Biomarker discovery is the focus in clinical proteomics, with the aim of finding proteins that can be used for early disease detection or diagnosis. Although blood is still the leading and most used biological specimen for biomarker discovery, urine has evolved in the past decade as an attractive alternative. This change is due to some factors such as the simplicity of urine samples as compared to plasma, and the fact that urine has proteins and peptides profiles that are unique and linked to different type of conditions [162].

In addition, urine provides several benefits by being non-invasive, the possibility to obtain large volumes and material at various hours, representing different biorhythms. Urine samples have over the years revealed several promising biomarker candidates. These biomarkers are especially promising for detecting, diagnosing, or monitoring various diseases, with particular attention paid to disorders of the urinary system [163,164].

Although several promising biomarker candidates have been identified, only a few have proceeded to the validation stages. One of the major targets of FDA-approved urine protein biomarkers is early diagnostics of acute kidney injury and bladder cancer diagnosis and monitoring. However, urinary proteomics is moving toward practical applications, but some challenges remain, especially in pre-analytical variables in urine sample handling. Addressing these variables requires more research and development of optimized standard protocols for the reliability and reproducibility of urinary proteomic analyses [163,164].

1.3.2 Urine proteome landscape

Proteins are one of the main constituents of urine, and the identified number of proteins is growing. The human urine proteome map is dynamic, showing the evolution of the urine proteome. Spahr *et al.* in 2001 briefly detected 124 urinary proteins [165]. Sun *et al.* in 2005 found 226 proteins [166], and the next year, Adachi *et al.* discovered 1543 proteins [167], which skyrocketed the number of proteins found in the human urine proteome. High-resolution mass spectrometry techniques have improved significantly. For example, Marimuthu *et al.* with 1823 protein identified and 2014 [168], Farrah *et al.* with 2491 protein [169]. Up to date, the most extensive coverage of urine proteome is by Zhao *et al.* 2017 [170]. in a detailed and comprehensive study with 3 dimensions of separation for the human urine proteome annotation delineate 6085 proteins in normal healthy urine samples, but only 2571 could be quantified.

1.3.3 Sample Treatment for MS-Based Urine proteomics.

Sample preparation is a crucial, yet often overlooked step prior to LC-MS/MS analysis, especially significant in the context of urinary proteomics. Urine exhibits a dynamic range of protein concentration comparable to plasma, featuring a variation of approximately 10 orders of magnitude in protein concentrations. Tackling this challenge within MS-based urine proteomics requires attention to three primary aspects: initial sample treatment, the processing and isolation of urine proteome, and protein digestion.

The large concentration range of proteins in liquid biopsies such as urine makes it difficult to standardize methods for medical purposes. As a rule, urine is subjected to one-shot analysis without further processing, using simple steps such as dilution or protein precipitation, or more complex methods such as solid-phase extraction (SPE) [171,172].

Filtration methods, which are complemented with centrifugation or proteome precipitation methodologies, are widely used to improve analysis and proteome coverage during

urinary proteome extraction and concentration. The simple sample preparation methods are quick; however, the resulting extracts may leave considerable sample matrix residues, which will affect the robustness and reproducibility of LC-MS/MS analyses [173].

Urinary proteins can also be precipitated by organic solvents and acids in the sample taken from urine, the supernatant thus obtained is isolated by centrifugation, and then, it can be evaporated and reconstituted to the required concentration [171]. Nevertheless, this approach may eliminate a part of proteins, which may affect the ability of the method to produce reliable and reproducible results.

Sorbents can also be handy and are used in solid-phase extraction (SPE) to retain analytes while matrix components are washed away before elution. SPE is quite effective but complex and time-consuming method usually used in drug and steroid hormone analysis in the serum and urine [174].

An emerging approach of ultrafiltration devices with non-binding membranes and molecular weight cut-off (MWCO) is viable for urine sample treatment for proteomics analyses. The mechanical separation and pre-concentration of proteins over the membrane by these devices enable peptide processing for LC-MS/MS analysis (Fig.1-6A) [175,176].

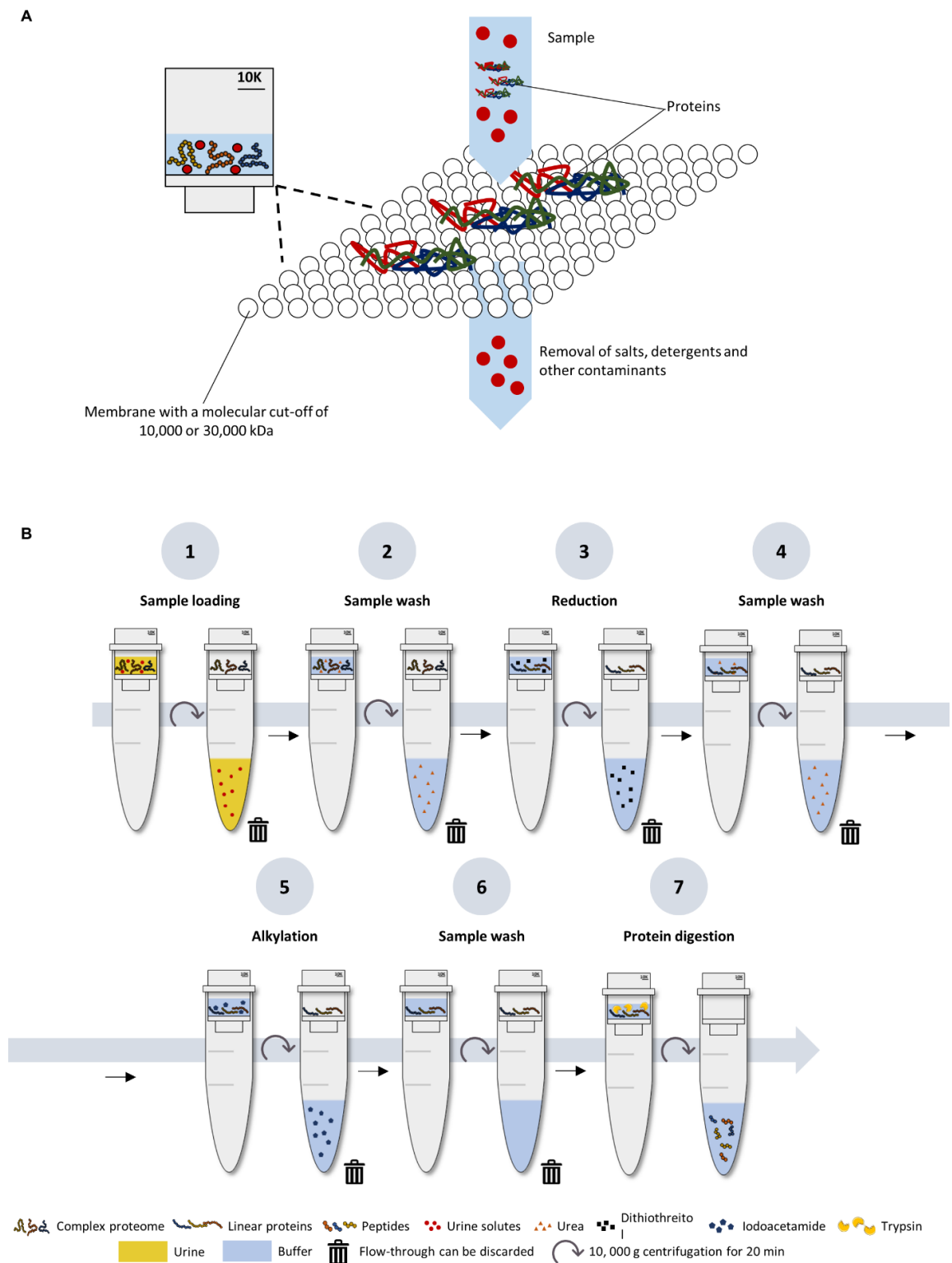


Figure 1-6. Filter Aided Sample Preparation (FASP). FASP allows for the conventional bottom-up proteomics approach workflow combining all the steps required to obtain peptides (1. Sample loading and concentration; 3. Reduction; 5. Alkylation; and 7. Protein digestion and in between 2,4,6. Sample cleaning) using the advantage of the built-in membrane with a molecular weight cut-off [175,176].

The same membranes are used in another proteomic technique, called filter-aided sample preparation (FASP) in the bottom approach (Fig.1-6B). Here, membranes with MWCO at 10kDa or 30kDa enable separation, desalting, reduction, alkylation, and digestion of proteins into peptides suitable for LC-MS/MS processing.

1.4 Personalised medicine-based proteomics

Precision medicine, as a growing area in healthcare, is designed to provide treatments based on unique people's molecular and biochemical phenotype. Proteomics-based precision medicine assumes an important place in oncology applications within this paradigm, especially in cancer treatment. The evolution in the field of mass spectrometry has provided insights into detecting specific proteins across different cancer types, achieving early diagnosis, prognosis, and treatment response prediction [177,178].

Mainly, mass spectrometry played a significant part in reconstructing proteomic maps of various subtypes of breast cancer, identifying the targets for treatment, and defining molecular subtypes for more precise treatment approaches. Nevertheless, proteomic integration into personalized medicine is not challenge-free; standardization, data interpretation and bioinformatics tools remain significant obstacles to the clinical application of proteomic data [179]. However, combined initiatives of technological innovation and cooperation in data exchange and processing gradually solve these issues. Therefore, proteomics is becoming increasingly critical in clinical medical practice, paving the way for the adoption of personalized medicine tailored to meet the unique needs of each patient [180].

1.4.1 The current extent of clinical protein-based diagnostics

Protein-based diagnostics, as diagnostic tools in clinical practice, allow us to understand the pathology of diseases and to properly administer personalized therapeutic regimens [181]. Clinical proteomics has dramatically increased our understanding of complex diseases such as cancer, diabetes, and cardiovascular diseases. This research area has also revealed new protein biomarkers, which have the potential for early disease diagnosis, prognosis, and treatment monitoring. Through the application of high-throughput technologies such as mass spectrometry and protein micro-arrays, researchers are now able to study hundreds to thousands of proteins at a time, diversifying the range and depth of clinical proteomic studies [182].

In addition to the development of clinical proteomics, attention has been paid to quality control with capillary measures to guarantee the accuracy and reproducibility of protein-based

diagnostics. Standardization and harmonization guidelines also play a critical role in suppressing inter-laboratory variances and ensuring comparability of the results under various clinical scenarios [183]. Also, the development of high-quality reference materials and assessment schemes has increased assurance of accuracy and precision of protein measurements which in turn has further increased the reliability of protein-based diagnostics.

The fusion of big data analytics and machine learning algorithms has given capacity to researchers for deriving meaningful information from large datasets that have come out of clinical proteomic initiatives. This computational methodology facilitates the discovery of stable protein signatures associated with certain disease states, leading to novel diagnostic panels that have improved predictive capabilities. Additionally, multimodal data integration, especially of proteomics and metabolomics, provides a comprehensive insight into disease pathways, acceleration of proteomic breakthroughs into clinically relevant outputs [184].

Although, development of protein-based clinical diagnostics is rapidly evolving, dedication to standardization and quality assurance remains of utmost importance. Stringent analytical validation procedures, compliance with regulatory frameworks, and continuous proficiency tests are crucial in the development and adoption of protein-based diagnostic assays in clinical diagnostics. These efforts not only improve confidence in the clinical relevance of protein biomarkers but also put clinical proteomics in a pivotal role in precision medicine whereby patient care and outcomes are improved, as well [185].

1.4.2 Clinical proteomics and towards personalised medicine

Clinical proteomics has a significant influence on predicting possible biomarkers for such types of cancers. Proteomic analysis of urine samples has become a promising non-invasive approach for the early diagnosis and monitoring of urothelial tract cancers, mainly bladder cancer. Through urine protein composition analysis, researchers can detect particular biomarkers signalling occurrence or development of cancer [186].

In the context of urine cancer, clinical proteomics primarily opens the door to personalized medicine. Proteomic profiling can provide a clinician with the molecular signature's characteristic of all subtypes of bladder cancer (Table 1-2), which, in turn, allows to modify a treatment plan for every patient [187]. This focused approach has high potential for the improvement of urinary tract cancer patients' outcomes.

In addition, the investigation of urinary proteomics can translate to new therapeutic approaches directed at the molecular pathways involved in the pathogenesis and progression of bladder cancer. Identification of important protein targets allows researchers to try to design

precise drugs that specifically disrupt the abnormal signalling pathways that promote cancer growth.

Table 1-2. Different biomarkers applied in clinical proteomics for bladder cancer.

Protein Name	Gene Name	Study reference
Angiotensinogen	AGT	[188]
Neuroblast differentiation-associated protein AHNAK	AHNAK	[189]
Angiopoietin-Like 4	ANGPTL4	[190]
Apolipoprotein A-I	APOA1	[191]
Apolipoprotein A-IV	APOA4	[192]
Apolipoprotein E	APOE	[193,194]
Carbonic Anhydrase 9	CA9	[195]
CD44 antigen	CD44	[196]
Cadherin-1	CDH1	[197,198]
Cystatin B	CSTB	[199]
Cullin-3	CUL3	[200]
Growth-regulated alpha protein	CXCL1	[201]
Deleted in Malignant Brain Tumors 1	DMBT1	[202]
Epiplakin	EPPK1	[189]
Fatty Acid-Binding Protein 5	FABP5	[203]
Gastrotropin	FABP6	[204]
Fibrinogen β chain precursor	FGB	[194]
Histone H2B type 1-K	H2BC12	[205]
Interleukin-2	IL-2	[206]
Interleukin-6	IL-6	[207]
Interleukin-8	IL-8	[208]
Kallikrein-6	KLK6	[209]
Leucine-rich α 2-glycoprotein	LRG1	[210]
Matrix Metalloproteinase-9	MMP-9	[211,212]
Myosin-14	MYH14	[189]
Olfactomedin-4	OLFM4	[189]
Alpha-1-acid glycoprotein 1	ORM1	[213]
Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase 2	PLOD2	[214]

Polymerase (RNA) I polypeptide E	POLR1E	[194]
Lithostathine-1-alpha	REG1A	[215]
Calgranulin A	S100A8	[216]
α 1-antitrypsin	SERPINA1	[194,205]
Topoisomerase 2A	TOP2A	[194]

1.4.3 Prescriptomics

Prescriptomics is a term proposed and applied by Capelo *et al*/2022 that revolutionizes the field of healthcare using information from omics to personalize prescriptions as per individual protein-pathway interactions [217].

The prescriptomics uses proteome that is expressed by a patient and quantifies, diagnosis, prognoses, and monitors diseases through liquid or solid biopsies such as urine, serum or plasma. It bridges the gap between diagnostics and therapeutics allowing personalized treatment through the analysis of patient-specific alterations in protein-protein networks. Alternatively, prescriptomics is the cornerstone of precision medicine, a branch of care aimed at providing individual treatments based on the patient's molecular profile, which eventually improves therapeutic efficacy and minimizes side effects.

Finally, prescriptomics, aims to revolutionize patient care by combining the use of electronic interfaces with drone sample collection and drug prescription, allowing for a prompt and individualized interventions to improve clinical outcomes

1.5 References

- [1] L. Dyrskj t, D.E. Hansel, J.A. Efstathiou, M.A. Knowles, M.D. Galsky, J. Teoh, D. Theodorescu, Bladder cancer, *Nat Rev Dis Primers* 9 (2023) 58. <https://doi.org/10.1038/s41572-023-00468-9>.
- [2] J. Mushtaq, R. Thuraija, R. Nair, Bladder cancer, *Surgery (Oxford)* 37 (2019) 529–537. <https://doi.org/10.1016/j.mpsur.2019.07.003>.
- [3] A.M. Kamat, N.M. Hahn, J.A. Efstathiou, S.P. Lerner, P.-U. Malmstr m, W. Choi, C.C. Guo, Y. Lotan, W. Kassouf, Bladder cancer, *The Lancet* 388 (2016) 2796–2810. [https://doi.org/10.1016/S0140-6736\(16\)30512-8](https://doi.org/10.1016/S0140-6736(16)30512-8).
- [4] A.A. Shokeir, Squamous cell carcinoma of the bladder: pathology, diagnosis and treatment, *BJU Int* 93 (2004) 216–220. <https://doi.org/10.1111/j.1464-410X.2004.04588.x>.

-
- [5] S. Antoni, J. Ferlay, I. Soerjomataram, A. Znaor, A. Jemal, F. Bray, Bladder cancer incidence and mortality: A global overview and recent trends, *Eur Urol* 71 (2017) 96–108. <https://doi.org/10.1016/j.eururo.2016.06.010>.
 - [6] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J Clin* 71 (2021) 209–249. <https://doi.org/10.3322/caac.21660>.
 - [7] International Agency for Research on Cancer, Global Cancer Observatory: Cancer Tomorrow. WHO, (2024). <https://gco.iarc.fr/today>.
 - [8] A. Bessa, R. Martin, C. Häggström, D. Enting, S. Amery, M.S. Khan, F. Cahill, H. Wylie, S. Broadhead, K. Chatterton, S. Malde, R. Nair, R. Thuraija, P. Kumar, A. Haire, S. Green, M. Northover, K. Briggs, M. Van Hemelrijck, Unmet needs in sexual health in bladder cancer patients: a systematic review of the evidence, *BMC Urol* 20 (2020) 64. <https://doi.org/10.1186/s12894-020-00634-1>.
 - [9] W. Kassouf, A. Aprikian, F. Saad, R.H. Breau, G. Kulkarni, D.M. Guttman, K. Bagshaw, J. Izawa, L. Eapen, A. Fairey, A. So, S. North, R. Rendon, S.S. Sridhar, F. Brimo, P. Chung, D. Drachenberg, Y. Fradet, N. Jacobsen, C. Morash, B. Shayegan, G. Gotto, A. Zlotta, N. Fleshner, D.R. Siemens, P.C. Black, Improving patient journey and quality of care: Summary from the second Bladder Cancer Canada-Canadian Urological Association- Canadian Urologic Oncology Group (BCC-CUA-CUOG) bladder cancer quality of care consensus meeting, *Canadian Urological Association Journal* 12 (2018) E281-97. <https://doi.org/10.5489/cuaj.5246>.
 - [10] E.Y.H. Wang, N. Pihlström, P.-U. Malmström, T. Gårdmark, Improved long-term outcome of patients with non-muscle invasive, low and intermediate risk bladder cancer between 1997 and 2014; a Swedish population-based study, *Scand J Urol* 56 (2022) 221–226. <https://doi.org/10.1080/21681805.2022.2062046>.
 - [11] J. Dobruch, M. Oszczudłowski, Bladder Cancer: current challenges and future directions, *Medicina (B Aires)* 57 (2021) 749. <https://doi.org/10.3390/medicina57080749>.
 - [12] A. Richters, K.K.H. Aben, L.A.L.M. Kiemeny, The global burden of urinary bladder cancer: an update, *World J Urol* 38 (2020) 1895–1904. <https://doi.org/10.1007/s00345-019-02984-4>.
 - [13] K. Chamie, M.S. Litwin, J.C. Bassett, T.J. Daskivich, J. Lai, J.M. Hanley, B.R. Konety, C.S. Saigal, Recurrence of high-risk bladder cancer: A population-based analysis, *Cancer* 119 (2013) 3219–3227. <https://doi.org/10.1002/cncr.28147>.
-

- [14] J. Ma, M. Roumiguie, T. Hayashi, Y. Kohada, A.R. Zlotta, S. Lévy, T. Matsumoto, T. Sano, P.C. Black, Long-term recurrence rates of low-risk non-muscle-invasive bladder cancer—How long is cystoscopic surveillance necessary?, *Eur Urol Focus* 10 (2024) 189–196. <https://doi.org/10.1016/j.euf.2023.06.012>.
- [15] P.-U. Malmström, C. Busch, B. Johan Norlén, Recurrence, progression and survival in bladder cancer, *Scand J Urol Nephrol* 21 (1987) 185–195. <https://doi.org/10.3109/00365598709180320>.
- [16] A.J. Edmondson, J.C. Birtwistle, J.W.F. Catto, M. Twiddy, The patients' experience of a bladder cancer diagnosis: a systematic review of the qualitative evidence, *Journal of Cancer Survivorship* 11 (2017) 453–461. <https://doi.org/10.1007/s11764-017-0603-6>.
- [17] H. Pham, H. Torres, P. Sharma, Mental health implications in bladder cancer patients: A review, *Urologic Oncology: Seminars and Original Investigations* 37 (2019) 97–107. <https://doi.org/10.1016/j.urolonc.2018.12.006>.
- [18] G. Ungerer, T. Anwar, M. Golzy, K.S. Murray, Living with bladder cancer: Self-reported changes in patients' functional and overall health status following diagnosis, *Eur Urol Open Sci* 20 (2020) 14–19. <https://doi.org/10.1016/j.euros.2020.06.007>.
- [19] A.D. Asimakopoulos, G. Colalillo, R. Telesca, A. Mauriello, R. Miano, S.M. Di Stasi, S. Germani, E. Finazzi Agrò, V. Petrozza, G. Caruso, A. Carbone, A.L. Pastore, A. Fuschi, T1 Bladder Cancer: Comparison of the Prognostic Impact of Two Substaging Systems on Disease Recurrence and Progression and Suggestion of a Novel Nomogram, *Front Surg* 8 (2021). <https://doi.org/10.3389/fsurg.2021.704902>.
- [20] A. Lopez-Beltran, L. Cheng, Stage T1 bladder cancer: diagnostic criteria and pitfalls, *Pathology* 53 (2021) 67–85. <https://doi.org/10.1016/j.pathol.2020.09.014>.
- [21] A. Aly, C. Johnson, Y. Doleh, V. Chirikov, M. Botteman, R. Shenolikar, A. Hussain, The real-world lifetime economic burden of urothelial carcinoma by stage at diagnosis., *J Clin Pathw* 6 (2020) 51–60.
- [22] M. Mossanen, J.L. Gore, The burden of bladder cancer care, *Curr Opin Urol* 24 (2014) 487–491. <https://doi.org/10.1097/MOU.0000000000000078>.
- [23] J. Leal, R. Luengo-Fernandez, R. Sullivan, J.A. Witjes, Economic burden of bladder cancer across the european union, *Eur Urol* 69 (2016) 438–447. <https://doi.org/10.1016/j.eururo.2015.10.024>.
- [24] M.R. Raspollini, E.M. Comperat, A. Lopez-Beltran, R. Montironi, A. Cimadamore, T. Tsuzuki, G.J. Netto, News in the classification of WHO 2022 bladder tumors, *Pathologica* (2023) 1–9. <https://doi.org/10.32074/1591-951X-838>.

-
- [25] Z. Chen, W. Ding, K. Xu, J. Tan, C. Sun, Y. Gou, S. Tong, G. Xia, Z. Fang, Q. Ding, The 1973 who classification is more suitable than the 2004 who classification for predicting prognosis in non-muscle-invasive bladder cancer, *PLoS One* 7 (2012) e47199. <https://doi.org/10.1371/journal.pone.0047199>.
 - [26] European Association of Urology, Non-muscle Invasive bladder cancer guideline, (2023). <https://uroweb.org/guidelines/non-muscle-invasive-bladder-cancer>.
 - [27] R.J. Sylvester, O. Rodríguez, V. Hernández, D. Turturica, L. Bauerová, H.M. Bruins, J. Bründl, T.H. van der Kwast, A. Brisuda, J. Rubio-Briones, M. Seles, A.E. Hentschel, V.R.M. Kusuma, N. Huebner, J. Cotte, L.S. Mertens, D. Volanis, O. Cussenot, J.D. Subiela Henríquez, E. de la Peña, F. Pisano, M. Pešl, A.G. van der Heijden, S. Herdegen, A.R. Zlotta, J. Hacek, A. Calatrava, S. Mannweiler, J. Bosschieter, D. Ashabere, A. Haitel, J.-F. Côté, S. El Sheikh, L. Lunelli, F. Algaba, I. Alemany, F. Soria, W. Runneboom, J. Breyer, J.A. Nieuwenhuijzen, C. Llorente, L. Molinaro, C.A. Hulsbergen-van de Kaa, M. Evert, L.A.L.M. Kiemeny, J. N'Dow, K. Plass, O. Čapoun, V. Soukup, J.L. Dominguez-Escrig, D. Cohen, J. Palou, P. Gontero, M. Burger, R. Zigeuner, A.H. Mostafid, S.F. Shariat, M. Rouprêt, E.M. Compérat, M. Babjuk, B.W.G. van Rhijn, European Association of Urology (EAU) prognostic factor risk groups for Non-muscle-invasive Bladder Cancer (NMIBC) incorporating the WHO 2004/2016 and WHO 1973 classification systems for grade: An update from the EAU NMIBC guidelines panel, *Eur Urol* 79 (2021) 480–488. <https://doi.org/10.1016/j.eururo.2020.12.033>.
 - [28] James D. Brierley, Mary K. Gospodarowicz, Christian Wittekind, *TNM Classification of Malignant Tumours*, 8th Edition, 2016.
 - [29] J. Tan, Y. Wang, L. Sun, S. Xu, C. Li, X. Jin, The origin and evolution of bladder cancer stem cells, *Front Cell Dev Biol* 10 (2022). <https://doi.org/10.3389/fcell.2022.950241>.
 - [30] H.-T. Li, C.E. Duymich, D.J. Weisenberger, G. Liang, Genetic and epigenetic alterations in bladder cancer, *Int Neurourol J* 20 (2016) S84-94. <https://doi.org/10.5213/inj.1632752.376>.
 - [31] C.D. Hurst, O. Alder, F.M. Platt, A. Droop, L.F. Stead, J.E. Burns, G.J. Burghel, S. Jain, L.J. Klimczak, H. Lindsay, J.-A. Roulson, C.F. Taylor, H. Thygesen, A.J. Cameron, A.J. Ridley, H.R. Mott, D.A. Gordenin, M.A. Knowles, Genomic subtypes of non-invasive bladder cancer with distinct metabolic profile and female gender bias in KDM6A mutation frequency, *Cancer Cell* 32 (2017) 701–715.e7. <https://doi.org/10.1016/j.ccell.2017.08.005>.
 - [32] C.D. Hurst, G. Cheng, F.M. Platt, M.A.A. Castro, N.S. Marzouka, P. Eriksson, E.V.I. Black, O. Alder, A.R.J. Lawson, S. V. Lindskrog, J.E. Burns, S. Jain, J.-A. Roulson, J.C. Brown, J. Koster,
-

- A.G. Robertson, I. Martincorena, L. Dyrskj t, M. H glund, M.A. Knowles, Stage-stratified molecular profiling of non-muscle-invasive bladder cancer enhances biological, clinical, and therapeutic insight, *Cell Rep Med* 2 (2021) 100472. <https://doi.org/10.1016/j.xcrm.2021.100472>.
- [33] J.J. Meeks, H. Al-Ahmadie, B.M. Faltas, J.A. Taylor, T.W. Flaig, D.J. DeGraff, E. Christensen, B.L. Woolbright, D.J. McConkey, L. Dyrskj t, Genomic heterogeneity in bladder cancer: challenges and possible solutions to improve outcomes, *Nat Rev Urol* 17 (2020) 259–270. <https://doi.org/10.1038/s41585-020-0304-1>.
- [34] Q. Li, A.W. Damish, Z. Frazier, D. Liu, E. Reznichenko, A. Kamburov, A. Bell, H. Zhao, E.J. Jordan, S.P. Gao, J. Ma, P.H. Abbosh, J. Bellmunt, E.R. Plimack, J.-B. Lazaro, D.B. Solit, D. Bajorin, J.E. Rosenberg, A.D. D’Andrea, N. Riaz, E.M. Van Allen, G. Iyer, K.W. Mouw, ERCC2 helicase domain mutations confer nucleotide excision repair deficiency and drive cisplatin sensitivity in muscle-invasive bladder cancer, *Clinical Cancer Research* 25 (2019) 977–988. <https://doi.org/10.1158/1078-0432.CCR-18-1001>.
- [35] A.G. Robertson, J. Kim, H. Al-Ahmadie, J. Bellmunt, G. Guo, A.D. Cherniack, T. Hinoue, P.W. Laird, K.A. Hoadley, R. Akbani, M.A.A. Castro, E.A. Gibb, R.S. Kanchi, D.A. Gordenin, S.A. Shukla, F. Sanchez-Vega, D.E. Hansel, B.A. Czerniak, V.E. Reuter, X. Su, B. de Sa Carvalho, V.S. Chagas, K.L. Mungall, S. Sadeghi, C.S. Pedamallu, Y. Lu, L.J. Klimczak, J. Zhang, C. Choo, A.I. Ojesina, S. Bullman, K.M. Leraas, T.M. Lichtenberg, C.J. Wu, N. Schultz, G. Getz, M. Meyerson, G.B. Mills, D.J. McConkey, J.N. Weinstein, D.J. Kwiatkowski, S.P. Lerner, Comprehensive molecular characterization of muscle-invasive bladder cancer, *Cell* 174 (2018) 1033. <https://doi.org/10.1016/j.cell.2018.07.036>.
- [36] J. Huan, P. Grivas, J. Birch, D.E. Hansel, Emerging roles for mammalian target of rapamycin (mTOR) complexes in bladder cancer progression and therapy, *Cancers (Basel)* 14 (2022) 1555. <https://doi.org/10.3390/cancers14061555>.
- [37] Y. Miyata, Y. Sagara, S. Kanda, T. Hayashi, H. Kanetake, Phosphorylated hepatocyte growth factor receptor/c-Met is associated with tumor growth and prognosis in patients with bladder cancer: correlation with matrix metalloproteinase–2 and –7 and E-cadherin, *Hum Pathol* 40 (2009) 496–504. <https://doi.org/10.1016/j.humpath.2008.09.011>.
- [38] A. Goriki, R. Seiler, A.W. Wyatt, A. Contreras-Sanz, A. Bhat, A. Matsubara, T. Hayashi, P.C. Black, Unravelling disparate roles of NOTCH in bladder cancer, *Nat Rev Urol* 15 (2018) 345–357. <https://doi.org/10.1038/s41585-018-0005-1>.
- [39] K. Saginala, A. Barsouk, J.S. Aluru, P. Rawla, S.A. Padala, A. Barsouk, Epidemiology of bladder cancer, *Medical Sciences* 8 (2020) 15. <https://doi.org/10.3390/medsci8010015>.

-
- [40] N.D. Freedman, Association between smoking and risk of bladder cancer among men and women, *JAMA* 306 (2011) 737. <https://doi.org/10.1001/jama.2011.1142>.
 - [41] M. Burger, J.W.F. Catto, G. Dalbagni, H.B. Grossman, H. Herr, P. Karakiewicz, W. Kassouf, L.A. Kiemeny, C. La Vecchia, S. Shariat, Y. Lotan, Epidemiology and risk factors of urothelial bladder cancer, *Eur Urol* 63 (2013) 234–241. <https://doi.org/10.1016/j.eururo.2012.07.033>.
 - [42] K. Golka, A. Wiese, G. Assennato, H.M. Bolt, Occupational exposure and urological cancer, *World J Urol* 21 (2004) 382–391. <https://doi.org/10.1007/s00345-003-0377-5>.
 - [43] C. Wang, H. Jiang, Meat intake and risk of bladder cancer: a meta-analysis, *Medical Oncology* 29 (2012) 848–855. <https://doi.org/10.1007/s12032-011-9985-x>.
 - [44] J.-W. Sun, L.-G. Zhao, Y. Yang, X. Ma, Y.-Y. Wang, Y.-B. Xiang, Obesity and risk of bladder cancer: A dose-response meta-analysis of 15 cohort studies, *PLoS One* 10 (2015) e0119313. <https://doi.org/10.1371/journal.pone.0119313>.
 - [45] T.N.P.C.D.U.O. Marianette T Inobaya Remigio M Olveda, A.G.P. Ross, Prevention and control of schistosomiasis: a current perspective, *Res Rep Trop Med* 5 (2014) 65–75. <https://doi.org/10.2147/RRTM.S44274>.
 - [46] C.D. Fankhauser, S. Waisbrod, C. Fierz, A.S. Becker, B. Kranzbühler, D. Eberli, T. Sulser, H. Mostafid, T. Hermanns, Diagnostic accuracy of ultrasonography, computed tomography, cystoscopy and cytology to detect urinary tract malignancies in patients with asymptomatic hematuria, *World J Urol* 39 (2021) 97–103. <https://doi.org/10.1007/s00345-020-03171-6>.
 - [47] E. Compérat, M.B. Amin, R. Cathomas, A. Choudhury, M. De Santis, A. Kamat, A. Stenzl, H.C. Thoeny, J.A. Witjes, Current best practice for bladder cancer: a narrative review of diagnostics and treatments, *The Lancet* 400 (2022) 1712–1721. [https://doi.org/10.1016/S0140-6736\(22\)01188-6](https://doi.org/10.1016/S0140-6736(22)01188-6).
 - [48] M. Babjuk, M. Burger, E.M. Compérat, P. Gontero, A.H. Mostafid, J. Palou, B.W.G. van Rhijn, M. Rouprêt, S.F. Shariat, R. Sylvester, R. Zigeuner, O. Capoun, D. Cohen, J.L.D. Escrig, V. Hernández, B. Peyronnet, T. Seisen, V. Soukup, European association of urology guidelines on non-muscle-invasive bladder cancer (TaT1 and carcinoma in situ) - 2019 update, *Eur Urol* 76 (2019) 639–657. <https://doi.org/10.1016/j.eururo.2019.08.016>.
 - [49] G.I. Russo, T.N. Sholklapper, A. Cocci, G. Broggi, R. Caltabiano, A.B. Smith, Y. Lotan, G. Morgia, A.M. Kamat, J.A. Witjes, S. Daneshmand, M.M. Desai, I.S. Gill, G.E. Cacciamani, Performance of Narrow Band Imaging (NBI) and Photodynamic Diagnosis (PDD) fluorescence imaging compared to White Light Cystoscopy (WLC) in detecting non-muscle
-

- invasive bladder cancer: A systematic review and lesion-level diagnostic meta-analysis, *Cancers (Basel)* 13 (2021) 4378. <https://doi.org/10.3390/cancers13174378>.
- [50] M.P. Hiorns, Imaging of the urinary tract: the role of CT and MRI, *Pediatric Nephrology* 26 (2011) 59–68. <https://doi.org/10.1007/s00467-010-1645-4>.
- [51] A. Chen, A. Liu, J. Liu, S. Tian, H. Wang, Y. Liu, Application of dual-energy spectral CT imaging in differential diagnosis of bladder cancer and benign prostate hyperplasia, *Medicine* 95 (2016) e5705. <https://doi.org/10.1097/MD.00000000000005705>.
- [52] H.C. Thoeny, M.-F. Bellin, E.-M. Comperat, G.N. Thalmann, Vesical imaging-reporting and data system (vi-rads): Added value for management of bladder cancer patients?, *Eur Urol* 74 (2018) 307–308. <https://doi.org/10.1016/j.eururo.2018.06.017>.
- [53] V. Panebianco, Y. Narumi, E. Altun, B.H. Bochner, J.A. Efstathiou, S. Hafeez, R. Huddart, S. Kennish, S. Lerner, R. Montironi, V.F. Muglia, G. Salomon, S. Thomas, H.A. Vargas, J.A. Witjes, M. Takeuchi, J. Barentsz, J.W.F. Catto, Multiparametric magnetic resonance imaging for bladder cancer: development of vi-rads (vesical imaging-reporting and data system), *Eur Urol* 74 (2018) 294–306. <https://doi.org/10.1016/j.eururo.2018.04.029>.
- [54] J.A. Witjes, M. Babjuk, J. Bellmunt, H.M. Bruins, T.M. De Reijke, M. De Santis, S. Gillesen, N. James, S. Maclellan, J. Palou, T. Powles, M.J. Ribal, S.F. Shariat, T. Van Der Kwast, E. Xylinas, N. Agarwal, T. Arends, A. Bamias, A. Birtle, P.C. Black, B.H. Bochner, M. Bolla, J.L. Boormans, A. Bossi, A. Briganti, I. Brummelhuis, M. Burger, D. Castellano, R. Cathomas, A. Chiti, A. Choudhury, E. Comp  rat, S. Crabb, S. Culine, B. De Bari, W. De Blok, P. J.L. De Visschere, K. Decaestecker, K. Dimitropoulos, J.L. Dominguez-Escrig, S. Fanti, V. Fonteyne, M. Frydenberg, J.J. Futterer, G. Gakis, B. Geavlete, P. Gontero, B. Grubm  ller, S. Hafeez, D.E. Hansel, A. Hartmann, D. Hayne, A.M. Henry, V. Hernandez, H. Herr, K. Herrmann, P. Hoskin, J. Huguet, B.A. Jerezczek-Fossa, R. Jones, A.M. Kamat, V. Khoo, A.E. Kiltie, S. Krege, S. Ladoire, P.C. Lara, A. Leliveld, E. Linares-Espin  s, V. L  gager, A. Lorch, Y. Lorient, R. Meijer, M.C. Mir, M. Moschini, H. Mostafid, A.-C. M  ller, C.R. M  ller, J. N'Dow, A. Necchi, Y. Neuzillet, J.R. Ouddens, J. Oldenburg, S. Osanto, W. J.G. Oyen, L. Pacheco-Figueiredo, H. Pappot, M.I. Patel, B.R. Pieters, K. Plass, M. Remzi, M. Retz, J. Richenberg, M. Rink, F. Rognmann, J.E. Rosenberg, M. Roupr  t, O. Rouvi  re, C. Salembier, A. Salminen, P. Sargos, S. Sengupta, A. Sherif, R.J. Smeenk, A. Smits, A. Stenzl, G.N. Thalmann, B. Tombal, B. Turkbey, S.V. Lauridsen, R. Valdagni, A.G. Van Der Heijden, H. Van Poppel, M.D. Vartolomei, E. Veskim  e, A. Vilaseca, F.A.V. Rivera, T. Wiegel, P. Wiklund, A. Williams, R. Zigeuner, A. Horwich, EAU-ESMO Consensus statements on the management of advanced and variant bladder cancer—An international collaborative

- multistakeholder effort under the auspices of the eau-esmo guidelines committees, *Eur Urol* 77 (2020) 223–250. <https://doi.org/10.1016/j.eururo.2019.09.035>.
- [55] C.-Z. Zhu, H.-N. Ting, K.-H. Ng, T.-A. Ong, A review on the accuracy of bladder cancer detection methods, *J Cancer* 10 (2019) 4038–4044. <https://doi.org/10.7150/jca.28989>.
- [56] L.H.C. Kim, M.I. Patel, Transurethral resection of bladder tumour (TURBT), *Transl Androl Urol* 9 (2020) 3056–3072. <https://doi.org/10.21037/tau.2019.09.38>.
- [57] M. Maas, T. Todenhöfer, P.C. Black, Urine biomarkers in bladder cancer — current status and future perspectives, *Nat Rev Urol* 20 (2023) 597–614. <https://doi.org/10.1038/s41585-023-00773-8>.
- [58] E. Tomiyama, K. Fujita, M. Hashimoto, H. Uemura, N. Nonomura, Urinary markers for bladder cancer diagnosis: A review of current status and future challenges, *International Journal of Urology* 31 (2024) 208–219. <https://doi.org/10.1111/iju.15338>.
- [59] L. Tran, J.-F. Xiao, N. Agarwal, J.E. Duex, D. Theodorescu, Advances in bladder cancer biology and therapy, *Nat Rev Cancer* 21 (2021) 104–121. <https://doi.org/10.1038/s41568-020-00313-1>.
- [60] R. Chou, J.L. Gore, D. Buckley, R. Fu, K. Gustafson, J.C. Griffin, S. Grusing, S. Selph, Urinary biomarkers for diagnosis of bladder cancer, *Ann Intern Med* 163 (2015) 922–931. <https://doi.org/10.7326/M15-0997>.
- [61] A. Schulz, J. Loloï, L. Pina Martina, A. Sankin, The development of non-invasive diagnostic tools in bladder cancer, *Onco Targets Ther* Volume 15 (2022) 497–507. <https://doi.org/10.2147/OTT.S283891>.
- [62] Y. Fradet, C. Lockhard, Performance characteristics of a new monoclonal antibody test for bladder cancer: ImmunoCyt trade mark., *Can J Urol* 4 (1997) 400–405.
- [63] M. SKACEL, M. FAHMY, J.A. BRAINARD, J.D. PETTAY, C. V. BISCOTTI, L.S. LIOU, G.W. PRO-COP, J.S. JONES, J. ULCHAKER, C.D. ZIPPE, R.R. TUBBS, Multitarget fluorescence in situ hybridization assay detects transitional cell carcinoma in the majority of patients with bladder cancer and atypical or negative urine cytology, *Journal of Urology* 169 (2003) 2101–2105. <https://doi.org/10.1097/01.ju.0000066842.45464.cc>.
- [64] B.R. Konety, T.S. Nguyen, R. Dhir, R.S. Day, M.J. Becich, W.M. Stadler, R.H. Getzenberg, Detection of bladder cancer using a novel nuclear matrix protein, BLCA-4., *Clin Cancer Res* 6 (2000) 2618–25.
- [65] M. Roupret, P. Gontero, S.R.C. McCracken, T. Dudderidge, J. Stockley, A. Kennedy, O. Rodriguez, C. Sieverink, F. Vanié, M. Allasia, J.A. Witjes, M. Colombel, R. Sylvester, F. Longo, E. Montanari, J. Palou, Diagnostic accuracy of mcm5 for the detection of

- recurrence in nonmuscle invasive bladder cancer followup: a blinded, prospective cohort, multicenter european study, *Journal of Urology* 204 (2020) 685–690. <https://doi.org/10.1097/JU.0000000000001084>.
- [66] H. Furuya, I. Pagano, K. Chee, T. Kobayashi, R.S. Wong, R. Lee, C.J. Rosser, Comparison of commercial ELISA kits, a prototype multiplex electrochemoluminescent assay, and a multiplex bead-based immunoassay for detecting a urine-based bladder-cancer-associated diagnostic signature, *Diagnostics* 9 (2019) 166. <https://doi.org/10.3390/diagnostics9040166>.
- [67] B. Cowan, E. Klein, K. Jansz, K. Westenfelder, T. Bradford, C. Peterson, D. Scherr, L.I. Karsh, B. Egerdie, A. Witjes, A. Trainer, R. Harris, B. Goldfarb, S. Flax, R. Kroeger, B. Boyd, J. Liao, S. Patel, J. Bridge, V. Reuter, N. Quigley, S. Brown, S. Zhao, M. Satya, M. Bates, I.M. Simon, S. Campbell, Y. Lotan, Longitudinal follow-up and performance validation of an mRNA-based urine test (Xpert® Bladder Cancer Monitor) for surveillance in patients with non-muscle-invasive bladder cancer, *BJU Int* 128 (2021) 713–721. <https://doi.org/10.1111/bju.15418>.
- [68] P. O’Sullivan, K. Sharples, M. Dalphin, P. Davidson, P. Gilling, L. Cambridge, J. Harvey, T. Toro, N. Giles, C. Luxmanan, C.F. Alves, H.-S. Yoon, V. Hinder, J. Masters, A. Kennedy-Smith, T. Beaven, P.J. Guilford, A multigene urine test for the detection and stratification of bladder cancer in patients presenting with hematuria, *Journal of Urology* 188 (2012) 741–747. <https://doi.org/10.1016/j.juro.2012.05.003>.
- [69] A. Wasserstrom, D. Frumkin, Z. Dotan, E. Bukin, T. Gadish, S. Hanuka, R. Knirsh, A.E. Darawsha, I. Leibovitch, MP13-15 molecular urine cytology – Bladder EpiCheck is a novel molecular diagnostic tool for monitoring of bladder cancer patients, *Journal of Urology* 195 (2016). <https://doi.org/10.1016/j.juro.2016.02.2496>.
- [70] J.-P. Roperch, C. Hennion, A novel ultra-sensitive method for the detection of FGFR3 mutations in urine of bladder cancer patients – Design of the Urodiag® PCR kit for surveillance of patients with non-muscle-invasive bladder cancer (NMIBC), *BMC Med Genet* 21 (2020) 112. <https://doi.org/10.1186/s12881-020-01050-w>.
- [71] K.E.M. van Kessel, W. Beukers, I. Lurkin, A. Ziel-van der Made, K.A. van der Keur, J.L. Boormans, L. Dyrskjöt, M. Márquez, T.F. Ørntoft, F.X. Real, U. Segersten, N. Malats, P.-U. Malmström, W. Van Criekinge, E.C. Zwarthoff, Validation of a DNA methylation-mutation urine assay to select patients with hematuria for cystoscopy, *Journal of Urology* 197 (2017) 590–595. <https://doi.org/10.1016/j.juro.2016.09.118>.

-
- [72] A. Feber, P. Dhami, L. Dong, P. de Winter, W.S. Tan, M. Martínez-Fernández, D.S. Paul, A. Hynes-Allen, S. Rezaee, P. Gurung, S. Rodney, A. Mehmood, F. Villacampa, F. de la Rosa, C. Jameson, K.K. Cheng, M.P. Zeegers, R.T. Bryan, N.D. James, J.M. Paramio, A. Freeman, S. Beck, J.D. Kelly, UroMark—A urinary biomarker assay for the detection of bladder cancer, *Clin Epigenetics* 9 (2017) 8. <https://doi.org/10.1186/s13148-016-0303-5>.
 - [73] X. Chen, J. Zhang, W. Ruan, M. Huang, C. Wang, H. Wang, Z. Jiang, S. Wang, Z. Liu, C. Liu, W. Tan, J. Yang, J. Chen, Z. Chen, X. Li, X. Zhang, P. Xu, L. Chen, R. Xie, Q. Zhou, S. Xu, D.L. Irwin, J.-B. Fan, J. Huang, T. Lin, Urine DNA methylation assay enables early detection and recurrence monitoring for bladder cancer, *Journal of Clinical Investigation* 130 (2020) 6278–6289. <https://doi.org/10.1172/JCI139597>.
 - [74] H.D. Pharo, M. Jeanmougin, E. Ager-Wick, H.M. Vedeld, A.K. Sørnbø, C. Dahl, L.K. Larsen, H. Honne, S. Brandt-Winge, M.-B. Five, S. Monteiro-Reis, R. Henrique, C. Jeronimo, K. Steven, R. Wahlqvist, P. Guldberg, G.E. Lind, BladMetrix: a novel urine DNA methylation test with high accuracy for detection of bladder cancer in hematuria patients, *Clin Epigenetics* 14 (2022) 115. <https://doi.org/10.1186/s13148-022-01335-2>.
 - [75] S. Zeng, Y. Ying, N. Xing, B. Wang, Z. Qian, Z. Zhou, Z. Zhang, W. Xu, H. Wang, L. Dai, L. Gao, T. Zhou, J. Ji, C. Xu, Noninvasive detection of urothelial carcinoma by cost-effective low-coverage whole-genome sequencing from urine-exfoliated cell DNA, *Clinical Cancer Research* 26 (2020) 5646–5654. <https://doi.org/10.1158/1078-0432.CCR-20-0401>.
 - [76] P.H. Avogbe, A. Manel, E. Vian, G. Durand, N. Forey, C. Voegelé, M. Zvereva, M.I. Hosen, S. Meziani, B. De Tilly, G. Polo, O. Lole, P. Francois, T.M. Delhomme, C. Carreira, S. Monteiro-Reis, R. Henrique, B. Abedi-Ardekani, G. Byrnes, M. Foll, E. Weiderpass, J. McKay, C. Jeronimo, G. Scelo, F. Le Calvez-Kelm, Urinary TERT promoter mutations as non-invasive biomarkers for the comprehensive detection of urothelial cancer, *EBioMedicine* 44 (2019) 431–438. <https://doi.org/10.1016/j.ebiom.2019.05.004>.
 - [77] C.A. Sieverink, R.P.M. Batista, H.J.M. Prazeres, J. Vinagre, C. Sampaio, R.R. Leão, V. Máximo, J.A. Witjes, P. Soares, Clinical validation of a urine test (uromonitor-v2®) for the surveillance of non-muscle-invasive bladder cancer patients, *Diagnostics* 10 (2020) 745. <https://doi.org/10.3390/diagnostics10100745>.
 - [78] M.-L. Eich, M.D.C. Rodriguez Pena, S.U. Springer, D. Taheri, A.C. Tregnago, D.C. Salles, S.M. Bezerra, I.W. Cunha, K. Fujita, D. Ertoy, T.J. Bivalacqua, C. Tomasetti, N. Papadopoulos, K.W. Kinzler, B. Vogelstein, G.J. Netto, Incidence and distribution of UroSEEK gene panel in a multi-institutional cohort of bladder urothelial carcinoma, *Modern Pathology* 32 (2019) 1544–1550. <https://doi.org/10.1038/s41379-019-0276-y>.
-

- [79] J.C. Dudley, J. Schroers-Martin, D. V. Lazzareschi, W.Y. Shi, S.B. Chen, M.S. Esfahani, D. Trivedi, J.J. Chabon, A.A. Chaudhuri, H. Stehr, C.L. Liu, H. Lim, H.A. Costa, B.Y. Nabet, M.L.Y. Sin, J.C. Liao, A.A. Alizadeh, M. Diehn, Detection and surveillance of bladder cancer using urine tumor DNA, *Cancer Discov* 9 (2019) 500–509. <https://doi.org/10.1158/2159-8290.CD-18-0825>.
- [80] S. Babu, D.C. Mockler, L. Roa-Peña, A. Szygalowicz, N.W. Kim, S. Jahanfard, S.S. Gholami, R. Moffitt, J.P. Fitzgerald, L.F. Escobar-Hoyos, K.R. Shroyer, Keratin 17 is a sensitive and specific biomarker of urothelial neoplasia, *Modern Pathology* 32 (2019) 717–724. <https://doi.org/10.1038/s41379-018-0177-5>.
- [81] N. Davis, Y. Mor, P. Idelevich, D. Terkieltaub, V. Ziv, A. Elkeles, S. Lew, E. Okon, M. Laufer, J. Ramon, D. Kedar, J. Baniel, O. Yossepowitch, A novel urine cytology stain for the detection and monitoring of bladder cancer, *Journal of Urology* 192 (2014) 1628–1632. <https://doi.org/10.1016/j.juro.2014.06.079>.
- [82] B. Ljungberg, L. Albiges, Y. Abu-Ghanem, K. Bensalah, S. Dabestani, S. Fernández-Pello, R.H. Giles, F. Hofmann, M. Hora, M.A. Kuczyk, T. Kuusk, T.B. Lam, L. Marconi, A.S. Merseburger, T. Powles, M. Staehler, R. Tahbaz, A. Volpe, A. Bex, European association of urology guidelines on renal cell carcinoma: The 2019 update, *Eur Urol* 75 (2019) 799–810. <https://doi.org/10.1016/j.eururo.2019.02.011>.
- [83] D. Lindgren, J. Sjölund, H. Axelson, Tracing renal cell carcinomas back to the nephron, *Trends Cancer* 4 (2018) 472–484. <https://doi.org/10.1016/j.trecan.2018.05.003>.
- [84] T.J. van Oostenbrugge, J.J. Fütterer, P.F.A. Mulders, Diagnostic imaging for solid renal tumors: A pictorial review, *Kidney Cancer* 2 (2018) 79–93. <https://doi.org/10.3233/KCA-180028>.
- [85] V.F. Muglia, A. Prando, Renal cell carcinoma: histological classification and correlation with imaging findings, *Radiol Bras* 48 (2015) 166–174. <https://doi.org/10.1590/0100-3984.2013.1927>.
- [86] H.K. Park, The metastasis pattern of renal cell carcinoma is influenced by histologic subtype, grade, and sarcomatoid differentiation, *Medicina (B Aires)* 59 (2023) 1845. <https://doi.org/10.3390/medicina59101845>.
- [87] R. Sharma, G. Kannourakis, P. Prithviraj, N. Ahmed, Precision medicine: An optimal approach to patient care in renal cell carcinoma, *Front Med (Lausanne)* 9 (2022). <https://doi.org/10.3389/fmed.2022.766869>.
- [88] S. Jorge, J.L. Capelo, W. LaFramboise, S. Satturwar, D. Korentzelos, S. Bastacky, G. Quiroga-Garza, R. Dhir, J.R. Wiśniewski, C. Lodeiro, H.M. Santos, Absolute quantitative

- proteomics using the total protein approach to identify novel clinical immunohistochemical markers in renal neoplasms, *BMC Med* 19 (2021) 196. <https://doi.org/10.1186/s12916-021-02071-9>.
- [89] A. Pandey, M. Mann, Proteomics to study genes and genomes, *Nature* 405 (2000) 837–846. <https://doi.org/10.1038/35015709>.
- [90] O.T. Schubert, H.L. Röst, B.C. Collins, G. Rosenberger, R. Aebersold, Quantitative proteomics: challenges and opportunities in basic and applied research, *Nat Protoc* 12 (2017) 1289–1294. <https://doi.org/10.1038/nprot.2017.040>.
- [91] R. Aebersold, J.N. Agar, I.J. Amster, M.S. Baker, C.R. Bertozzi, E.S. Boja, C.E. Costello, B.F. Cravatt, C. Fenselau, B.A. Garcia, Y. Ge, J. Gunawardena, R.C. Hendrickson, P.J. Hergenrother, C.G. Huber, A.R. Ivanov, O.N. Jensen, M.C. Jewett, N.L. Kelleher, L.L. Kiessling, N.J. Krogan, M.R. Larsen, J.A. Loo, R.R. Ogorzalek Loo, E. Lundberg, M.J. MacCoss, P. Mallick, V.K. Mootha, M. Mrksich, T.W. Muir, S.M. Patrie, J.J. Pesavento, S.J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schlüter, S. Sechi, S.A. Slavoff, L.M. Smith, M.P. Snyder, P.M. Thomas, M. Uhlén, J.E. Van Eyk, M. Vidal, D.R. Walt, F.M. White, E.R. Williams, T. Wohlschläger, V.H. Wysocki, N.A. Yates, N.L. Young, B. Zhang, How many human proteoforms are there?, *Nat Chem Biol* 14 (2018) 206–214. <https://doi.org/10.1038/nchembio.2576>.
- [92] S. Al-Amrani, Z. Al-Jabri, A. Al-Zaabi, J. Alshekaili, M. Al-Khabori, Proteomics: Concepts and applications in human medicine, *World J Biol Chem* 12 (2021) 57–69. <https://doi.org/10.4331/wjbc.v12.i5.57>.
- [93] Q. Zhang, C. Ma, M. Gearing, P.G. Wang, L.-S. Chin, L. Li, Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer’s disease, *Acta Neuropathol Commun* 6 (2018) 19. <https://doi.org/10.1186/s40478-018-0524-2>.
- [94] S.J. Rhee, D. Shin, D. Shin, Y. Song, E.-J. Joo, H.Y. Jung, S. Roh, S.-H. Lee, H. Kim, M. Bang, K.Y. Lee, J. Lee, J. Kim, Y. Kim, Y. Kim, Y.M. Ahn, Network analysis of plasma proteomes in affective disorders, *Transl Psychiatry* 13 (2023) 195. <https://doi.org/10.1038/s41398-023-02485-4>.
- [95] S.R. Shuken, An introduction to mass spectrometry-based proteomics, *J Proteome Res* 22 (2023) 2151–2171. <https://doi.org/10.1021/acs.jproteome.2c00838>.
- [96] L.R. Pade, K.E. Stepler, E.P. Portero, K. DeLaney, P. Nemes, Biological mass spectrometry enables spatiotemporal ‘omics: From tissues to cells to organelles, *Mass Spectrom Rev* 43 (2024) 106–138. <https://doi.org/10.1002/mas.21824>.

- [97] A. Mateus, M.M. Savitski, I. Piazza, The rise of proteome-wide biophysics, *Mol Syst Biol* 17 (2021). <https://doi.org/10.15252/msb.202110442>.
- [98] D.J. Geiszler, D.A. Polasky, F. Yu, A.I. Nesvizhskii, Detecting diagnostic features in MS/MS spectra of post-translationally modified peptides, *Nat Commun* 14 (2023) 4132. <https://doi.org/10.1038/s41467-023-39828-0>.
- [99] T.R. Keeney, C. Bock, L. Gold, S. Kraemer, B. Lollo, M. Nikrad, M. Stanton, A. Stewart, J.D. Vaught, J.J. Walker, Automation of the somalogic proteomics assay: A platform for biomarker discovery, *JALA: Journal of the Association for Laboratory Automation* 14 (2009) 360–366. <https://doi.org/10.1016/j.jala.2009.05.003>.
- [100] A. Petrera, C. von Toerne, J. Behler, C. Huth, B. Thorand, A. Hilgendorff, S.M. Hauck, Multipatform approach for plasma proteomics: complementarity of Olink proximity extension assay technology to mass spectrometry-based protein profiling, *J Proteome Res* 20 (2021) 751–762. <https://doi.org/10.1021/acs.jproteome.0c00641>.
- [101] X. Wang, K.C. Yip, A. He, J. Tang, S. Liu, R. Yan, Q. Zhang, R. Li, Plasma olink proteomics identifies CCL20 as a novel predictive and diagnostic inflammatory marker for preeclampsia, *J Proteome Res* 21 (2022) 2998–3006. <https://doi.org/10.1021/acs.jproteome.2c00544>.
- [102] X.-H. Bao, B.-F. Chen, J. Liu, Y.-H. Tan, S. Chen, F. Zhang, H.-S. Lu, J.-C. Li, Olink proteomics profiling platform reveals non-invasive inflammatory related protein biomarkers in autism spectrum disorder, *Front Mol Neurosci* 16 (2023). <https://doi.org/10.3389/fnmol.2023.1185021>.
- [103] A. Joshi, M. Mayr, In aptamers they trust, *Circulation* 138 (2018) 2482–2485. <https://doi.org/10.1161/CIRCULATIONAHA.118.036823>.
- [104] D.E. Haslam, J. Li, S.T. Dillon, X. Gu, Y. Cao, O.A. Zeleznik, N. Sasamoto, X. Zhang, A.H. Eliassen, L. Liang, M.J. Stampfer, S. Mora, Z. Chen, K.L. Terry, R.E. Gerszten, F.B. Hu, A.T. Chan, T.A. Libermann, S.N. Bhupathiraju, Stability and reproducibility of proteomic profiles in epidemiological studies: comparing the Olink and SOMAscan platforms, *Proteomics* 22 (2022). <https://doi.org/10.1002/pmic.202100170>.
- [105] G.H. Eldjarn, E. Ferkingstad, S.H. Lund, H. Helgason, O.Th. Magnusson, K. Gunnarsdottir, T.A. Olafsdottir, B. V. Halldorsson, P.I. Olason, F. Zink, S.A. Gudjonsson, G. Sveinbjornsson, M.I. Magnusson, A. Helgason, A. Oddsson, G.H. Halldorsson, M.K. Magnusson, S. Saevarsdottir, T. Eiriksdottir, G. Masson, H. Stefansson, I. Jonsdottir, H. Holm, T. Rafnar, P. Melsted, J. Saemundsdottir, G.L. Norddahl, G. Thorleifsson, M.O. Ulfarsson, D.F. Gudbjartsson, U. Thorsteinsdottir, P. Sulem, K. Stefansson, Large-scale plasma proteomics

- comparisons through genetics and disease associations, *Nature* 622 (2023) 348–358. <https://doi.org/10.1038/s41586-023-06563-x>.
- [106] J.H. Gross, *Mass Spectrometry*, Springer International Publishing, Cham, 2017. <https://doi.org/10.1007/978-3-319-54398-7>.
- [107] Edmond de Hoffmann, Vincent Stroobant, *Mass Spectrometry: Principles and Applications*, 3rd Edition, 2007.
- [108] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science* (1979) 246 (1989) 64–71. <https://doi.org/10.1126/science.2675315>.
- [109] C.M. Whitehouse, R.N. Dreyer, Masamichi Yamashita, J.B. Fenn, Electrospray interface for liquid chromatographs and mass spectrometers, *Anal Chem* 57 (1985) 675–679. <https://doi.org/10.1021/ac00280a023>.
- [110] Michael Karas, Franz Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal Chem* 60 (1988) 2299–2301. <https://doi.org/10.1021/ac00171a028>.
- [111] F.W. McLafferty, Mass spectrometry across the sciences, *Proceedings of the National Academy of Sciences* 105 (2008) 18088–18089. <https://doi.org/10.1073/pnas.0800784105>.
- [112] B.T. Chait, Mass spectrometry: Bottom-up or top-down?, *Science* (1979) 314 (2006) 65–66. <https://doi.org/10.1126/science.1133987>.
- [113] K.A. Brown, J.A. Melby, D.S. Roberts, Y. Ge, Top-down proteomics: challenges, innovations, and applications in basic and clinical research, *Expert Rev Proteomics* 17 (2020) 719–733. <https://doi.org/10.1080/14789450.2020.1855982>.
- [114] A.D. Catherman, O.S. Skinner, N.L. Kelleher, Top down proteomics: Facts and perspectives, *Biochem Biophys Res Commun* 445 (2014) 683–693. <https://doi.org/10.1016/j.bbrc.2014.02.041>.
- [115] Y. Zhang, B.R. Fonslow, B. Shan, M.-C. Baek, J.R. Yates, Protein Analysis by Shotgun/Bottom-up Proteomics, *Chem Rev* 113 (2013) 2343–2394. <https://doi.org/10.1021/cr3003533>.
- [116] E.J. Dupree, M. Jayathirtha, H. Yorkey, M. Mihasan, B.A. Petre, C.C. Darie, A critical review of bottom-up proteomics: the good, the bad, and the future of this field, *Proteomes* 8 (2020) 14. <https://doi.org/10.3390/proteomes8030014>.
- [117] J.R. Wiśniewski, C. Wegler, P. Artursson, Multiple-enzyme-digestion strategy improves accuracy and sensitivity of label- and standard-free absolute quantification to a level

- that is achievable by analysis with stable isotope-labeled standard spiking, *J Proteome Res* (2018) acs.jproteome.8b00549. <https://doi.org/10.1021/acs.jproteome.8b00549>.
- [118] S.N. Thomas, D. French, P.J. Jannetto, B.A. Rappold, W.A. Clarke, Liquid chromatography–tandem mass spectrometry for clinical diagnostics, *Nature Reviews Methods Primers* 2 (2022) 96. <https://doi.org/10.1038/s43586-022-00175-x>.
- [119] J.J. Pitt, Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry., *Clin Biochem Rev* 30 (2009) 19–34.
- [120] J. Šesták, D. Moravcová, V. Kahle, Instrument platforms for nano liquid chromatography, *J Chromatogr A* 1421 (2015) 2–17. <https://doi.org/10.1016/j.chroma.2015.07.090>.
- [121] D.T. -T. Nguyen, D. Guilleme, S. Rudaz, J. Veuthey, Fast analysis in liquid chromatography using small particle size and high pressure, *J Sep Sci* 29 (2006) 1836–1848. <https://doi.org/10.1002/jssc.200600189>.
- [122] R.D. Mittal, Tandem mass spectroscopy in diagnosis and clinical research, *Indian Journal of Clinical Biochemistry* 30 (2015) 121–123. <https://doi.org/10.1007/s12291-015-0498-9>.
- [123] S. Banerjee, S. Mazumdar, 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. <https://doi.org/10.1155/2012/282574>.
- [124] D.J. Ryan, J.M. Spraggins, R.M. Caprioli, Protein identification strategies in MALDI imaging mass spectrometry: a brief review, *Curr Opin Chem Biol* 48 (2019) 64–72. <https://doi.org/10.1016/j.cbpa.2018.10.023>.
- [125] C.S. Ho, C.W.K. Lam, M.H.M. Chan, R.C.K. Cheung, L.K. Law, L.C.W. Lit, K.F. Ng, M.W.M. Suen, H.L. Tai, Electrospray ionisation mass spectrometry: principles and clinical applications., *Clin Biochem Rev* 24 (2003) 3–12.
- [126] E. Boel, R. Koekoekx, S. Dedroog, I. Babkin, M.R. Vetrano, C. Clasen, G. Van den Mooter, Unraveling particle formation: From single droplet drying to spray drying and electrospraying, *Pharmaceutics* 12 (2020) 625. <https://doi.org/10.3390/pharmaceutics12070625>.
- [127] John C. Lindon, George E. Tranter, David Koppenaal, *Encyclopedia of spectroscopy and spectrometry*, 3rd Edition, 2016.
- [128] C. Li, S. Chu, S. Tan, X. Yin, Y. Jiang, X. Dai, X. Gong, X. Fang, D. Tian, Towards higher sensitivity of mass spectrometry: A perspective from the mass analyzers, *Front Chem* 9 (2021). <https://doi.org/10.3389/fchem.2021.813359>.

-
- [129] A.M. Haag, Mass analyzers and mass spectrometers, in: 2016: pp. 157–169. https://doi.org/10.1007/978-3-319-41448-5_7.
- [130] F. Meier, S. Beck, N. Grassl, M. Lubeck, M.A. Park, O. Raether, M. Mann, Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device, *J Proteome Res* 14 (2015) 5378–5387. <https://doi.org/10.1021/acs.jproteome.5b00932>.
- [131] H.I. Stewart, D. Grinfeld, A. Giannakopoulos, J. Petzoldt, T. Shanley, M. Garland, E. Denisov, A.C. Peterson, E. Damoc, M. Zeller, T.N. Arrey, A. Pashkova, S. Renuse, A. Hakimi, A. Kühn, M. Biel, A. Kreutzmann, B. Hagedorn, I. Colonius, A. Schütz, A. Stefes, A. Dwivedi, D. Mourad, M. Hoek, B. Reitemeier, P. Cochems, A. Kholomeev, R. Ostermann, G. Quiring, M. Ochmann, S. Möhring, A. Wagner, A. Petker, S. Kanngiesser, M. Wiedemeyer, W. Balschun, D. Hermanson, V. Zabrouskov, A.A. Makarov, C. Hock, Parallelized acquisition of orbitrap and astral analyzers enables high-throughput quantitative analysis, *Anal Chem* 95 (2023) 15656–15664. <https://doi.org/10.1021/acs.analchem.3c02856>.
- [132] M. van Bentum, M. Selbach, An introduction to advanced targeted acquisition methods, *Molecular & Cellular Proteomics* 20 (2021) 100165. <https://doi.org/10.1016/j.mcpro.2021.100165>.
- [133] C.A. Sherwood, A. Eastham, L.W. Lee, J. Risler, H. Mirzaei, J.A. Falkner, D.B. Martin, Rapid optimization of mrm-ms instrument parameters by subtle alteration of precursor and product m/z targets, *J Proteome Res* 8 (2009) 3746–3751. <https://doi.org/10.1021/pr801122b>.
- [134] P. Picotti, R. Aebersold, Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions, *Nat Methods* 9 (2012) 555–566. <https://doi.org/10.1038/nmeth.2015>.
- [135] N. Rauniyar, Parallel reaction monitoring: A targeted experiment performed using high resolution and high mass accuracy mass spectrometry, *Int J Mol Sci* 16 (2015) 28566–28581. <https://doi.org/10.3390/ijms161226120>.
- [136] V. Davies, J. Wandy, S. Weidt, J.J.J. van der Hooft, A. Miller, R. Daly, S. Rogers, Rapid development of improved data-dependent acquisition strategies, *Anal Chem* 93 (2021) 5676–5683. <https://doi.org/10.1021/acs.analchem.0c03895>.
- [137] F. Meier, A.-D. Brunner, S. Koch, H. Koch, M. Lubeck, M. Krause, N. Goedecke, J. Decker, T. Kosinski, M.A. Park, N. Bache, O. Hoerning, J. Cox, O. Räther, M. Mann, Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass
-

- Spectrometer, *Molecular & Cellular Proteomics* 17 (2018) 2534–2545. <https://doi.org/10.1074/mcp.TIR118.000900>.
- [138] A. Doerr, DIA mass spectrometry, *Nat Methods* 12 (2015) 35–35. <https://doi.org/10.1038/nmeth.3234>.
- [139] L.K. Pino, S.C. Just, M.J. MacCoss, B.C. Searle, Acquiring and analyzing data independent acquisition proteomics experiments without spectrum libraries, *Molecular & Cellular Proteomics* 19 (2020) 1088–1103. <https://doi.org/10.1074/mcp.P119.001913>.
- [140] V. Demichev, L. Szyrwił, F. Yu, G.C. Teo, G. Rosenberger, A. Niewianda, D. Ludwig, J. Decker, S. Kaspar-Schoenefeld, K.S. Lilley, M. Mülleder, A.I. Nesvizhskii, M. Ralser, dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts, *Nat Commun* 13 (2022) 3944. <https://doi.org/10.1038/s41467-022-31492-0>.
- [141] Oliver M Bernhardt, Nathalie Selevsek, Ludovic Gillet, Oliver Rinner, Spectronaut: a fast and efficient algorithm for MRM-like processing of data independent acquisition (SWATH-MS) data, in: Conference: 60th American Society for Mass Spectrometry Conference 2012, n.d.
- [142] V. Marquioni, F.M.F. Nunes, M.T.M. Novo-Mansur, Protein identification by database searching of mass spectrometry data in the teaching of proteomics, *J Chem Educ* 98 (2021) 812–823. <https://doi.org/10.1021/acs.jchemed.0c00853>.
- [143] P. Wang, S.R. Wilson, Mass spectrometry-based protein identification by integrating de novo sequencing with database searching, *BMC Bioinformatics* 14 (2013) S24. <https://doi.org/10.1186/1471-2105-14-S2-S24>.
- [144] S. Anand, M. Samuel, C.-S. Ang, S. Keerthikumar, S. Mathivanan, Label-based and label-free strategies for protein quantitation, in: 2017: pp. 31–43. https://doi.org/10.1007/978-1-4939-6740-7_4.
- [145] X. Chen, Y. Sun, T. Zhang, L. Shu, P. Roepstorff, F. Yang, Quantitative proteomics using isobaric labeling: A practical guide, *Genomics Proteomics Bioinformatics* 19 (2021) 689–706. <https://doi.org/10.1016/j.gpb.2021.08.012>.
- [146] A. Thompson, N. Wölmer, S. Koncarevic, S. Selzer, G. Böhm, H. Legner, P. Schmid, S. Kienle, P. Penning, C. Höhle, A. Berfelde, R. Martinez-Pinna, V. Farztdinov, S. Jung, K. Kuhn, I. Pike, TMTpro: Design, synthesis, and initial evaluation of a proline-based isobaric 16-plex tandem mass tag reagent set, *Anal Chem* 91 (2019) 15941–15950. <https://doi.org/10.1021/acs.analchem.9b04474>.
- [147] Z. Wang, K. Kavdia, K.K. Dey, V.R. Pagala, K. Kodali, D. Liu, D.G. Lee, H. Sun, S.R. Chepyala, J.-H. Cho, M. Niu, A.A. High, J. Peng, High-throughput and deep-proteome profiling by

- 16-plex tandem mass tag labeling coupled with two-dimensional chromatography and mass spectrometry, *Journal of Visualized Experiments* (2020). <https://doi.org/10.3791/61684>.
- [148] W. Zhu, J.W. Smith, C.-M. Huang, Mass spectrometry-based label-free quantitative proteomics, *J Biomed Biotechnol* 2010 (2010) 1–6. <https://doi.org/10.1155/2010/840518>.
- [149] L. Arike, L. Peil, Spectral counting label-free proteomics, in: 2014: pp. 213–222. https://doi.org/10.1007/978-1-4939-0685-7_14.
- [150] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, M. Mann, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein, *Molecular & Cellular Proteomics* 4 (2005) 1265–1272. <https://doi.org/10.1074/mcp.M500061-MCP200>.
- [151] J. Cox, M.Y. Hein, C.A. Lubner, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, Termed MaxLFQ, *Molecular & Cellular Proteomics* 13 (2014) 2513–2526. <https://doi.org/10.1074/mcp.M113.031591>.
- [152] J.R. Wiśniewski, Label-free and standard-free absolute quantitative proteomics using the “total protein” and “proteomic ruler” approaches, in: 2017: pp. 49–60. <https://doi.org/10.1016/bs.mie.2016.10.002>.
- [153] J.R. Wiśniewski, M.Y. Hein, J. Cox, M. Mann, A “proteomic ruler” for protein copy number and concentration estimation without spike-in standards, *Molecular & Cellular Proteomics* 13 (2014) 3497–3506. <https://doi.org/10.1074/mcp.M113.037309>.
- [154] J.R. Wiśniewski, D. Rakus, Multi-enzyme digestion FASP and the ‘total protein approach’-based absolute quantification of the escherichia coli proteome, *J Proteomics* 109 (2014) 322–331. <https://doi.org/10.1016/j.jprot.2014.07.012>.
- [155] A. Macklin, S. Khan, T. Kislinger, Recent advances in mass spectrometry based clinical proteomics: applications to cancer research, *Clin Proteomics* 17 (2020) 17. <https://doi.org/10.1186/s12014-020-09283-w>.
- [156] D. Vitko, W.-F. Chou, S. Nouri Golmaei, J.-Y. Lee, C. Belthangady, J. Blume, J.K. Chan, G. Flores-Campuzano, Y. Hu, M. Liu, M.A. Marispini, M.G. Mora, S. Ramaswamy, P. Ranjan, P.B. Williams, R.J.X. Zawada, P. Ma, B.E. Wilcox, timsTOF HT improves protein identification and quantitative reproducibility for deep unbiased plasma protein biomarker discovery, *J Proteome Res* 23 (2024) 929–938. <https://doi.org/10.1021/acs.jproteome.3c00646>.

- [157] L.R. Heil, E. Damoc, T.N. Arrey, A. Pashkova, E. Denisov, J. Petzoldt, A.C. Peterson, C. Hsu, B.C. Searle, N. Shulman, M. Riffle, B. Connolly, B.X. MacLean, P.M. Remes, M.W. Senko, H.I. Stewart, C. Hock, A.A. Makarov, D. Hermanson, V. Zabrouskov, C.C. Wu, M.J. MacCoss, Evaluating the performance of the astral mass analyzer for quantitative proteomics using data-independent acquisition, *J Proteome Res* 22 (2023) 3290–3300. <https://doi.org/10.1021/acs.jproteome.3c00357>.
- [158] Y. LI, D. XUN, L. LI, B. WANG, J. LV, H. LIU, L. ZHU, F. MA, X. CHEN, S. TIAN, C. DING, Deep dive on the proteome of human body fluids: a valuable data resource for biomarker discovery, *Cancer Genomics - Proteomics* 18 (2021) 549–568. <https://doi.org/10.21873/cgp.20280>.
- [159] S. Hu, J.A. Loo, D.T. Wong, Human body fluid proteome analysis, *Proteomics* 6 (2006) 6326–6353. <https://doi.org/10.1002/pmic.200600284>.
- [160] N.S. Yee, Liquid biopsy: A biomarker-driven tool towards precision oncology, *J Clin Med* 9 (2020) 2556. <https://doi.org/10.3390/jcm9082556>.
- [161] V. Caputo, F. Ciardiello, C.M. Della Corte, G. Martini, T. Troiani, S. Napolitano, Diagnostic value of liquid biopsy in the era of precision medicine: 10 years of clinical evidence in cancer, *Explor Target Antitumor Ther* (2023) 102–138. <https://doi.org/10.37349/etat.2023.00125>.
- [162] É. Csősz, G. Kalló, B. Márkus, E. Deák, A. Csutak, J. Tőzsér, Quantitative body fluid proteomics in medicine — A focus on minimal invasiveness, *J Proteomics* 153 (2017) 30–43. <https://doi.org/10.1016/j.jprot.2016.08.009>.
- [163] S. Thomas, L. Hao, W.A. Ricke, L. Li, Biomarker discovery in mass spectrometry-based urinary proteomics, *Proteomics Clin Appl* 10 (2016) 358–370. <https://doi.org/10.1002/prca.201500102>.
- [164] S. Decramer, A.G. de Peredo, B. Breuil, H. Mischak, B. Monsarrat, J.-L. Bascands, J.P. Schanstra, Urine in clinical proteomics, *Molecular & Cellular Proteomics* 7 (2008) 1850–1862. <https://doi.org/10.1074/mcp.R800001-MCP200>.
- [165] C.S. Spahr, M.T. Davis, M.D. McGinley, J.H. Robinson, E.J. Bures, J. Beierle, J. Mort, P.L. Courchesne, K. Chen, R.C. Wahl, W. Yu, R. Luethy, S.D. Patterson, Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I. Profiling an unfractionated tryptic digest, *Proteomics* 1 (2001) 93–107. [https://doi.org/10.1002/1615-9861\(200101\)1:1<93::AID-PROT93>3.0.CO;2-3](https://doi.org/10.1002/1615-9861(200101)1:1<93::AID-PROT93>3.0.CO;2-3).

-
- [166] W. Sun, F. Li, S. Wu, X. Wang, D. Zheng, J. Wang, Y. Gao, Human urine proteome analysis by three separation approaches, *Proteomics* 5 (2005) 4994–5001. <https://doi.org/10.1002/pmic.200401334>.
 - [167] J. Adachi, C. Kumar, Y. Zhang, J. V Olsen, M. Mann, The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins, *Genome Biol* 7 (2006) R80. <https://doi.org/10.1186/gb-2006-7-9-r80>.
 - [168] A. Marimuthu, Robert.N. O'Meally, R. Chaerkady, Y. Subbannayya, V. Nanjappa, P. Kumar, D.S. Kelkar, S.M. Pinto, R. Sharma, S. Renuse, R. Goel, R. Christopher, B. Delanghe, Robert.N. Cole, H.C. Harsha, A. Pandey, A Comprehensive map of the human urinary proteome, *J Proteome Res* 10 (2011) 2734–2743. <https://doi.org/10.1021/pr2003038>.
 - [169] T. Farrah, E.W. Deutsch, G.S. Omenn, Z. Sun, J.D. Watts, T. Yamamoto, D. Shteynberg, M.M. Harris, R.L. Moritz, State of the human proteome in 2013 as viewed through peptideatlas: Comparing the kidney, urine, and plasma proteomes for the biology- and disease-driven human proteome project, *J Proteome Res* 13 (2014) 60–75. <https://doi.org/10.1021/pr4010037>.
 - [170] M. Zhao, M. Li, Y. Yang, Z. Guo, Y. Sun, C. Shao, M. Li, W. Sun, Y. Gao, A comprehensive analysis and annotation of human normal urinary proteome, *Sci Rep* 7 (2017) 3024. <https://doi.org/10.1038/s41598-017-03226-6>.
 - [171] V. Thongboonkerd, K.R. Mcleish, J.M. Arthur, J.B. Klein, Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation, *Kidney Int* 62 (2002) 1461–1469. <https://doi.org/10.1111/j.1523-1755.2002.kid565.x>.
 - [172] T.K. Sigdel, C.D. Nicora, S.-C. Hsieh, H. Dai, W.-J. Qian, D.G. Camp, M.M. Sarwal, Optimization for peptide sample preparation for urine peptidomics, *Clin Proteomics* 11 (2014) 7. <https://doi.org/10.1186/1559-0275-11-7>.
 - [173] S. Tkáčiková, I. Talian, J. Sabo, Optimisation of urine sample preparation for shotgun proteomics, *Open Chem* 18 (2020) 850–856. <https://doi.org/10.1515/chem-2020-0150>.
 - [174] Z.-M. Li, K. Kannan, Determination of 19 steroid hormones in human serum and urine using liquid chromatography-tandem mass spectrometry, *Toxics* 10 (2022) 687. <https://doi.org/10.3390/toxics10110687>.
 - [175] J.R. Wiśniewski, Filter aided sample preparation – A tutorial, *Anal Chim Acta* 1090 (2019) 23–30. <https://doi.org/10.1016/j.aca.2019.08.032>.
 - [176] J.R. Wiśniewski, Filter-aided sample preparation for proteome analysis, in: 2018: pp. 3–10. https://doi.org/10.1007/978-1-4939-8695-8_1.
-

- [177] M. Su, Z. Zhang, L. Zhou, C. Han, C. Huang, E.C. Nice, Proteomics, personalized medicine and cancer, *Cancers (Basel)* 13 (2021) 2512. <https://doi.org/10.3390/cancers13112512>.
- [178] A.B. Chambliss, D.W. Chan, Precision medicine: from pharmacogenomics to pharmacoproteomics, *Clin Proteomics* 13 (2016) 25. <https://doi.org/10.1186/s12014-016-9127-8>.
- [179] S. Doll, F. Gnad, M. Mann, The case for proteomics and phospho-proteomics in personalized cancer medicine, *Proteomics Clin Appl* 13 (2019). <https://doi.org/10.1002/prca.201800113>.
- [180] L. Uechi, S. Vasudevan, D. Vilenski, S. Branciamore, D. Frankhouser, D. O'Meally, S. Meshinchi, G. Marcucci, Y.-H. Kuo, R. Rockne, N. Kravchenko-Balasha, Transcriptome free energy can serve as a dynamic patient-specific biomarker in acute myeloid leukemia, *NPJ Syst Biol Appl* 10 (2024) 32. <https://doi.org/10.1038/s41540-024-00352-6>.
- [181] M.H. Roehrl, V.B. Roehrl, J.Y. Wang, Proteome-based pathology: the next frontier in precision medicine, *Expert Rev Precis Med Drug Dev* 6 (2021) 1–4. <https://doi.org/10.1080/23808993.2021.1854611>.
- [182] M. Neagu, M. Bostan, C. Constantin, Protein microarray technology: Assisting personalized medicine in oncology (Review), *World Acad Sci J* (2019). <https://doi.org/10.3892/wasj.2019.15>.
- [183] A.K. Füzerý, J. Levin, M.M. Chan, D.W. Chan, Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges, *Clin Proteomics* 10 (2013) 13. <https://doi.org/10.1186/1559-0275-10-13>.
- [184] M. Mann, C. Kumar, W.-F. Zeng, M.T. Strauss, Artificial intelligence for proteomics and biomarker discovery, *Cell Syst* 12 (2021) 759–770. <https://doi.org/10.1016/j.cels.2021.06.006>.
- [185] H. Alkhatib, J. Conage-Pough, S. Roy Chowdhury, D. Shian, D. Zaid, A.M. Rubinstein, A. Sonnenblick, T. Peretz-Yablonsky, A. Granit, E. Carmon, I.N. Kohale, J.C. Boughey, M.P. Goetz, L. Wang, F.M. White, N. Kravchenko-Balasha, Patient-specific signaling signatures predict optimal therapeutic combinations for triple negative breast cancer, *Mol Cancer* 23 (2024) 17. <https://doi.org/10.1186/s12943-023-01921-9>.
- [186] A.G. Birhanu, Mass spectrometry-based proteomics as an emerging tool in clinical laboratories, *Clin Proteomics* 20 (2023) 32. <https://doi.org/10.1186/s12014-023-09424-x>.
- [187] L.B. Carvalho, J.L. Capelo, C. Lodeiro, R. Dhir, L.C. Pinheiro, H. López-Fernández, G. Martins, M. Medeiros, F. Díaz, H.M. Santos, Pathway-guided monitoring of the disease course in bladder cancer with longitudinal urine proteomics, *Communications Medicine* 3 (2023) 8. <https://doi.org/10.1038/s43856-023-00238-4>.

-
- [188] T. Yoshida, H. Kinoshita, K. Fukui, T. Matsuzaki, K. Yoshida, T. Mishima, M. Yanishi, Y. Komai, M. Sugi, T. Inoue, T. Murota, T. Matsuda, Prognostic impact of renin-angiotensin inhibitors in patients with bladder cancer undergoing radical cystectomy, *Ann Surg Oncol* 24 (2017) 823–831. <https://doi.org/10.1245/s10434-016-5534-3>.
 - [189] H. Lee, K. Kim, J. Woo, J. Park, H. Kim, K.E. Lee, H. Kim, Y. Kim, K.C. Moon, J.Y. Kim, I.A. Park, B.B. Shim, J.H. Moon, D. Han, H.S. Ryu, Quantitative Proteomic Analysis Identifies AHNAK (Neuroblast Differentiation-associated Protein AHNAK) as a Novel Candidate Biomarker for Bladder Urothelial Carcinoma Diagnosis by Liquid-based Cytology, *Molecular & Cellular Proteomics* 17 (2018) 1788–1802. <https://doi.org/10.1074/mcp.RA118.000562>.
 - [190] H.-Y. Hsieh, Y.-C. Jou, C.-L. Tung, Y.-S. Tsai, Y.-H. Wang, C.-L. Chi, R.-I. Lin, S.-K. Hung, Y.-M. Chuang, S.-F. Wu, C. Li, C.-H. Shen, M.W.Y. Chan, C.-D. Hsu, Epigenetic silencing of the dual-role signal mediator, ANGPTL4 in tumor tissues and its overexpression in the urothelial carcinoma microenvironment, *Oncogene* 37 (2018) 673–686. <https://doi.org/10.1038/onc.2017.375>.
 - [191] H. Li, C. Li, H. Wu, T. Zhang, J. Wang, S. Wang, J. Chang, Identification of Apo-A1 as a biomarker for early diagnosis of bladder transitional cell carcinoma, *Proteome Sci* 9 (2011) 21. <https://doi.org/10.1186/1477-5956-9-21>.
 - [192] V. Soukup, O. Capoun, M. Pesl, L. Vavrova, R. Sobotka, K. Levova, T. Hanus, T. Zima, M. Kalousova, The significance of calprotectin, CD147, APOA4 and DJ-1 in non-invasive detection of urinary bladder carcinoma, *Neoplasma* 66 (2019) 1019–1023. https://doi.org/10.4149/NEO_2019_190124N74.
 - [193] L. Ren, J. Yi, W. Li, X. Zheng, J. Liu, J. Wang, G. Du, Apolipoproteins and cancer, *Cancer Med* 8 (2019) 7032–7043. <https://doi.org/10.1002/cam4.2587>.
 - [194] M. Lindén, U. Segersten, M. Runeson, K. Wester, C. Busch, U. Pettersson, S.B. Lind, P.-U. Malmström, Tumour expression of bladder cancer-associated urinary proteins, *BJU Int* 112 (2013) 407–415. <https://doi.org/10.1111/j.1464-410X.2012.11653.x>.
 - [195] A.-P. Xiang, X.-N. Chen, P.-F. Xu, S.-H. Shao, Y.-F. Shen, Expression and prognostic value of carbonic anhydrase IX (CA-IX) in bladder urothelial carcinoma, *BMC Urol* 22 (2022) 120. <https://doi.org/10.1186/s12894-022-01074-9>.
 - [196] S. Liu, Z. Liu, A. Shang, J. Xun, Z. Lv, S. Zhou, C. Liu, Q. Zhang, Y. Yang, CD44 is a potential immunotherapeutic target and affects macrophage infiltration leading to poor prognosis, *Sci Rep* 13 (2023) 9657. <https://doi.org/10.1038/s41598-023-33915-4>.
-

- [197] M. Gurhan Balci, M. Tayfur, Loss of E-cadherin expression in recurrent non-invasive urothelial carcinoma of the bladder, 2018. www.ijcep.com/ (accessed February 2, 2021).
- [198] N. Pećina-Šlaus, Tumor suppressor gene E-cadherin and its role in normal and malignant cells, *Cancer Cell Int* 3 (2003) 17. <https://doi.org/10.1186/1475-2867-3-17>.
- [199] A.S. Feldman, J. Banyard, C.-L. Wu, W.S. McDougal, B.R. Zetter, Cystatin B as a tissue and urinary biomarker of bladder cancer recurrence and disease progression, *Clinical Cancer Research* 15 (2009) 1024–1031. <https://doi.org/10.1158/1078-0432.CCR-08-1143>.
- [200] L. Grau, J.L. Luque-Garcia, P. González-Peramato, D. Theodorescu, J. Palou, J.M. Fernandez-Gomez, M. Sánchez-Carbayo, A quantitative proteomic analysis uncovers the relevance of CUL3 in bladder Cancer aggressiveness, *PLoS One* 8 (2013) e53328. <https://doi.org/10.1371/journal.pone.0053328>.
- [201] M. Miyake, S. Hori, Y. Morizawa, Y. Tatsumi, Y. Nakai, S. Anai, K. Torimoto, K. Aoki, N. Tanaka, K. Shimada, N. Konishi, M. Toritsuka, T. Kishimoto, C.J. Rosser, K. Fujimoto, CXCL1-Mediated interaction of cancer cells with tumor-associated macrophages and cancer-associated fibroblasts promotes tumor progression in human bladder cancer, *Neoplasia* 18 (2016) 636–646. <https://doi.org/10.1016/j.neo.2016.08.002>.
- [202] Y. Dodurga, C.B. Avci, S. Yilmaz, Z.O. Dogan, Z. Kesen, C. Tataroglu, N.L. Satioglu-Tufan, T. Bushra, C. Gunduz, Evaluation of deleted in malignant brain tumors 1 (DMBT1) gene expression in bladder carcinoma cases: preliminary study, *Biomarkers* 16 (2011) 610–615. <https://doi.org/10.3109/1354750X.2011.620627>.
- [203] W. George Warren, M. Osborn, A. Yates, K. Wright, S.E. O’Sullivan, The emerging role of fatty acid binding protein 5 (FABP5) in cancers, *Drug Discov Today* 28 (2023) 103628. <https://doi.org/10.1016/j.drudis.2023.103628>.
- [204] C.-H. Lin, H.-H. Chang, C.-R. Lai, H.-H. Wang, W.-C. Tsai, Y.-L. Tsai, C.-Y. Changchien, Y.-C. Cheng, S.-T. Wu, Y. Chen, Fatty acid binding protein 6 inhibition decreases cell cycle progression, migration and autophagy in bladder cancers, *Int J Mol Sci* 23 (2022) 2154. <https://doi.org/10.3390/ijms23042154>.
- [205] S.-Y. Lin, C.-H. Chang, H.-C. Wu, C.-C. Lin, K.-P. Chang, C.-R. Yang, C.-P. Huang, W.-H. Hsu, C.-T. Chang, C.-J. Chen, Proteome profiling of urinary exosomes identifies alpha 1-antitrypsin and H2B1K as diagnostic and prognostic biomarkers for urothelial carcinoma, *Sci Rep* 6 (2016) 34446. <https://doi.org/10.1038/srep34446>.
- [206] R. Kaempfer, L. Gerez, H. Farbstein, L. Madar, O. Hirschman, R. Nussinovich, A. Shapiro, Prediction of response to treatment in superficial bladder carcinoma through pattern of

- interleukin-2 gene expression., *Journal of Clinical Oncology* 14 (1996) 1778–1786. <https://doi.org/10.1200/JCO.1996.14.6.1778>.
- [207] M.-F. Chen, P.-Y. Lin, C.-F. Wu, W.-C. Chen, C.-T. Wu, IL-6 expression regulates tumorigenicity and correlates with prognosis in bladder cancer, *PLoS One* 8 (2013) e61901. <https://doi.org/10.1371/journal.pone.0061901>.
- [208] K. Qu, J. Gu, Y. Ye, S.B. Williams, C.P. Dinney, X. Wu, A. Kamat, High baseline levels of interleukin-8 in leukocytes and urine predict tumor recurrence in non-muscle invasive bladder cancer patients receiving bacillus Calmette–Guerin therapy: A long-term survival analysis, *Oncoimmunology* 6 (2017) e1265719. <https://doi.org/10.1080/2162402X.2016.1265719>.
- [209] K. Zhao, M. Gao, M. Lin, KLK6 functions as an oncogene and unfavorable prognostic factor in bladder urothelial carcinoma, *Dis Markers* 2022 (2022) 1–14. <https://doi.org/10.1155/2022/3373851>.
- [210] C. Camilli, A.E. Hoeh, G. De Rossi, S.E. Moss, J. Greenwood, LRG1: an emerging player in disease pathogenesis, *J Biomed Sci* 29 (2022) 6. <https://doi.org/10.1186/s12929-022-00790-6>.
- [211] J. Kudelski, A. Tokarzewicz, M. Gudowska-Sawczuk, B. Mroczko, P. Chłosta, M. Bruczko-Goralewska, P. Mitura, G. Młynarczyk, The significance of matrix metalloproteinase 9 (MMP-9) and metalloproteinase 2 (MMP-2) in urinary bladder cancer, *Biomedicines* 11 (2023) 956. <https://doi.org/10.3390/biomedicines11030956>.
- [212] F.C. Zeng, S. Cen, Z.Y. Tang, X.L. Kang, Elevated matrix metalloproteinase-9 expression may contribute to the pathogenesis of bladder cancer, *Oncol Lett* 11 (2016) 2213–2222. <https://doi.org/10.3892/ol.2016.4187>.
- [213] F. Li, Z. Yu, P. Chen, G. Lin, T. Li, L. Hou, Y. Du, W. Tan, The increased excretion of urinary orosomucoid 1 as a useful biomarker for bladder cancer, *Am J Cancer Res* 6 (2016) 331–40. <http://www.ncbi.nlm.nih.gov/pubmed/27186407> (accessed February 2, 2021).
- [214] K. Miyamoto, N. Seki, R. Matsushita, M. Yonemori, H. Yoshino, M. Nakagawa, H. Enokida, Tumour-suppressive miRNA-26a-5p and miR-26b-5p inhibit cell aggressiveness by regulating PLOD2 in bladder cancer, *Br J Cancer* 115 (2016) 354–363. <https://doi.org/10.1038/bjc.2016.179>.
- [215] J. Geng, J. Fan, P. Wang, Z.J. Fang, G.W. Xia, H.W. Jiang, G. Chen, Q. Ding, REG1A predicts recurrence in stage Ta/T1 bladder cancer, *European Journal of Surgical Oncology (EJSO)* 35 (2009) 852–857. <https://doi.org/10.1016/j.ejso.2008.12.007>.

- [216] S.-K. Kim, E.-J. Kim, S.-H. Leem, Y.-S. Ha, Y.-J. Kim, W.-J. Kim, Identification of S100A8-correlated genes for prediction of disease progression in non-muscle invasive bladder cancer, *BMC Cancer* 10 (2010) 21. <https://doi.org/10.1186/1471-2407-10-21>.
- [217] J.L. Capelo, C. Lodeiro, R. Dhir, N. Kravchenko-Balasha, H.M. Santos, Prescriptomics: the next frontier in medicine, *J Integr OMICS* 12 (2022). <https://doi.org/10.5584/ji-omics.v12i1.211>.

OBJECTIVES

There is growing evidence that longitudinal proteomics analysis can anticipate diseases progression and response to therapy. This project aims to join and integrate different areas of expertise in bioanalytical chemistry, mass-spectrometry, bioinformatics, biology and medicine to empower personal health in bladder cancer patients by facilitating diagnostic and prognostic of disease course.

Fig.2-1 highlights the rationale and thought process of the PhD project, outlining and combining all the objectives with the chapters of present doctoral thesis.

Specifically, the project aims to:

- i. Improve the current standard method of urine sample treatment and collection, optimizing it for clinical application and shipping easy between facilities.
- ii. Develop a fast and high throughput novel urine proteome approach that can easily be applied to the medical field.
- iii. Understand the optimal mass spectrometry data treatment for clinical applications.
- iv. Identify common protein phenotypes associated with BCa progression, recurrence, and response to treatment.
- v. Pinpointing dysregulated proteins and their biological pathways that can be targeted for therapeutic intervention and used to surveillance patients' response to treatment.
- vi. Further extend and validated the method using a different urinary tract type cancer, renal carcinoma.

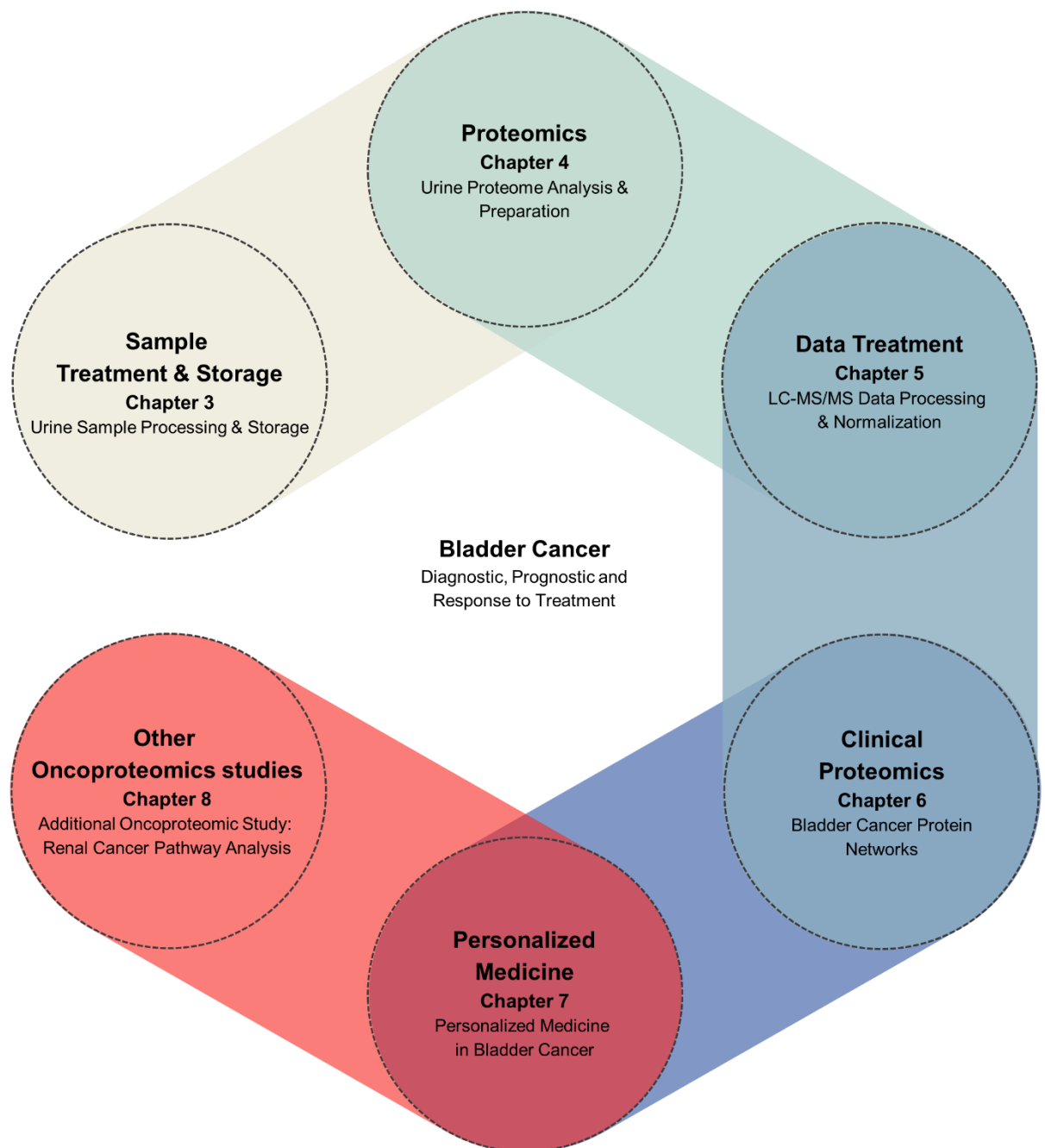


Figure 2-1. Rationale and Integration of the doctoral thesis objectives with the different chapters of the thesis.

Furthermore, the identification of specific alterations in protein dynamics and its functional impact will cast light in the mechanisms of BCa progression and response to treatment, which can be particularly interesting to select potential targets for new drugs improve survival rate in BCa patients. In addition, the knowledge retrieved from this study aims to:

- (i) Support research for BCa (non-communicable disease);
- (ii) Reduce the high disease burden and societal costs of BCa;
- (iii) Ensure healthy lives in all ages.

SNAP-HEATED FREEZE-FREE PRESERVATION AND PROCESSING OF THE URINE PROTEOME USING THE COMBINATION OF STABILIZER-BASED TECHNOLOGY AND FILTER AIDED SAMPLE PREPARATION

Luís B. Carvalho, José Luis Capelo Martínez, Carlos Lodeiro, Jacek R. Wiśniewski, Hugo M. Santos

Analytica Chimica Acta 1076 (2019) 82-90

Doi: 10.1016/j.aca.2019.05.051

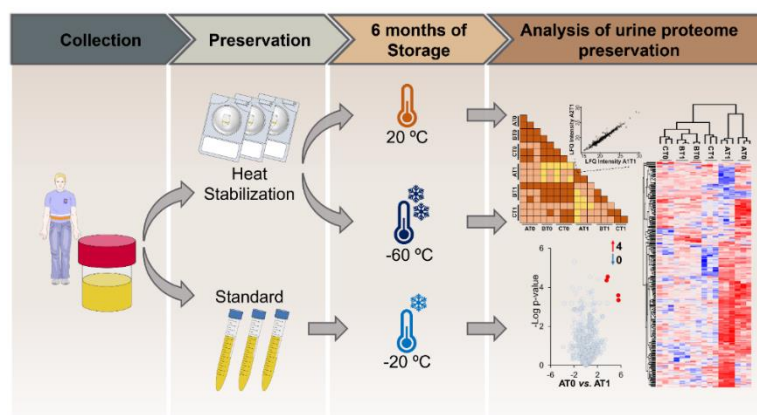


Figure 3-1. Graphical abstract.

3.1 Abstract

We present a method to preserve and process urine proteins for proteomic analysis in a filter aided sample preparation (FASP) format. The method combines concentration of urine proteins in ultrafiltration devices, their thermal stabilization, allowing long term storage of the samples, and filter aided sample preparation. Proteomes of four different urines were preserved during 48 h and 6 months using (i) the classic freeze preservation at -20°C , (ii) snap-heated freeze-free preservation at laboratory temperature (20°C) and (iii) snap-heated preservation at -60°C . The three storage methods can effectively preserve the urine proteome for at least 6 months without significant alterations. Abundances of more than 500 proteins and specially 24 selected -cleared or -approved protein assayed in serum or plasma were found similar within the three preservation methods assessed. The new method here proposed dramatically simplifies the conditions for preserving the urine proteome for biobanks in terms of space and storage, including lowering the risks of sample degradation caused by malfunction of the freezer. Furthermore, the shipping of large number of samples can be made without the need of freezing. The application of the FASP format to isolate and preserve the proteins facilitates long-term storage and processing of proteome of urine samples.

3.2 Introduction

Human body fluids such as urine, serum, plasma, cerebrospinal fluid, saliva, and tears have characteristic protein profiles that can be used to unravel diseases [1,2]. Of the fluids mentioned above, serum and plasma are still the most used for biomarker discovery. However, urine is gaining momentum due to a number of unique advantages, such as that large sample volumes can be collected in a non-invasive and therefore painless manner, it can be collected at any time and at any place [3,4]. Urine contains proteins of which circa 70% are from the urinary tract, while the remaining 30% corresponds to blood proteins [5]. Urinary proteins can originate from (i) glomerular filtration of blood proteins, (ii) secretion of proteins by the renal tubular epithelial cells, (iii) shedding of cells along the urinary tract and (iv) exosome secretion [6]. Some reliable biomarkers present in urine has been identified as candidates to diagnostic and prognostic of some diseases [7,8]. Such biomarkers are mainly related to diseases of the urinary system, and some have been approved and also adopted by the FDA qualification program [9]. By now, the proteomics community is intensively working on finding new biomarkers for a number of diseases.

In urine-based proteomics biomarker discovery, second-morning urine is becoming the standard sample to be collected for biomarker-based studies. Such urine is collected during mid-stream due to the low probability of protein contamination from overgrown bacteria and epithelial cells [10]. During collection, urine preservation is often done using bacterial growth inhibitors [11]. However, the use of protease inhibitors is no longer recommended [12].

The in-depth study of the urine proteome by mass spectrometry is time consuming due to the time required to complete the analysis, usually 90 min gradients, plus the subsequent bioinformatic data interrogation. The impossibility of analysing the proteome of many urine samples in short periods of time makes mandatory the need of urine preservation and storage. Thus, after collection, the conditions for the stabilization and storage of urine proteomes in the short, medium and long term are critical because urine can degrade. Currently, the gold standard methodology for urine storage for proteomics applications was defined by the Human Kidney and Urine Proteome Project (HKUPP) [13] and by the European Urine and Kidney Proteomics (EuroKUP) [14] as follows: after collection, it is recommended urine centrifugation to remove cell debris and then to storage it at -20 °C or -80 °C [15,16]. Such recommendations present some drawbacks, as there is the need for expensive freezers which, in addition to the price, must be connected to emergency energy supply systems to prevent the effects of electric power failures. Furthermore, repeated freezing and thawing cycles have a detrimental effect in the sample. Thus, it has been shown urine proteome alterations after just 5 freeze-thaw cycles [17,18]. Finally, to store a large amount of urine samples is space troubling.

Of the many different sample treatment methods available to deal with complex proteomes, the Filter-Aided Sample Preparation method, FASP [19], is one of the most versatile bottom-up proteomic approaches, outperforming another in-solution digestion procedures [20]. Thus, FASP consists in a protein extraction and subsequent digestion approach that uses an ultrafiltration weight cut-off molecular membrane inserted in a microtube. Such device facilitates removal of salts, detergents and other contaminants from the protein sample using centrifugation [21]. It was first designed to extract and digest proteins from tissues [22] and cell samples [23], however it has been further applied to a large variety of biological samples including urine [24]. In addition, large scale sample handling is possible as the application of FASP in 96 well plates has also been shown [25]. To date, the use of membranes to separate and preserve the urine proteome has been cited in literature few times [26,27].

The Stabilizer™ T1 is a portable, benchtop instrument that uses rapid conductive heating to preserve biological samples as closely as possible to their original state. This technology has

proven to be very useful in the preservation of the proteome complex samples such as tissues and blood [28,29].

In this work we describe a new approach to preserve the urinary proteome at constant laboratory temperature (20 °C). The novel method combines urine proteome separation using the FASP membrane followed by proteome thermal snap using the Stabilizer®. In brief, the urinary proteome was separated and concentrated using the filter aided sample preparation method, and then the proteome attached to the FASP membrane was snap-heated by a thermal shock treatment provided by a rapid conductive heating that results in irreversible denaturation of proteins, permanent elimination of all enzymatic activity and also in the inactivation of viral and bacterial pathogens. We demonstrate in this work that by using a volume as low as 1.2 ml of urine, the proteome of urine samples can be preserved and stored in the FASP membrane during at least 6 months at a constant temperature of 20 °C. Furthermore, the handling of the proteome for subsequent label-free quantification analysis by high resolution mass spectrometry is facilitated as the standard FASP-protocol can be followed. Therefore, sampling large volume of urine for proteomic analysis, and urine storage and preservation in large volume containers with the aid of freezers is no longer necessary, at least for a six month period.

3.3 Experimental Section

3.3.1 Reagents

All reagents used were HPLC grade or higher. Albumin from bovine serum (BSA), iodoacetamide (IAA), urea, formic acid (FA), tris base and Bradford reagent were purchased from Sigma-Aldrich (Basel, Switzerland). Ammonium bicarbonate (AmBic) and hydrochloric acid solution 37% (HCl) were purchased from Fluka (Basel, Switzerland). Acetonitrile (ACN) and boric acid were purchased from Panreac (Barcelona, Spain). 1,4-Dithio-DL-threitol (DTT) was purchased from Alfa Aesar (Haverhill, Massachusetts, USA). Pierce™ trypsin protease MS grade, trifluoroacetic acid (TFA) and pierce™ quantitative colorimetric peptide assay was purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Ethanol 96% was purchased from Carlo Erba Reagents (Val-de-Reuil, France).

3.3.2 Apparatus

A biological safety cabinet from Alpina (Konin, Poland) and a centrifuge MPW-350 from MPW Med. Instruments (Warsaw, Poland) were used throughout the urine sample treatment.

The instrument Stabilizer™ T1 and maintainor® tissue card from Denator (Möndal, Sweden) were used in the urinary proteome heat stabilization. The membrane vivacon 500 10,000 MWCO Hydrosart from Sartorius (Goettingen, Germany) and prism™ refrigerated microcentrifuge from Labnet (New Jersey, USA) were used in the FASP methodology. A vacuum concentrator centrifuge model UNIVAPO 150 ECH Speed Vac and a vacuum pump model UNIJET II (Munich, Germany) were used for sample drying. A mini-incubator from Labnet (New Jersey, USA) was used for protein reduction and digestion. Vortex models ELMI CM70M-09 from Sky-Line (Southern California, USA) and VX-200 Vortex Mixer from Labnet (New Jersey, USA) were used in sample mix. CLARIOstar® High Performance Monochromator Multimode from BMG LABTECH (Germany) was used for Bradford and total peptide assays. Acquisition of mass spectrometry data was done using an UHR-QqTOF IMPACT HD from Bruker Daltonics (Bremen, Germany). Chromatographic separation of peptides were 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 × 150mm (Thermo Fisher Scientific).

3.3.3 Urine Specimens and Collection

Fig.3-2 depicts the sample handling and sample treatment workflow. The urine samples were collected according the HUKPP guideline. Briefly, fresh second-morning mid-stream urines were collected from four healthy male volunteers in the same age group (24 ± 1 years old). All the volunteers provided a signed consent for the usage of their urine in this study. The urine samples were collected into 100 mL urine collection tube containing 37 mg of boric acid. The urine was divided into two 50 mL tubes and were then centrifuged at 1000 g for 10 min to remove cell debris. After centrifugation, the supernatants were extracted and divided into 10 mL aliquots.

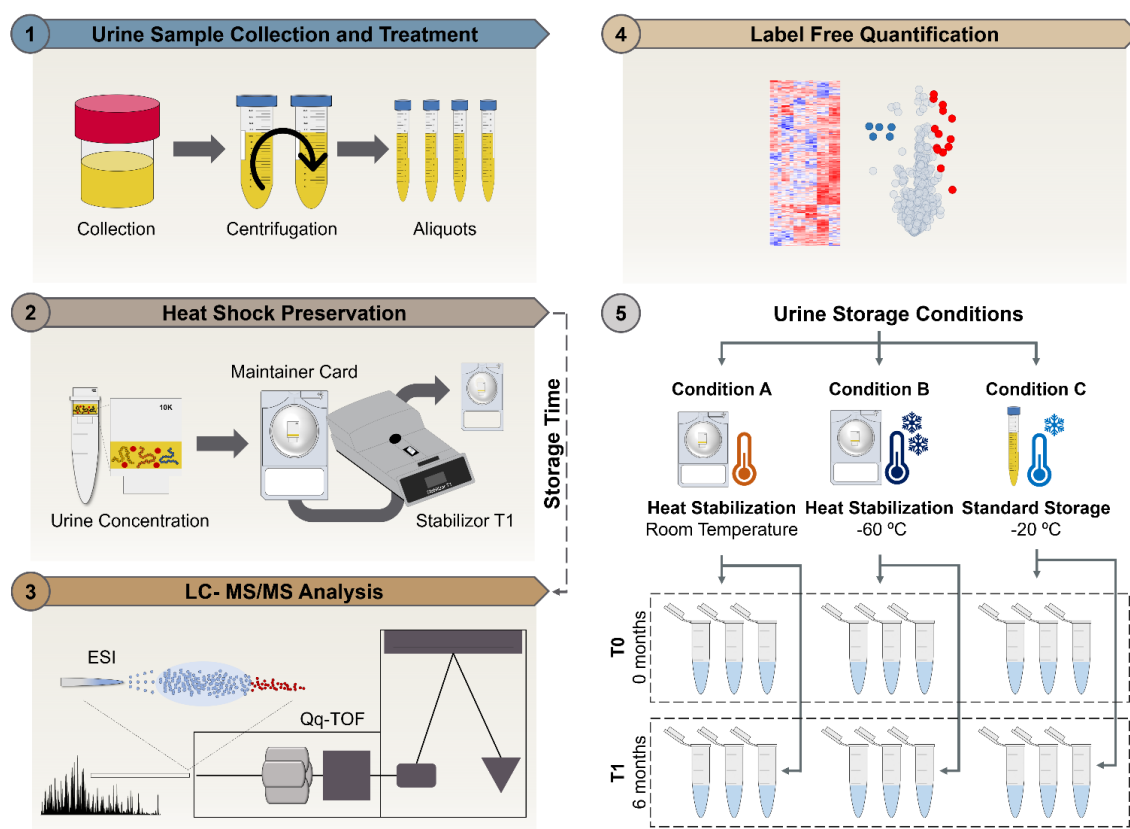


Figure 3-2. Method Workflow chart. 1. First, urine is collected, then centrifuged to avoid cell debris and then aliquoted and 2. stored following one of these 3 conditions: (A) using the ultrafiltration membrane, the heat shock preservation and stored to laboratory temperature (20 °C); (B) using the membrane, the heat shock preservation and stored to -60°; (C) standard storage at -20 °C. 3 & 4. After 48 h or 6 months all stored samples are treated following the FASP protocol for protein cleavage and then samples are quantified by label-free high-resolution mass spectrometry. 5. Urine storage conditions and proteome comparisons done in this work. The proteomes were quantitatively compared to find out differences between the storage conditions within the same period of storage, 48 h or 6 months, and between periods of storage, 48h versus 6 months.

3.3.4 Urinary Proteome Extraction

Urine sample concentration was processed in a Vivacon 500, 10,000 molecular weight cut-off hydrosart membrane. Before sample loading, 100 µL of 70% ethanol solution was added to the vivacon and then centrifugated at 14,000 g for 15 minutes. A urine volume of 400 µL was then loaded into the filter device and then was centrifuged during 15 minutes at 14,000 g. This step was repeated three times, and a total of 1.2 mL urine was passed through the membrane. In the last centrifugation, the time of centrifugation was 25 min to ensure that the membrane was dried and there was no urine on the top of the membrane (Fig.3-6 A). The proteins remained on the membrane.

3.3.5 Heat Shock Preservation

Urinary proteome present in the membrane was preserved by a heat treatment using the bench instrument stabilizer™ T1. The membrane containing the urine proteome was stored in a stabilizer™ T1 maintainer® tissue card that then undergoes inside the stabilizer™ T1 equipment where is applied a rapid conductive heating to preserve the biological samples (conditions A and B). The preservation is processed under a custom program with a controlled pressure and the following parameters: 30 seconds of sample treatment time, cavity vacuum pressure set to 5mbar and heater position to 2.1 mm with a temperature of 95 °C (Fig.3-6 B–D). For control purposes, urine samples were stored at -20 °C without heating treatment (condition C).

3.3.6 Urine Storage

Urine sample storage and preservation were processed in 3 different conditions, as it follows: (A) Urine proteome separation using ultrafiltration and heat stabilization followed by storage at laboratory room temperature (RT, 20 °C), (B) Heat stabilization and storage at -60 °C and (C) and for control purposes, freezing at -20 °C with no heating stabilization. When urine samples in condition C were thawed a precipitate was observed. In order to resolubilize the precipitate, urine sample pH was adjusted to approximately 8 using 1 M tris-HCl. After storage, the total protein of each urine was quantified by a Bradford protein assay.

3.3.7 Protein Digestion Using Filter Aided Sample Preparation

The preserved and stored urinary proteins were analyzed at two different time points: (i) T0, 48 hours of storage and (ii) T1, 6 months of storage. After storage, the urinary proteins were digested using the FASP method with a few modifications [19]. Briefly, FASP, was used as the methodology to first separate, preserve and then digest the proteome of the urines. The protein present in the membrane were washed with 200 µL of 8 M urea 25 mM ammonium bicarbonate (AmBic) solution to remove low-weight material and then centrifuged for 20 min at 14,000 g. Next, the proteins were reduced by adding 200 µL of 50 mM dithiothreitol in 8 M urea and 25 mM AmBic and incubated for 60 minutes at 37 °C. Then, a centrifugation of 20 minutes at 14,000 g was done, and then, the sample was alkylated during 45 min in the dark by the addition of 100 µL 50mM iodoacetamide in 8M urea and 25mM AmBic solution. Subsequently, a centrifugation of 20 minutes at 14,000 g was processed and then the sample was washed twice with 200 µL 25mM AmBic. Finally, protein digestion was initiated by the addition

of 100 μL 1:30 trypsin in 12.5 mM AmBic solution. Protein digestion proceeded overnight (approximately 14 hours) at 37 °C and the peptides were collected with a centrifugation of 20 minutes at 14,000 g. The peptides were transferred to a 500 μL microtube and were dried and stored at -20 °C.

3.3.8 Nano-LC-MS/MS

Peptides were resuspended in 100 μL of 3% (v/v) acetonitrile containing 0.1% (v/v) aqueous formic acid (FA). Then, samples were homogenized by 5 minutes on vortex followed by 10 minutes on an ultrasonic bath at 100% ultrasonic amplitude. Afterwards, samples were quantified using a pierce™ quantitative colorimetric peptide assay. Following the previous step, 4 μL containing 400 ± 6 ng of peptides were loaded onto a trap column (Acclaim PepMap100, 5 μm , 100 Å, 300 μm i.d. $\times$ 5 mm) and desalted for 5 min from 3% to 5% B (B: 90% acetonitrile 0.08% FA) at a flow rate of 15 $\mu\text{L min}^{-1}$. Then the peptides were separated using an analytical column (Acclaim™ PepMap™ 100 C18, 2 μm , 0.075 mm i.d $\times$ 150 mm) with a linear gradient at 300 nL min⁻¹ (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) acetonitrile and 0.08% (v/v) FA) 5-90 min from 5% to 35% of mobile phase B, 90-100 min linear gradient from 35% to 95% of mobile phase B, 100-110 95% B. Chromatographic separation was carried out at 35 °C. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8–32Hz), 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.

3.3.9 Protein Identification and Quantification

LC-MS/MS data were analyzed using Data Analysis 4.2 software (Bruker). Relative label-free quantification was carried out using MaxQuant software V1.6.0.16. All raw files were processed in a single run with default parameters [30,31]. Database searches were performed using Andromeda search engine with the UniProt-SwissProt Human Uniprot Proteome database as a reference and a contaminants database of common contaminants [32]. Data processing was performed using Perseus (version 1.6.2.2) with default settings [33]. Protein group LFQ intensities were log₂-transformed and the quantitative profiles were filtered for missing values with

the following settings: min valid number 2 in at least one group and values greater than 0. To overcome the obstacle of missing LFQ values, missing values were imputed using default parameter, with $\delta = 0.3$ and down shift = 1.8 [33]. Log ratios were calculated as the difference in average log₂ LFQ intensity values between the tested conditions (two-tailed, Student's t-test, FDR 0.05 and S0 of 1.5).

3.4 Results & Discussion

3.4.1 Quantitative comparison between different storage conditions at T0 (48h)

To assess the potential differences between the conditions of storage, the proteins levels of the urine samples were compared at T0 using label-free mass spectrometry-based protein quantification. The results presented in Fig.3-7 (volcano plots of the storage comparisons at T0) show that within 48 h the differences among the three conditions of preservation were minimal for the urines assessed. Red dots represent proteins with levels found statistically higher than expected between conditions (two-tailed, Student's t-test, FDR 0.05, S0 of 1.5, $n=3$). Blue dots represent proteins with levels found statistically lower than expected between conditions (two-tailed Student's t-test, FDR 0.05, S0 of 1.5, $n=3$). The number of proteins quantified was between 388 and 560, from which, less than 1% were found statistically significant (two-tailed, Student's t-test, FDR 0.05, S0 of 1.5, $n=3$), with the exception of condition A versus B in urine 1, where 20 proteins over a total of 560 were found statistically significant (two-tailed, Student's t-test, FDR 0.05, S0 of 1.5, $n=3$).

3.4.2 Quantitative comparison between different storage conditions at T1 (6 months)

The same comparison as for T0 was done to compare the storage conditions after 6 months of preservation. Volcano plots regarding the storage comparisons at T1 (Table 3-1) show that less than 5% proteins were found statistically significant (up or down, two-tailed Student's t-test, FDR 0.05, S0 of 1.5, $n=3$) for most of the urine and conditions assessed, with the exception of condition A versus condition C in urine number 4. However, the number of proteins found different was lower than 15%.

3.4.3 Quantitative T0 versus T1 comparison for each storage condition

The variation of the levels of the common proteins with storage conditions and storage time was assessed using Pearson's correlation plots with all the technical replicates as shown in Fig.3-3 A. The Pearson's correlation coefficient was higher than 90% for the 87% of the comparisons done, thus showing the excellent preservation for a period of six months provided by each one of the different conditions tested in this work.

Once it was verified that the urine proteomes were preserved with similar performance between methods either at T0 or at T1 regardless the storage condition chosen, the variability in preservation for each method within the time period of 6 months was investigated next. Volcano plots from the urines 1, 2, 3 and 4 in Fig.3-3 B show that proteome preservation was successfully achieved within 6 months for the three storage conditions tested. Indeed, preservation conditions performed in a similar way, resulting in less than 2% statistically significant proteins (up or down, two-tailed Student's T-test, FDR 0.05, SO of 1.5, n=3) for all urines and conditions with the unique exception of condition A in urine 1 which was lower than 6%.

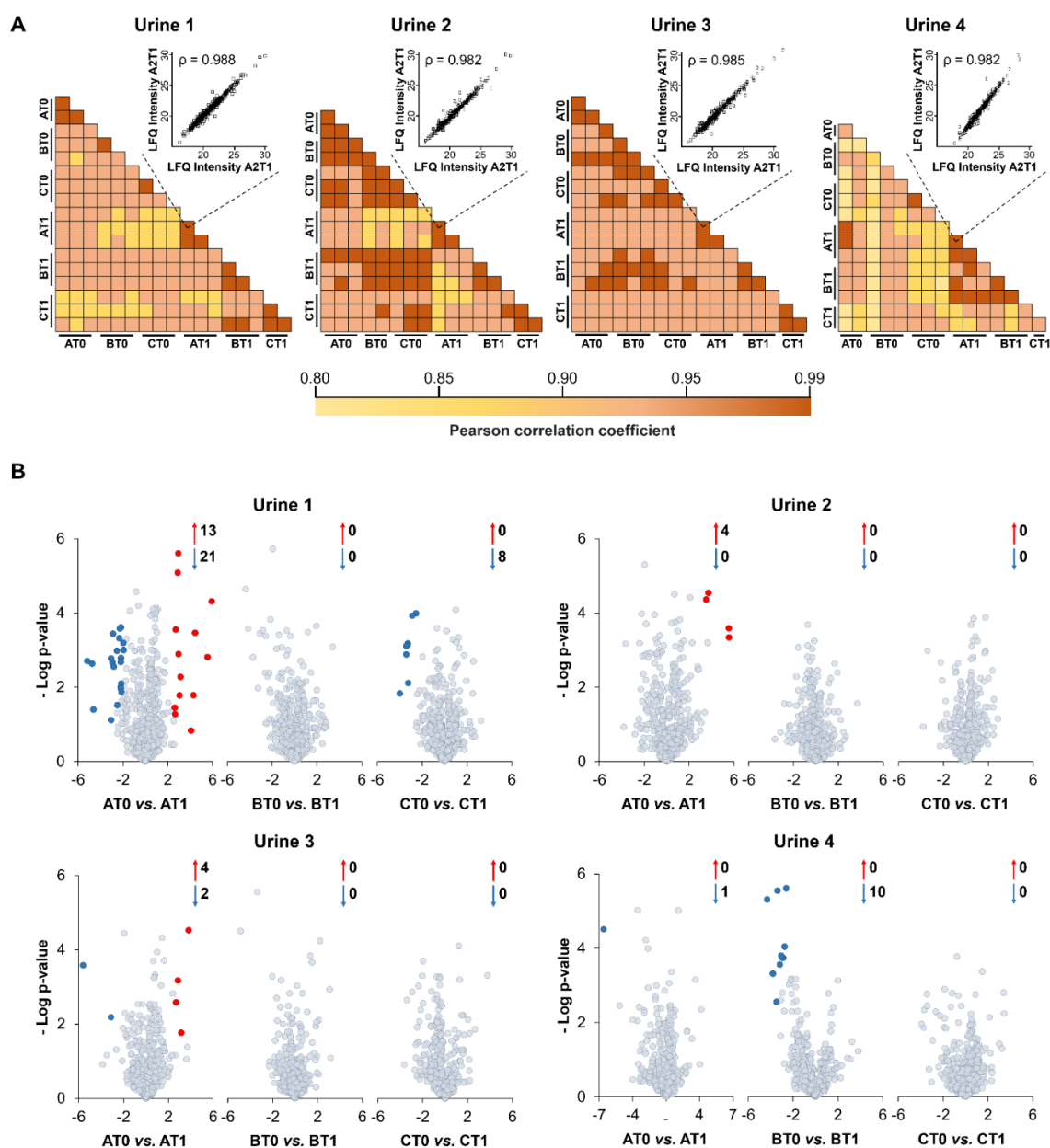


Figure 3-3. Quantative proteomic analysis for each storage condition. A. Color-coded Pearson's correlation values showing the linearity of protein levels between conditions of storage and/or time of storage. **B.** Volcano plots showing the variation in protein levels with storage condition and storage time. T0: storage time 48h. T1: storage time 6 months. Red and blue arrows indicate the statistically significant features (two-tailed Student's t-test, FDR 0.05 and SO of 1.5, $n=3$).

3.4.4 Hierarchical cluster analysis of each storage methods and time

Fig.3-4 depicts the hierarchical clusters obtained when all the LFQ data series for each urine obtained at different storage conditions and storage time are compared. The LFQ data of each preservation method for T0 and T1 was compared using an ANOVA test (Permutation-based FDR, FDR 0.05, S0 of 0 [34,35]) and the significant proteins were normalized by Z-scores. The Z-score offers information about how many standard deviations differ from the mean. Once this data for each technical replicate is selected as explained above, they are grouped by storage conditions and storage time. Then, all are compared by unsupervised clustering. The hierarchical clusters shown in Fig.3-4 suggest that data obtained with condition of storage A (Stabilizer T1 and room temperature) tend to group together whilst data obtained with condition B (Stabilizer T1 and -60 °C) and condition C (20 °C) tend to group together. This second group cannot be grouped correctly for urines 1, 2 and 3. In urine 4, conditions A and B are mixed. The aforementioned results obtained with the hierarchical cluster analysis are in agreement with the information retrieved from the volcano plots, which address that the three methods perform equally in terms of preservation.

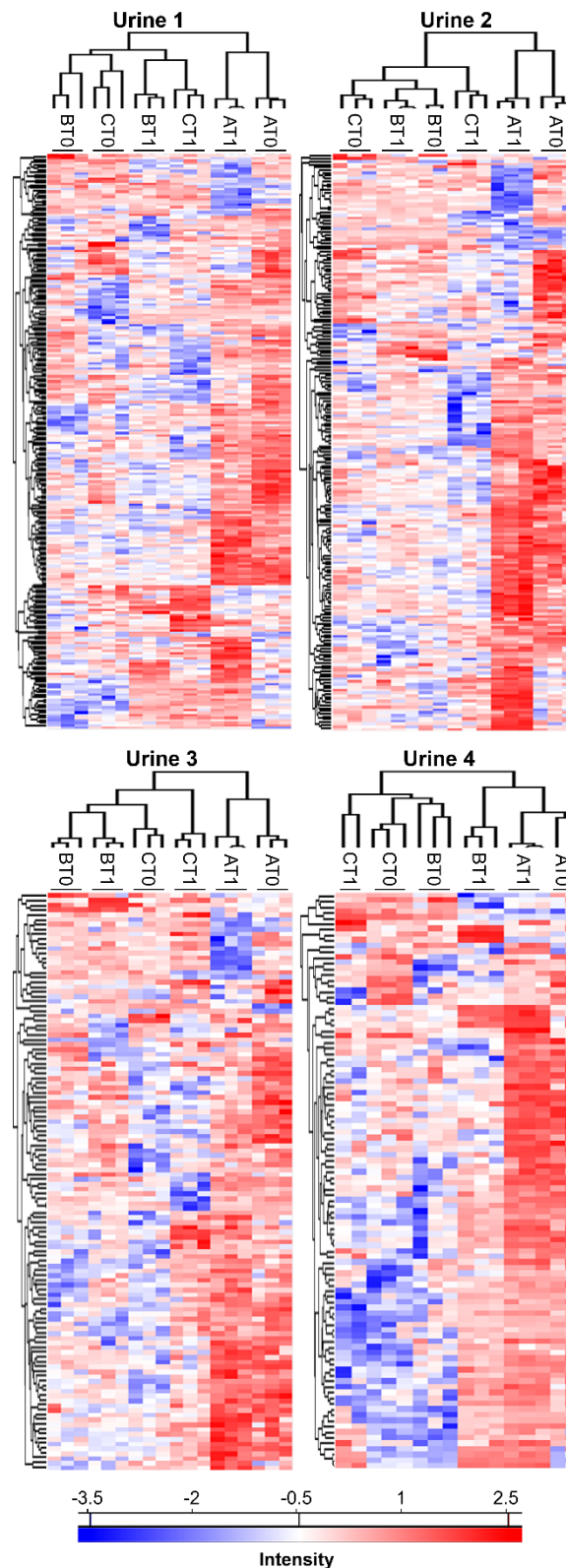


Figure 3-4. Cluster analysis presenting the variation of proteins levels between the different conditions of storage and different time of storage. The profiles of preservation for conditions A present higher similarities among them and cluster together whiles the profiles for samples preserved at low temperature, condition B and condition C, cluster together.

3.4.5 Proof-of-concept: preservation of (i) acute kidney injury urinary biomarkers; (ii) kidney tubular injury; and (iii) FDA -cleared or -approved protein assayed in serum or plasma

To further investigate the preservation conditions studied in this work we decided to check the status of different urinary biomarkers of (i) Acute Kidney Injury (AKI), namely neutrophil gelatinase-associated lipocalin and cystatin-C; (ii) kidney tubular injury, namely clustering and osteopontin; and (iii) other 20 FDA -cleared or -approved protein assayed in serum or plasma [36]. Thus, a total of 24 biomarkers was assessed. Further details about the selected biomarkers, are shown in Fig.3-5 and Table 3-1.

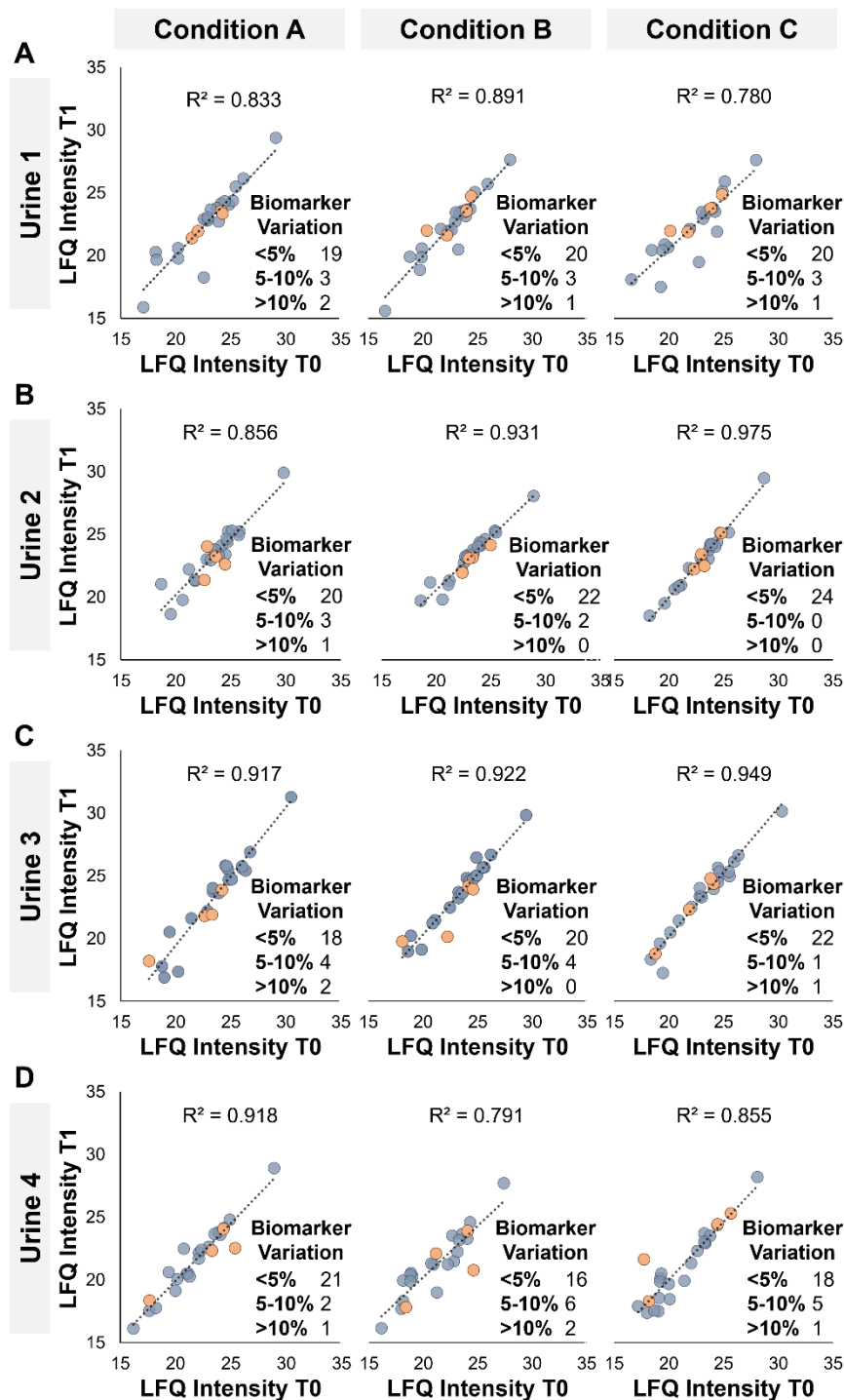


Figure 3-5. Comparison of LFQ intensities for 24 selected -cleared or -approved protein assayed in serum or plasma between T0 (48 h) and T1 (6 months) for the 3 conditions tested in this study. It can be seen that most of the biomarkers present variations lower than 5% between T0 and T1. Serum and plasma biomarkers: Albumin, α -1-acid glycoprotein, α -1-antitrypsin, α -2-HS-glycoprotein, α -2-macroglobulin, amylase, antithrombin III, apolipoprotein A1, β -2 microglobulin, biotinidase, ceruloplasmin, complement C3, fibrinogen, fibronectin, her-2/neu protein, lactoferrin, plasminogen, prostatic acid phosphatase, pyruvate kinase, thyroxine binding globulin. Urinary biomarkers: clusterin, cystatin C, neutrophil gelatinase-associated lipocalin, osteopontin.

Over the last decade, different biomarkers have been studied and identified to diagnose AKI disease [37], and thus it can be found in literature the effect of urine storage at -70 °C on the preservation of different AKI biomarkers. As an example, Liu et al. [38], have found in a study about biomarker preservation at -70 °C that there was no significant impact of storage time over a median of 18 months on urine KIM-1, NGAL, IL-18 or L-FABP biomarkers in hospitalized samples; but a statistically significant effect towards a decrease over time was noted for KIM-1 in outpatient samples. In our case, the number of biomarkers with LFQ variation lower than 5% between the analysis done after 48 h and after 6 months (considering the three conditions evaluated) was around 20 ± 2 ($n = 24$, range 16 - 24) as depicted in Fig.3-5. The number of biomarkers with a variation comprised between 5% and 10% was 3 ± 2 ($n = 24$, range 0 - 6) and the number of biomarkers with a variation between 10% and 20% was 1 ± 1 ($n = 24$, range 0 - 2). The aforementioned results are a good indicative that the three conditions assessed in this study works well. Interestingly, some of the biomarkers assessed in this proof of concept are typically used in studies done with plasma samples. Furthermore, we show here that such biomarkers can be found also in urine and that they are well preserved at least for six months. As an example, for the NGAL protein our result is in agreement with the one reported by Liu et al. [38] and by van de Vrie et al. [39] who claim that the levels of this biomarker remain substantially the same at -70 °C and -80 °C, respectively, for 6 months. The variation for this biomarker was lower than 5% in all the comparisons done by us as shown in Table 3-1, thus confirming that preservation at low temperatures works well, as it also works the preservation using the snap-heat chock filter aided method. Further literature is in agreement with us, as Collins et al. [40] reported albumin preservation during 6 months at -20 °C with no significant changes; and Giensen et al. [41] reported albumin and NGAL preservation at room temperature (18 °C) and at -20 °C with the addition of boric acid with no changes for 1 week.

3.5 Conclusions

The work here presented shows clearly that the use of heat stabilization in combination with the FASP approach allows preservation and further processing of the entire urine proteome within six months without the need of storing the sample in a freezer. Such a novel approach is going to change the way urine proteome is preserved as space and conditions needed for storage are dramatically simplified. The risk of sample degradation caused by misfunction of freezers is also overcome, decreasing the cost of sample storage. Furthermore, the use of FASP approach to preconcentrate the urinary proteins simplifies storage and transport

of samples. The approach presented in this work opens new perspectives in sample preservation as it would be interesting to test for how long samples can be preserved at room temperature using heat stabilization. Finally, although this method has been applied to urine, it can be extended to any liquid biopsy such as serum and plasma. Therefore, it is expected that this study will open new avenues for other types of liquid biopsies.

3.6 Supplementary Material

3.6.1 Supplementary data

The mass spectrometry proteomics raw data have been deposited to the ProteomeXchange Consortium [42] via the PRIDE partner repository with the data set PXD012880.

3.6.2 Supplementary figures & tables

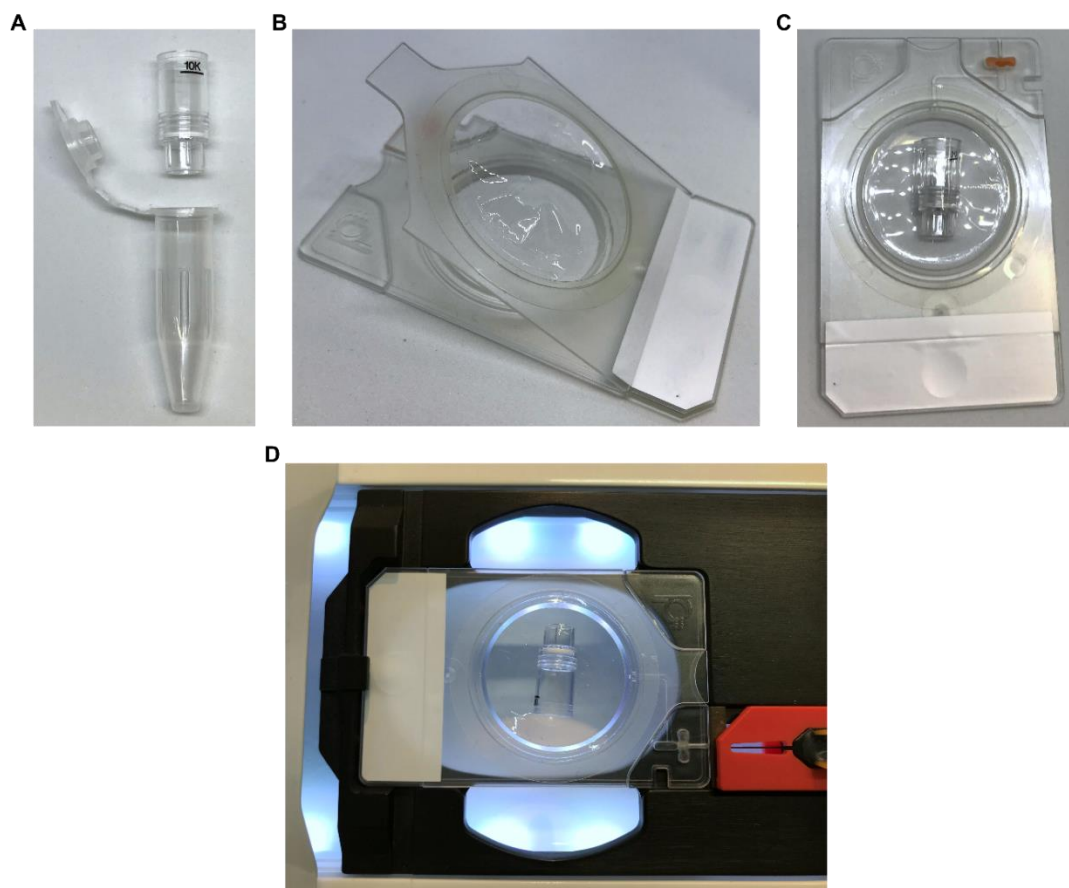


Figure 3-6. System for urinary proteome preservation and storage. (A) Vivacon 500, 10,000 MWCO used for urinary proteome precontraction. (B) Stabilizer™ T1 maintainer tissue card, (C) Stabilizer™ T1 maintainer tissue card with Vivacon 500, 10,000 MWCO containing the urinary proteome that then undergoes inside the stabilizer™ T1 for heat treatment (D).

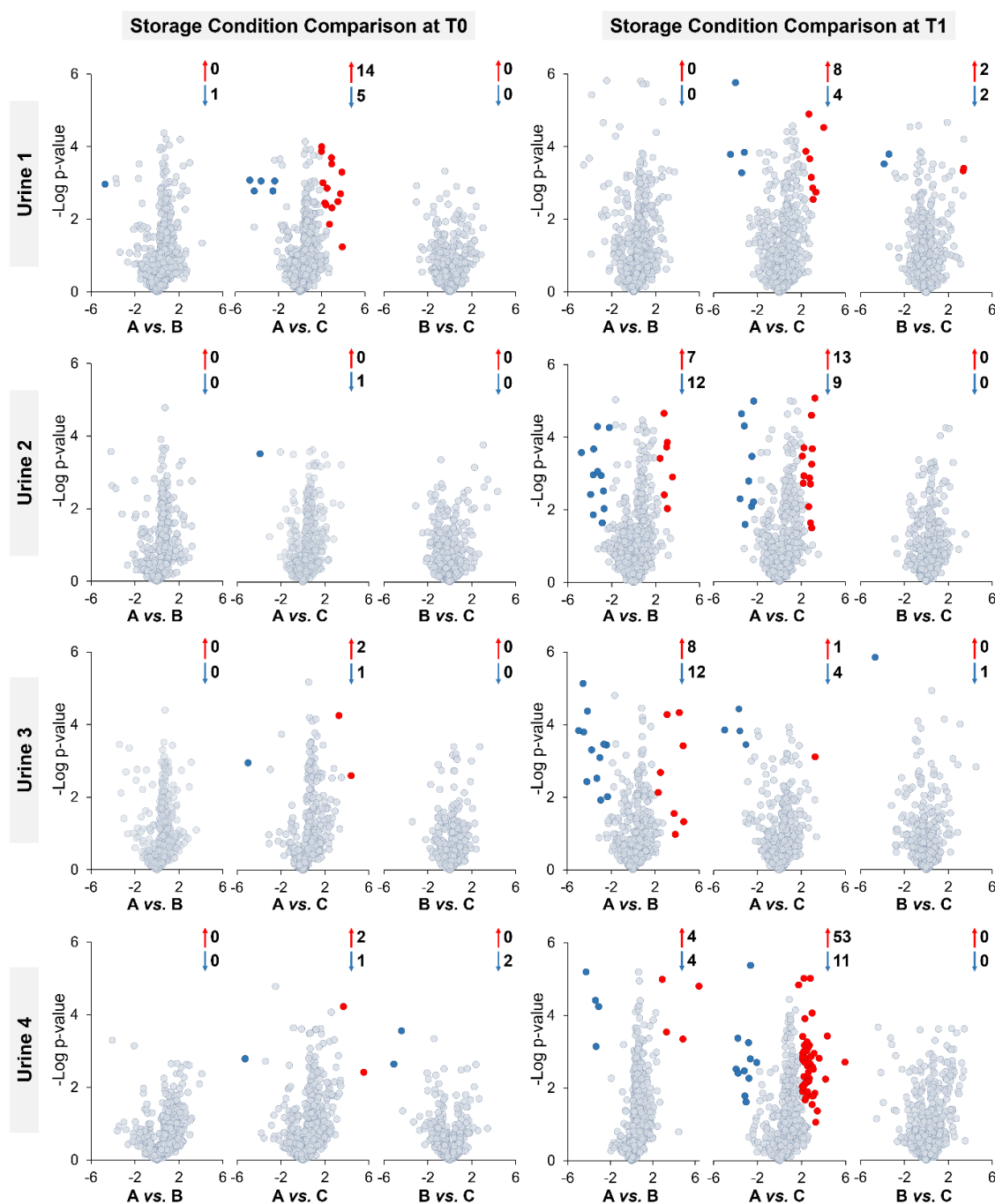


Figure 3-7. Volcano plots showing the variation in protein levels with storage condition in different times points of storage. T0: storage time 48h. T1: storage time 6 months. Red and blue arrows indicate the statistically significant features (two-tailed Student's t-test, FDR 0.05 and S0 of 1.5, n=3).

Table 3-1. LFQ variation of 24 FDA approved biomarkers in the urine preserved for 6 months. Variation of the 24 FDA approved biomarkers analysed. For all the different storage conditions accessed in the study (A, B and C), the preservation of each of the 24 biomarkers were evaluated during 6 months of storage and compared in two different time points (T0, T1). Overall, the LFQ variations (%) were above 5% for 21 of the biomarkers. Thus, confirming that the differences were minor between the storage conditions and both conditions equally preserved the urine proteome. Variation (%) was calculated using the following formula: $\frac{X_{t=0} - X_{t=6}}{Y} \times 100$, where $X_{t=0}$ and $X_{t=6}$ is the mean of 3 technical replicates for 48 h and 6 months of storage respectively. Y correspond to the mean of $X_{t=0}$ and $X_{t=6}$.

24 FDA approved biomarkers				
Type of sample	Biomarker name	LFQ variation after 6 months of storage		
		Condition A (%, n = 3)	Condition B (%, n = 3)	Condition C (%, n = 3)
Blood and serum	Albumin	1	2	1
	α -1-Acid glycoprotein	4	3	4
	α -1-Antitrypsin	3	2	2
	α -2-HS-glycoprotein	0	2	2
	α -2-Macroglobulin	3	2	4
	Amylase	1	1	1
	Antithrombin III	5	4	5
	Apolipoprotein A1	2	3	1
	β -2 Microglobulin	9	7	7
	Biotinidase	6	6	3
	Ceruloplasmin	2	1	2
	Complement C3	1	2	3
	Fibrinogen	3	1	2
	Fibronectin	1	2	1
	her-2/neu protein	1	1	1
	Lactoferrin	4	2	6
	Plasminogen	5	5	3
	Prostatic acid phosphatase	2	1	1
	Pyruvate kinase	4	4	4

	Thyroxine binding globulin	2	3	1
	Clusterin	2	1	1
	Cystatin C	3	6	8
Urine	Neutrophil gelatinase-associated lipocalin	3	4	1
	Osteopontin	7	6	2

3.7 Contributions & Acknowledgements

3.7.1 Author Contributions

J.L.C., C.L. and H.M.S. designed the experimental work and provide financial support. L.B.C. performed the laboratorial work under the supervision of H.M.S.. L.B.C., J.L.C. and H.M.S. drafted the manuscript. J.L.C., C.L.E., W.J.R. and H.M.S. revisited the drafted version, corrected it and made valuable suggestions. All authors have given approval to the final version of the manuscript.

3.7.2 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).

3.8 References

- [1] D.M. Good, V. Thongboonkerd, J. Novak, J.-L. Bascands, J.P. Schanstra, J.J. Coon, A. Dominiczak, H. Mischak, Body fluid proteomics for biomarker discovery: Lessons from the past hold the key to success in the future, *J. Proteome Res.* 6 (2007) 4549–4555. doi:10.1021/pr070529w.
- [2] J. Bergquist, M. Palmblad, M. Wetterhall, P. Håkansson, K.E. Markides, Peptide mapping of proteins in human body fluids using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Mass Spectrom. Rev.* 21 (2002) 2–15. doi:10.1002/mas.10016.

-
- [3] É. Csősz, G. Kalló, B. Márkus, E. Deák, A. Csutak, J. Tőzsér, Quantitative body fluid proteomics in medicine — A focus on minimal invasiveness, *J. Proteomics*. 153 (2017) 30–43. doi:10.1016/J.JPROT.2016.08.009.
- [4] S. Thomas, L. Hao, W.A. Ricke, L. Li, Biomarker discovery in mass spectrometry-based urinary proteomics, *Proteomics - Clin. Appl.* 10 (2016) 358–70. doi:10.1002/prca.201500102.
- [5] R. Pieper, C.L. Gatlin, A.M. McGrath, A.J. Makusky, M. Mondal, M. Seonarain, E. Field, C.R. Schatz, M.A. Estock, N. Ahmed, N.G. Anderson, S. Steiner, Characterization of the human urinary proteome: A method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots, *Proteomics*. 4 (2004) 1159–1174. doi:10.1002/pmic.200300661.
- [6] V. Thongboonkerd, P. Malasit, Renal and urinary proteomics: Current applications and challenges, *Proteomics*. 5 (2005) 1033–1042. doi:10.1002/pmic.200401012.
- [7] S. Decramer, A. Gonzalez de Peredo, B. Breuil, H. Mischak, B. Monsarrat, J.-L. Bascands, J.P. Schanstra, Urine in clinical proteomics, *Mol. Cell. Proteomics*. 7 (2008) 1850–62. doi:10.1074/mcp.R800001-MCP200.
- [8] M. Harpole, J. Davis, V. Espina, Current state of the art for enhancing urine biomarker discovery, *Expert Rev. Proteomics*. 13 (2016) 609–626. doi:10.1080/14789450.2016.1190651.
- [9] FDA, List of qualified biomarkers, (n.d.). <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/ucm535383.htm> (accessed January 10, 2019).
- [10] V. Thongboonkerd, Practical points in urinary proteomics, *J. Proteome Res.* 6 (2007) 3881–3890. doi:10.1021/PR070328S.
- [11] V. Thongboonkerd, P. Saetun, Bacterial overgrowth affects urinary proteome analysis: Recommendation for centrifugation, temperature, duration, and the use of preservatives during sample collection, *J. Proteome Res.* 6 (2007) 4173–4181. doi:10.1021/PR070311+.
- [12] P. Havanapan, V. Thongboonkerd, Are protease inhibitors required for gel-based proteomics of kidney and urine?, *J. Proteome Res.* 8 (2009) 3109–3117. doi:10.1021/pr900015q.
- [13] T. Yamamoto, R.G. Langham, P. Ronco, M.A. Knepper, V. Thongboonkerd, Towards standard protocols and guidelines for urine proteomics: A report on the human kidney and urine proteome project (HKUPP) symposium and workshop - 6 October 2007, Seoul, Korea and 1 November 2007, San Francisco, CA, USA, *Proteomics*. 8 (2008) 2156–2159. doi:10.1002/pmic.200800138.
-

- [14] A. Vlahou, J. Schanstra, J. Frokiaer, M. El Nahas, G. Spasovski, H. Mischak, B. Domon, G. Allmaier, E. Bongcam-Rudloff, T. Attwood, Establishment of a european network for urine and kidney proteomics, *J. Proteomics*. 71 (2008) 490–492. doi:10.1016/J.JPROT.2008.06.009.
- [15] HKUPP, Standard protocol for urine collection and storage, (n.d.). [http://www.hkupp.org/Urine collection Documents.htm](http://www.hkupp.org/Urine%20collection%20Documents.htm) (accessed January 7, 2019).
- [16] C.E. Thomas, W. Sexton, K. Benson, R. Sutphen, J. Koomen, Urine collection and processing for protein biomarker discovery and quantification, *Cancer Epidemiol. Biomarkers Prev.* 19 (2010) 953–959. doi:10.1158/1055-9965.EPI-10-0069.
- [17] D.H. Jackson, R.E. Banks, Banking of clinical samples for proteomic biomarker studies: A consideration of logistical issues with a focus on pre-analytical variation, *Proteomics - Clin. Appl.* 4 (2010) 250–270. doi:10.1002/prca.200900220.
- [18] Y. Zhang, Y. Luo, H. Lu, N. Wang, Y. Shen, R. Chen, P. Fang, H. Yu, C. Wang, W. Jia, Effect of freeze/thaw cycles on several biomarkers in urine from patients with kidney disease, *Bio-preserv. Biobank*. 13 (2015) 144–146. doi:10.1089/bio.2014.0033.
- [19] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat. Methods*. 6 (2009) 359–362. doi:10.1038/nmeth.1322.
- [20] J.R. Wiśniewski, Quantitative evaluation of filter aided sample preparation (FASP) and multienzyme digestion FASP protocols, *Anal. Chem.* 88 (2016) 5438–5443. doi:10.1021/acs.analchem.6b00859.
- [21] J.R. Wiśniewski, D.F. Zielinska, M. Mann, Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method, *Anal. Biochem.* 410 (2011) 307–309. doi:10.1016/J.AB.2010.12.004.
- [22] J.R. Wiśniewski, P. Ostasiewicz, M. Mann, High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers, *J. Proteome Res.* 10 (2011) 3040–3049. doi:10.1021/pr200019m.
- [23] A.H. Bryk, J.R. Wiśniewski, Quantitative analysis of human red blood cell proteome, *J. Proteome Res.* 16 (2017) 2752–2761. doi:10.1021/acs.jproteome.7b00025.
- [24] M. Zhao, M. Li, Y. Yang, Z. Guo, Y. Sun, C. Shao, M. Li, W. Sun, Y. Gao, A comprehensive analysis and annotation of human normal urinary proteome, *Sci. Rep.* 7 (2017) 3024. doi:10.1038/s41598-017-03226-6.
- [25] Y. Yu, M.-J. Suh, P. Sikorski, K. Kwon, K.E. Nelson, R. Pieper, Urine sample preparation in 96-well filter plates for quantitative clinical proteomics, *Anal. Chem.* 86 (2014) 5470–5477. doi:10.1021/ac5008317.

- [26] L. Jia, X. Liu, L. Liu, M. Li, Y. Gao, Urinemem, a membrane that can store urinary proteins simply and economically, makes the large-scale storage of clinical samples possible, *Sci. China Life Sci.* 57 (2014) 336–339. doi:10.1007/s11427-013-4582-1.
- [27] J.R. Jesus, H.M. Santos, H. López-Fernández, C. Lodeiro, M.A.Z. Arruda, J.L. Capelo, Ultrasonic-based membrane aided sample preparation of urine proteomes, *Talanta*. 178 (2018) 864–869. doi:10.1016/J.TALANTA.2017.09.078.
- [28] C.B. Rountree, C.A. Van Kirk, H. You, W. Ding, H. Dang, H.D. VanGuilder, W.M. Freeman, Clinical application for the preservation of phospho-proteins through in-situ tissue stabilization, *Proteome Sci.* 8 (2010) 61. doi:10.1186/1477-5956-8-61.
- [29] D. Blessborn, K. Sköld, D. Zeeberg, K. Kaewkhao, O. Sköld, M. Ahnoff, Heat stabilization of blood spot samples for determination of metabolically unstable drug compounds, *Bioanalysis*. 5 (2013) 31–9. doi:10.4155/bio.12.294.
- [30] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372. doi:10.1038/nbt.1511.
- [31] S. Tyanova, T. Temu, A. Carlson, P. Sinitcyn, M. Mann, J. Cox, Visualization of LC-MS/MS proteomics data in MaxQuant, *Proteomics*. 15 (2015) 1453–1456. doi:10.1002/pmic.201400449.
- [32] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J. V. Olsen, M. Mann, Andromeda: A peptide search engine integrated into the MaxQuant environment, *J. Proteome Res.* 10 (2011) 1794–1805. doi:10.1021/pr101065j.
- [33] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data, *Nat. Methods*. 13 (2016) 731–740. doi:10.1038/nmeth.3901.
- [34] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, *Proc. Natl. Acad. Sci.* 98 (2001) 5116–5121. doi:10.1073/pnas.091062498.
- [35] S. Tyanova, J. Cox, Perseus: A bioinformatics platform for integrative analysis of proteomics data in cancer research, in: *Methods Mol. Biol.*, 1711th ed., Humana Press, New York, NY, 2018: pp. 133–148. doi:10.1007/978-1-4939-7493-1_7.
- [36] N.L. Anderson, The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum, *Clin. Chem.* 56 (2010) 177–185. doi:10.1373/clinchem.2009.126706.
- [37] J. Vanmassenhove, R. Vanholder, E. Nagler, W. Van Biesen, Urinary and serum biomarkers for the diagnosis of acute kidney injury: An in-depth review of the literature, *Nephrol. Dial. Transplant.* 28 (2013) 254–273. doi:10.1093/ndt/gfs380.

- [38] K.D. Liu, E.D. Siew, W.B. Reeves, J. Himmelfarb, A.S. Go, C. Hsu, M.R. Bennett, P. Devarajan, T.A. Ikizler, J.S. Kaufman, P.L. Kimmel, V.M. Chinchilli, C.R. Parikh, Storage time and urine biomarker levels in the ASSESS-AKI study, *PLoS One*. 11 (2016) e0164832. doi:10.1371/journal.pone.0164832.
- [39] M. van de Vrie, J.K. Deegens, J. van der Vlag, L.B. Hilbrands, Effect of long-term storage of urine samples on measurement of kidney injury molecule 1 (KIM-1) and neutrophil gelatinase-associated lipocalin (NGAL), *Am. J. Kidney Dis.* 63 (2014) 573–576. doi:10.1053/J.AJKD.2013.10.010.
- [40] A.C. Collins, M. Sethi, F.A. MacDonald, D. Brown, G.C. Viberti, Storage temperature and differing methods of sample preparation in the measurement of urinary albumin, *Diabetologia*. 36 (1993) 993–7. <http://www.ncbi.nlm.nih.gov/pubmed/8243884>.
- [41] C. Giesen, J.C. Lieske, The influence of processing and storage conditions on renal protein biomarkers, *Clin. J. Am. Soc. Nephrol.* 11 (2016) 1726–1728. doi:10.2215/CJN.08800816.
- [42] J.A. Vizcaíno, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J.A. Dienes, Z. Sun, T. Farrah, N. Bandeira, P.-A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.-J. Kraus, J.P. Albar, S. Martinez-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination, *Nat. Biotechnol.* 32 (2014) 223–226. doi:10.1038/nbt.2839.

ULTRASONIC-BASED FILTER AIDED SAMPLE PREPARATION AS THE GENERAL METHOD TO SAMPLE PREPARATION IN PROTEOMICS

Luís B. Carvalho, José Luis Capelo Martínez, Carlos Lodeiro, Jacek R. Wiśniewski, Hugo M. Santos

Analytical Chemistry 2020, 92, 9164–9171

Doi: 10.1021/acs.analchem.0c01470

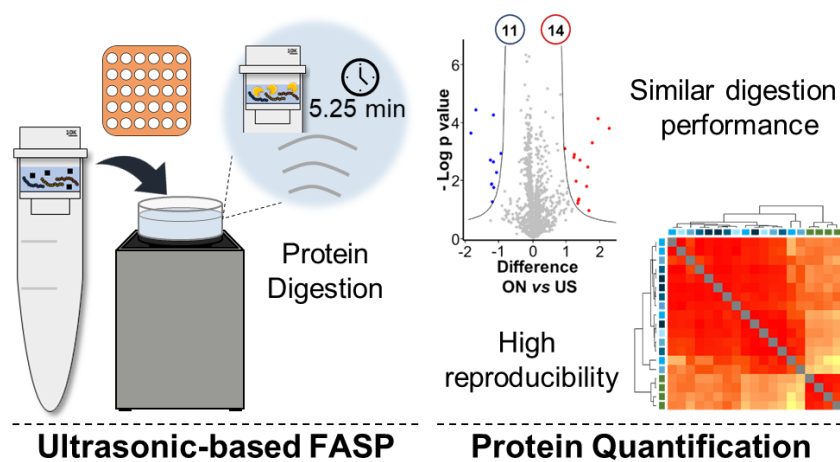


Figure 4-1. Graphical abstract.

4.1 Abstract

We propose a new high-throughput ultrafast method for large scale proteomics approaches by speeding the classic filter aided sample preparation protocol, FASP, from overnight to a 2.5 h. Thirty-six samples can be treated in 2.5 hours, and the method is scalable to 96-well plate-based pipelines. After a modification of the FASP-tube, the steps of protein reduction, protein alkylation, and protein digestion of complex proteomes are done in just 5.25 min each one under the effects of an ultrasonic field (7 cycles: 30s on and 15s off). The new method was compared with the standard overnight digestion FASP protocol, and no statistical differences were found for more than 92.4%, 92%, and 93,3% of the proteins identified studying the proteome of *E. coli*, mouse brain, and mouse liver tissue samples, respectively. Furthermore, the successful relative label-free quantification of four spiked proteins in *E. coli* samples, namely BSA, β -lactoglobulin, α -casein, and α -lactalbumin, was achieved, either using the ultrasonic-based FASP protocol or the classic overnight one. The new US-FASP method matches the analytical minimalism rules as time, cost, sample requirement, reagent consumption, energy requirements and production of waste products are reduced to a minimum while maintaining high sample throughput in a robust manner as all the advantages of the filter aided sample preparation protocol are maintained.

4.2 Introduction

Mass spectrometry-based bottom-up protein quantification of complex proteomes relies on extensively handling and time-consuming approaches, all of them share several steps, including proteome extraction and solubilization, fractionation of the proteome, to reduce complexity, reduction and alkylation of protein disulfide bonds, and proteome digestion using enzymes [1].

The protein digestion step is the most time-consuming one in the pipeline for proteome-wide quantification, as the standard protocols often rely on overnight incubations lasting up to 18 h [1,2]. Therefore, attempts to reduce such time-consuming steps are reported in the literature. Thus, it has been demonstrated that the enzymatic digestion of proteins can be accelerated through an array of different tools, including microwave energy, infrared energy, heating, electrical fields, ultrasonic energy and immobilized enzymes [2,3]. In this context, ultrasonic energy as a way to speed digestion of complex proteomes for protein identification from overnight to some minutes was first reported in 2005 by Ferrer et al.[4]. Subsequent

developments of this methodology showed that ultrasonic energy can also accelerate from tens of minutes to some few the reduction and alkylation of proteins' disulfide bonds [5–11]. Furthermore, it has been shown recently that ultrasonic energy can also be successfully applied in the format of a 96-well plate, thus making it possible to digest 96 samples in just 4 min[12]. As the methodology was applied for protein identification, it remains unknown if such an approach can also be applied for protein quantification by mass spectrometry.

Filter aided sample preparation (FASP) is a frequently used sample treatment method in proteomics [13,14]. The method uses ultrafiltration membrane for protein handling and isolation of clean peptide fraction. In brief, in-solution complex protein mixtures are separated from other components using an ultrafiltration device where a membrane, with a certain cut-off, is used to retain the proteome content, while allowing the removal of smaller contaminant molecules and interfering buffers. Then, using such membranes as a solid support, proteins are first reduced, then alkylated, and finally digested. Furthermore, there is an array of membranes types and cut-offs making this method extremely versatile for proteomics. The FASP protocol offers a convenient way to handle the different steps of complicated workflows for proteome-wide quantification.

In the present work, ultrasonic energy is introduced as a tool to speed sample preparation in the standard FASP pipeline, and this new approach is named US-FASP. The ultrafiltration membrane in FASP acts as a proteomic reactor for detergent removal and buffer exchange, while ultrasonic energy is used to speed chemical reactions and protein digestion. Overall, the treatment time is reduced from overnight to 2.5 h, allowing 36 samples to be processed simultaneously, facilitating high throughput sample preparation.

4.3 Experimental Section

4.3.1 Reagents

All reagents used were HPLC grade or electrophoresis grade. Bovine serum albumin (BSA), α -lactalbumin from bovine milk, β -lactoglobulin from bovine milk, α -casein from bovine milk, iodoacetamide (IAA), urea, phosphate buffered saline (PBS), formic acid (FA), tris base and Bradford reagent were purchased from Sigma-Aldrich (Basel, Switzerland). Ammonium bicarbonate (AmBic) and hydrochloric acid solution 37% (HCl) were purchased from Fluka (Basel, Switzerland). Sodium dodecyl sulfate (SDS), acetonitrile (ACN) and boric acid were purchased from Panreac (Barcelona, Spain). 1,4-Dithio-DL-threitol (DTT) was purchased from Alfa Aesar

(Haverhill, Massachusetts, USA). Pierce™ trypsin protease MS grade, trifluoroacetic acid (TFA) and pierce™ quantitative colorimetric peptide assay were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA).

4.3.2 Apparatus

Ultrasonic processor Branson SLPe digital sonifier with a probe of 1 mm was utilized for tissue homogenization. The membrane vivacon 500 10,000 and 30,000 MWCO Hydrosart from Sartorius (Goettingen, Germany) and prism™ refrigerated microcentrifuge from Labnet (New Jersey, USA) were used in the standard FASP and ultrasound FASP methodology. A mini incubator from Labnet (New Jersey, USA) was used for protein reduction and digestion in the standard FASP method. Microplate Horn Assembly equipped with a water recirculation system and operating with the Q700 system (20 kHz) from QSonica (Newtown, CT, USA) was used to accelerate the steps of reduction, alkylation and digestion in the ultrasound FASP. A custom-built wood plate vivacon support (10.5 cm x 9.9 cm x 9.9 cm) was used in the ultrasound FASP steps. A vacuum concentrator centrifuge model UNIVAPO 150 ECH Speed Vac and a vacuum pump model UNIJET II (Munich, Germany) were used for sample drying. Vortex models ELMi CM70M-09 from SkyLine (Southern California, USA) and VX-200 Vortex Mixer from Labnet (New Jersey, USA) were used in sample mix. CLARIOstar® High Performance Monochromator Multimode from BMG LABTECH (Germany) was used for Bradford and total peptide assays. Acquisition of mass spectrometry data was done using an UHR-QqTOF IMPACT HD from Bruker Daltonics (Bremen, Germany). Chromatographic separation of peptides were carried out using an Ultimate 3000 nLC nano-system equipped with a trap-column Acclaim PepMap100, 5 µm, 100 Å, 300 µm i.d. x 5 mm (Thermo Fisher Scientific) and an analytical column Acclaim™ PepMap™ 100 C18, 2µm, 0.075mm i.d x 150mm (Thermo Fisher Scientific).

4.3.3 E. coli proteome

E. coli samples were prepared using the ReadyPrep™ E. coli lysate sample from BioRad, following manufactures' instructions with minor modifications. Briefly, 500 µL of 3 M Urea 25 mM ammonium bicarbonate (AmBic) buffer solution were added to the E. coli extracts and incubated for five min at room temperature, following by five min sonication using an ultrasonic bath at 100% ultrasonic amplitude, UA, 35 kHz ultrasonic frequency, UF, to ensure that the dried E. coli protein extracts were solubilized. Next, several aliquots of 30 µL were prepared, followed by storage at - 60 °C until further use. Furthermore, in some experiments to access

the different methods reproducibility, five different known concentrations of bovine serum albumin (BSA), β -lactoglobulin, α -casein, and α -lactalbumin were spiked to the *E. coli* proteome.

4.3.4 Mouse liver and brain proteomes

Fresh frozen mouse liver and brain were obtained from Patricell Limited. Tissue samples were weight and washed with cold 0.1 M phosphate buffer solution (PBS) 4 °C to remove traces of blood. Next, the tissues were homogenized in a lysis buffer 0.1 M Tris-HCl, pH 7.8, 0.05 M DTT 2% SDS (w/v) using a Potter-Elvehjem tissue grinder. The ratio of tissue mass to lysis buffer was 1:10 (g/mL). The homogenates were then sonicated for 2 min on ice (30 s on 10 s off ultrasonic time, UT, 50% UA, 40 kHz UF, 1 mm diameter probe) using a Branson SLPe digital sonifier, followed by incubation at 95 °C in a water bath for five min. After cooling to room temperature, tissue lysates were clarified by centrifugation at 16 000g for 10 min. As final step the supernatant was transferred into aliquots of 1 mL and stored at -60 °C until further use.

4.3.5 Total protein and peptide quantification

The total protein present in the *E. coli*, mouse brain and mouse liver were quantified by Bradford protein assay. After digestion, peptides were quantified using the Thermo Fisher Scientific pierce™ quantitative colorimetric peptide assay following manufacturer's instructions.

4.3.6 Filter aided sample preparation (FASP)

The standard overnight digestion FASP, ON-FASP, depicted in Fig.4-2 A, was used to process the *E. coli*, mouse brain and liver proteome with a few minor modifications [12–15]. A total of 100 μ g of *E. coli* protein was loaded in a membrane with a 10,000 molecular weight cut-off (MWCO). The same amount of mouse brain and liver protein were loaded in membrane with a 30,000 molecular weight cut-off (MWCO). The proteins present in the membrane were washed with 200 μ L of 8 M urea 25 mM AmBic solution and then centrifuged for 20 min at 14,000 x g. *E. coli* samples were reduced by adding 200 μ L of 50 mM dithiothreitol (DTT) in 8 M urea and 25 mM AmBic and incubated for 60 minutes at 37 °C. Mouse brain and liver proteomes skipped the step of reduction as the lysis buffer had this already incorporated. Then, a centrifugation of 20 minutes at 14,000 x g was done, and the sample was alkylated during 45 min in the dark by the addition of 100 μ L 50 mM iodoacetamide (IAA) in 8 M urea and 25 mM AmBic solution. Subsequently, a centrifugation of 20 minutes at 14,000 x g was processed and then the sample was washed twice with 200 μ L 25 mM AmBic. Finally, protein

digestion was initiated by the addition of 100 μL 1:30 trypsin in 12.5 mM AmBic solution. Protein digestion was proceeded overnight (approximately 16 hours) at 37 °C and then the peptides were collected by 20 min centrifugation at 14,000 x g. To ensure that all the peptides were extracted, 100 μL of 3% (v/v) acetonitrile and 0.1% (v/v) formic acid were added followed by a centrifugation of 20 minutes at 14,000 x g. This previous step was repeated one more time and then, the peptides were transferred to a 500 μL microtube, dried and stored at -20 °C until further peptide quantification (SI 3) and analysis by Nano-LC-MS/MS (SI4).

4.3.7 Ultrasonic-based filter aided sample preparation (US-FASP)

The US-FASP depicted in Fig.4-2 A was used to process the protein *E. coli*, mouse brain, and mouse liver proteome and compared to the standard FASP described above. The same workflow, including cleaning steps, as described before, was used, however, ultrasound energy was applied to speed up the reduction, alkylation, and digestion steps. Briefly, the proteins were reduced by adding 200 μL of 50 mM dithiothreitol (DTT) prepared in 8 M urea and 25 mM AmBic. Tissue samples skipped the step of reduction as this was already done during the tissue homogenization using the lysis buffer. Then, it was applied ultrasound energy using the ultrasonic microplate horn assembly for 5.25 min (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). Afterward, centrifugation for 20 minutes at 14,000 x g was done, following by protein alkylation by the addition of 100 μL 50 mM iodoacetamide (IAA) prepared in 8 M urea and 25 mM AmBic solution. The alkylation step was speeded using the ultrasonic microplate horn assembly during 5.25 min (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). Finally, 100 μL of 1:30 trypsin in 12.5 mM AmBic solution was added and the protein digestion was processed using the ultrasonic microplate horn assembly for 5.25 min (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). Peptides were dried and stored at -20 °C until further peptide quantification (SI 3) and analysis by Nano-LC-MS/MS (SI 4).

4.3.8 Nano-LC-MS/MS

LC-MS/MS analysis was carried out using a Ultimate 3000 nano LC system coupled to an Impact HD (Bruker Daltonics) with a CaptiveSpray nanoBooster using acetonitrile as dopant. Peptides were resuspended in 100 μL of 3% (v/v) acetonitrile containing 0.1% (v/v) aqueous formic acid (FA). Then, samples were homogenized by 5 min on vortex followed by 10 min on an ultrasonic bath at 100% UA, 35 kHz UF. Afterwards, samples were quantified using a pierce™ quantitative colorimetric peptide assay. Following the previous step, 5 μL containing 750 ± 5

ng of peptides were loaded onto a trap column (Acclaim PepMap100, 5 μm , 100 $\text{\AA}$, 300 μm i.d. $\times$ 5mm) and desalted for 5 min from 3% to 5% B (B: 90% acetonitrile 0.08% FA) at a flow rate of 15 $\mu\text{L}\cdot\text{min}^{-1}$. Then the peptides were separated using an analytical column (Acclaim™ Pep-Map™ 100C18, 2 μm , 0.075mm i.d $\times$ 150 mm) with a linear gradient at 300 $\text{nL}\cdot\text{min}^{-1}$ (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) acetonitrile and 0.08% (v/v) FA) 5-90 min from 5% to 35% of mobile phase B, 90-100min linear gradient from 35% to 95% of mobile phase B, 100-110 95% B. Chromatographic separation was carried out at 35 °C. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8-32 Hz), cycle time 3.0 s, 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 was 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.

4.3.9 Label-free protein quantification and statistical analysis

Relative label-free quantification was carried out using MaxQuant software V1.6.0.16. All raw files were processed in a single run with default parameters [16,17]. Database searches were performed using Andromeda search engine with the UniProt - Proteome_UP000000625 which contains proteins from *E. coli* and the sequences from the spiked proteins that were added to the fasta file [18]. Data processing was performed using Perseus (version 1.6.5.0) with default settings [19]. Protein group LFQ intensities were log₂-transformed and the quantitative profiles were filtered for missing values with the following settings: min valid percentage of 70% in at least one group and values greater than 0. To overcome the obstacle of missing LFQ values, missing values were imputed using the parameters, with = 0.5 and down shift = 1.8 [19]. Log ratios were calculated as the difference in average log₂ LFQ intensity values between the tested conditions (two-tailed, Student's t-test, permutation-based FDR 0.05 and S0 of 0.1).

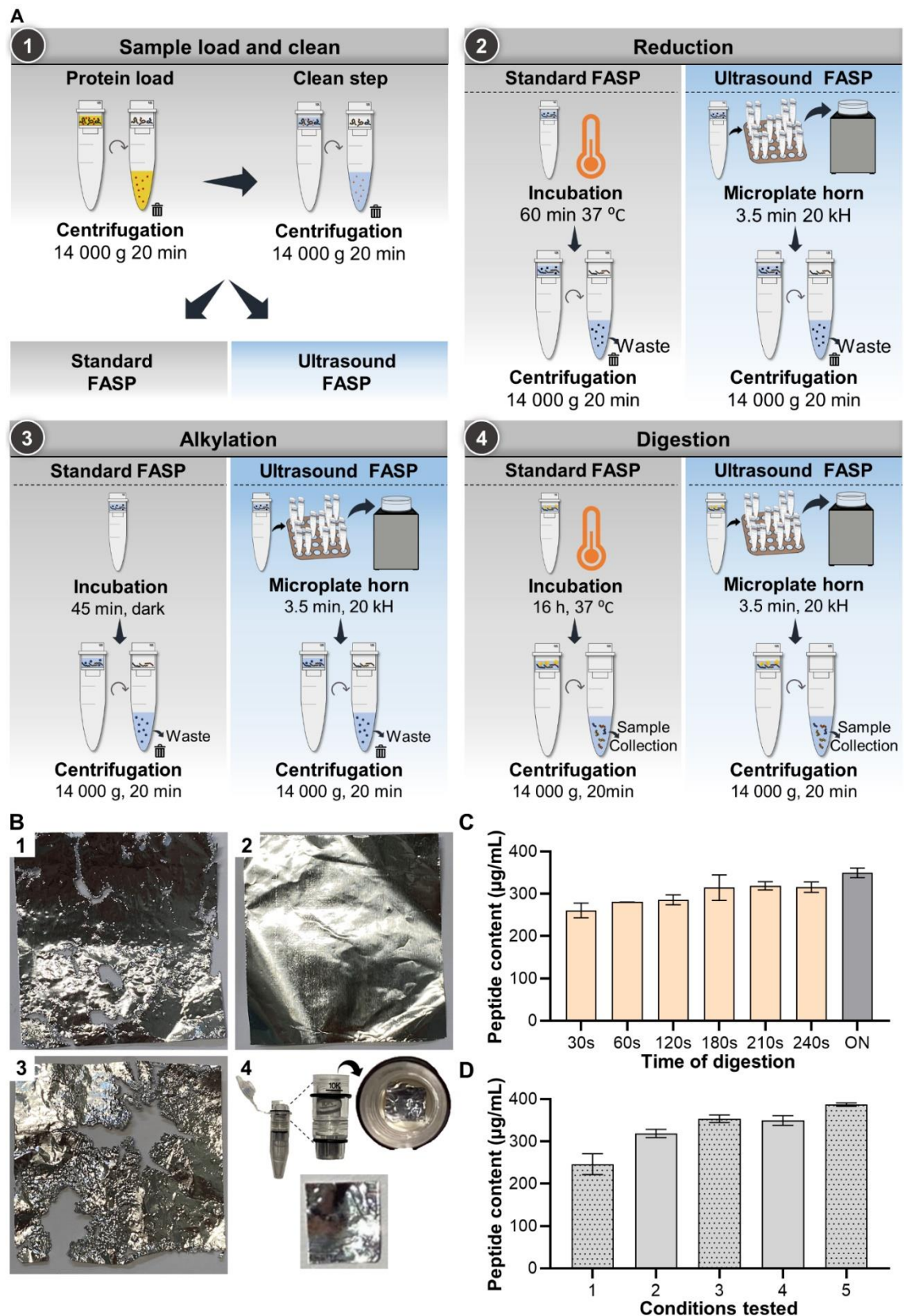


Figure 4-2. Ultrasonic-based filter aided sample preparation workflow and optimization. A. Different methods, standard overnight digestion FASP (ON-FASP) and ultrasound FASP (US-FASP), was used to process *E. coli*, mouse brain and liver proteomes. Ultrasound energy was used to speed up the reduction, alkylation and digestion FASP

steps. **B.** Effect of ultrasound energy in 5 cm x 5 cm aluminum foils. 1. 30 s ultrasonic time, UT, in an ultrasound bath at 100% ultrasonic amplitude, UA, and 35 kHz ultrasonic frequency, UF. 2. 30 s UT in an ultrasound bath at 100% UA, and 130 kHz UF. 3. 30 s UT in an ultrasonic microplate horn assembly at 25% UA, and 20 kHz UF. 4. 5.25 min UT in an ultrasonic microplate horn assembly at 25% UA, and 20 kHz UF (0.5 cm x 0.5 cm aluminum foil inside the FASP tube). **C.** Digestion time optimization. Different total elapsed ultrasound digestion times (30 s, 60 s, 180 s, 210 s, 240 s) were accessed using the ultrasonic microplate horn assembly at 25% UA, and 20 kHz UF and compared with the ON-FASP. **D.** Ultrasound in the reduction, alkylation and digestion steps. Bars with gray color with and without dots, represent different batches of *E. coli*. Different conditions using the ultrasonic microplate horn assembly were done (1. no reduction and alkylation step, use of ultrasound energy in the digestion; 2. standard alkylation step and use of ultrasound energy in the reduction and digestion step; 3. use of ultrasound in the reduction, alkylation and digestion step) and compared with the ON-FASP (4, 5).

4.4 Results & Discussion

4.4.1 Initial considerations about variables affecting ultrasonic performance

Previous research performed in our laboratory has recently shown that the ultrasonic microplate horn assembly is the ultimate ultrasonic tool to speed complex proteomics pipelines as 96 samples can be digested in four minutes, and volumes as low as 10 µl can be handled [12]. Thus, the analytical minimalism rules, as explained by Halls are accomplished [12,20]. We have also shown that high intensity focused ultrasonic energy has two effects over enzymatic kinetics, first speeding the kinetics and then, after 5 min, inactivating the enzyme [21]. Therefore, ultrasonic energy was not applied in this work for longer than 5 min. We have observed that ultrasonic amplitudes higher than 25% led to the rapid deterioration of the microplate horn assembly surface, but we demonstrated that for in-solution digestion of complex proteomes, such as *E. coli* lysates and tissue extracts, a 25% ultrasonic amplitude is adequate to speed the overnight proteome digestion process up to 4 min (intervals of 30 s on and 15 s off). Consequently, the ultrasonic amplitude chosen was 25% [12].

4.4.2 Effect of Ultrasonic frequency over the FASP membrane

The implosion of ultrasonic nano-bubbles generated by high-intensity ultrasound produced at frequencies between 20 kHz and 40 kHz creates shock waves with disruptive effects [22–24]. However, at higher frequencies, this disruptive effect is less intense. Fig.4-2 B1 and Fig.4-2 B2 show two aluminum foils (5 x 5 cm) that were exposed to an ultrasonic field generated in an ultrasonic bath. Fig.4-2 B1 clearly shows the disruptive effects over the aluminum

foil of ultrasound generated at a frequency of 35 kHz whilst Fig.4-2 B2 shows that at a frequency of 130 kHz there is no disruption. The ultrasonic well plate works at a frequency of 20 kHz, thus having a disruptive effect, that could potentially affect the membrane of the FASP tube. The difference in cavitation effects between the ultrasonic bath and the microplate horn is visualized in Fig.4-2 B3, where it may be seen microplate horn's disruptive effect more intense when the following conditions were used to deliver ultrasound: aluminum foil 5 x 5 cm, 30 s UT, 20 kHz UF, 25% UA. Therefore, the first task was to investigate the integrity of the FASP membrane. To this end, the effect of ultrasound energy over the FASP membrane, was tested by introducing an aluminum foil in the FASP tube just over the membrane as depicted in Fig.4-2 B4 (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). It was found that the microtube walls have a protective effect, diminishing the disruptive effects of ultrasound, as no disruption was observed nor in the aluminum foil neither in the FASP membrane (Fig.4-6 and Fig.4-7). This fact was later confirmed by the results obtained for protein identification and quantification as shown below. Moreover, no membrane polymers were observed in the mass spectra. However, during the first experiments with ultrasound, it was observed water inside the FASP tubes after some tens of seconds of ultrasonication. To avoid this effect, it was necessary to develop a small modification on the FASP tube design consisting in the addition of two O-rings, as video 1-SI shows. Finally, once it was verified that (i) no degradation of the FASP membranes was observed under the effects of an ultrasonic field and that (ii) no water entered inside the FASP tube during ultrasonication, the next step was to assess the acceleration of the FASP protocol using ultrasonic energy.

4.4.3 Effect of time of ultrasonication in the FASP-based digestion protocol

The workflow of complex proteomes can be divided into three main steps. Reduction, to break disulfide bonds; alkylation to inactivate sulfur groups thus hampering disulfide bonds of forming again and finally protein cleavage using trypsin, also known as protein digestion step. The latter is by far the step that consumes more time as it is generally done overnight (18 h). Therefore, our first efforts were oriented to shorten this step. To this end, a series of *E. coli* samples were digested overnight following the standard FASP. Other series of the same *E. coli* were also digested under the effects of an ultrasonic field during different times, comprising the range 0.5 to 4 min (reduction and alkylation were done with no US applied for both sets of samples. If the success of the digestion is measured as the concentration of peptides obtained

with the overnight digestion protocol, Fig.4-2 C shows that after 3 min. of digestion under the effects of an ultrasonic field, the peptide content of the ultrasonically digested sample is $91 \pm 8\%$ of the sample digested overnight. Therefore, an ultrasonication time of 5.25 min was settled as optimum for the digestion protocol (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF).

4.4.4 Effect of ultrasonication in reduction and alkylation steps

Once it was verified that the overnight digestion process in the FASP membrane can be accelerated to 5.25 minutes, it was investigated if the alkylation and digestion processes could be also shortened. To this end, a series of experiments were designed in which *E. coli* extracts were submitted to reduction, alkylation and digestion, with and without the effects of an ultrasonic field (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). First, it was investigated whether the process could be done applying ultrasound only in the digestion step, skipping alkylation and reduction steps. The results of this experiment are shown in Fig.4-2 D1, and it may be seen that 60% of the expected peptide concentration is obtained. Next, we apply ultrasound only in the reduction and digestion steps, skipping the application of ultrasound in the alkylation process. However, results were not satisfactory as only 80% of the expected peptides were obtained (Fig.4-2 D2). Finally, when all the steps, reduction, alkylation and digestion, were done under the effects of an ultrasonic field the concentration of peptides was more than 90% of the expected amount (Fig.4-2 D3). Thus, the set of experiments described above showed that ultrasonic energy conveniently speed each step of the proteomic workflow (Fig.4-2 D4 and Fig.4-2 D5).

Finally, in further experiments, and as best conditions, for each step assessed (reduction, alkylation and digestion), 5.25 min of ultrasonication time was applied in periods of 30 s followed by resting periods of 15 s seconds (total time of 5.25 min UT, with a 25% UA and at an UF of 20 kHz).

4.4.5 Mass spectrometry label free-based protein quantification: ultrasonic-based FASP versus classic overnight FASP

Fig.4-3 shows the comparison of ultrasonic-based FASP protocol, US-FASP, versus classic overnight FASP, ON-FASP, protocol using *E. coli* samples and label free-based mass spectrometry (n=2 different batches for each protocol run in two different days, with two technical replicates each batch, resulting in a total of four samples for each method). The two conditions

tested resulted in 11653 unique peptides and 1053 proteins, Fig.4-3 A. Of the unique peptides identified 8583 (73.7%) were found common for both methods, from which (i) 7022 peptides (81.8%) had no missed cleavages, (ii) 1432 peptides (16.7%) had one missed cleavage and (iii) 125 unique peptides (1.5%) had two missed cleavages. Of the 1053 identified proteins, 995 (94.5%) were found common for both methods, Fig.4-3 A. The second step was to quantify differences in digestion efficiency across the detected proteome. These differences are shown in Fig.4-3 B. This volcano plot shows that only 25 proteins are differentially digested, representing 2% of the detected proteome (two-tailed Student's t-test, FDR 0.05 and S0 of 0.1, n=8). This result shows the similarity of both proteomic pipelines (Fig.4-8). Furthermore, Fig.4-3 C shows that the majority of the 2% of proteins found with different digestion efficiencies are of a molecular weight below 50 kDa. Interestingly, when these 25 proteins are used to generate a hierarchical clustering, Fig.4-3 D, it becomes evident that such proteins are preferentially digested either by the US-FASP or ON-FASP. This observation is in agreement with results reported previously⁹, [12].

The reproducibility for each method was assessed as the difference between two batches of the same sample done in different days, and it is shown in Fig.4-3 E and Fig.4-3 F. These volcano plots show that the interdaily variability of both methods is negligible (two-tailed Student's t-test, FDR 0.05 and S0 of 0.1, n=4 each method). When the 1053 protein LFQ values obtained after finishing each experimental pipeline are compared using the median-based normalization, the distribution is centered around 1, as shown in Fig.4-3 G, further reflecting the few differences between both US- and ON-sample treatments.

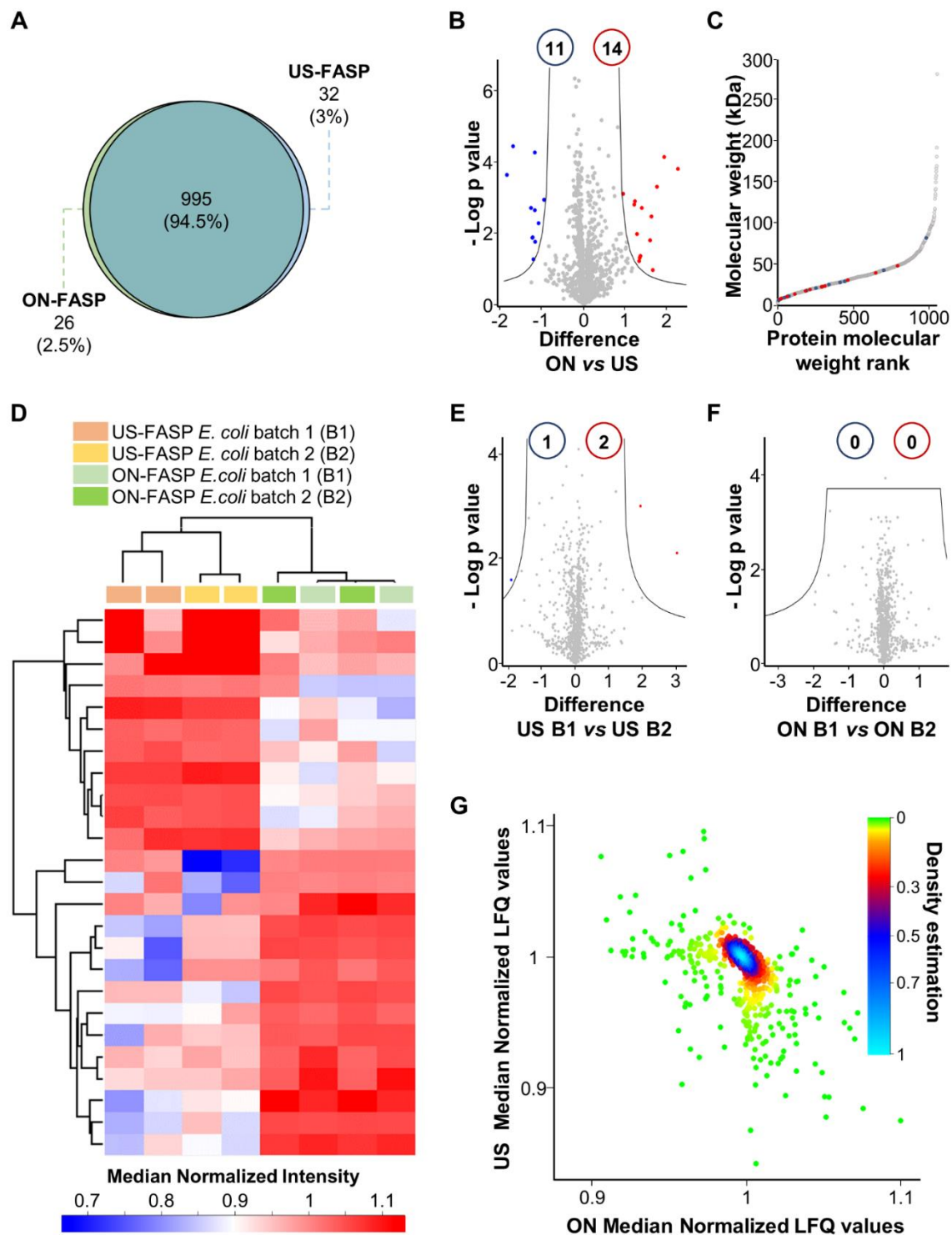


Figure 4-3. *E. coli* proteome digestion using the US-FASP and ON-FASP. **A.** Venn diagram of the identified peptides and proteins in the US-FASP and ON-FASP. Different proteins identified in each method with a filter of 70% in at least one group. **B.** Volcano plot showing the variation in protein levels in the comparison ON-FASP vs. US-FASP. Proteins were normalized by the median of each protein and a two-tailed Student's t-test (FDR 0.05 and S0 of 0.1, $n=8$) was used for statistical analysis. Blue and red color dots represent the statistically significant proteins (numbers inside the colored circles) while the grey dots represent the non-statistically significant proteins, respectively. **C.** Protein molecular weight rank in the methods of digestion ON-FASP and US-FASP. **D.** Hierarchical clustering with

average linkage using the statically significant proteins (two-tailed Student's t-test, FDR 0.05 and SO of 0.1, n=8). **E&F.** Batch effect in each method. Protein levels were normalized by the median of each protein and statistically significant changes were detected using a two-tailed Student's t-test (FDR 0.05 and SO of 0.1, n=4). **G.** Correlation plot and density from the comparison of ON-FASP vs. US-FASP. Correlation plot of the proteins with median-based normalization shows a distribution centered around 1, reflecting a lower variation between the tested methods.

4.4.6 Performance of Ultrasonic-based FASP

In order to verify the homogeneity of the ultrasonic effects inside the microplate horn, *E. coli* samples were distributed randomly and were spiked at different levels with four proteins, as depicted in Fig.4-4 A and 4-4 B. Then the samples were overnight or ultrasonically digested as depicted in Fig.4-4 C. When the complete set of *E. coli* proteins of each sample is identified and quantified and then used to cluster the samples, the reproducibility as expressed by the Pearson correlation value is higher than 0.985 (Fig.4-4 D), showing that the digestion of the *E. coli* proteomes is done with the same efficiency in all the positions and thus concluding that the effects of the ultrasound energy are homogeneously distributed in the entire microplate horn. Moreover, when the ultrasonic treated samples are compared with the overnight treated ones (Fig.4-2 D), the Pearson correlations are higher than 0.945, again indicating that for a few proteins both methods perform slightly different but for the majority both perform equally.

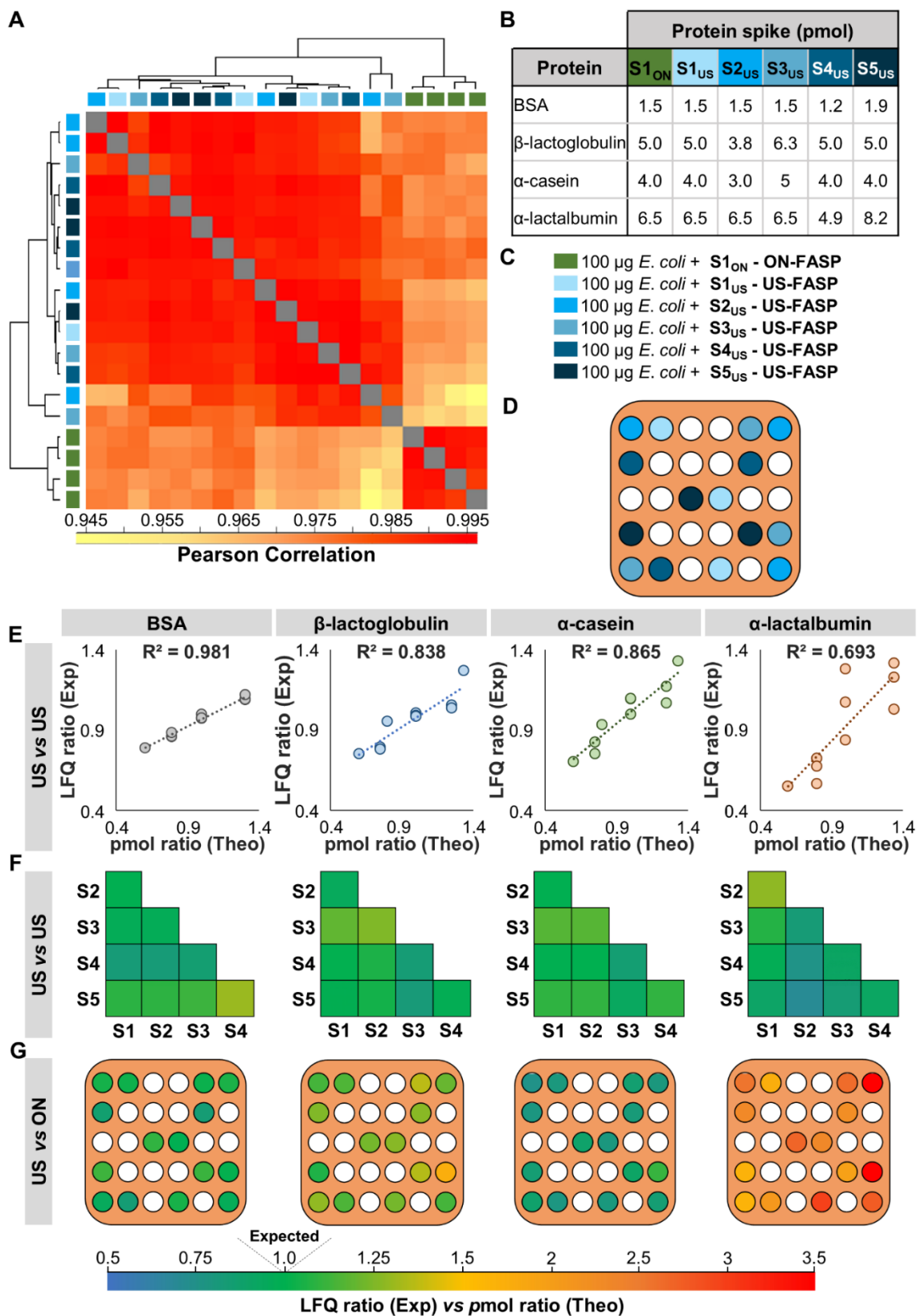


Figure 4-4. Ultrasonic space effectiveness in US-FASP. **A.** Protein spikes and *E. coli* used in each condition. **B.** Position of the different samples in the ultrasonication plate. **C.** *E. coli* proteome and spiked proteins processed with ON-FASP and US-FASP. **D.** Cluster by Pearson correlation of the *E. coli* samples with the different protein spikes using

the ON-FASP and US-FASP. **E, F & G.** Comparison between the theoretical pmol ratio (Theo) and the LFQ values obtained ration (Exp). The protein spikes were compared as followed: **E & F.** US-FASP vs. US-FASP and **G.** US-FASP vs. ON-FASP. The difference was obtained using the following formulas (i) to (v).

4.4.7 Ultrasonic-based FASP Label-free quantification of a set of spiked proteins in *E. coli* samples

In order to get a deeper insight into the potentialities of the proposed US-FASP method, *E. coli* samples were spiked with four known proteins, namely BSA, β -lactoglobulin, α -casein and α -lactalbumin. The known protein amount ratios (Fig.4-4 A) between samples were compared with their respective mass spectrometer signal ratios (LFQs ratio) using the formulas:

$$pmol\ ratio\ (Theo) = \frac{Spike_{sx}}{Spike_{sy}} \quad (i)$$

$$LFQ\ ratio\ (Exp) = \frac{\bar{X}\ LFQ_{sx}}{\bar{X}\ LFQ_{sy}} \quad (ii)$$

Where x and y can be samples S1, S2, S3 S4 or S5, where $Spike_{sx}$ and $Spike_{sy}$ refer to the spikes added to the samples as they are presented in Fig.4-4 A, and where LFQ_{sx} and LFQ_{sy} refer to the respective mass spectrometry data of samples x and y . The quantification results are depicted in Fig.4-4 E, where it may be seen that for proteins BSA, β -lactoglobulin and α -casein the spiked ratios match close to the LFQ ratios, what it is reflected in the regression coefficients close to 1. For the case of α -lactalbumin, the match is worst, with a regression coefficient of circa 0.7.

In an ideal situation, where the proteins are efficiently digested and where the LFQs are proportional to the protein concentration, the ratio using the formula:

$$Ratio = \frac{LFQ\ ratio\ (Exp)}{pmol\ ratio\ (Theo)} \quad (iii)$$

should be 1. When this ratio is calculated for the different positions inside the microplate horn, as shown in Fig.4-4 F it can be seen that the effects of ultrasonic energy are homogeneously distributed all over the surface, being the best results obtained for BSA and the worst for α -lactalbumin. Thus, for BSA protein the differences between the pmol ratios and the LFQs ratios were up to 6% ($n=10$ range 1% to 13.5%); for the β -lactoglobulin were 20% ($n=10$, range 15% to 22%), and for α -casein were 15% ($n=10$. Range 6% to 19%). For α -lactalbumin the results were worst, and the difference was 56% ($n=10$, range 35% to 91%).

4.4.8 Ultrasonic-based FASP versus classic FASP Label-free quantification of a set of spiked proteins in *E. coli* samples

The final insight was done comparing the values of the US-FASP versus ON-FASP for the four spiked proteins as follows:

$$\rho mol\ ratio\ (Theo) = \frac{Spike_{S1\ ON}}{Spike_{Sx\ US}} \quad (iv)$$

$$LFQ\ ratio\ (Exp) = \frac{\bar{X}\ LFQ_{S1\ ON}}{LFQ_{Sx\ US}} \quad (v)$$

Where $Spike_{Sx\ US}$ can be ultrasonic treated samples S1, S2, S3, S4 or S5, where the $Spike_{S1\ ON}$ refers to the sample treated overnight (see Fig.4-4 C). Where $LFQ_{Sx\ US}$ refers to the mass spectrometry data of ultrasonic treated samples S1, S2, S3 S4 or S5, and where $LFQ_{S1\ ON}$ refers to the mass spectrometry data obtained for the sample treated overnight.

Then the ratio, using the formula (iii), was calculated. This ratio should ideally be 1. It can be seen in Fig. 4-4 G that proteins BSA, β -lactoglobulin and α -casein are digested at the same level either using the ultrasonic or the overnight method. For the case of the α -lactalbumin, the digestion was better for the overnight method, almost 1.5 times better.

4.4.9 Application to brain and liver tissue samples

The optimum sample treatment conditions were also applied to tissue samples. To this end, brain and liver tissue samples were submitted to the protein extraction explained in the experimental section and, of each protein tissue extract, four subsamples were made. These subsamples were divided in two groups of two samples each. Samples were digested with the optimized conditions in different positions inside the microplate horn in different days as depicted in Fig.4-2. Results are shown in Fig.4-5. The first interesting fact is that no differences were found neither in the number of proteins nor in their quantification values when the analysis was done in different days (Fig.4-5 A and Fig.4-5 D). Moreover, the same data also reflects that the position where the digestion is done inside the microplate horn does not affect the digestion efficiency (Fig.4-5 B and 4-5 E). Furthermore, the number of common proteins identified for each tissue type was almost the same regardless the sample treatment used, 1134 (93.3%) for the brain tissue and 1097 (92%) for the liver tissue (Fig.4-5 C and Fig.4-5 F, respectively). In addition, the number of peptides identified was also similar for each method and each tissue type (Fig.4-5 H). Finally, the data obtained also allows a nice classification of both tissues (Fig.4-5 G).

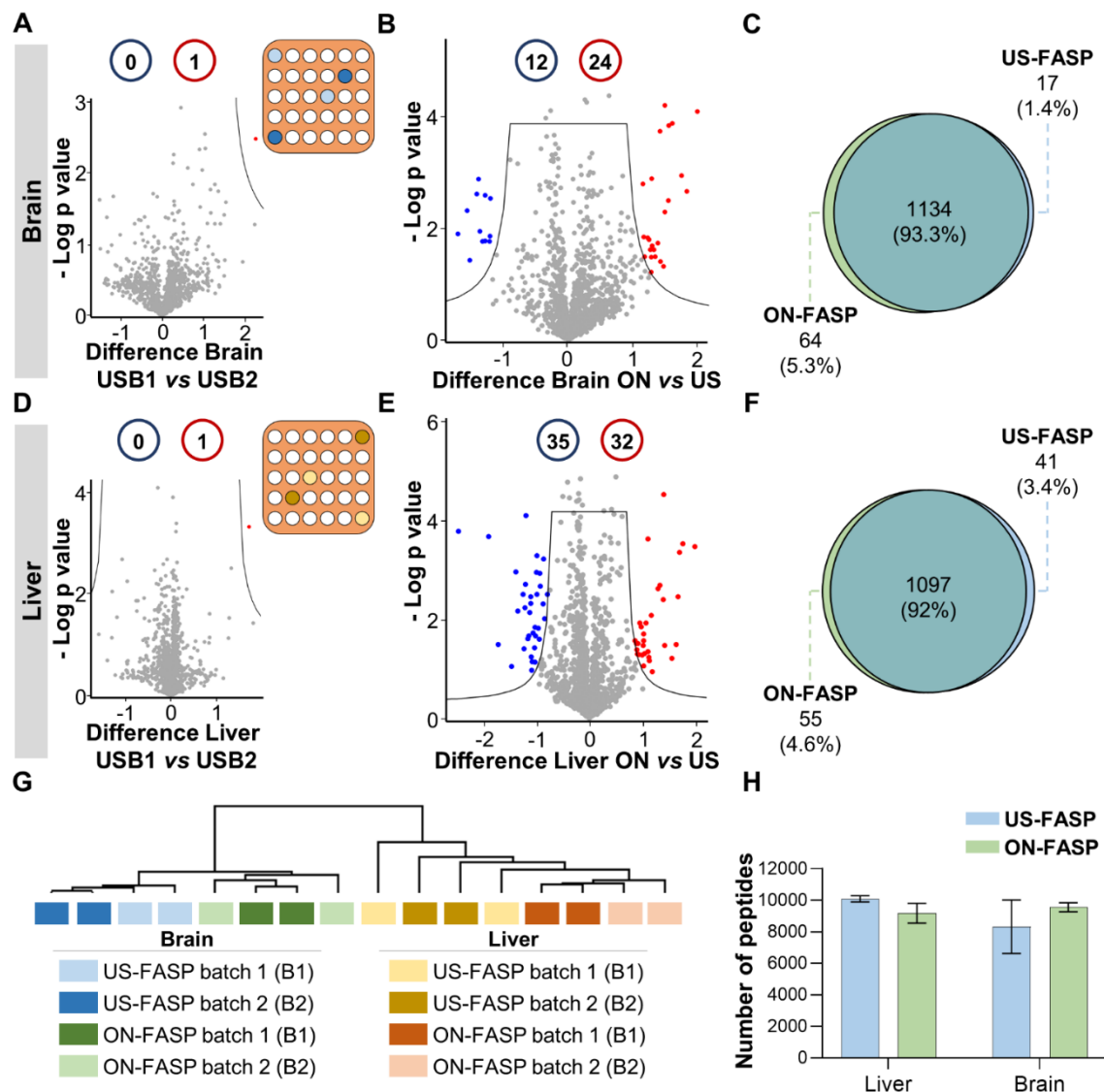


Figure 4-5. Mouse liver and brain proteome digestion using the US-FASP. **A, B & C.** Quantitative proteomics analysis using the US-FASP and ON-FASP methods in brain mouse tissues. **(D, E and F)** Quantitative proteomics analysis using the US-FASP and ON-FASP methods in liver mouse tissues. **A & D.** Batch and plate position effect in the US-FASP method. Volcano plot comparing each batch of tissue with random position in the ultrasonication plate. Protein levels were normalized by the median of each protein and the different quantified proteins were obtained using a two-tailed Student's t-test (FDR 0.05 and S0 of 0.1, $n=4$). Blue and red color dots represent the statistically significant proteins (numbers inside the colored circles) while the grey dots represent the non-statistically significant proteins, respectively. **B & E.** Protein quantitative levels comparison between the US-FASP and ON-FASP method. Volcano plot comparing the protein levels in each method. Protein levels were normalized by the median of each protein and the different quantified proteins were obtained using a two-tailed Student's t-test (FDR 0.05 and S0 of 0.1, $n=8$). **C & F.** Venn diagram with the proteins identified by US-FASP and ON-FASP method. Different proteins identified in each method with a filter of 70% presence in at least one group. **G.** Cluster of the mouse brain and liver tissues processed by ON-FASP and US-FASP. Protein identified with a filter of 70 % presence in at least one group were normalized by Z-score and a multi-sample test, ANOVA (FDR 0.05, $n=16$) were plotted in a hierarchal

clustering. **H.** Number of peptides identified in each tissue type with the different FASP methods used. ON-FASP method is represented by the green color bars and US-FASP by the blue color bars.

4.5 Conclusions

A novel ultrasonic-assisted filter aided sample preparation protocol has been developed for large scale proteomics approaches, allowing high sample throughput and reducing treatment time from 18 h to 2.5 h for a total of 36 samples. The method is scalable to 96 samples using 96-well plates, and virtually applicable to similar digestion platforms such as S-Trap [25]. We foresee this method as the universal one for proteomics pipelines as it paves the way to clinical proteomics in hospitals because it is of easy handling as the operator needs no special skills. Also, this method matches perfectly the new high-resolution mass spectrometers that allow fast chromatographic runs in 15 min, thus allowing to measure the proteome of about 100 samples a day per spectrometer. Further, shortening of the sample processing time can be achieved by skipping the cysteine alkylation step [26].

4.6 Supplementary material

4.6.1 Supplementary data

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [27] via the PRIDE [28] partner repository with the dataset identifier PXD018360.~

4.6.2 Supplementary figures

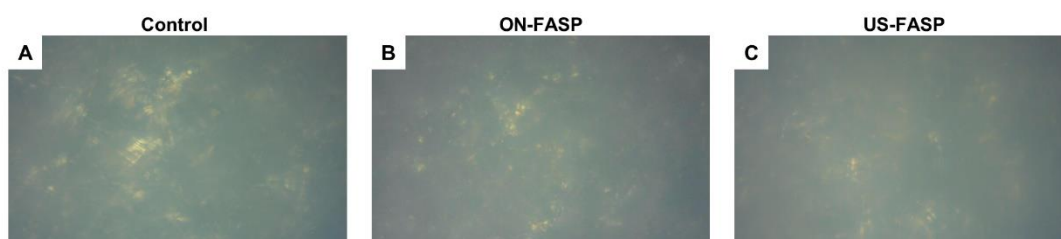


Figure 4-6. Original optical microscopy images (40x) from the FASP membranes using the conditions (A) Control, (B) overnight digestion and (C) ultrasonic digestion.

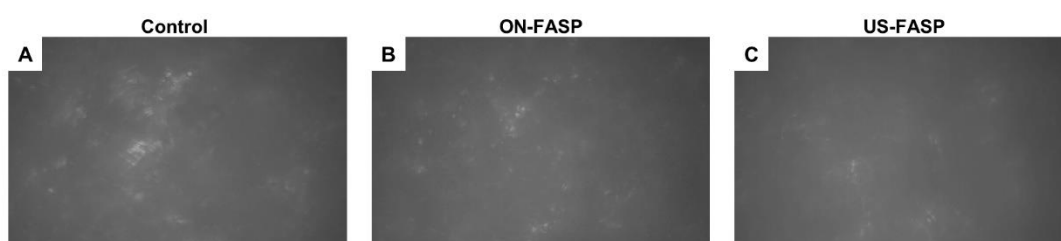


Figure 4-7. Color adjusted optical microscopy images (40x) from the FASP membranes using the conditions (A) Control, (B) overnight digestion and (C) ultrasonic digestion. The images on top are the original images from the microscope, and in the images at the bottom color has been adjusted.

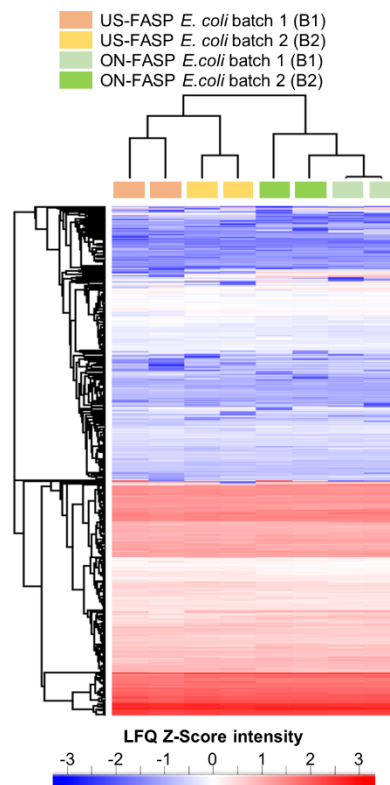


Figure 4-8. Hierarchical clustering of the *E. coli* proteome processed with US-FASP and ON-FASP. 1053 proteins identified in each method (US-FASP and ON-FASP) with a filter of 70% presence in at least one group. Protein label free intensities normalized by Z-score were plotted in a hierarchical clustering using average as linkage.

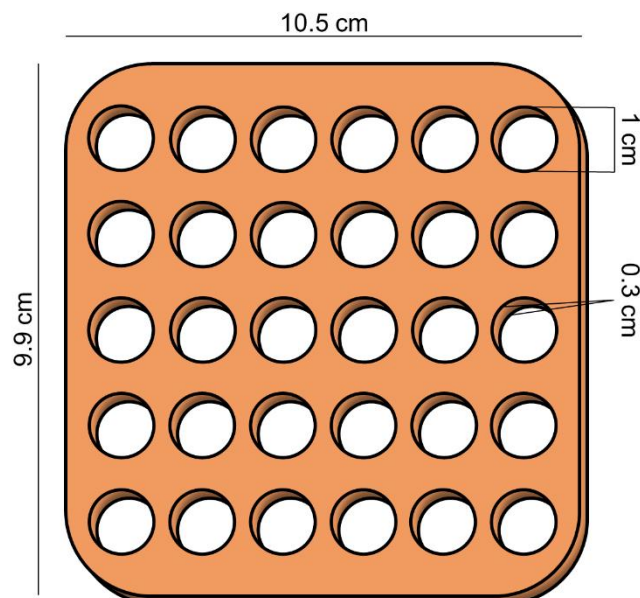


Figure 4-9. Custom-built wood plate vivacon support floater.

4.7 Contributions & Acknowledgements

4.7.1 Author Contributions

All authors have given approval to the final version of the manuscript. J.L.C., C.L. and H.M.S. designed the experimental work and provide financial support. L.B.C. performed the laboratorial work under the supervision of H.M.S. J.L.C. and H.M.S. drafted the manuscript. J.L.C., C.L.E., J.R.W. and H.M.S. revisited the drafted version, corrected it and made valuable suggestions. All authors have given approval to the final version of the manuscript.

4.7.2 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. and L. B. C. acknowledge the funding provided by FCT/MEC within the framework of FCT Investigator Program (IF/00007/2015) and the PhD grant (SFRH/BD/144222/2019), respectively. The authors acknowledge Eng. F. Silvestre from FCT FabLab (powered by USA Embassy) for his help in the manufacturing of the wood plate vivacon support floater, Fig.4-9.

4.8 References

- [1] D. López-Ferrer, B. Cañas, J. Vázquez, C. Lodeiro, R. Rial-Otero, I. Moura, J.L. Capelo, Sample treatment for protein identification by mass spectrometry-based techniques, *TrAC - Trends in Analytical Chemistry*. 25 (2006) 996–1005. <https://doi.org/10.1016/j.trac.2006.05.015>.
- [2] J.L. Capelo, R. Carreira, M. Diniz, L. Fernandes, M. Galesio, C. Lodeiro, H.M. Santos, G. Vale, Overview on modern approaches to speed up protein identification workflows relying on enzymatic cleavage and mass spectrometry-based techniques, *Anal Chim Acta*. 650 (2009) 151–159. <https://doi.org/10.1016/j.aca.2009.07.034>.
- [3] X. Shen, L. Sun, Systematic Evaluation of Immobilized Trypsin-Based Fast Protein Digestion for Deep and High-Throughput Bottom-Up Proteomics, *Proteomics*. 18 (2018). <https://doi.org/10.1002/pmic.201700432>.

-
- [4] D. López-Ferrer, J.L. Capelo, J. Vázquez, Ultra fast trypsin digestion of proteins by high intensity focused ultrasound, *J Proteome Res.* 4 (2005) 1569–1574. <https://doi.org/10.1021/pr050112v>.
- [5] R. Rial-Otero, R.J. Carreira, F.M. Cordeiro, A.J. Moro, L. Fernandes, I. Moura, J.L. Capelo, Sonoreactor-based technology for fast high-throughput proteolytic digestion of proteins, *Journal of Proteome Research.* 6 (2007) 909–912. <https://doi.org/10.1021/pr060508m>.
- [6] H.M. Santos, R. Rial-Otero, L. Fernandes, G. Vale, M.G. Rivas, I. Moura, J.L. Capelo, Improving sample treatment for in-solution protein identification by peptide mass fingerprint using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Journal of Proteome Research.* 6 (2007) 3393–3399. <https://doi.org/10.1021/pr0702518>.
- [7] M. Galesio, D. V. Vieira, R. Rial-Otero, C. Lodeiro, I. Moura, J.L. Capelo, 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, *Journal of Proteome Research.* 7 (2008) 2097–2106. <https://doi.org/10.1021/pr700850w>.
- [8] H.M. Santos, C. Mota, C. Lodeiro, I. Moura, I. Isaac, J.L. Capelo, An improved clean sonoreactor-based method for protein identification by mass spectrometry-based techniques, *Talanta.* 77 (2008) 870–875. <https://doi.org/10.1016/j.talanta.2008.07.048>.
- [9] J.R. Jesus, H.M. Santos, H. López-Fernández, C. Lodeiro, M.A.Z. Arruda, J.L. Capelo, Ultrasonic-based membrane aided sample preparation of urine proteomes, *Talanta.* 178 (2018) 864–869. <https://doi.org/10.1016/j.talanta.2017.09.078>.
- [10] G. Martins, J. Fernández-Lodeiro, J. Djafari, C. Lodeiro, J.L. Capelo, H.M. Santos, Label-free protein quantification after ultrafast digestion of complex proteomes using ultrasonic energy and immobilized-trypsin magnetic nanoparticles, *Talanta.* 196 (2019) 262–270. <https://doi.org/10.1016/j.talanta.2018.12.066>.
- [11] S. Jorge, J.L. Capelo, W. LaFramboise, R. Dhir, C. Lodeiro, H.M. Santos, Development of a robust ultrasonic-based sample treatment to unravel the proteome of OCT-embedded solid tumor biopsies, *Journal of Proteome Research.* 18 (2019) 2979–2986. <https://doi.org/10.1021/acs.jproteome.9b00248>.
- [12] S. Jorge, J.E. Araújo, F.M. Pimentel-Santos, J.C. Branco, H.M. Santos, C. Lodeiro, J.L. Capelo, Unparalleled sample treatment throughput for proteomics workflows relying on ultrasonic energy, *Talanta.* 178 (2018) 1067–1076. <https://doi.org/10.1016/j.talanta.2017.07.079>.
-

- [13] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nature Methods*. 6 (2009) 359–362. <https://doi.org/10.1038/nmeth.1322>.
- [14] J.R. Wiśniewski, Filter Aided Sample Preparation – A tutorial, *Analytica Chimica Acta*. 1090 (2019) 23–30. <https://doi.org/10.1016/j.aca.2019.08.032>.
- [15] L.B. Carvalho, J.L. Capelo-Martínez, C. Lodeiro, J.R. Wiśniewski, H.M. Santos, Snap-heated freeze-free preservation and processing of the urine proteome using the combination of stabilizer-based technology and filter aided sample preparation, *Analytica Chimica Acta*. 1076 (2019) 82–90. <https://doi.org/10.1016/j.aca.2019.05.051>.
- [16] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nature Biotechnology*. 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.
- [17] S. Tyanova, T. Temu, A. Carlson, P. Sinitcyn, M. Mann, J. Cox, Visualization of LC-MS/MS proteomics data in MaxQuant, *PROTEOMICS*. 15 (2015) 1453–1456. <https://doi.org/10.1002/pmic.201400449>.
- [18] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J. V. Olsen, M. Mann, Andromeda: A peptide search engine integrated into the MaxQuant environment, *Journal of Proteome Research*. 10 (2011) 1794–1805. <https://doi.org/10.1021/pr101065j>.
- [19] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data, *Nature Methods*. 13 (2016) 731–740. <https://doi.org/10.1038/nmeth.3901>.
- [20] D.J. Halls, Analytical minimalism applied to the determination of trace elements by atomic spectrometry. Invited lecture, *Journal of Analytical Atomic Spectrometry*. 10 (1995) 169–175. <https://doi.org/10.1039/ja9951000169>.
- [21] M. Galesio, J. Loureno, D. Madeira, M. Diniz, J.L. Capelo, Unravelling the role of ultrasonic energy in the enhancement of enzymatic kinetics, *Journal of Molecular Catalysis B: Enzymatic*. 74 (2012) 9–15. <https://doi.org/10.1016/j.molcatb.2011.08.001>.
- [22] K.S. Suslick, W.L. Nyborg, ULTRASOUND: Its chemical, physical and biological effects, *The Journal of the Acoustical Society of America*. 87 (1990) 919–920. <https://doi.org/10.1121/1.398864>.
- [23] T.J. Mason, Sonochemistry: Uses of ultrasound in chemistry and related disciplines, in: 1996: pp. 25–54. https://doi.org/10.1007/978-1-4613-1243-7_2.
- [24] J.-Luis. Capelo-Martínez, *Ultrasound in chemistry: analytical applications*, Wiley-VCH, 2009.

- [25] K.R. Ludwig, M.M. Schroll, A.B. Hummon, Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies, *J Proteome Res.* 17 (2018) 2480–2490. <https://doi.org/10.1021/acs.jproteome.8b00235>.
- [26] J.R. Wiśniewski, K. Zettl, M. Pilch, B. Rysiewicz, I. Sadok, ‘Shotgun’ proteomic analyses without alkylation of cysteine, *Analytica Chimica Acta.* 1100 (2020) 131–137. <https://doi.org/10.1016/j.aca.2019.12.007>.
- [27] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: Improving support for quantification data, *Nucleic Acids Res.* 47 (2019) D442–D450. <https://doi.org/10.1093/nar/gky1106>.
- [28] E.W. Deutsch, N. Bandeira, V. Sharma, Y. Perez-Riverol, J.J. Carver, D.J. Kundu, D. García-Seisdedos, A.F. Jarnuczak, S. Hewapathirana, B.S. Pullman, J. Wertz, Z. Sun, S. Kawano, S. Okuda, Y. Watanabe, H. Hermjakob, B. Maclean, M.J. Maccoss, Y. Zhu, Y. Ishihama, J.A. Vizcaíno, The ProteomeXchange consortium in 2020: Enabling “big data” approaches in proteomics, *Nucleic Acids Res.* 48 (2020) D1145–D1152. <https://doi.org/10.1093/nar/gkz984>.

NORMALIZATION METHODS IN MASS SPECTROMETRY-BASED ANALYTICAL PROTEOMICS: A CASE STUDY BASED ON RENAL CELL CARCINOMA DATASETS

Luis B. Carvalho, Pedro A.D. Teigas-Campos, Susana Jorge; Michele Protti, Laura Mercolini, Rajiv Dhir, Jacek R. Wiśniewski, Carlos Lodeiro, Hugo M. Santos, and José L. Capelo

Talanta 266 (2024) 124953

Doi: 10.1016/j.talanta.2023.124953

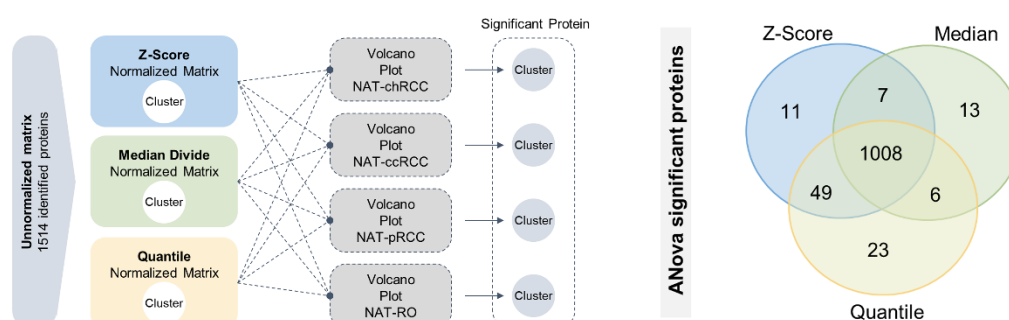


Figure 5-1. Graphical Abstract

5.1 Abstract

Normalization is a crucial step in proteomics data analysis as it enables data adjustment and enhances comparability between datasets by minimizing multiple sources of variability, such as sampling, sample handling, storage, treatment, and mass spectrometry measurements. In this study, we investigated different normalization methods, including Z-score normalization, median divide normalization, and quantile normalization, to evaluate their performance using a case study based on renal cell carcinoma datasets. Our results demonstrate that when comparing datasets by pairs, both the Z-score and quantile normalization methods consistently provide better results in terms of the number of proteins identified and quantified as well as in identifying statistically significant up or down-regulated proteins. However, when three or more datasets are compared at the same time the differences are found to be negligible.

5.2 Introduction

Some analytical chemists encounter significant challenges in analyzing large-scale proteomics data using bioinformatics tools, often because these topics are not sufficiently developed in undergraduate or post-graduate programs. As a result, analytical chemists may need to acquire additional skills and knowledge to effectively handle these types of data. From an analytical chemistry point of view, the so-called shot-gun-based proteomics workflows are characterized by multiple sources of variability, including sampling, sample handling, sample storage, sample treatment and mass spectrometry measurements [1,2]. For instance, analytical chemists know that large-scale and longitudinal studies often encounter a source of variation associated with the difference in storage time and ageing of the samples [2,3]. Samples are typically collected and prepared at different times during the study, introducing an environmental factor that may affect the data. This factor may be compounded by other variables, such as the collection of samples in different centres or countries, which may be subjected to varying numbers of freeze-thaw cycles. In addition, the sample collection protocol may be applied with different levels of attention by different operators, which can lead to unwanted differences in the samples. Furthermore, transporting the samples to the laboratory where they will be stored or analyzed can also introduce noise into the data, as the transportation conditions may vary, potentially altering the samples differently. Finally, in a typical shot-gun proteomics workflow, samples are collected and treated in the analytical laboratory before being stored or investigated using mass spectrometry. The samples are then cleaved with various

enzymes depending on the experimental strategy, and the resulting cleavage proteome composed of thousands of peptides is subjected to chromatography-mass spectrometry analysis. The noise generated during the pipeline of work, from sampling to mass spectrometry, makes shot-gun proteomics challenging to compare samples, extract meaningful information and derive accurate conclusions. The ultimate tools to overcome this challenge of variability in shot-gun proteomics workflows are the so-called normalization methods [4–6].

Normalization is a method used to adjust the data obtained from different samples, platforms, or batches to remove unwanted variability and enhance comparability between the datasets. It involves applying a mathematical correction to the data to reduce the effects of systematic technical variation and other sources of bias. The choice of normalization method depends on the nature of the data and the research question at hand [7–9]. However, this choice is not as simple as it seems at first glance. The importance of selecting an appropriate normalization method is critical to obtain well-grounded results. Therefore, selecting the optimal way to normalize data has important implications, as it significantly impacts downstream analyses, delivering different results depending on the normalization method selected [1,10,11]. Although several studies have evaluated different proteome normalization approaches, primarily focusing on their ability to enhance data repeatability, there is no consensus about which normalization method must be used, or whether the use is conditioned by the number of groups or the number of samples in each group [10,12–15].

There are several normalisation methods, but the most popular ones include Z-score normalization, median divide normalization, and quantile normalization. The z-score normalization method creates an induced normal matrix by subtracting to each protein in each sample their respective median across all samples in the dataset and then dividing each protein value by the standard deviation of the respective protein across all samples. This process forces the dataset to follow a normal distribution, which can improve the accuracy of downstream statistical analyses. By normalizing the data in this way, the z-score normalization method can address variations in the data, making it a valuable tool for data normalization in proteomic studies [10]. The median divide normalization approach involves dividing all results by the median, which reduces the impact of bias caused by equipment variations and makes the data more comparable to other assays. This method is particularly useful when dealing with large datasets that are collected over time and across different batches, as it reduces the variation in the data introduced by the technical aspects of the assay. The median divide normalization method is a reliable approach for comparing protein expression levels between different samples and provides an accurate representation of the overall distribution of the data. However, this method

may not be appropriate for datasets with extreme values or when most values are close to zero [7]. The quantile normalization involves substituting the value of each data point in a sample with the mean of the corresponding quantile, thereby modifying the data distribution of each sample to be identical. This normalization method is particularly useful for comparing samples with different distributions. By standardizing the distribution of each sample, quantile normalization can improve the accuracy and reproducibility of downstream analyses, such as differential expression analysis and pathway analysis [6,16,17].

In a previous work [18], we applied Z-Score normalization and the Total Protein Approach (TPA) to proteomic data derived from renal carcinoma solid biopsies. Thus, near adjacent tissue (NAT), Papillary Renal Cell Carcinoma (pRCC), Clear Cell Renal Carcinoma (ccRCC), Chromophobe Renal Cell Carcinoma (chRCC) and Renal Oncocytoma (RO), samples were interrogated with the main aim of obtaining new immunohistochemical markers for diagnosis and prognosis. The best biomarkers found via mass spectrometry, namely perilipin-2 (PLIN2), beta-tubulin III (TUBB3), lysosomal-associated membrane protein-1 (LAMP1) and hexokinase-1 (HK1) were later successfully validated using tissue microarrays.

In the present work, the complete mass spectrometry datasets obtained for the different types of renal carcinomas and NAT are interrogated using bioinformatic pipelines differing only in the selected normalization methods, namely Z-score, median divide and quantal normalization. Selecting these specific tumour types allows for a comprehensive analytical assessment of the normalization methods in a clinically relevant context. Through this evaluation, we aim to provide insights into the selection process of appropriate normalization techniques that can enhance the accuracy and reproducibility of proteomic analyses in tumour research. The results obtained were compared with those previously validated using tissue microarrays.

5.3 Experimental Section

5.3.1 Patients and sample collection

The experimental workflow developed to interrogate the renal carcinoma and the NAT tissue samples, along with the extended study design and patient sampling collection is described elsewhere [19], furthermore the validation of the biomarkers found using tissue microarrays is also described elsewhere [18] A brief patient summary is provided in Table 5-1.

5.3.2 Mass spectrometry raw data

This study used frozen tissue biopsies of human renal tissues from four different tumour types (ccRCC, chRCC, pRCC, and RO) and a control group (NAT). The samples were properly treated and analysed by mass spectrometry to identify and quantify the protein expression levels [19]. Mass spectrometry raw data was accessed in ProteomeXchange Consortium [20] using the Proteomics IDentications Database (PRIDE) [21] with the unique identifier PXD023296.

5.3.3 Identification and quantification of proteins

Protein identification and relative label-free quantification (LFQ) was newly performed used the previous described raw data using the precursor signal intensity method and delayed normalization, MaxLFQ, on MaxQuant software V2.0.3.0 [12]. All raw files were processed in a single run with default parameters [22,23]. Database searches were performed using peptide search engine Andromeda against the human UniProt UP000005640_9606 and UniProt UP000005640_9606_additional database[24]. Searches were configured with cysteine carbamidomethylation as a fixed modification and *N*-terminal acetylation and methionine oxidation as variable modifications. We set the false discovery rate (FDR) to 0.01 for protein and peptide levels with a minimum length of seven amino acids. The FDR was determined by searching a reverse database. Trypsin enzyme specificity was set as *C*-terminal to arginine and lysine with a maximum allowance of two missed cleavages.

5.3.4 Bioinformatics data processing

Bioinformatic data processing was conducted using Perseus V1.6.0 [25,26] as explained in Fig.5-2B. Briefly, the data processing can be roughly divided into stages: (i) Elimination of protein groups that were reversely identified or only identified by site from the list of proteins; (ii) Annotation of sample condition and technical replicates; (iii) Transformation of LFQ intensities into Log(2x) value to lessen the impact of outliers and to understand the protein changes proportional across the conditions; and finally (iv) Proteins were filtered so that they were at least present in 70% of the samples in at least one of the created groups and the remaining missing values were imputed according with the respective sample normal distribution with a width of 0.3 and down shift of 1.8.

5.3.5 Bioinformatic data analysis and normalization comparisons

Bioinformatic data analysis was carried out in Perseus V1.6.0 [25,26] as explained in Fig.5-2C in which the analytical comparison of three different types of commonly used normalizations was conducted (Z-Score normalization, Median divide normalization and Quantile normalization). Z-Score and Median divide normalization were obtained using the standard tools in Perseus while Quantile normalization was acquired by using the R package limma add-on for Perseus [27,28]. To access the impact of normalization, the NAT (control group) data set of protein expression was compared with the renal tumours using the three different types of normalization. The number of proteins whose levels were found to be statistically different between the group comparison were accessed by using a volcano plot analysis (Student's t test, FDR 0.05 and S_0 of 0.1). Finally, all the datasets were compared together for each type of normalization using a Multisample Test ANOVA with a FDR of 0.05. All comparison resulting in statistically different proteins were used respectively in a hierarchical clustering analysis with an average linkage, no constraint, pre-processing with k-means, and Euclidean distance. The proteomic data was also processed (i) by comparing the NAT group of samples against two, three or four tumour types at the same time, as shown in Fig.5-2 and (ii) by comparing each tumour condition and the control group at a time, as shown in Fig.5-7. This approach was adopted to ensure that the results would be consistent across different subtypes of renal tumour and would be useful for researchers focusing on precision medicine.

These analytical approaches allowed for a comprehensive evaluation of the normalization methods in terms of their ability to accurately identify significant proteins and their impact on downstream analyses, such as differential expression analysis and pathway analysis. Enrichment pathways analysis was carried out using the software platform Cytoscape v3.9.1 [29] and the application StringApp [30] v2.0.0 for protein-protein analysis and biochemical pathway analysis. Different types of databases for enrichment pathway analysis (GO biological process [31], The Kyoto Encyclopedia of Genes and Genomes - KEGG [32], Reactome pathways [33]) were considered to access the impact of normalization on downstream analysis.

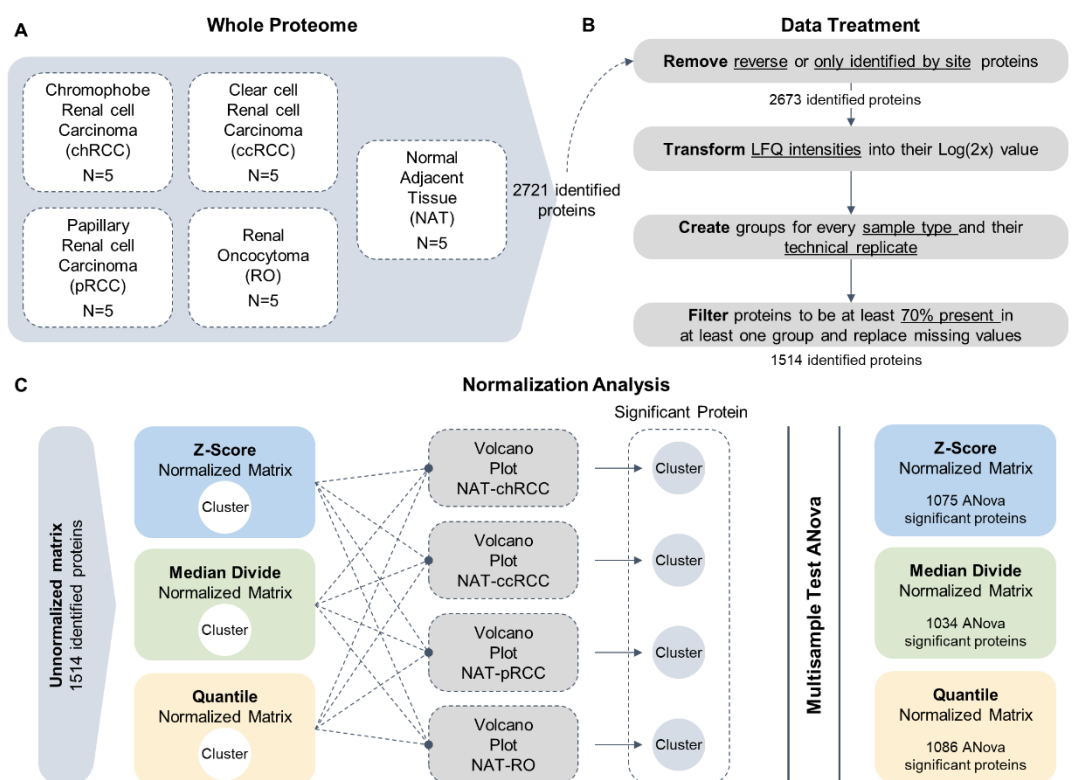


Figure 5-2. Schematic representation of the data process applied to the proteomic data when comparing Neat Adjacent Tissue, NAT (control group) against multiple renal tumour types. **A.** Type and number of tumours (description of patients is given in Table 5-1). **B.** Data treatment for the proteins quantified using the five datasets. **C.** Types of normalization made: Z-Score, median divide, Quantile and types of comparisons done: Volcano plots, Cluster analysis and Multisample Test Anova.

5.4 Results & Discussion

The data pre-processing steps were standardized across all normalization methods. A detailed workflow of the pre-processing steps is illustrated in Fig.5-2 and Fig.5-7.

5.4.1 Comparing each type of renal carcinoma against NAT

After comparing the datasets in pairs, we tested three normalization methods. Our findings revealed that both z-score normalization and quantile normalization produced similar results in terms of the number of statistically significant differentially expressed proteins throughout all comparisons, as shown in Fig.5-3A-2D. However, quantile normalization consistently identified a slightly higher number of proteins, depending on the carcinoma compared against the near adjacent tissue, NAT. On the other hand, the median divide method consistently produced about 37% fewer differentially expressed proteins for all the pairs compared, as depicted

in Fig.5-3. This trend was also observed for the down- and upregulated proteins as shown in Fig.5-8.

In a previous study utilizing the same raw data, our team employed mass spectrometry to identify four potential biomarkers: PLIN2, TUBB3, LAMP1, and HK1. The Z-score and Total Protein Approach (TPA) were utilized for this purpose, and these markers are depicted in Fig.5-8. TPA is a label-free mass spectrometry method developed by Wiśniewski et al. [34,35] that enables the measurement of absolute expression levels of numerous proteins without the need for standards.

Using TPA, we were able to highlight these potential biomarkers, and their validity was subsequently confirmed through immunohistochemical tissue arrays [18]. Interestingly, in our current study, we observed statistically significant differential expression for the same proteins across all three normalization methods employed, with similar p-values. This suggests that when a protein exhibits high significance, it will be consistently identified as significant regardless of the specific normalization method used.

These findings emphasize the robustness and reliability of the identified biomarkers and support the notion that they are biologically relevant. It further reinforces the notion that the choice of normalization method does not significantly impact the detection of highly significant proteins.

Furthermore, we compared the top 100 proteins with the highest log P-values across the three normalization methods in all two-group comparisons. We found that the z-score and quantile normalization methods had 94 proteins in common, while the median method shared 86 and 84 proteins with the z-score and the quantile methods, respectively. These results suggest that the z-score and quantile normalization methods may be more consistent with each other in identifying highly differentially expressed proteins, while the median method may identify a slightly different set of proteins.

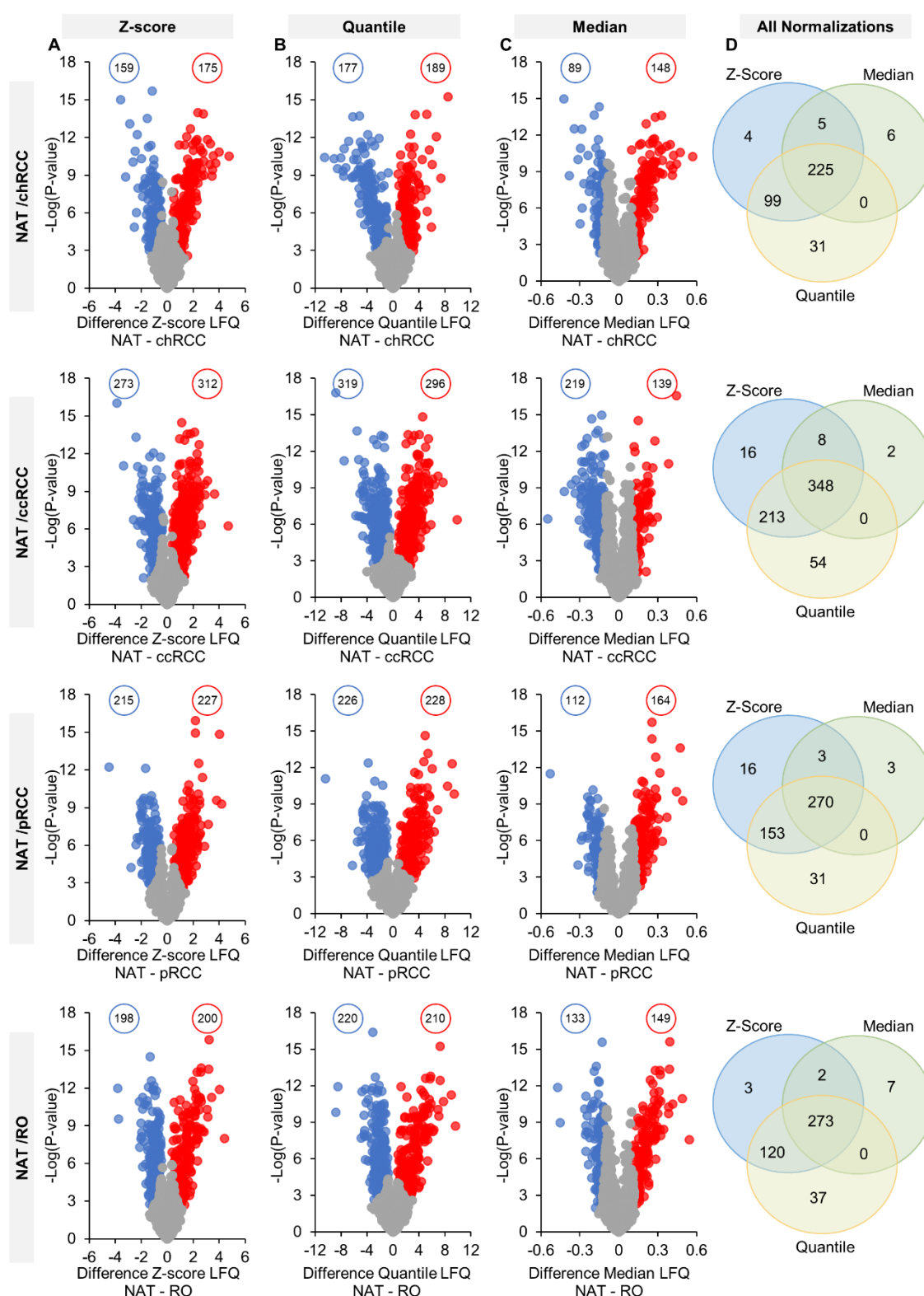


Figure 5-3. Comparison of the numbers of proteins found statistically differentially expressed using each normalization method when comparing the near adjacent tissue (NAT) against each renal tumour type. Each column represents the volcano plots obtained using one normalization method: **A.** Z-score; **B.** Quantile; **C.** Median. Each row shows the comparison by pairs (Volcano plot, according to Student's t test, FDR 0.05 and S_0 of 0.1, red dots: over-expressed proteins in NAT; blue dots overexpressed proteins in tumours). chRCC: chromophobe renal cell

carcinoma; ccRCC: clear cell renal cell carcinoma; pRCC papillary renal cell carcinoma; RO: Renal oncocytoma. **D:** column summarizes the common and uncommon proteins found among pair comparisons for each normalization method.

5.4.2 Comparison of the two, three and four renal carcinomas at a time against NAT

Our next step was to compare the tumours datasets in groups of 2, 3 and 4 against NAT. Proteins for further processing were selected if they were found in any dataset in at least 70% of the replicates. Focusing on the results obtained comparing the four neoplasia's datasets against NAT (Fig.5-4A) it can be concluded that when the comparison is not made in pairs, the numbers of proteins found to be statistically different is almost the same (i.e., 1075, 1086 and 1034 for the Z-score, the quantile and the median divide, respectively). In addition, 1008 of all the proteins were identified with any normalization method. The same patterns were repeated comparing three or two tumours against NAT, Fig.5-4A respectively.

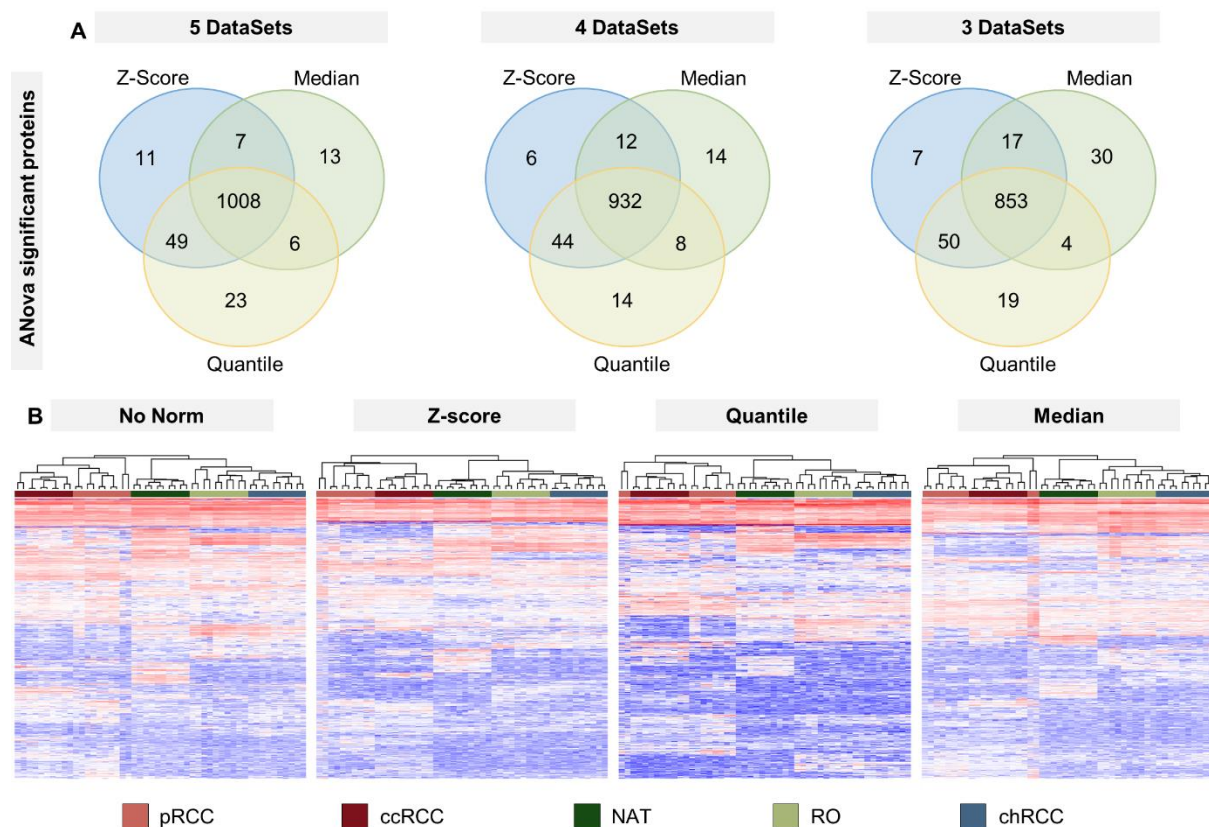


Figure 5-4. Proteins found statistically differentially expressed using the comparison of: **A:** 5 datasets: All tumours against NAT; 4 datasets chRCC , ccRCC and RO tumours against NAT and 3 datasets: ccRCC and RO against NAT. **B:** Hierarchical clusters were performed by using proteins that were found to be statistically different using each

normalization method when comparing all tumours against NAT (average linkage, no constraint, pre-processing with k-means, and Euclidean distance, $n=25$).

These results suggest that the choice of normalization method did not have a substantial impact on the number of differentially expressed proteins identified when datasets are not compared in pairs, as explained in the previous section. This is due to the fact that the MaxQuant statistics varied differently, from comparing pairs to comparing three or more datasets, so the differences were not on the same scale as those observed when comparing datasets in pairs, as shown in the previous section.

To cluster the samples, we exploited the differentially expressed proteins identified by each normalization method using the five datasets. As noted in Fig.5-4B, we found that three distinct clusters were formed regardless of the normalization method. The tumours pCRR and ccRCC were grouped together in the first cluster, while RO and chRCC formed a distinct cluster in the second. The third one was NAT tissue. Notably, we observed that the proteomic expressions of RO and chRCC tumours were more similar to NAT than to pRCC and ccRCC.

When the differentially expressed proteins were analysed to extract those whose levels make them unique for just one type of carcinoma, the number of proteins was 178, regardless of the normalization process, as depicted in Fig.5-5A. Furthermore, we also found that 174 out of these 178 biomarkers were present in our previous work [18], including the four biomarkers validated by us using immunohistochemistry. These results suggest that the choice of normalization method has a negligible impact on the identification of specific biomarkers.

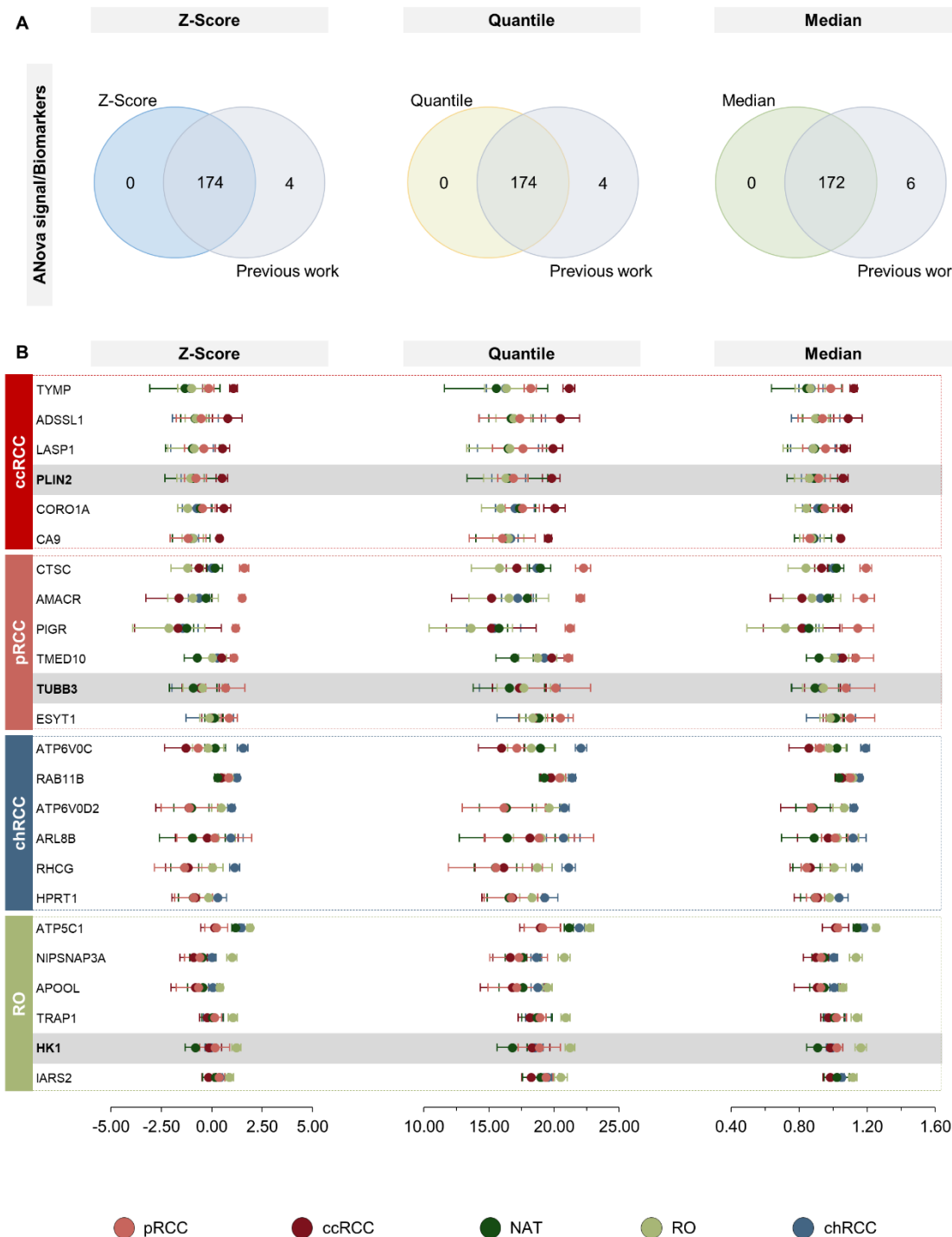


Figure 5-5. Comparison of candidate biomarkers for each renal tumour subtype across all different type of data normalization. **A.** Biomarkers found in this work for each normalization method versus those found in our previous

work [218]. **B.** Biomarkers proposed in our previous work (Fig.5-8) found with each normalization method in the present work. Note that for all of them the standard deviations of the normalized LFQ values overlap.

5.4.3 Comparison of the best biomarkers for *Immunohistochemistry*

In a previous work we recommended several proteins as new potential biomarkers for immunohistochemistry (Fig.5-9). These proteins were first identified using mass spectrometry (LFQs) and then with TPA. Of these proteins, those represented in grey in Fig.5-5B were selected in the previous work because the standard deviations did not overlap. These proteins were later validated against tissue arrays [18].

When we used the levels obtained through each normalization method for each protein to identify statistically significant proteins, we obtained two interesting results. Firstly, the three normalization methods delivered almost the same results, but they differed substantially from those obtained using TPA as described above. As shown in Fig.5-5B, proteins PLIN2, TUBB3, and HK1 would have never been identified as the best potential biomarkers because the standard deviations overlap each other (Fig.5-5). This unexpected result highlights the need to re-evaluate how data is treated and the importance of TPA in biomarker discovery.

5.4.4 Exploring the impact of normalization methods on identifying main pathways affected in renal tumours

To determine whether the identified dysregulated pathways changed as a function of the selected normalization method, we used the set of proteins found to be dysregulated when the five different datasets were compared. Then, through this data set, using an ANOVA test, we compared the proteins differentially expressed between NAT and chRCC.

Fig.5-6 shows the 10 most important (as per p-value) dysregulated pathways for (i) GO biological process, (ii) KEGG, and (iii) Reactome pathways. The results show that for the GO biological processes the first 7 pathways affected are the same, with no variations nor in the p-value, neither in the number of proteins involved, thus showing no differences in the normalization methods. For the KEGG, the Z-score and the quantile normalizations were found more consistent between them than when compared to the median, as they have the same top-10 pathways. For the Reactome pathways, the three normalization methods were found to provide similar results.

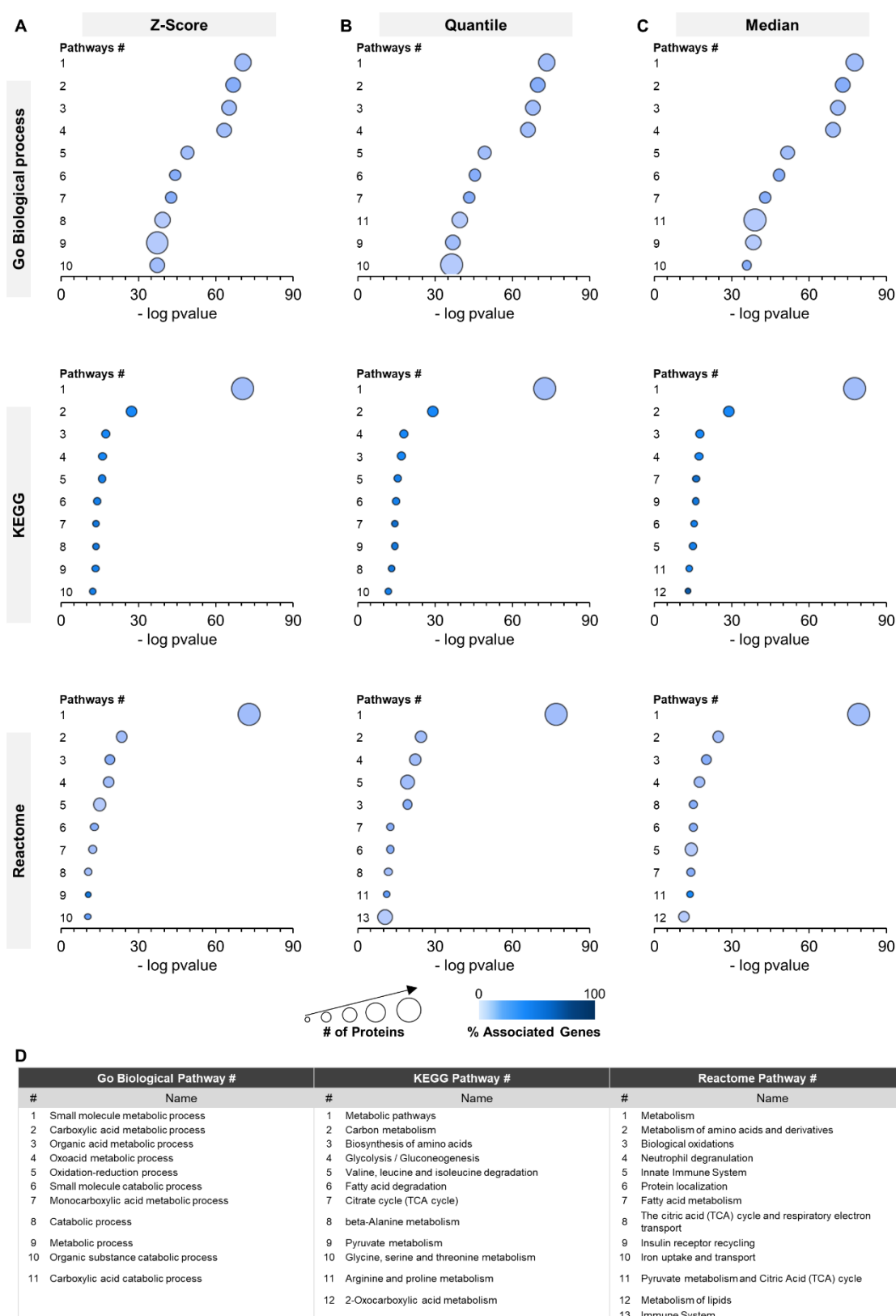


Figure 5-6. Comparison of Top-10 pathways differentially expressed when chRCC is compared against NAT as per the proteins differentially expressed found with each normalization method. **A.** Z-score; **B.** Quantile; **C.** Median. Databases used for pathway enrichment analysis were Go biological processes, KEGG and Reactome.

5.5 Conclusions

When comparing two groups of datasets, such as healthy versus diseased samples, the Z-score normalization and the quantile methods tend to yield a higher number of differentially expressed proteins compared to the median normalization method. Thus, in these cases, we recommend using the Z-score normalization method. However, when comparing a larger number of datasets groups (3, 4, or 5), the differences in terms of proteins and biochemical pathways statistically different among the three normalization methods are negligible, and any can be selected. Interestingly, our findings show that four biomarkers, previously identified using the z-score normalization (LFQs) in combination with the TPA approach, and later validated with immunohistochemistry [18], could not have been easily highlighted using any of the normalisation methods assessed in this study. This discovery opens new avenues of research in biomarker discovery as it emphasises the TPA approach as a powerful tool in biomarker discovery.

5.6 Supplementary material

5.6.1 Supplementary data

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [21] partner repository with the dataset identifier PXD023296.

5.6.2 Supplementary figures & tables

Table 5-1. Description of human kidney biopsies used in the study. RCC: renal cell carcinoma; NAT: normal adjacent tissue; chRCC: chromophobe renal cell carcinoma; RO: renal oncocytoma; pRCC: papillary renal cell carcinoma; ccRCC: clear cell renal cell carcinoma. Patient data collection and sample treatment was performed on a previous study [19].

Biopsy	Age (Years)	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
O15	55	Female	RCC	RO
P16	51	Female	RCC	pRCC
P17	66	Male	RCC	pRCC
P18	75	Male	RCC	pRCC
P19	87	Male	RCC	pRCC
P20	69	Male	RCC	pRCC
CC21	50	Male	RCC	ccRCC
CC22	68	Male	RCC	ccRCC
CC23	60	Male	RCC	ccRCC
CC24	68	Female	RCC	ccRCC
CC25	58	Male	RCC	ccRCC
CC26	42	Female	RCC	ccRCC
CC27	58	Female	RCC	ccRCC

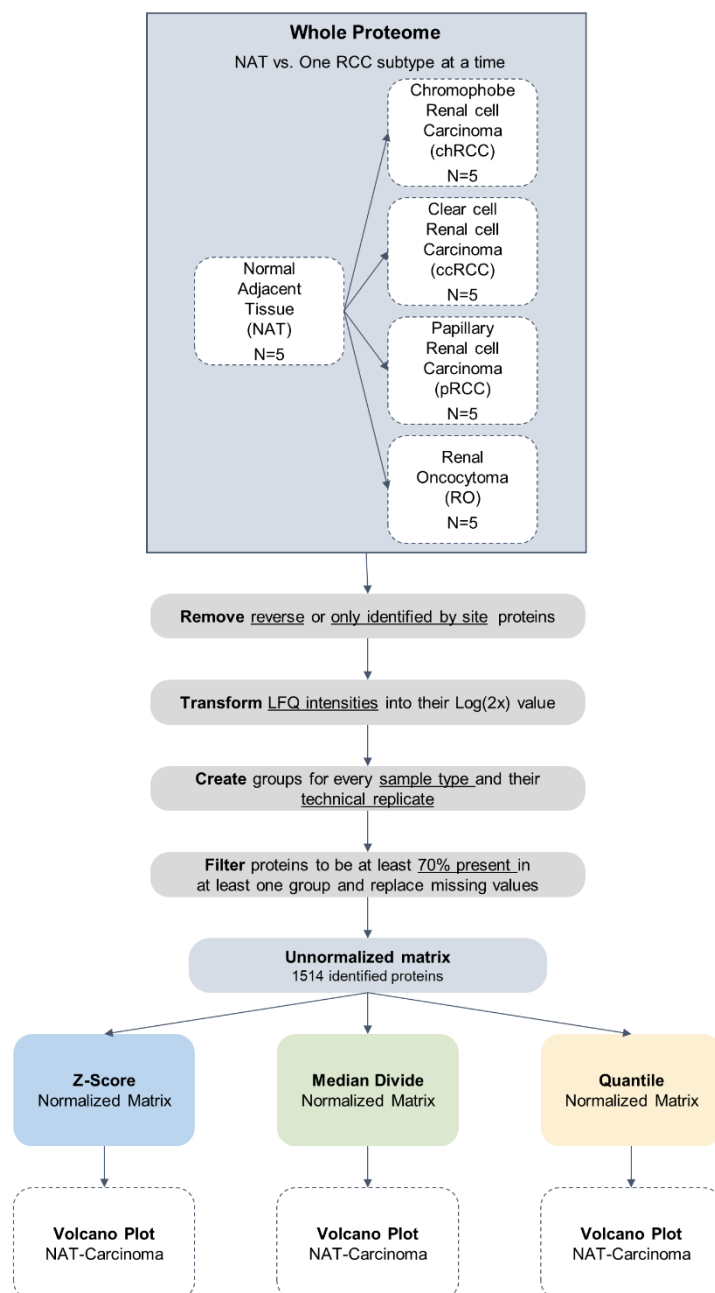
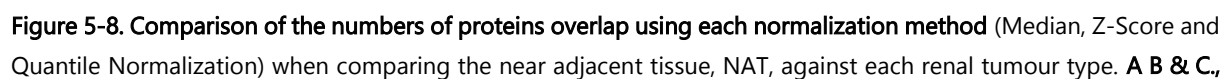


Figure 5-7. Workflow of the data process applied to the proteomics data when comparing by pairs the Neat Adjacent Tissue, NAT (control group) against the other renal tumour types.



column summarizes the common and uncommon proteins found among pair comparisons for each normalization method for all the statistically different dysregulated, upregulated and downregulated proteins.

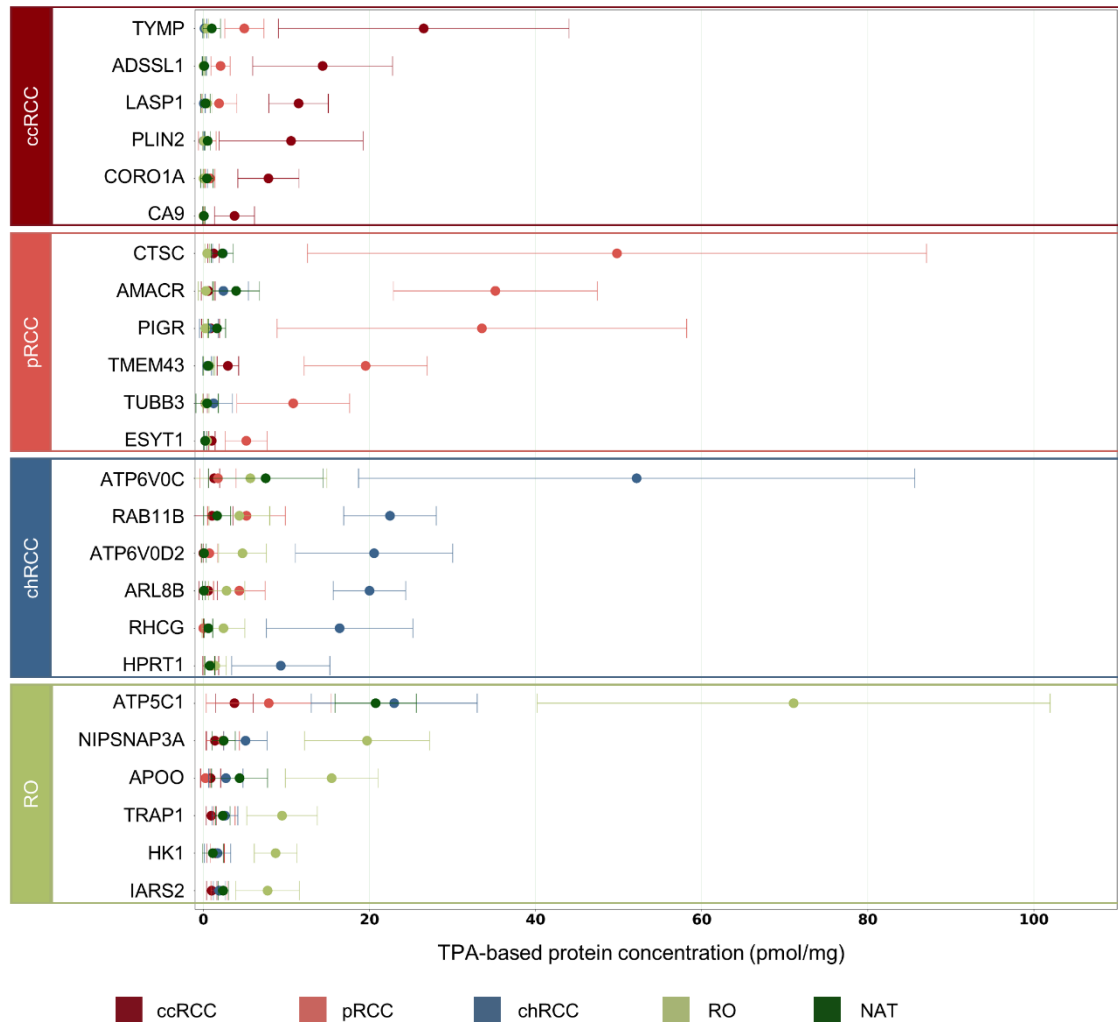


Figure 5-9. TPA-based concentrations of potential biomarkers for each renal carcinoma tumour subtype. Absolute protein concentration expressed in pmol/mg of total protein was calculated with the TPA method [35]. Image reproduced from paper [18].

5.7 Contributions & Acknowledgements

5.7.1 Author contributions

J.L.C., H.M.S., and L.B.C. designed the bioinformatics pipelines. L.B.C. and P.A.D.T.C. performed the bioinformatics analysis. J.L.C drafted the first manuscript. P.A.D.T.C., H. M. S., L.B.C., R.D., J.W., C.L., L.M. and M.P. read the final draft, corrected it and made valuable suggestions. R.D. provided the solid biopsies, and J.W. helped with the total protein approach. S.J. made the proteomics sample treatment and H.M.S. acquire the raw LC-MS/MS data.

5.7.2 Acknowledgements

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry - LAQV which is financed by national funds from FCT/MCTES, *Fundação para a Ciência e a Tecnologia* and *Ministério da Ciência, Tecnologia e Ensino Superior*, through the projects UIDB/50006/2020 and UIDP/50006/202). H. M. S. acknowledges LAQV (LA/P/0008/2020) funded by FCT/MCTES for his research contract. L. B. C. thanks the FCT/MEC for the FCT PhD grant 2019 (SFRH/BD/144222/2019). S. J. thanks FCT/MEC (Portugal) for her PhD grant reference SFRH/BD/120537/2016. This project utilized the University of Pittsburgh Hillman Cancer Center shared resource facilities (Cancer Genomics Facility and The Health Science Tissue Bank) supported in part by award P30CA047904 (Dr. R. Dhir).

5.8 References

- [1] E. Dubois, A.N. Galindo, L. Dayon, O. Cominetti, Assessing normalization methods in mass spectrometry-based proteome profiling of clinical samples, *Biosystems*. 215–216 (2022) 104661. <https://doi.org/10.1016/j.biosystems.2022.104661>.
- [2] B.J.A. Mertens, Transformation, normalization, and batch effect in the analysis of mass spectrometry data for omics studies, in: *Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry*, Springer International Publishing, Cham, 2017: pp. 1–21. https://doi.org/10.1007/978-3-319-45809-0_1.
- [3] A.M. De Livera, M. Sysi-Aho, L. Jacob, J.A. Gagnon-Bartsch, S. Castillo, J.A. Simpson, T.P. Speed, Statistical methods for handling unwanted variation in metabolomics data, *Anal Chem*. 87 (2015) 3606–3615. <https://doi.org/10.1021/ac502439y>.
- [4] A. Chawade, E. Alexandersson, F. Levander, Normalyzer: A tool for rapid evaluation of normalization methods for omics data sets, *J Proteome Res*. 13 (2014) 3114–3120. <https://doi.org/10.1021/pr401264n>.
- [5] Y. V Karpievitch, A.R. Dabney, R.D. Smith, Normalization and missing value imputation for label-free LC-MS analysis, *BMC Bioinformatics*. 13 (2012) S5. <https://doi.org/10.1186/1471-2105-13-S16-S5>.

- [6] B.M. Bolstad, R.A. Irizarry, M. Åstrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, *Bioinformatics*. 19 (2003) 185–193. <https://doi.org/10.1093/bioinformatics/19.2.185>.
- [7] S.J. Callister, R.C. Barry, J.N. Adkins, E.T. Johnson, W. Qian, B.-J.M. Webb-Robertson, R.D. Smith, M.S. Lipton, Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics, *J Proteome Res.* 5 (2006) 277–286. <https://doi.org/10.1021/pr050300l>.
- [8] K. Kultima, A. Nilsson, B. Scholz, U.L. Rossbach, M. Fälth, P.E. Andrén, Development and evaluation of normalization methods for label-free relative quantification of endogenous peptides, *Molecular & Cellular Proteomics*. 8 (2009) 2285–2295. <https://doi.org/10.1074/mcp.M800514-MCP200>.
- [9] T. Välikangas, T. Suomi, L.L. Elo, A systematic evaluation of normalization methods in quantitative label-free proteomics, *Brief Bioinform.* (2016) bbw095. <https://doi.org/10.1093/bib/bbw095>.
- [10] B.-J.M. Webb-Robertson, M.M. Matzke, J.M. Jacobs, J.G. Pounds, K.M. Waters, A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors, *Proteomics*. 11 (2011) 4736–4741. <https://doi.org/10.1002/pmic.201100078>.
- [11] A. Chawade, M. Sandin, J. Teleman, J. Malmström, F. Levander, Data processing has major impact on the outcome of quantitative label-free LC-MS analysis, *J Proteome Res.* 14 (2015) 676–687. <https://doi.org/10.1021/pr500665j>.
- [12] J. Cox, M.Y. Hein, C.A. Lubner, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Molecular & Cellular Proteomics*. 13 (2014) 2513–2526. <https://doi.org/10.1074/mcp.M113.031591>.
- [13] B. Zhang, L. Käll, R.A. Zubarev, DeMix-Q: Quantification-centered data processing workflow, *Molecular & Cellular Proteomics*. 15 (2016) 1467–1478. <https://doi.org/10.1074/mcp.O115.055475>.
- [14] A. Pursiheimo, A.P. Vehmas, S. Afzal, T. Suomi, T. Chand, L. Strauss, M. Poutanen, A. Rokka, G.L. Corthals, L.L. Elo, Optimization of statistical methods impact on quantitative proteomics data, *J Proteome Res.* 14 (2015) 4118–4126. <https://doi.org/10.1021/acs.jproteome.5b00183>.
- [15] D.L. Tabb, L. Vega-Montoto, P.A. Rudnick, A.M. Variyath, A.-J.L. Ham, D.M. Bunk, L.E. Kilpatrick, D.D. Billheimer, R.K. Blackman, H.L. Cardasis, S.A. Carr, K.R. Clauser, J.D. Jaffe, K.A.

Kowalski, T.A. Neubert, F.E. Regnier, B. Schilling, T.J. Tegeler, M. Wang, P. Wang, J.R. Whiteaker, L.J. Zimmerman, S.J. Fisher, B.W. Gibson, C.R. Kinsinger, M. Mesri, H. Rodriguez, S.E. Stein, P. Tempst, A.G. Paulovich, D.C. Liebler, C. Spiegelman, Repeatability and reproducibility in proteomic identifications by liquid chromatography–tandem mass spectrometry, *J Proteome Res.* 9 (2010) 761–776. <https://doi.org/10.1021/pr9006365>.

[16] B. Bolstad, preprocessCore: A collection of pre-processing functions. R package version 1.62.0, (2023).

[17] Y. Zhao, L. Wong, W.W. Bin Goh, How to do quantile normalization correctly for gene expression data analyses, *Sci Rep.* 10 (2020) 15534. <https://doi.org/10.1038/s41598-020-72664-6>.

[18] S. Jorge, J.L. Capelo, W. LaFramboise, S. Satturwar, D. Korentzelos, S. Bastacky, G. Quiroga-Garza, R. Dhir, J.R. Wiśniewski, C. Lodeiro, H.M. Santos, Absolute quantitative proteomics using the total protein approach to identify novel clinical immunohistochemical markers in renal neoplasms, *BMC Med.* 19 (2021) 196. <https://doi.org/10.1186/s12916-021-02071-9>.

[19] S. Jorge, J.L. Capelo, W. LaFramboise, R. Dhir, C. Lodeiro, H.M. Santos, Development of a robust ultrasonic-based sample treatment to unravel the proteome of OCT-embedded solid tumor biopsies, *J Proteome Res.* 18 (2019) 2979–2986. <https://doi.org/10.1021/acs.jproteome.9b00248>.

[20] E.W. Deutsch, A. Csordas, Z. Sun, A. Jarnuczak, Y. Perez-Riverol, T. Ternent, D.S. Campbell, M. Bernal-Llinares, S. Okuda, S. Kawano, R.L. Moritz, J.J. Carver, M. Wang, Y. Ishihama, N. Bandeira, H. Hermjakob, J.A. Vizcaíno, The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition, *Nucleic Acids Res.* 45 (2017) D1100–D1106. <https://doi.org/10.1093/nar/gkw936>.

[21] Y. Perez-Riverol, J. Bai, C. Bandla, D. García-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D.J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, J.A. Vizcaíno, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, *Nucleic Acids Res.* 50 (2022) D543–D552. <https://doi.org/10.1093/nar/gkab1038>.

[22] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat Biotechnol.* 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.

[23] S. Tyanova, T. Temu, A. Carlson, P. Sinitcyn, M. Mann, J. Cox, Visualization of LC-MS/MS proteomics data in MaxQuant, *Proteomics.* 15 (2015) 1453–1456. <https://doi.org/10.1002/pmic.201400449>.

- [24] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J. V. Olsen, M. Mann, Andromeda: A peptide search engine integrated into the MaxQuant environment, *J Proteome Res.* 10 (2011) 1794–1805. <https://doi.org/10.1021/pr101065j>.
- [25] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data, *Nat Methods.* 13 (2016) 731–740. <https://doi.org/10.1038/nmeth.3901>.
- [26] S. Tyanova, J. Cox, Perseus: A bioinformatics platform for integrative analysis of proteomics data in cancer research, in: 2018: pp. 133–148. https://doi.org/10.1007/978-1-4939-7493-1_7.
- [27] S. Yu, D. Ferretti, J.P. Schessner, J.D. Rudolph, G.H.H. Borner, J. Cox, Expanding the perseus software for omics data analysis with custom plugins, *Curr Protoc Bioinformatics.* 71 (2020). <https://doi.org/10.1002/cpbi.105>.
- [28] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.* 43 (2015) e47–e47. <https://doi.org/10.1093/nar/gkv007>.
- [29] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- [30] N.T. Doncheva, J.H. Morris, J. Gorodkin, L.J. Jensen, Cytoscape StringApp: Network analysis and visualization of proteomics data, *J Proteome Res.* 18 (2019) 623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>.
- [31] The Gene Ontology Resource: 20 years and still GOing strong, *Nucleic Acids Res.* 47 (2019) D330–D338. <https://doi.org/10.1093/nar/gky1055>.
- [32] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic Acids Res.* 45 (2017) D353–D361. <https://doi.org/10.1093/nar/gkw1092>.
- [33] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C.D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D’Eustachio, The reactome pathway knowledgebase, *Nucleic Acids Res.* 46 (2018) D649–D655. <https://doi.org/10.1093/nar/gkx1132>.
- [34] J.R. Wiśniewski, M.Y. Hein, J. Cox, M. Mann, A “proteomic ruler” for protein copy number and concentration estimation without spike-in standards, *Molecular & Cellular Proteomics.* 13 (2014) 3497–3506. <https://doi.org/10.1074/mcp.M113.037309>.

[35] J.R. Wiśniewski, Label-free and standard-free absolute quantitative proteomics using the “total protein” and “proteomic ruler” approaches. *Methods in Enzymology*, 585, 2017, Pages 49-60, <https://doi.org/10.1016/bs.mie.2016.10.002>.

BIOCHEMICAL NETWORK ANALYSIS OF PROTEIN-PROTEIN INTERACTIONS TO FOLLOW-UP T1 BLADDER CANCER PATIENTS

Luís B. Carvalho, José Luis Capelo Martínez, Carlos Lodeiro, Rafael Bento, Rajiv Dhir, Jeremiah J. Morrissey, Luis Campos Pinheiro, Mariana Medeiros, Hugo M. Santos

Journal of Proteomics 278 (2023) 104865

Doi: 10.1016/j.jprot.2023.104865

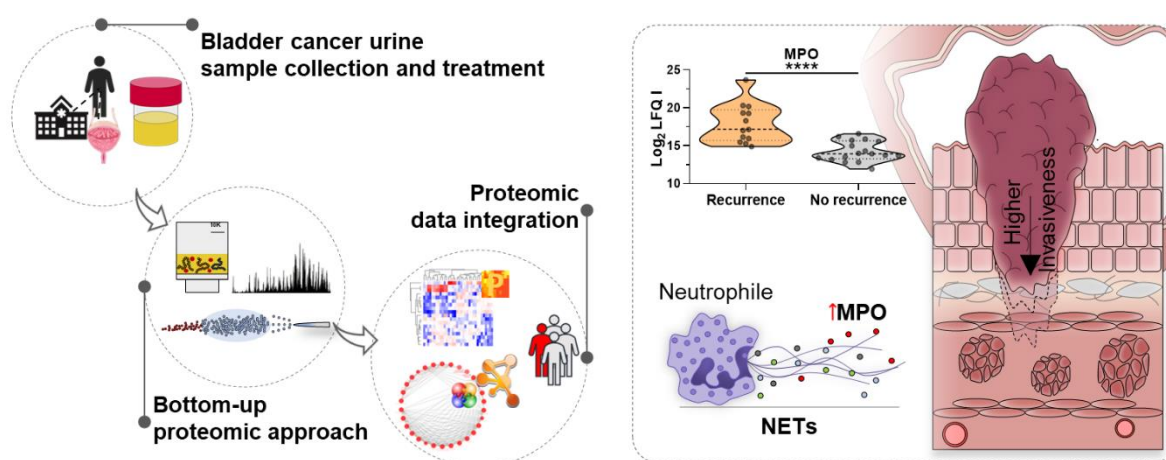


Figure 6-1. Graphical abstract.

6.1 Abstract

Bladder cancer (BCa) is a prevalent disease with a high risk of aggressive recurrence in T1-stage patients. Despite the efforts to anticipate recurrence, a reliable method has yet to be developed. In this work, we employed high-resolution mass spectrometry to compare the urinary proteome of T1-stage BCa patients with recurring versus non-recurring disease to uncover actionable clinical information predicting recurrence. All patients were diagnosed with T1-stage bladder cancer between the ages of 51 and 91, and urine samples were collected before medical intervention. Our results suggest that the urinary myeloperoxidase to cubilin ratio could be used as a new tool for predicting recurrence and that dysregulation of the inflammatory and immune systems may be a key driver of disease worsening. Furthermore, we identified neutrophil degranulation and neutrophil extracellular traps (NETs) as key pathways in the progression of T1-stage BCa. We propose that proteomics follow-up of the inflammatory and immune systems may be useful for monitoring the effectiveness of therapy.

6.2 Introduction

Precision medicine, which tailors treatment to the individual patient, is becoming increasingly important. One way to achieve this is through the use of personalized proteogenomics, which looks at the unique proteins and genetic information of each patient rather than grouping and pooling data from multiple patients. [1]. This approach allows for more accurate diagnosis and treatment, as it considers each patient's unique characteristics. Precision medicine can retrieve unprecedented levels of biological information by combining robust fast-sample treatment protocols, high-resolution mass spectrometry, and bioinformatics [2]. Furthermore, the use of mass spectrometry-based multiple reaction monitoring has made possible the large-scale validation of candidates as it has never been seen before, thus ensuring the correctness and applicability of the clinical information retrieved from each patient [3].

BCa is the ninth most common neoplasm, second in incidence and mortality of the genitourinary system [4]. Literature shows that BCa has a high incidence, progression, and recurrence rates. Therefore, it is extremely important to monitor BCa patients, especially those with high-risk tumors, which require frequent and invasive examination procedures [5]. Urinary cytology and cystoscopy are the current gold standard tools to diagnose BCa. However, cystoscopy is an invasive examination with a high burthen for the patient due to pain, bleeding, urinary tract infections and other complications. On the other hand, Cytology is a non-invasive

test that can identify tumor cells shed in the urine. It is simple to use and inexpensive, but it has low sensitivity and low diagnostic efficiency [6]. Much effort has been made to diagnose and prognose BCa patients using DNA, RNA, or proteins, focusing on solid or liquid biopsies (blood-derived samples) [7]. Despite the valuable information the urinary proteome carries, this liquid biopsy remains to be further explored to disclose new features about BCa.

The clinical outcome of individuals with T1-stage BCa with a high incidence of recurrence and progression is highly variable. The aetiology behind T1-stage BCa aggressiveness remains unclear [8]. Early and accurate diagnosis of T1-stage BCa that will recur or progress is now a significant challenge in treating patients with BCa.

In this work, the urinary proteome of T1-stage BCa with and without recurrence was interrogated via high-resolution mass spectrometry and bioinformatics to unveil prognosis signatures. Herein, we present the myeloperoxidase to cubilin ratio as a potential indicator of recurrence, and we propose to flag the course of BCa patients following the changes in the information related to the immune and inflammatory systems encoded in the urinary proteome.

6.3 Experimental Section

6.3.1 Patients

Mid-stream second void urine specimens were collected from 30 patients, all male, diagnosed with T1-stage BCa, ranging between 51 to 91 years. Urine samples were collected before the medical intervention, and T1-stage BCa diagnosis was assessed by pathological examination. Exclusion criteria: (1) unclear BCa diagnosis, (2) urinary cancer history, (3) HIV affected, and (4) organ transplant receivers or in recent chemo/radiotherapy (last six months). Further information is given in Table 6-1.

6.3.2 Urine collection and preservation

Mid-stream second void morning urine samples (30 ml) were collected in centrifuge tubes containing 38 mg of boric acid (Sigma-Aldrich) to prevent bacterial growth. Urine samples were placed on ice and were processed within the first two hours after collection. Samples were centrifuged for 20 min at 5000 g to remove cell debris and divided into 10 mL aliquots, and these aliquots were stored at -80 °C until further processing. The urinary proteome was concentrated by ultrafiltration from 10 mL to a final volume of c.a 300 µL using VivaSpin 15R

(cut-off 10kDa, Sartorius) at 5000 g for 20 min. The concentrated urinary proteomes were quantified for total protein using the Bradford protein assay method with Bradford reagent (Sigma-Aldrich) and Bovine Serum Albumin (BSA, Sigma-Aldrich) to perform the standard curve. Each concentrated urine sample and BSA standard were measured in duplicate. (See Fig.6-2 flow chart for further information).

6.3.3 Filter-Fided sample preparation.

Filter Aided Sample Preparation (FASP) was used for urinary proteome digestion using a modified protocol [9]. Two technical replicates of each urine sample were prepared. Protein amounts ranging from 50 to 100 µg of were diluted with milli-Q water to a final volume of 200 µL, and then 200 µL of 2 M urea (Sigma-Aldrich), 75 mM Tris-HCl pH 8.8 (Sigma-Aldrich), 100 mM NaCl (Sigma-Aldrich), 0.02% SDS (Sigma-Aldrich) was added. The resulting proteome solution was loaded onto centrifugation filters of 500 µL with a 10,000 molecular weight cut-off (MWCO). The urinary proteome onto the FASP membrane was washed with 200 µL of (8 M urea+25 mM ammonium bicarbonate (AmBic)) solution and then centrifuged for 20 min at 14,000 x g. Then, the disulfide bonds within the proteins were reduced by adding 200 µL of 50 mM dithiothreitol (DTT) prepared in an (8 M urea+25 mM AmBic) and then incubated for 60 minutes at 37 °C. Then, the FASP membranes were centrifuged for 20 minutes at 14,000 x g. Next, the samples were alkylated for 45 min in the dark by adding 100 µL 50 mM iodoacetamide (IAA) prepared in (8 M urea+25 mM AmBic solution). After reduction, IAA was removed by centrifugation for 20 minutes at 14,000 x g, followed by two washings with 200 µL 25 mM AmBic. Finally, protein digestion was done by adding 100 µL of 1:30 trypsin solution prepared in 12.5 mM AmBic. Protein digestion was proceeded overnight (approximately 16 hours) at 37 °C, and the peptides were collected by centrifugation for 20 minutes at 14,000 x g. To ensure that all the peptides were extracted, 100 µL of 3% (v/v) acetonitrile and 0.1% (v/v) formic acid was added, followed by centrifugation for 20 minutes at 14,000 x g. This previous step was repeated one more time, and then the peptides were transferred to a 500 µL microtube, dried, and stored at -20 °C until further analysis by Nano-LC-MS/MS.

6.3.4 LC-MS/MS

LC-MS/MS analysis was performed using an Ultimate 3000 nano LC system in coupled to an Impact HD (Bruker Daltonics) with a CaptiveSpray nanoBooster using acetonitrile as dopant. Peptides were resuspended in 100 µL of 3% (v/v) acetonitrile containing 0.1% (v/v)

aqueous formic acid (FA). The samples were homogenized by 5 min on vortex followed by 10 min on an ultrasonic bath at 100% ultrasonic amplitude, 35 kHz. Afterward, samples were quantified using a Pierce™ quantitative colorimetric peptide assay (Thermo Fisher Scientific). Following the previous step, 3 µL containing 563 ng of peptides were loaded onto a trap column (Acclaim PepMap100, 5 µm, 100 Å, 300 µm i.d. × 5mm) and desalted for 5 min from 3% to 5% B (B: 90% acetonitrile 0.08% FA) at a flow rate of 15 µL.min⁻¹. Then the peptides were separated using an analytical column (Acclaim™ PepMap™ 100C18, 2 µm, 0.075 mm i.d x 150 mm) with a linear gradient at 300 nL.min⁻¹ (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) acetonitrile and 0.08% (v/v) FA) 5-90 min from 5% to 35% of mobile phase B, 90-100min linear gradient from 35% to 95% of mobile phase B, 100-110 95% B. Chromatographic separation was carried out at 35 °C. The MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8-32 Hz), cycle time 3.0 s, with active exclusion (precursors were excluded from precursor selection for 0.5 min after the 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 was 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.

6.3.5 Bioinformatics

Relative label-free quantification (LFQ) was carried out using MaxQuant software V1.6.0.16. All raw files were processed in a single run with default parameters [10]. Database searches were performed using the Andromeda search engine against the UniProt UP000005640_9606 database (20,600 sequences; 11,395,157 residues, downloaded on April 27 2021). Data processing was performed using Perseus software V1.6.15.0 using default settings using the workflow depicted in Fig.6-2. Protein group LFQ intensities were log2-transformed, and the quantitative profiles were filtered for missing values with the following settings: min valid percentage of 50% in at least one group and values greater than 0. To overcome the obstacle of missing LFQ values, missing values were imputed from the total matrix with width = 0.5 and down-shift = 1.8 [11]. Log ratios were calculated as the difference in average log2 LFQ intensity values between the tested conditions (two-tailed, Student's t-test, permutation-based FDR 0.05, and S0 of 0.1). Perseus was also used to obtain clusters, using average linkage, no constraint, pre-process with k-means, and Euclidean distance between column trees. Protein complex networks were integrated and visualized using Cytoscape V3.8.2 with the

applications ClueGo V2.5.8, CluepediaV1.5.8, and StringApp V1.6.0. Reactome and biological process go terms were used as ontologies databases [12].

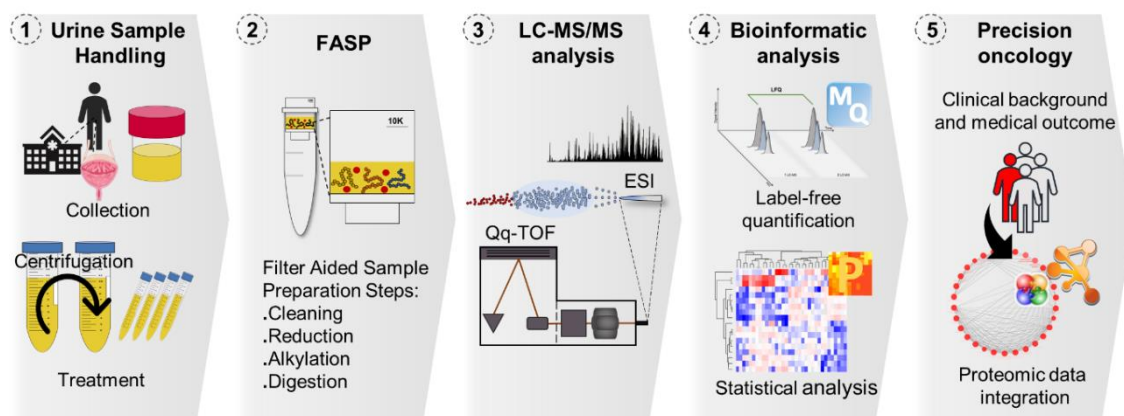


Figure 6-2. Proteomics and bioinformatics workflow. 1. Urine sample collection pre-treatment and preservation. 2. Filter-aided sample preparation (FASP). 3. LS-MS/MS analysis. 4. Bioinformatic and statistical analysis 5. Data interpretation and correlation with a clinical background.

6.3.6 MPO in vitro Enzyme-Linked Immunosorbent Assay

Quantitative measurement of urinary MPO was performed using an in vitro ELISA kit from Abcam (ab119605) following manufacturer instructions. The plate was read for absorbance at 450 nm using a CLARIOstar® High-Performance Monochromator Multimode from BMG LAB-TECH within 30 min after adding the stop solution.

6.4 Results & Discussion

The clinical information of the patient cohort used in the study is presented in Table 6-1. The comparison of the urine proteome of T1-stage BCa patients that had experienced recurrence (RG) versus patients that had not (NRG) is explained in a comprehensive flow-chart in Fig.6-5. This comparison is presented in the format of a volcano plot in Fig.6-6. A set of 197 proteins differentially expressed between both groups was found. The bioinformatics analysis of the set of 197 proteins, focusing on both single proteins and biochemical networks, showed that (i) myeloperoxidase (MPO) was up-regulated (12.6 fold change and p -value = 0.0028) in the RG group and (ii) cubilin (CUBN) was up-regulated (3.6 fold change and p -value = 0.0289) in the NRG group. The variation in the levels of both proteins for each patient is presented in Fig.6-3A, whilst the CUBN/MPO ratio is presented in Fig.6-3B. The orthogonal validation of the MPO levels in urine was done using ELISA as shown in Fig.6-3G. The set of 197 proteins dysregulated between both groups of patients was used to obtain the biochemical networks

dysregulated between RG and NRG. Such networks are shown in Fig.6-3H and I, displaying the protein-protein interaction and network analysis focusing on the inflammation process. Finally, the most interesting protein-protein interactions and biological functions addressed by 197 statistically significant proteins are integrated comprehensively in Fig.6-4.

6.4.1 CUBN to MPO ratio

As the bladder is in direct contact with the urine, we hypothesized that the urine proteome of BCa patients should contain critical and diagnostic/prognostic information about the disease. So far, we have studied the differences in the urinary proteome of BCa patients with and without a background of recurrence and/or progression, as this stratification is significant for medical treatment, and there is no current way to make it. The groups are called RG and NRG, respectively. The comparison is presented as a volcano plot in Fig.6-6. In this comparison, the myeloperoxidase (MPO) was found on the proteins most up-regulated in the RG group. On the other hand, the same data showed cubilin (CUBN), and lipopolysaccharide-binding protein 2 (LRP2) proteins down-regulated in the same group.

The MPO is a hemoprotein expressed by polymorphonuclear neutrophils, which is secreted during leucocyte degranulation in a cellular process that releases antimicrobial cytotoxic or mediator molecules as a defense response to invading microorganisms. More precisely, MPO catalyzes the conversion of H_2O_2 to hypochlorous acid, an efficient intracellular bacterial killer. In addition, it inactivates protease inhibitors, thus allowing lytic enzymes released from neutrophils to degrade tissue matrix and foreign materials facilitating tumor growth [13].

Furthermore, MPO has been proposed recently as a marker of poor prognosis in the lung [14] and ovarian carcinomas [15]. However, to the best of our knowledge, no studies report MPO levels as a prognosis biomarker for BCa carcinoma. Our data in Fig.6-3A shows that those patients experiencing recurrence and progression presented the highest levels of MPO. In addition to MPO, the RG presented down-regulation of CUBN and LRP2. Remarkably, the down-regulation of CUBN has already been noted as a poor prognosis biomarker for renal cell carcinoma [16]. Thus, the lower the CUBN level, the worst the prognosis for renal carcinoma. Our data address the same trend also for BCa. It is worth noting that the group with recurrence and progression has the lowest levels of CUBN (see Fig.6-3A, B, and C). Therefore, an index based on the MPO to CUBN ratio can be established such that patients with ratios equal to or lower than one present recurrence and/or progression.

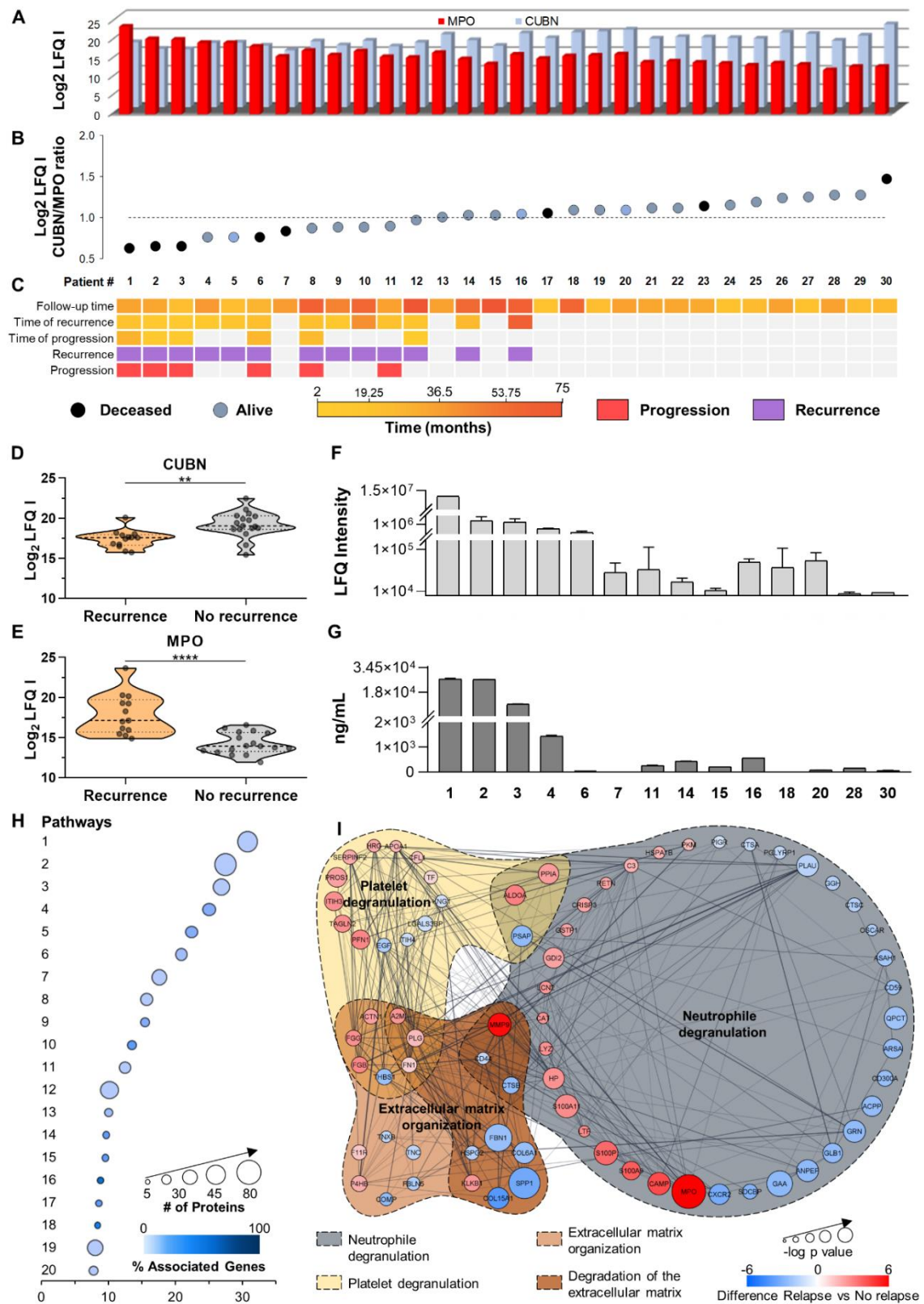


Figure 6-3. T1 BCa patients MPO/CUBN levels and protein-protein pathway analysis. **A.** Relative label-free quantification of CUBN and MPO within the T1 patients. Log₂-transformed Label-free quantification intensity (Log₂ LFQ I). **B.** CUBN/MPO ratio correlation with patient survival. **C.** Patients with T1 clinical and medical follow-up information.

D & E. Violin plot showing the relative expression of MPO and CUBN by the protein label-free quantification (LFQ) levels. Expression of those proteins is shown in the group of patients with recurrence and no recurrence. Student's t-test was used for statistical analysis. ** $p \leq 0.01$; **** $p \leq 0.0001$. **F.** Label-free quantification of myeloperoxidase and **G.** Validation of mass spectrometry data using myeloperoxidase (MPO) protein by ELISA (ng/mL). X-axes represent the patient number. **H.** The top 20 pathways were obtained with the differential protein expression between the RG and NRG groups. Pathways were obtained by the software platform cystoscape using StringApp with the Reactome pathway database. Pathways: (1) Innate immune system; (2) Immune system; (3) Neutrophil degranulation; (4) Regulation of insulin-like growth factor (IGF) transport and uptake by Insulin-like growth factor binding proteins (IGFBPs); (5) Post-translational protein phosphorylation; (6) Platelet degranulation; (7) Hemostasis; (8) Platelet activation, signaling and aggregation; (9) Regulation of complement cascade; (10) Formation of fibrin clot (clotting cascade); (11) Extracellular matrix organization; (12) Metabolism of proteins; (13) Integrin cell surface interactions; (14) Common pathway of fibrin clot formation; (15) Intrinsic pathway of fibrin clot formation; (16) Terminal pathway of complement; (17) Plasma lipoprotein assembly; (18) Chylomicron assembly; (19) Post-translational protein modification; (20) Degradation of the extracellular matrix. **I.** Protein-protein interaction and pathway analysis related to immune responses focus on the neutrophile action.

6.4.2 Biochemical networks

The study of the differences between the biochemical networks of the RG and NRG groups using the 197 proteins differentially expressed between them were analyzed, and the first 20 pathways with the lowest p-value were selected (Fig.6-3H). Most of them are linked to immune and inflammatory responses. Therefore, we decided to assess the pathways that markedly differentiate RG and NRG groups, namely neutrophile degranulation and degradation of extracellular matrix pathways (Fig.6-3I).

It is well known that the tumorigenesis process comprises the activation of several biochemical pathways, among which the ones related to invasion, proliferation, and inflammation are critical in tumor progression. Accordingly, our results address such pathways overexpressed in the recurrence group and are presented comprehensively in Fig.6-4. Thus, our results suggest that the first line of defense, the neutrophils, are recruited and deployed near the tumor to destroy the tumoral cells. The neutrophils mature into the tumor-associated neutrophils, TAN-N1 and TAN-N2 [17]. The TAN-N1 Type has pro-inflammatory and antitumor effects, while the TAN-N2 type has anti-inflammatory and pro-tumor effects. The TAN-N2 type tries to destroy the tumoral cells by releasing their cellular content rich in proteases, histones, and DNA, forming the so-called neutrophil extracellular traps, NETs, that are thought to coat tumor cells, shielding them from immune attack [18]. Pointing this process to occur, we found that matrix metalloproteinase 9 (MMP9) was up-regulated in the RG group. This protein and MMP8 are proteases that, in conjunction with other enzymes released during the NETs formation, are known to degrade the extracellular matrix, collagen, and gelatin [19]. This

extracellular degradation also increases the concentration of fibrin-related proteins, such as fibrinogen beta and gamma (FGB and FGG), that we found in the RG group. These proteins further promote tumor invasion [20].

Furthermore, several proteins released during NETs formation have been associated with a poor prognosis in other cancer types [21,22]. Belonging to this group, we have found proteins Neutrophil Gelatinase-Associated Lipocalin (LNC2), Cathelicidin antimicrobial peptide CAMP, and MPO up-regulated in the RG group (see Fig.6-4). The last one, MPO, is a notable finding because here we have found it markedly up-regulated in the patients with RG in agreement with results reported in other cancer types [14,15]. As mentioned above, MPO is involved in forming H_2O_2 to kill cancer cells via reactive oxygen species (ROS)-induced apoptosis. This also leads to an increment in the ROS and a higher number of mutations in the cancer cells. This may promote, to some extent, for the tumor to evolve and change expression patterns to survive by escaping the immune system [23]. The cancer cells respond to this attack by over-expressing proteins that promote protection against ROS, such as catalase (CAT) and beta-globulin (HBB), which we also found in the RG group [24].

Finally, we have found the inflammatory response unbalanced for RG patients. Thus, whilst the serpin protein family involved in controlling the inflammatory response is down-regulated (SERPINA5), the S100 family, which promotes inflammation, is overexpressed (S100P, S100A9). Remarkably, S100A9 is a calcium- and zinc-binding protein that plays a prominent role in regulating inflammatory processes and the immune response [25]. It can induce neutrophil chemotaxis and adhesion and increase the bactericidal activity of neutrophils by promoting phagocytosis via activation of SYK, PI3K/AKT, and ERK1/2 pathways. Also, it can induce the degranulation of neutrophils by a MAPK-dependent mechanism [26].

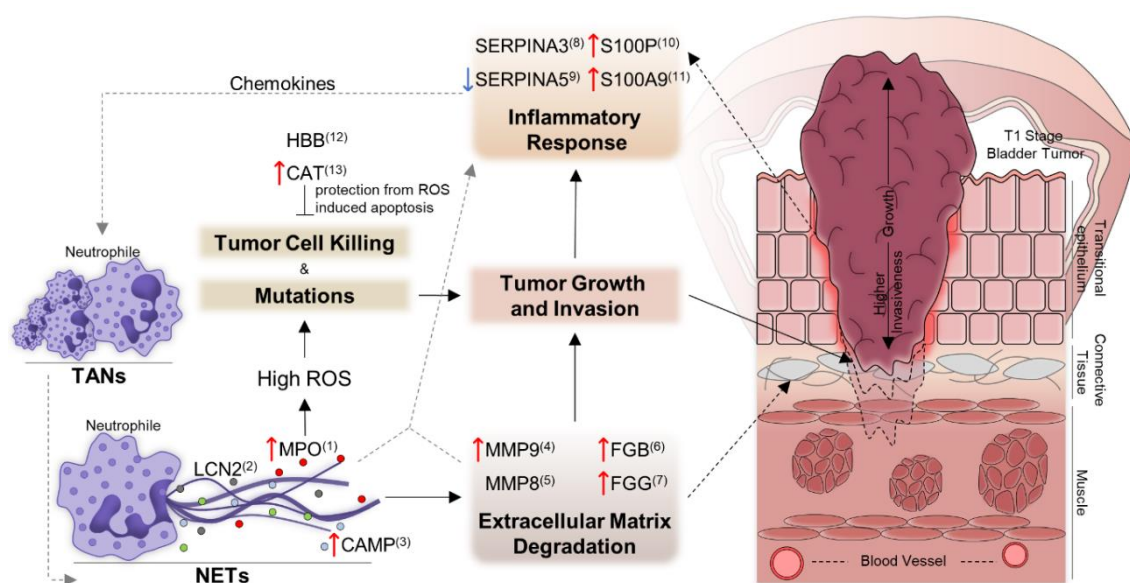


Figure 6-4. Data integration for T1 BCa stage patients. Gene name, protein name and protein ID: (1) Myeloperoxidase (MPO, P05164-2); (2) Neutrophil gelatinase-associated lipocalin (LCN2, X6R8F3); (3) Cathelicidin antimicrobial peptide (CAMP, J3KNB4); (4) Matrix metalloproteinase-9 (MMP9, P14780); (5) Matrix metalloproteinase 8 (MMP8, P22894); (6) Fibrinogen beta chain (FGB, P02675); (7) Fibrinogen gamma chain (FGG, P02679-2); (8) Alpha-1-antichymotrypsin (SERPINA 3, P01011); (9) Plasma serine protease inhibitor (SERPINA 5, P05154); (10) Protein S100-P (S100P, P25815); (11) Protein S100-A9 (S100A9, P06702); (12) Hemoglobin subunit beta (HBB, P68871); (13) Catalase (P04040, CAT).

In summary, our data show the presence of an inflammatory microenvironment that facilitates immune cell-matrix cross-talk, and promotes chronic inflammation. This state of inflammation further recruits more neutrophils and immune cells into the tumor microenvironment resulting in a cycle that the tissue cannot resolve. In agreement with the literature [27,28], our urinary proteome-based data suggest that the deregulation of inflammatory processes leads to an accelerated BCa progression and a worse outcome for patients with BCa leading to recurrence and tumor growth. On the other hand, when the inflammatory process is balanced, the BCa seems to evolve slowly and with a better prognosis for the patient. In support of our conclusions, it has been recently reported that NET formation in chronic inflammation leads to radio-resistance in a mouse heterotopic model of invasive BCa [29].

6.5 Conclusions

We conclude that inflammation dysregulation is an essential driver of BCa recurrence and progression and that anti-inflammatory or immune checkpoint therapies should be further explored in treating patients with BCa. Furthermore, consider the MPO to CUBN ratio a promising candidate for a urine prognosis tool in BCa. A limitation of this study is its small patient

sample, which may not fully capture the variability of T1-stage bladder cancer. Another issue is that some patients' prognoses were not evaluated within the time frame of this study. Further validation with larger patient groups is necessary.

6.6 Supplementary material

6.6.1 Supplementary data

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [30] partner repository with the data set number PXD026784.

6.6.2 Supplementary figures & tables

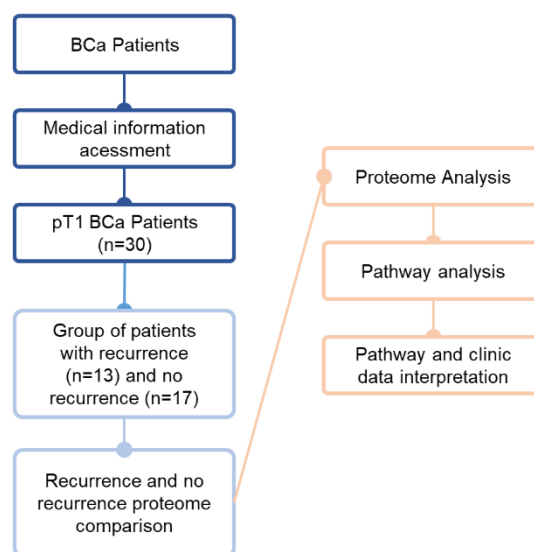


Figure 6-5. Workflow of the experimental design.

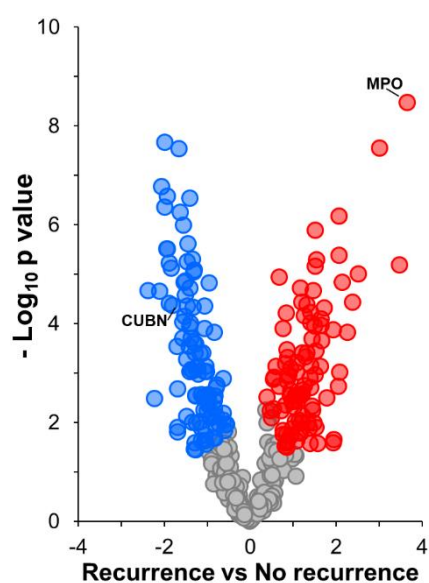


Figure 6-6. Volcano plots showing statistically significant changes in protein levels in the group of T1-stage bladder cancer patients that had experienced recurrence, (RG), versus patients that had not, (NRG).

Table 6-1. List of patients with the respective medical information.

Patient	Complete resection	Intersection of Muscular layer	Stage	Grade	Gender	Age (years)	Smoker	Follow up time (months)	BCG Treatment	Mitomycin Treatment	Recurrence	Time of recurrence (months)	Progress	Time of progression (months)	Status (Alive)
1	Yes	Yes	pt1	High Grade	M	61	Yes	26	Maintenance therapy	No	Yes	6	Yes	16	No
2	No	Yes	pt1	High Grade	M	91	-	24	No	-	Yes	2	Yes	2	No
3	Yes	No	pt1	High Grade	M	81	-	9	No	No	Yes	2	Yes	2	No
4	No	Yes	pt1	High Grade	M	74	No	15	Induction	No	Yes	8	Yes	13	No
5	Yes	Yes	pt1	High Grade	M	77	Yes	33	Induction	No	Yes	3	No	-	Yes
6	Yes	Yes	pt1	High Grade	M	75	Yes	9	Induction	No	Yes	5	No	-	Yes
7	Yes	Yes	pt1	High Grade	M	75	Yes	31	Maintenance therapy	No	No	-	No	-	No
8	Yes	Yes	pt1	High Grade	M	60	Yes	69	Maintenance therapy	No	Yes	8	Yes	11	Yes
9	Yes	Yes	pt1	High Grade	M	81	Yes	47	Maintenance therapy	No	Yes	3	No	-	Yes
10	No	Yes	pt1	High Grade	M	78	Yes	28	Induction	No	Yes	7	Yes	15	Yes
11	No	No	pt1	High Grade	M	76	Yes	65	Maintenance therapy	No	Yes	39	No	-	Yes
12	Yes	Yes	pt1	High Grade	M	77	Yes	27	Maintenance therapy	No	No	-	No	-	Yes
13	Yes	Yes	pt1	High Grade	F	74	-	75	Induction	-	Yes	5	No	-	Yes
14	Yes	Yes	pt1	High Grade	M	75	Yes	9	Maintenance therapy	-	No	-	No	-	Yes
15	Yes	Yes	pt1	High Grade	M	65	Yes	69	Maintenance therapy	No	Yes	69	No	-	Yes
16	Yes	Yes	pt1	High Grade	M	56	-	6	Maintenance therapy	-	No	-	No	-	No
17	Yes	Yes	pt1	High Grade	M	74	No	63	Maintenance therapy	No	Yes	12	No	-	Yes
18	Yes	No	pt1	High Grade	M	68	Yes	28	Maintenance therapy	No	No	-	No	-	Yes
19	Yes	Yes	pt1	High Grade	M	76	-	24	Induction	Yes	No	-	No	-	Yes
20	Yes	Yes	pt1	High Grade	M	66	-	22	-	-	No	-	No	-	No
21	No	Yes	pt1	High Grade	M	78	Yes	75	No	No	No	-	No	-	Yes
22	Yes	No	pt1	High Grade	M	74	Yes	66	Maintenance therapy	No	No	-	No	-	Yes
23	Yes	Yes	pt1	High Grade	M	51	Yes	7	Induction	-	No	-	No	-	Yes
24	Yes	Yes	pt1	High Grade	M	83	No	24	No	No	No	-	No	-	Yes
25	Yes	Yes	pt1	High Grade	M	58	Yes	27	No	Yes	No	-	No	-	Yes
26	Yes	Yes	pt1	High Grade	M	66	No	11	Induction	-	No	-	No	-	Yes
27	Yes	Yes	pt1	High Grade	M	73	Yes	27	Maintenance therapy	No	No	-	No	-	Yes
28	Yes	Yes	pt1	High Grade	M	77	Yes	7	-	-	No	-	No	-	Yes
29	Yes	Yes	pt1	High Grade	M	61	Yes	10	Induction	No	No	-	No	-	Yes
30	Yes	Yes	pt1	High Grade	M	79	-	12	-	-	No	-	No	-	No

6.7 Contributions & Acknowledgements

6.7.1 Author Contributions

JLC, CL, LCP RD, and HMS designed the experimental work and provided financial support. LBC performed the laboratory work and data analysis under the supervision of HMS. RB contributed to a database study. JLC, HMS, and LBC drafted the manuscript. JLC, CLE, LCP, RD, MM, HMS, JJM. and LBC revisited the drafted version, corrected it, and made valuable suggestions. LCP and MM managed patient interventions and provided samples and medical data.

6.7.2 Acknowledgments

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). The authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV, which is financed by national funds from FCT/MCTES (UIDB/50006/2020 and UIDP/50006/2020). H. M. S. acknowledges the Associate Laboratory for Green Chemistry-LAQV (LA/P/0008/2020) funded by Fundação para a Ciência e a Tecnologia, I.P for his research contract. L. B. C. thanks the Fundação para a Ciência e a Tecnologia for his Ph.D. grant 2019 (SFRH/BD/144222/2019).

6.8 References

- [1] Ho D, Quake SR, McCabe ERB, Chng WJ, Chow EK, Ding X, et al. Enabling Technologies for Personalized and Precision Medicine. *Trends Biotechnol* 2020;38:497–518.
- [2] Rodriguez H, Zenklusen JC, Staudt LM, Doroshow JH, Lowy DR. The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. *Cell* 2021;184:1661–70. <https://doi.org/10.1016/j.cell.2021.02.055>.
- [3] Kontostathi G, Makridakis M, Bitsika V, Tsolakos N, Vlahou A, Zoidakis J. Development and validation of multiple reaction monitoring (MRM) assays for clinical applications. *Methods Mol. Biol.*, vol. 1959, Humana Press Inc.; 2019, p. 205–23. https://doi.org/10.1007/978-1-4939-9164-8_14.

- [4] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209–49. <https://doi.org/10.3322/caac.21660>.
- [5] Zhu CZ, Ting HN, Ng KH, Ong TA. A review on the accuracy of bladder cancer detection methods. *J Cancer*. 2019 Jul 8;10(17):4038–4044. <https://doi.org/10.7150/jca.28989>.
- [6] Witjes JA, Bruins HM, Cathomas R, Compérat EM, Cowan NC, Gakis G, et al. European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 Guidelines. *Eur Urol* 2021;79:82–104. <https://doi.org/10.1016/j.eururo.2020.03.055>.
- [7] Kamoun A, de Reyniès A, Allory Y, Sjö Dahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. *Eur Urol* 2020;77:420–33. <https://doi.org/10.1016/j.eururo.2019.09.006>.
- [8] Jordan B, Meeks JJ. T1 bladder cancer: current considerations for diagnosis and management. *Nat Rev Urol* 2018 161 2018;16:23–34. <https://doi.org/10.1038/s41585-018-0105-y>.
- [9] Carvalho LB, Capelo-Martínez JL, Lodeiro C, Wiśniewski JR, Santos HM. Snap-heated freeze-free preservation and processing of the urine proteome using the combination of stabilizer-based technology and filter-aided sample preparation. *Anal Chim Acta* 2019;1076:82–90. <https://doi.org/10.1016/j.aca.2019.05.051>.
- [10] Tyanova S, Temu T, Carlson A, Sinitcyn P, Mann M, Cox J. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* 2015;15:1453–6. <https://doi.org/10.1002/pmic.201400449>.
- [11] Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 2016;13:731–40. <https://doi.org/10.1038/nmeth.3901>.
- [12] Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res* 2019;18:623–32. https://doi.org/10.1021/ACS.JPROTEOME.8B00702/SUPPL_FILE/PR8B00702_SI_002.XLSX.
- [13] Klebanoff SJ, Kettle AJ, Rosen H, Winterbourn CC, Nauseef WM. Myeloperoxidase: a front-line defender against phagocytosed microorganisms. *J Leukoc Biol* 2013;93:185–98. <https://doi.org/10.1189/JLB.0712349>.
- [14] Rymaszewski AL, Tate E, Yimbessalu JP, Gelman AE, Jarzembowski JA, Zhang H, et al. The Role of Neutrophil Myeloperoxidase in Models of Lung Tumor Development. *Cancers* 2014, Vol 6, Pages 1111–1127 2014;6:1111–27. <https://doi.org/10.3390/CANCERS6021111>.

-
- [15] Castillo-Tong DC, Pils D, Heinze G, Braicu I, Sehouli J, Reinthaller A, et al. Association of myeloperoxidase with ovarian cancer. *Tumor Biol* 2014;35:141–8. <https://doi.org/10.1007/s13277-013-1017-3>.
 - [16] Gremel G, Djureinovic D, Niinivirta M, Laird A, Ljungqvist O, Johannesson H, et al. A systematic search strategy identifies cubilin as independent prognostic marker for renal cell carcinoma. *BMC Cancer* 2017;17:9. <https://doi.org/10.1186/s12885-016-3030-6>.
 - [17] Mukaida N, Sasaki SI, Baba T. Two-faced roles of tumor-associated neutrophils in cancer development and progression. *Int J Mol Sci* 2020;21:3457. <https://doi.org/10.3390/ijms21103457>.
 - [18] Jaillon S, Ponzetta A, Di Mitri D, Santoni A, Bonecchi R, Mantovani A. Neutrophil diversity and plasticity in tumour progression and therapy. *Nat Rev Cancer* 2020;20:485–503. <https://doi.org/10.1038/s41568-020-0281-y>.
 - [19] Masucci MT, Minopoli M, Carriero MV. Tumor Associated Neutrophils. Their Role in Tumorigenesis, Metastasis, Prognosis and Therapy. *Front Oncol* 2019;9:1146. <https://doi.org/10.3389/fonc.2019.01146>.
 - [20] Mosesson MW. Fibrinogen and fibrin structure and functions. *J. Thromb. Haemost.*, vol. 3, John Wiley & Sons, Ltd; 2005, p. 1894–904. <https://doi.org/10.1111/j.1538-7836.2005.01365.x>.
 - [21] Tohme S, Yazdani HO, Al-Khafaji AB, Chidi AP, Loughran P, Mowen K, et al. Neutrophil extracellular traps promote the development and progression of liver metastases after surgical stress. *Cancer Res* 2016;76:1367–80. <https://doi.org/10.1158/0008-5472.CAN-15-1591>.
 - [22] Eruslanov EB. Phenotype and function of tumor-associated neutrophils and their subsets in early-stage human lung cancer. *Cancer Immunol Immunother* 2017;66:997–1006. <https://doi.org/10.1007/s00262-017-1976-0>.
 - [23] Aggarwal V, Tuli HS, Varol A, Thakral F, Yerer MB, Sak K, et al. Role of reactive oxygen species in cancer progression: Molecular mechanisms and recent advancements. *Biomolecules* 2019;9:735. <https://doi.org/10.3390/biom9110735>.
 - [24] Smith PS, Zhao W, Robbins M. Knocking down catalase leads to increased ROS production and proliferation in glioma cells. *Cancer Res* 2006;66.
 - [25] Wang S, Song R, Wang Z, Jing Z, Wang S, Ma J. S100A8/A9 in inflammation. *Front Immunol* 2018;9:1298. <https://doi.org/10.3389/fimmu.2018.01298>.
 - [26] Simard J-C, Simon M-M, Tessier PA, Girard D. Damage-Associated Molecular Pattern S100A9 Increases Bactericidal Activity of Human Neutrophils by Enhancing Phagocytosis. *J Immunol* 2011;186:3622–31. <https://doi.org/10.4049/jimmunol.1002956>.
-

- [27] Sui X, Lei L, Chen L, Xie T, Li X, Sui X, et al. Inflammatory microenvironment in the initiation and progression of bladder cancer. *Oncotarget* 2017;8:93279–94. <https://doi.org/10.18632/ONCOTARGET.21565>.
- [28] Arantes-Rodrigues R, Pinto-Leite R, Ferreira R, Neuparth MJ, Pires MJ, Gaivão I, et al. Meloxicam in the treatment of in vitro and in vivo models of urinary bladder cancer. *Biomed Pharmacother* 2013;67:277–84. <https://doi.org/10.1016/J.BIOPHA.2013.01.010>.
- [29] Shinde-Jadhav S, Mansure JJ, Rayes RF, Marcq G, Ayoub M, Skowronski R, et al. Role of neutrophil extracellular traps in radiation resistance of invasive bladder cancer. *Nat Commun* 2021;12:1–14. <https://doi.org/10.1038/s41467-021-23086-z>.
- [30] Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res* 2019;47:D442–50. <https://doi.org/10.1093/nar/gky1106>.

Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

PATHWAY-GUIDED MONITORING OF THE DISEASE COURSE IN BLADDER CANCER WITH LONGITUDINAL URINE PROTEOMICS

Luís B. Carvalho, J. L. Capelo, Carlos Lodeiro, Rajiv Dhir, Luis Campos Pinheiro, Hugo López-Fernández, Gonçalo Martins, Mariana Medeiros, Fernando Díaz & Hugo M. Santos,

Nature Communications Medicine 3, 8 (2023)
Doi: 10.1038/s43856-023-00238-4

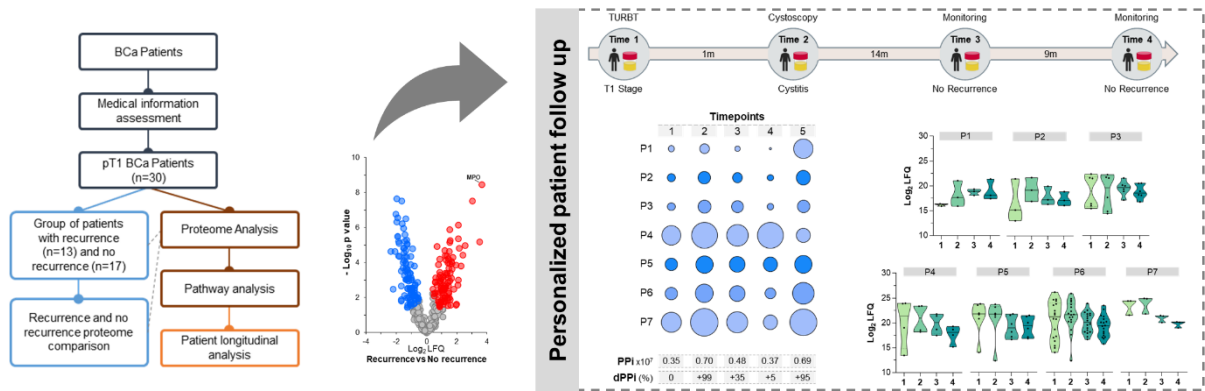


Figure 7-1. Graphical abstract.

7.1 Abstract

Monitoring bladder cancer over time requires invasive and costly procedures. Less invasive approaches are required using readily available biological samples such as urine. In this study, we demonstrate a method for longitudinal analysis of the urine proteome to monitor the disease course in patients with bladder cancer. We compared the urine proteomes of patients who experienced recurrence and/or progression ($n = 13$) with those who did not ($n = 17$). We identified differentially expressed proteins within various pathways related to the hallmarks of cancer. The variation of such pathways during the disease course was determined using our differential personal pathway index (dPPi) calculation, which could indicate disease progression and the need for medical intervention. Seven hallmark pathways are used to develop the dPPi. We demonstrate that we can successfully longitudinally monitor the disease course in bladder cancer patients through a combination of urine proteomic analysis and the dPPi calculation, over a period of 62 months. Using the information contained in the patient's urinary proteome, the dPPi reflects the individual's course of bladder cancer, and helps to optimise the use of more invasive procedures such as cystoscopy.

7.2 Introduction

Bladder cancer (BCa) is among the worst neoplasms because of its high incidence and mortality [1,2] and because cystoscopy is necessary to diagnose BCa. If BCa is confirmed, the patient undergoes surgery, then the patient must undergo a control cystoscopy every three months [3,4]. Cystoscopy is an expensive intervention that requires a surgical block and a minimum of two physicians and one nurse. However, an even worse problem is the detrimental effect on the patient's state of mind, as cystoscopy is highly invasive and is performed through the urethra. Furthermore, some patients experience a rapid recurrence; thus, the control intervention is performed too late, and transurethral resection of the bladder tumour is needed. Therefore, new approaches to diagnose BCa and follow-up are needed.

The biochemical changes that occur at the protein level in an individual after they develop a disease provides information, and with this information, we can adjust therapy, provide follow up medical care and make precision medicine truly personalized [5]–[6]. The concept of performing longitudinal profiling with individual proteomes has been described in the literature [7–9]; however, to date, there is no easy way to achieve this in a simple, straightforward and robust manner. The possibility of following the course of BCa using the urinary proteome

has not yet been described in the literature. To overcome this gap, we explored the evolution of the urinary proteome of BCa patients using high-resolution mass spectrometry-based proteomics in conjunction with advanced bioinformatics tools. Thus, we introduced, as far as we know, a new concept named the differential personal pathway index (dPPI), which uses the variation in the expression of selected biochemical pathways, which is calculated using a large number of urine proteins that are linked to Hanahan and Weinberg's biological hallmarks of cancer (HWhc [10,11]). Thus, based on variation in the levels of tens of signature urine proteins, the dPPI reflects both the response of the individual's urine proteome to whatever course the disease takes and any medical care that ensues. The variation reflects the status of the biochemical pathways to which the proteins correspond, and these pathways are specifically chosen to combine multiple sources of diagnostic information. Thus, in this work, we showed how to monitor the evolution of BCa in a comprehensive yet straightforward and personalized manner using the patient's urine proteome. Essentially, the more positive the dPPI is, the more severe the course of BCa; the more negative the dPPI is, the better the clinical outlook. Therefore, the dPPI flags the need for medical intervention at an early stage. This work further demonstrates this concept using six patients diagnosed with T1-BCa who were followed for 62 months.

7.3 Experimental Section

7.3.1 Patients

Human mid-stream second void morning urine specimens were collected from patients with a pT1 bladder cancer diagnosis that was confirmed by pathological examination. All patients were informed about the study and provided a signed informed consent according to the policies by the Central Lisbon Hospital Center Ethics Committee. The approval ethics committee number is 669/2018. The patients that were enrolled in the study were selected based on the following criteria: (a) Inclusion—clear bladder cancer diagnosis, and (b) Exclusion—no records of urinary cancer history, no HIV, no organ transplant, and no recent chemo/radiotherapy. We compared the proteome of 30 patients from two groups—those who had experienced BCa recurrence and/or progression (a total of 13) and those who had not (a total of 17). Fig.7-2A depicts the steps from proteome analysis to pathway, and clinical data interpretation, while the medical information for the cohorts of patients is presented in Fig.7-2B. Furthermore, we followed six patients for up to 62 months.

7.3.2 Sample treatment

Urine samples were collected in 50 mL centrifuge tubes (DB Falcon) that contained 38 mg of boric acid (Sigma–Aldrich) to prevent bacterial growth [12]. Samples with haematuria were not included in our study. The urine samples were centrifuged at $5,000 \times g$ for 20 min to remove cell debris, and 10 mL aliquots of the resulting supernatants were stored at -60°C until further use. An aliquot of urine (10 mL) was concentrated by centrifugal ultrafiltration using a VivaSpin 15R (10 kDa MWCO, Sartorius) at $5,000 \times g$ for 20 min. Finally, the concentrated urinary proteomes were quantified via a 96 well plate Bradford protein assay (Sigma–Aldrich) using bovine serum albumin to generate the calibration curve.

7.3.3 Filter-aided sample preparation (FASP)

Filter Aided Sample Preparation (FASP) was used to perform urinary proteome digestion, following a modified FASP method [5,13–15]. Briefly, different amounts of protein ranging from 50 and 100 μg were diluted with H_2O to obtain 200 μL and an additional 200 μL of 8 M urea (Sigma–Aldrich), 75 mM Tris-HCl (Sigma–Aldrich), 100 mM NaCl (Sigma–Aldrich), and 0.02% SDS (Sigma–Aldrich). Then this protein solution was loaded onto a VivaSpin 500 (10 kDa MWCO, Sartorius) and centrifuged for 20 min at $14,000 \times g$. Each sample was processed in duplicate. The proteins in the membrane were washed with 200 μL of 8 M urea 25 mM AmBic solution and then centrifuged for 20 min at $14,000 \times g$. Protein disulfide bonds were reduced by adding 200 μL of 50 mM dithiothreitol (DTT) in 8 M urea and 25 mM AmBic and incubated for 60 min at 37°C . Then, centrifugation for 20 min at $14,000 \times g$ was performed, and the sample was alkylated for 45 min in the dark by adding 100 μL 50 mM iodoacetamide (IAA) in 8 M urea and 25 mM AmBic solution. Subsequently, centrifugation for 20 min at $14,000 \times g$ was performed, and then the samples were washed twice with 200 μL of 25 mM AmBic. Finally, protein digestion was performed by adding 100 μL of trypsin solution (1:30 trypsin to protein ratio) prepared in 12.5 mM AmBic. Protein digestion was performed overnight (approximately 16 hours) at 37°C , and then the peptides were collected by 20 min centrifugation at $14,000 \times g$, followed by two additional membrane washing steps with 100 μL of 3% (v/v) acetonitrile that contained 0.1% (v/v) aqueous formic acid and centrifugation of 20 min at $14,000 \times g$. Finally, peptides were transferred to a 500 μL microtube, dried and stored at -20°C until further analysis by Nano-LC–MS/MS.

7.3.4 LC-MS/MS

LC-MS /MS analysis was carried out using an Ultimate 3000 nano-LC system coupled to an Impact HD (Bruker Daltonics) with a CaptiveSpray nanoBooster that used acetonitrile as a dopant. The protein digests were resuspended in 100 μL of 3% (v/v) acetonitrile containing 0.1% (v/v) aqueous formic acid (FA) and were sonicated for 10 min using an ultrasonic bath at 100% ultrasonic amplitude, and 35 kHz ultrasonic frequency. Afterwards, the samples were quantified using a Pierce™ quantitative colorimetric peptide assay. Then, 3 μL of protein digest containing 563 ng of peptides was loaded onto a trap column (Acclaim PepMap100, 5 μm , 100 Å, 300 μm i.d. $\times$ 5 mm) and desalted for 5 min from 3% to 5% B (B: 90% acetonitrile 0.08% FA) at a flow rate of 15 $\mu\text{L}\cdot\text{min}^{-1}$. The peptides were separated using an analytical column (Acclaim™ PepMap™ 100C18, 2 μm , 0.075 mm i.d $\times$ 150 mm) with a linear gradient at 300 $\text{nL}\cdot\text{min}^{-1}$ (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) acetonitrile and 0.08% (v/v) FA) 5-90 min from 5% to 35% mobile phase B, 90-100 min linear gradient from 35% to 95% of mobile phase B, 100-110 min 95% B. Chromatographic separation was carried out at 35°C. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8-32 Hz) with a cycle time of 3.0 s and active exclusion (precursors were excluded from the precursor selection for 0.5 min after the acquisition of the 1 MS/MS spectrum; the intensity threshold for fragmentation was 2,500 counts). As, the active exclusion was set to one, we reconsidered the precursor and whether the intensity of a precursor increased by a factor of three; this mass was taken from the 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. The mass spectrometry proteomics data were deposited in the ProteomeXchange Consortium [16] via the PRIDE [17] partner repository with the data set identifier PXD025139. To confirm myeloperoxidase label-free quantification levels, validation was performed using a commercially available ELISA kit (ab136943 - Myeloperoxidase (MPO) Human ELISA kit from Abcam). The method was carried out according to the standard protocol described by the manufacturer.

7.3.5 Bioinformatics and statistical analyses

Relative label-free quantification was performed using the precursor signal intensity method and delayed normalization, MaxLFQ, on MaxQuant software V2.0.3.0[18]. All raw files were processed in a single run with default parameters [19,20]. Database searches were performed using Andromeda against the human UniProt UP000005640_9606 database (20,600 sequences; 11,395,157 residues, downloaded on April 27, 2021) [21]. Searches were configured

with cysteine carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidation as variable modifications. We set the false discovery rate (FDR) to 0.01 for protein and peptide levels with a minimum length of seven amino acids for peptides. The FDR was determined by searching a reverse database. Enzyme specificity was set as C-terminal to arginine and lysine as expected using trypsin. A maximum of two missed cleavages were allowed. Data processing was performed using Perseus (version 1.6.15.0) with default settings [22]. Protein group LFQ intensities were log₂-transformed, and the quantitative profiles were filtered for missing values with the following settings: min valid percentage of 50% in at least one group and values greater than 0. To overcome the obstacle introduced by missing LFQ values, the missing values were imputed using the parameters, with = 0.5 and downshift = 1.8. The list of differentially expressed proteins was obtained with a two-tailed Student's t test between the two groups (permutation-based FDR 0.05 and S0 of 0.1), and then the log ratios were calculated as the difference in average log₂ LFQ intensity values between the tested conditions. Perseus was also used to obtain clusters, using average linkage, no constraint, preprocess with k-means and Euclidean distance between column trees. GraphPad Prism 8.2.1. was employed for graphic construction and statistical analysis. Cytoscape V3.8.2 and the application ClueGo V2.5.8 was used for pathway enrichment analysis. Reactome and biological process GO terms were used as ontologu databases.

7.3.6 dPPI calculation

The dPPI is calculated from the numerical output of proteomics analysis of urine at different time points, as follows:

$$dPPI = \frac{\sum_{nx=1}^{nx=i} Protnx LFQ_{ty} - \sum_{nx=1}^{nx=i} Protnx LFQ_{t1}}{\sum_{nx=1}^{nx=i} Protnx LFQ_{t1}} \quad (i)$$

where $\sum_{nx=1}^{nx=i} Protnx LFQ_{ty}$ accounts for the sum of all label-free quantitation (LFQ) values of proteins of interest at a given time t_y , while $\sum_{nx=1}^{nx=i} Protnx LFQ_{t1}$ accounts for all LFQ measurements of the same proteins at the time of disease onset t_1 .

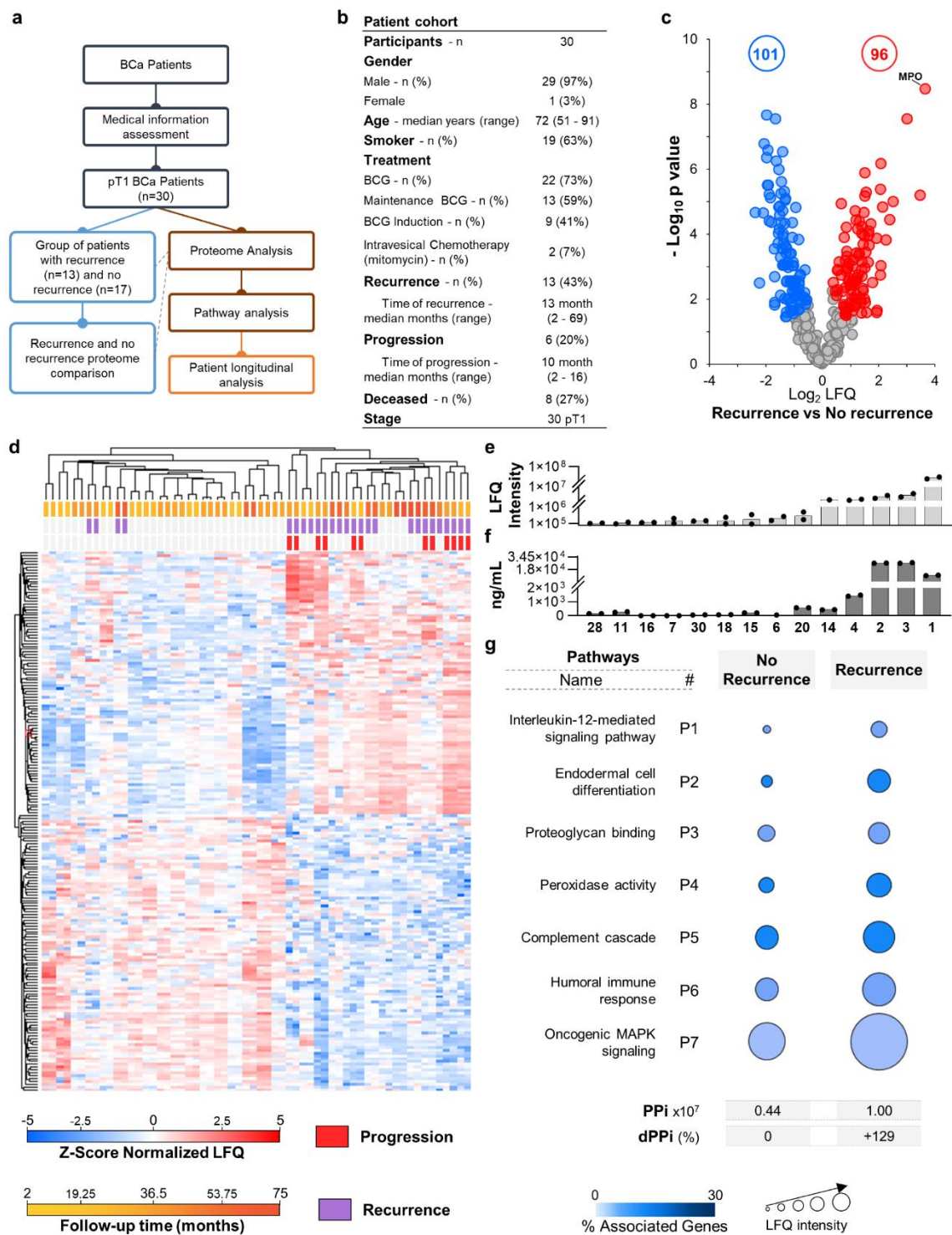


Figure 7-2. Proteome analysis of 30 pT1 bladder cancer patients. A Workflow of the experimental design. **B** Clinical information of the patient cohort used in the study. **C** Volcano plots showing statistically significant changes in protein levels in T1-BCa patients with recurrences and/or progression (n=13 patients, two technical replicates per patient) compared to the group with nor recurrence (n=17 patients, two technical replicates per patient), according to Student's t test (FDR 0.05 and S_0 of 0.1). Dots represent proteins that show statistically significant increases (red), decreases (blue) nonstatistically significant changes (grey). The total numbers of proteins that exhibited statistically significant differential expression are shown in the corresponding-coloured circles. **D** Hierarchical clustering of the

T1-BCa patients into two groups, as expected, recurrence and no recurrence (average linkage, no constraint, pre-processing with k-means, and Euclidean distance, $n=30$ patients, two technical replicates per patient). **E & F.** Validation of mass spectrometry data using myeloperoxidase (MPO) protein label-free quantification (LFQ) and ELISA (ng/mL). Bars represent the average of two technical replicates. The MPO protein was recently proposed to be a marker of poor prognosis in lung [27] and ovarian [28] carcinomas. To the best of our knowledge, no studies report MPO overexpression as a prognostic tool for BCa carcinoma. **G.** Pathways selected for calculating the Personal Pathway Index, after applying the guidelines provided by Hanahan and Weinberg [10,11] to the proteins showing statistically significant differential expression derived from the volcano plot in **C** (Supplementary Data 1). Differentially expressed proteins were grouped according to their functions, and networks were analysed in ClueGo (v 2.5.8), and Cytoscape (v 3.8.2). The LFQ average for all proteins involved in the pathway is expressed as a circle area plot. The maximum number of proteins identified in one single urine of our set was 494. The number of unique proteins considering all the samples of our set was about 754. For our studies, we have chosen the proteins that are found common in at least 50% of the urines of each group. This way, the number of proteins used in Fig.7-2 (recurrence versus no recurrence) was 380.

7.4 Results & Discussion

7.4.1 Selection of pathway hallmarks of cancer

The differentially expressed proteins between the two groups ($n = 197$, Fig.7-2C) were submitted to clue-go analysis (cystoscope) to select the biochemical pathways involved in recurrence and/or progression. The 197 differentially expressed proteins distinguished 93.5% of patients with recurrence and 100% of patients with progression, as shown in Fig.7-2D. The statistically significant pathways were cross-checked against the HWhc [10,11], and seven pathways were selected. These pathways involve 35 related proteins (Fig.7-2G. and Table 7-1). These 35 proteins were used to calculate the dPPi as defined in (I). Four of these 35 proteins have been described in literature as potential biomarkers for bladder cancer (highlighted in red color in Table 7-1) [23–25]. Two of these four proteins namely fibrinogen β chain (FGB) [25,26] and fibrinogen γ chain (FGG) [25] have been described with increased expression in urine of BCa patients. In addition, we found the protein S100A9 which has been described as a biomarker candidate in plasma of BCa patients [23]. Protein myeloperoxidase (MPO) was found to exhibit the largest overexpression in patients with recurrence (Fig.7-2C). MPO has been recently proposed to be a marker of poor prognosis in lung [27] and ovarian [28] carcinomas. Therefore, MPO was used for quality control to validate the mass spectrometry data using ELISA as an orthogonal method (Fig.7-2E and Fig.7-2F). Next, we investigated how the status of the seven pathways changed over time by following six patients with BCa for up to 40 months. A comprehensive flow chart of the process is provided in Fig.7-6.

Table 7-1. Pathways and proteins used on the longitudinal follow up of bladder cancer patients. Red color highlights the potential biomarkers for bladder cancer describe in literature.

Path-way Number	GOID	Name	Ontology Source	Associated Proteins Found
P1	GO:0035722	Interleukin-12-mediated signaling pathway	GO_BiologicalProcess-EBI-Uni-Prot-GOA-ACAP-ARAP_08.05.2020_00h00	[LCP1, PPIA, S100A9]
P2	GO:0035987	Endodermal cell differentiation	GO_BiologicalProcess-EBI-Uni-Prot-GOA-ACAP-ARAP_08.05.2020_00h00	[FN1, MMP8, MMP9]
P3	GO:0043394	Proteoglycan binding	GO_MolecularFunction-EBI-Uni-Prot-GOA-ACAP-ARAP_08.05.2020_00h00	[AZU1, CFH, CTSB, FN1, HRG, LCP1, NID1]
P4	GO:0004601	Peroxidase activity	GO_BiologicalProcess-EBI-Uni-Prot-GOA-ACAP-ARAP_08.05.2020_00h00	[CAT, HBA1, HBB, HBD, MPO]
P5	R-HSA:166658	Complement cascade	REACTOME_Pathways_08.05.2020	[C3, C4A, C9, CD55, CFH, F2]
P6	GO:0006959	Humoral immune response	GO_BiologicalProcess-EBI-Uni-Prot-GOA-ACAP-ARAP_08.05.2020_00h00	[A2M, APCS, AZU1, C3, C4A, C9, CAMP, CD55, CFH, F2, FGB , HPX, HRG, IGHG1, IGKV2D-30, KRT1, LCN2, LTF, LYZ, PRTN3, S100A9]
P7	R-HSA:6802957	Oncogenic MAPK signaling	REACTOME_Pathways_08.05.2020	[FGB , FGG , FN1]

7.4.2 Longitudinal follow-up of BCa patients

In the following section, we will discuss three of the six patients in detail, and the other three are discussed in the Figs.7-8, 7-9 and 7-10.

During the diagnosis period, Patient I presented with symptoms that matched BCa. One reference urine sample was taken before urinary and bladder exploration was performed (see Fig.7-3A for a comprehensive analysis timeline). Then, the urine proteome was extracted and analysed via mass spectrometry, and the PPI was calculated as explained in (I) and is presented in Fig.7-3D. Physicians diagnosed this patient with T1-stage BCa and performed a transurethral

resection of the bladder tumour, TURBT. Then, one month later, a second analysis of the urinary proteome revealed a difference of +35% between the PPI at time point 1 and at time point 2 (Fig.7-3D, PPI₂ – PPI₁), suggesting that the patient had not improved. In fact, all the hallmark pathways were obtained with higher values. This prompted us to call the patient and perform a medical inspection of the bladder, and the patient was diagnosed with cystitis; thus, no surgical intervention was needed. Another advantage of this type of analysis is that the mass spec LFQ values of individual protein biomarkers can also be used to obtain additional information about the course of the disease (Fig.7-3C). For instance, APOA4, a urinary candidate biomarker for the presence of BCa, exhibited lower value at sampling point 2 which is a good indicator in terms of cancer evolution. Increased levels of APOA4 in urine have been linked to bladder cancer [29]. The same applies to the biomarker CDH1, in which an increment is also a good signal. On the other hand, MMP9 is a biomarker linked to inflammation, and MMP9 exhibited an augmented value at sampling point 2, which is a negative indicator that is not necessarily related to cancer [30].

In conclusion, the evolution of these biomarkers is consistent with cystitis rather than a recurrence or progression. Further information supporting the consistency of this diagnostic is the complex MMP9/NGAL described augmented in urine of children with acute cystitis [30]. At sampling point 2, both proteins are augmented. This result shows that dPPI is useful as an alert tool to follow individual variations in the urine proteome caused by disease. Another cystoscopy three months later was performed and revealed no evidence of recurrence. Fourteen months later, a third urine sample was monitored again, showing a dPPI of -80% relative to time 1 (Fig.7-3D3), indicating the key immunological and inflammatory responses were down-regulated (Fig.7-3E), which thus reflected the excellent progression of this patient, as it was confirmed by medical analysis. Nine months later, a fourth urine analysis revealed an increase in the dPPI index from -80% to -61%, with a clear increase in the MAPK oncogenic signalling pathway. This pathway has been reported to be involved in BCa development [31], and over-expression of MAPK signalling proteins has been linked to a mutation in the P53 gene and poor prognosis [32]. Furthermore, the variations in APOA4 and CDH1 indicated that the cancer progression had worsened. Additionally, the MMP9 variation showed an increase in the inflammatory response. Altogether, it was concluded that the patient was worsening. A new urine analysis was requested to confirm this evolution before medical intervention was performed.

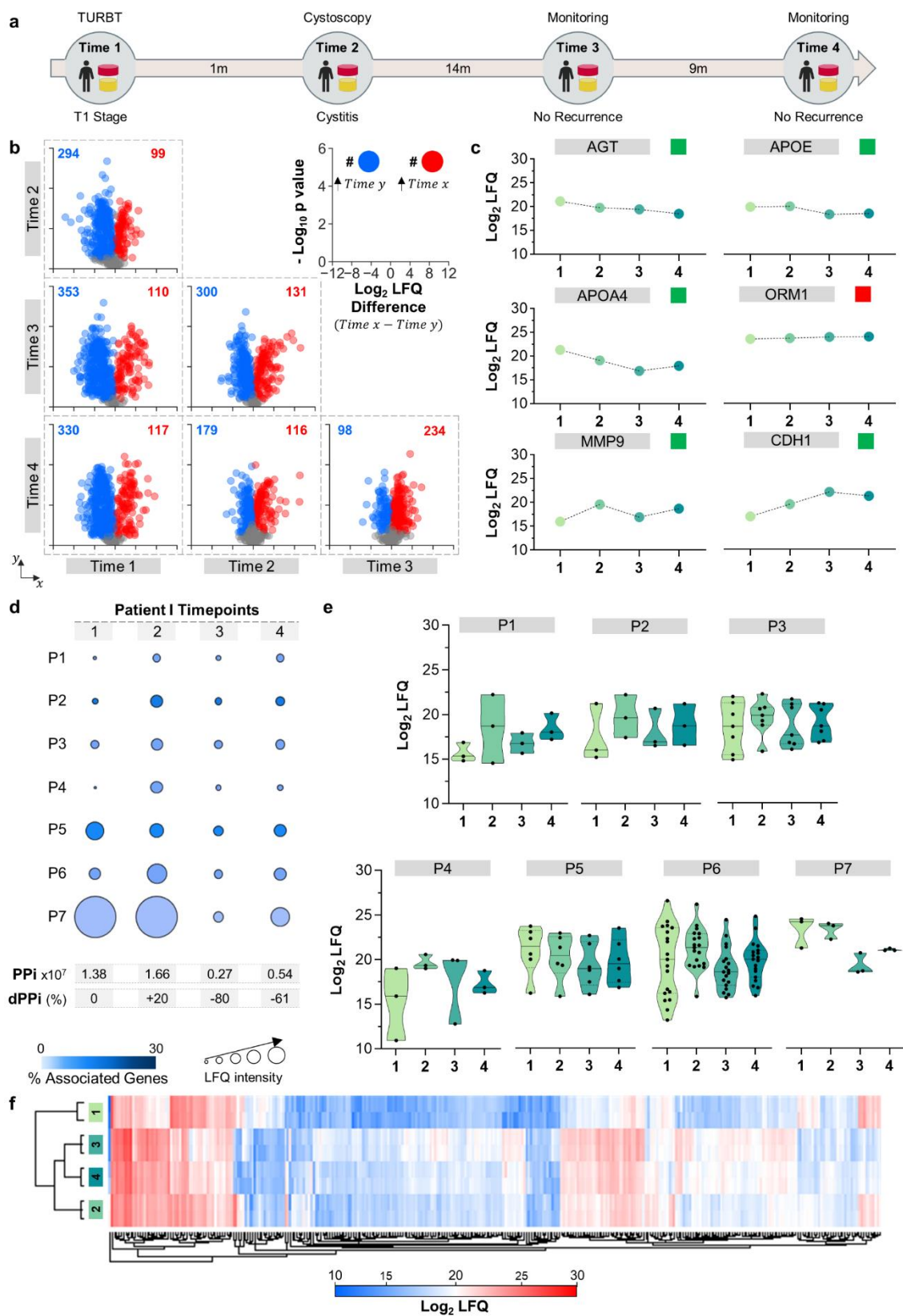


Figure 7-3. Longitudinal analysis of the urinary proteome of patient I. **A** Timeline of the disease course in Patient I shows the intervals in months (m) between medical interventions and urine sampling (time 1 to time 4). TURBT: transurethral resection of bladder tumour; Re-TURBT: repeated transurethral resection of bladder tumour. **B**.

Volcano plot matrix showing proteome changes in the urine samples at time points 1-4. The dots represent the proteins that showed a statistically significant increase (red), decrease (blue) or nonstatistically significant changes (grey) according to the Student's *t* test (FDR 0.05 and S0 of 0.1). **C.** Variation in the following known protein biomarkers for bladder cancer: angiotensinogen (AGT)[33], apolipoprotein E (APOE)[34], matrix metalloproteinase-9 (MMP9)[35], apolipoprotein A-IV (APOA4)[29], alpha-1-acid glycoprotein 1 (ORM1)[36] and cadherin-1 (CDH1)[37,38]. The green and red squares indicate whether the variation in the urine of Patient I matched or not, respectively, with trends reported for each marker in the literature. Dots represent the average of two biological samples with two technical replicates each. **D.** Personal pathway index (PPI) at each time point and estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; Pa: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **E.** Distribution and density variation protein LFQ values (including two biological replicates) at each sampling point for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **F.** Hierarchical clustering of the three urinary proteomes of Patient I. Protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

Patient II is an interesting case due to their frequent recurrences (Fig.7-4). The first proteome analysis of the urine was performed 64 months after the onset of BCa. Then, the PPI for this urine, which is presented in Fig.7-4D1, was calculated. At this point, the patient experienced a TURBT intervention. Two months later, a new urine analysis revealed a dPPI variation of +99%, (Fig.7-4D2), indicating that the patient's course was worsening. It is worth noting the increase in the expression of the MAPK pathway (P7, Fig.7-4D2). Based on these data, the patient underwent a urological assessment that revealed another tumour recurrence, confirming that the alert from the dPPI approach is valuable. Another TURBT was performed, and BCa was documented.

After 35 months, a routine surveillance assessment was performed, and the urine proteome analysis revealed a dPPI of +35% relative to the time one reference value. The patient thus underwent another cystoscopy evaluation, which agreed with the dPPI alert and demonstrated the recurrence of BCa. Consequently, a new TURBT was performed. A fourth urine proteomic analysis was completed 15 months later (Fig.7-4D4), which showed a change in the dPPI value of +5%, suggesting that the tumour had not evolved positively. Our prediction was again confirmed by cystoscopy as recurrence was diagnosed. Finally, ten months later an increase of +95% in the dPPI suggested that the disease was worsening. According to our medical intervention advice, the patient underwent one cystoscopy that confirmed our result. The patient was submitted to a TURBT.

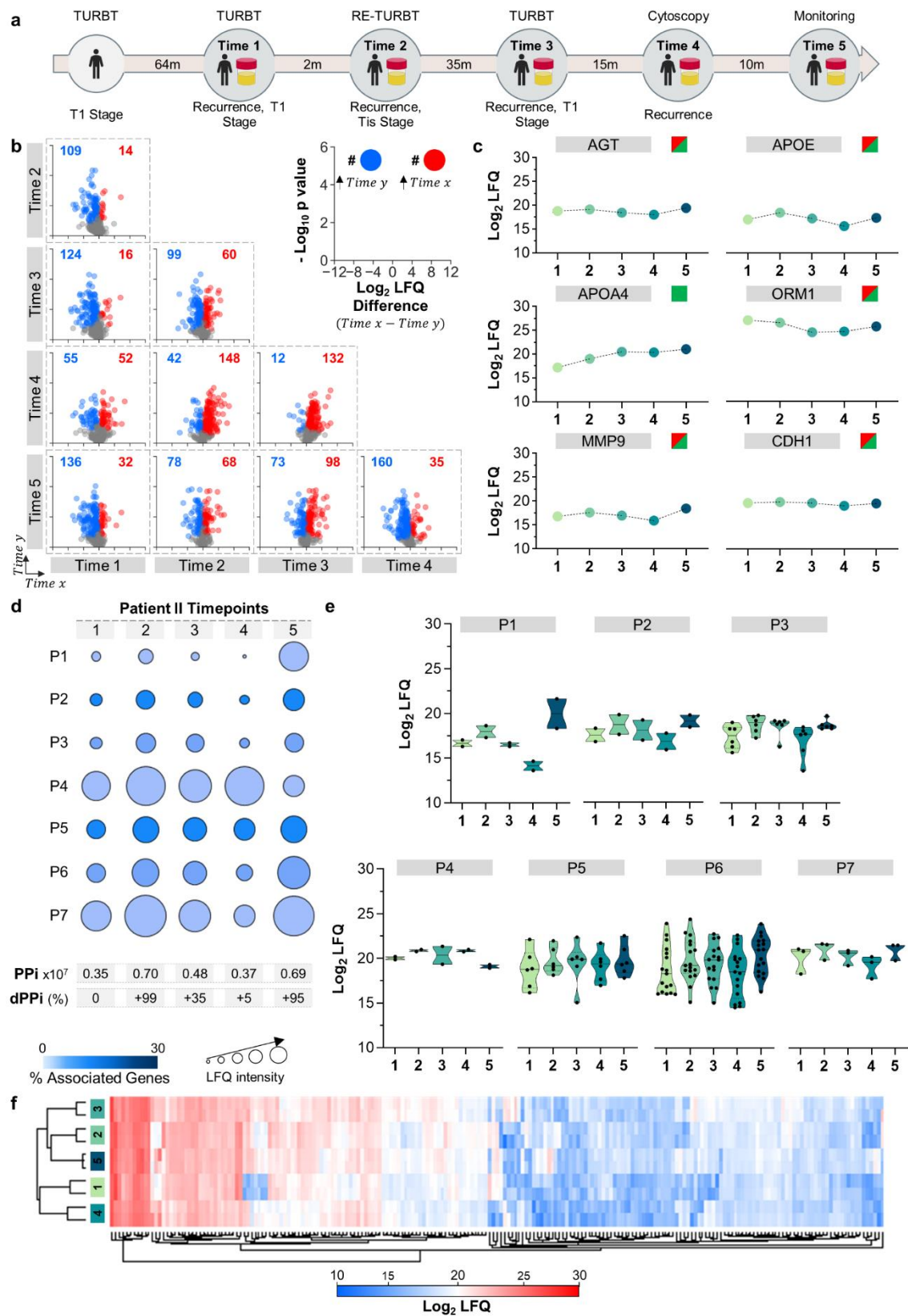


Figure 7-4. Longitudinal analysis of the urinary proteome of patient II. **A** Timeline of disease course in Patient II showing the intervals in months (m) between medical interventions and urine sampling (time 1- time 5). TURBT: transurethral resection of bladder tumour; Re-TURBT: repeated transurethral resection of bladder tumour. **B**.

Volcano plot matrix showing proteome changes in urine collected at time points 1-5. The dots represent proteins showing statistically significant increases (red), decreases (blue) or nonstatistically significant changes (grey) according to the Student's *t* test (FDR 0.05 and *S*₀ of 0.1). **C.** Variation in the known protein biomarkers for bladder cancer, including angiotensinogen (AGT)[33], apolipoprotein E (APOE)[34], matrix metalloproteinase-9 (MMP9)[35], apolipoprotein A-IV (APOA4)[29], alpha-1-acid glycoprotein 1 (ORM1)[36] and cadherin-1 (CDH1) [37,38]. The green and red squares indicate whether the variation in the urine of Patient II matched or not, respectively, with trends reported for each marker in the literature. Dots represent the average of two biological samples with two technical replicates each. **D.** Personal pathway index (PPI) at each time point and the estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **E.** Distribution and density variation protein LFQ values (including two biological replicates) at each sampling point for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **F.** Hierarchical clustering of the four urinary proteomes for Patient II. The protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

Patient III (Fig.7-5) presented with symptoms of BCa, and a TURBT was performed because the cystoscopy results were positive. Histopathological analysis documented stage T1 BCa. The patient's urine sample was collected prior to medical intervention, and the dPPI was calculated, as it is shown in Fig.7-5D1. Afterwards, the patient underwent three more TURBT procedures within the following four months. A second urine sample was taken one month after the fourth TURBT was performed (time point 2). The dPPI augmented a +36% (Fig.7-5D2), presenting seven pathways that were increased. Again, the value of the dPPI approach was illustrated by the variation observed in the MAPK signalling pathway (P7 in Fig.7-5D).

Our diagnosis alert was confirmed as the medical diagnosis indicated that the disease had progressed to the T2-stage, and a radical cystectomy was performed. The urine proteome was analysed again eight and eleven months after cystectomy, as shown in Fig.7-5D3 and 7-5D4, respectively. The expression levels of all the pathways were found to be lower after resection of the bladder (Fig.7-5D). The dPPI decreased 63% and 79% between onset and time points three and four, respectively. The latter is consistent with the recovery experienced by this patient. Currently, this patient is in good health. The profile of the four proteomes since the onset of disease were compared (Fig.7-5F) and clearly show that the urines were clustered into two groups (before and after radical cystectomy). Three further patient examples are presented in the supplementary Fig. 7-8, 7-9 and 7-10.

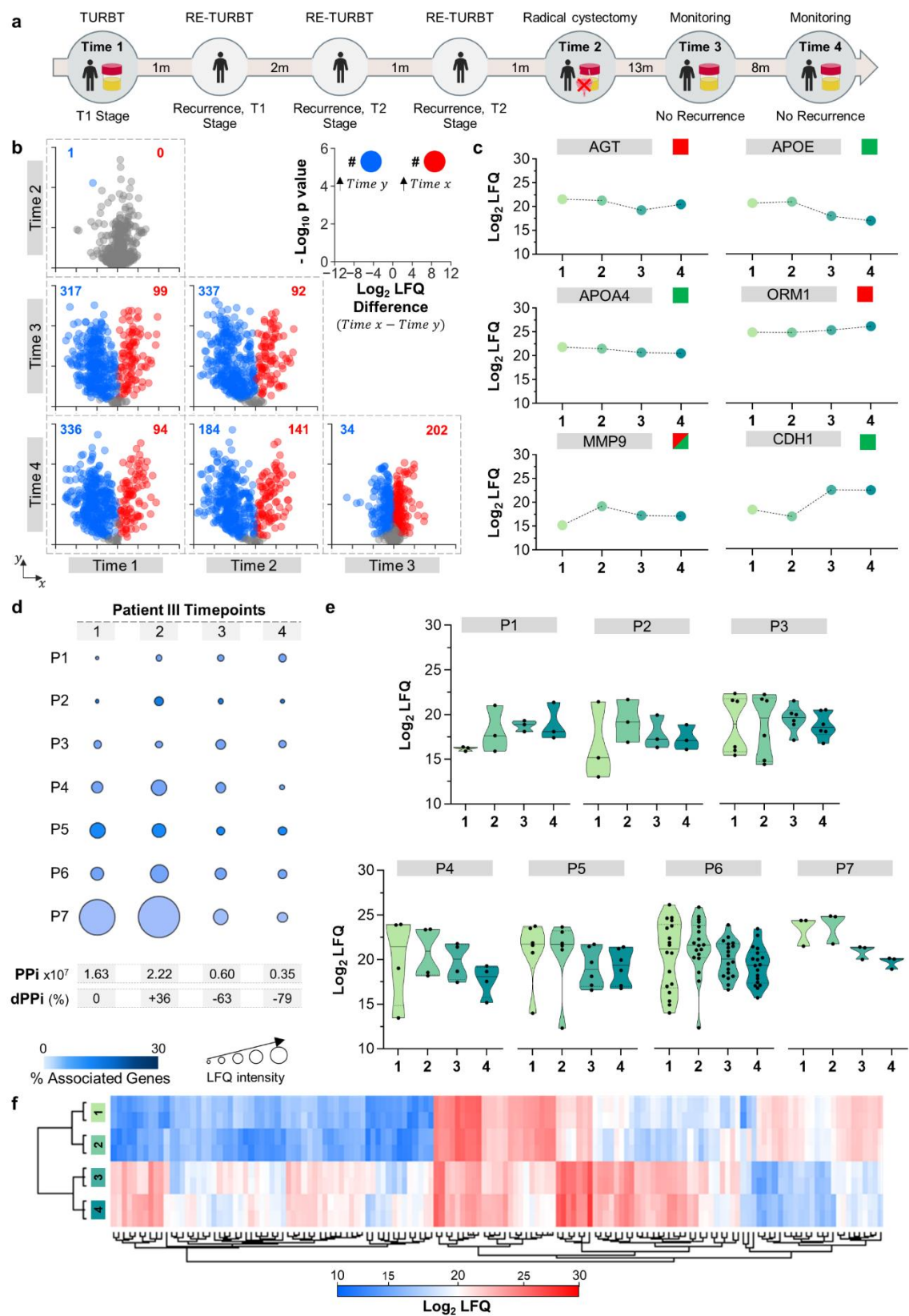


Figure 7-5. Longitudinal analysis of the urinary proteome of patient III. **A.** Timeline of the disease course in Patient III showing intervals in months (m) between medical interventions and urine sampling (time 1 to time 4). TURBT: transurethral resection of bladder tumour; Re-TURBT: repeated transurethral resection of bladder tumour. **B.**

Volcano plot matrix showing proteome changes in urine collected at time points 1–4. The dots represent proteins showing statistically significant increases (red), decreases (blue) or nonstatistically significant changes (grey) at the different time points according to Student's t test (FDR 0.05 and S_0 of 0.1). **C.** Variation in the following known protein biomarkers for bladder cancer: angiotensinogen (AGT)[33], apolipoprotein E (APOE)[34], matrix metalloproteinase-9 (MMP9)[35], apolipoprotein A-IV (APOA4)[29], alpha-1-acid glycoprotein 1 (ORM1)[36] and cadherin-1 (CDH1)[37,38]. The green and red squares indicate whether the variation in the urine of Patient III matched or not, respectively, with trends reported for each marker in the literature. Dots represent the average of two biological samples with two technical replicates each. **D.** Personal pathway index (PPI) for each time point and the estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **E.** Distribution and density variation protein LFQ values (including two biological replicates) at each sampling point for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines.. **F.** Hierarchical clustering of the four urinary proteomes of Patient III. The protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

Table 7-2. The dPPI index column shows the index calculated as explained in the methods section. The dPPI intervention column shows whether medical intervention is necessary or not (Yes/No). The medical procedure column indicates the medical intervention performed for the patient, while the outcome column highlights the diagnosis given by the physicians. The column named dPPI intervention true? Confirms whether the intervention that was recommended by the dPPI index was necessary or not (Yes/No). Note that patient numbers 4,5 and 6 are described in the supplementary Fig. 7-8, 7-9 and 7-10.

Patient	Time point	dPPI index	dPPI Intervention ?	Medical procedure	Outcome	dPPI intervention true?
1	2	20	Yes	Cystoscopy	Cystitis	Yes
1	3	-80	No	Medical Monitoring*	No tumor	Yes
1	4	-61	No	Medical Monitoring*	No tumor	Yes
2	2	99	Yes	Cystoscopy/RE-TURBT	Tis stage recurrence	Yes
2	3	35	Yes	Cystoscopy/TURBT	T1 stage recurrence	Yes
2	4	5	Yes	Cystoscopy	Recurrence (no AP exam)	Yes
2	5	95	Yes	Medical Monitoring*	Recurrence (no AP exam)	Yes
3	2	36	Yes	Cystoscopy/RE-TURBT	T2 Stage recurrence	Yes
3	3	-63	No	Medical Monitoring*	No recurrence	Yes
3	4	-79	No	Medical Monitoring*	No Recurrence	Yes
4	2	-80	No	Medical Monitoring*	No Recurrence	Yes
4	3	-56	No	Medical Monitoring*	No Recurrence	Yes
5	2	-3	Yes	Cystoscopy/RE-TURBT	Ta stage recurrence	Yes
5	3	-2	Yes	Medical Monitoring*	Nonneoplastic Bladder Mass	Yes
6	2	-41	No	Cystoscopy	No recurrence	Yes

6	3	-7	No	Medical Monitoring*	No Recurrence	Yes
---	---	----	----	---------------------	---------------	-----

Table 7-2 shows the dPPi index for each time point for each patient in a comprehensive manner. First, the urine was monitored, the dPPi was calculated, and an alert for intervention was set as yes or no. Next, the patient was examined by the physician’s team through a medical procedure as described in the table. The diagnostic provided was then compared with the alert provided by the dPPi. As shown in Table 7-2, 100% of the alerts (yes or no) were accurate.

In this work, we demonstrated how using the dPPi algorithm to assess the urine proteome is a robust tool to (i) monitor the course of BCa disease and (ii) guide clinical decisions regarding intervention. As the disease progresses, the levels of selected cancer hallmark pathways change. This variation is used to monitor the course of BCa disease in a simple, straightforward, and robust manner. This study would benefit from a large cohort of patients and will need to be replicated on a larger scale to validate the reported conclusions.

7.5 Conclusions

Bladder cancer must be closely monitored for progression, but this requires expensive and invasive procedures such as cystoscopy. Less invasive procedures using readily available samples such as urine are needed. Here, we present an approach that measures the levels of various proteins in the urine. We compare protein levels at different points during the disease course in patients with bladder cancer, and show this helps to flag disease recurrence and the need for medical intervention. Our approach could help clinicians to determine which patients require more invasive testing and treatment.

7.6 Supplementary material

7.6.1 Supplementary data

The mass spectrometry proteomics data that support the findings of this study have been deposited in ProteomeXchange Consortium [16] via the PRIDE [17] partner with the PXD025139 accession codes.

7.6.2 Supplementary figures

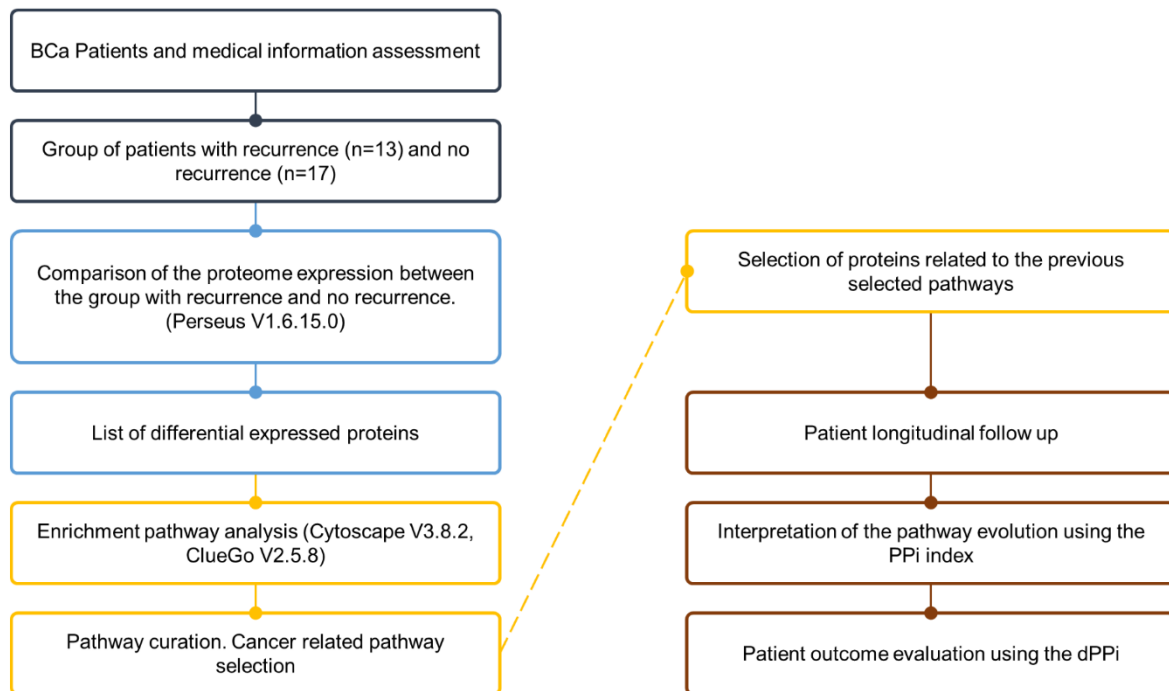


Figure 7-6. Bioinformatics analysis workflow. List of differentially expressed proteins was obtained with a Student's t test (FDR 0.05 and S0 of 0.1). Reactome and biological process GO terms for the enrichment pathway analysis were used as ontologies databases. The curation of pathways was carried out based on the enrichment pathways analysis from ClueGo. Pathways with 60% or more up-regulated proteins and lower p-values were selected. Furthermore, those pathways were crosscheck with a bibliography to select the ones correlated with Hanahan and Weinberg biological hallmarks of cancer [10,11] as shown in as shown in Table 7-1.

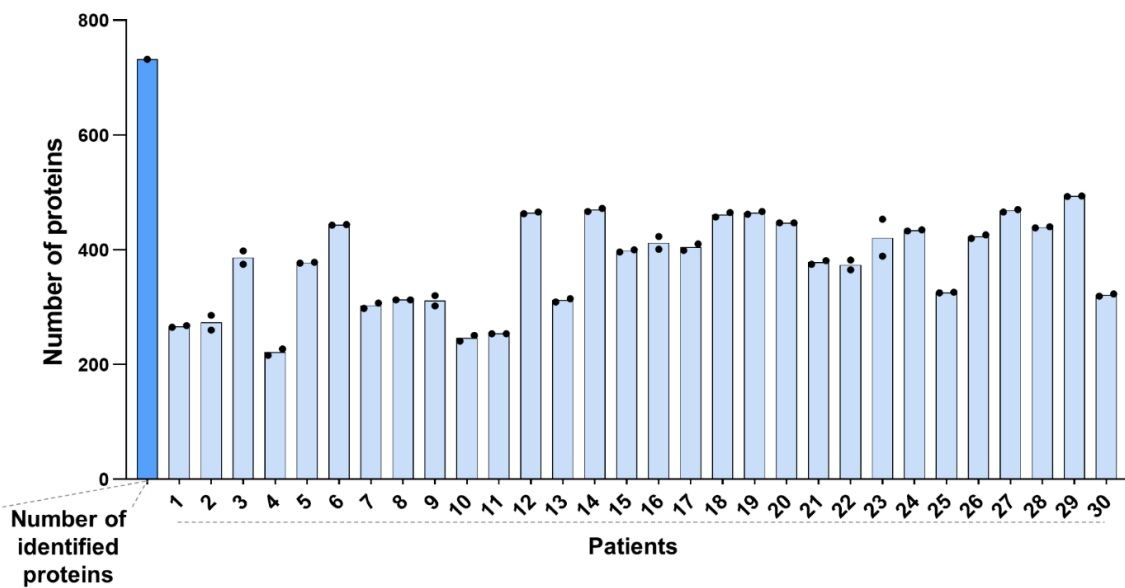


Figure 7-7. Proteomics data summary with the number of identified proteins for each patient. Bars represent the average of two technical replicates.

7.6.3 Supplementary discussion

Patient IV was diagnosed with T1 stage bladder cancer (BCa) and was submitted to the standard TURBT procedure. One reference urine sample was taken at the disease onset before the medical intervention, as shown in Fig.7-8A, where a comprehensive clinical outcome timeline for 25 months, including the sampling time points, is shown. Control urine was taken ten days after surgical intervention, and the lower values for all the pathways are reflected in a dPPi value of -80, indicating an amelioration period. However, 25 months later, all the pathways presented higher levels, with a dPPi value of -56. The increase in the MAPK pathway and the increment in the levels of the APOA4 and APOE biomarkers indicate a worsening of the disease course. Also, the cluster shows that the sampling point three patient's proteomics profile (25 months) is similar to the one present in sampling point one, Fig.7-8F. Under our request, the patient is waiting for cystoscopy revision.

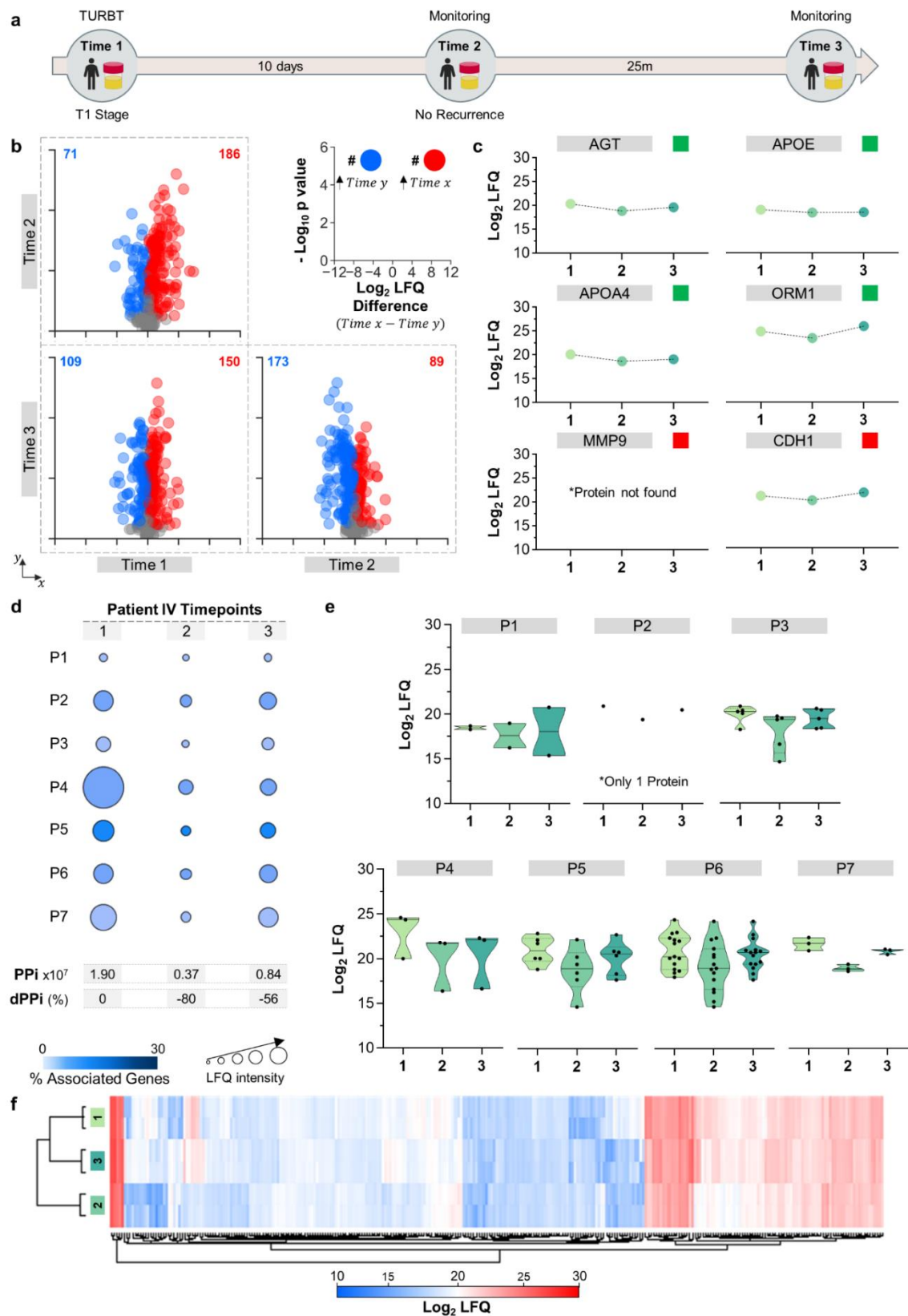


Figure 7-8. Longitudinal analysis of the urinary proteome of patient IV. **A.** Timeline of disease course in Patient IV showing the intervals in months (m) between medical interventions and urine sampling (time 1 to time 3). TURBT: transurethral resection of bladder tumour. **B.** Volcano plot matrix showing proteome changes in the urine collected

at timepoints 1-3. The dots represent the proteins that showed statistically significant increases (red), decreases (blue) or nonstatistically significant changes (grey) according to Student's t test (FDR 0.05 and S0 of 0.1). **C.** Variation in the known protein biomarkers for bladder cancer: angiotensinogen (AGT) [29], apolipoprotein E (APOE) [30], matrix metalloproteinase-9 (MMP9) [31], apolipoprotein A-IV (APOA4) [32], alpha-1-acid glycoprotein 1 (ORM1) [33] and cadherin-1 (CDH1) [34,35]. The green and red squares indicate whether the biomarker variation in the urine of Patient V matched or not, respectively, with trends reported in the literature (considering only the first two sampling points). Dots represent the average of two biological samples with two technical replicates each. **D.** Personal pathway index (PPI) at each time point and the estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **E.** Distribution and density variation protein LFQ values (including two biological replicates) at each sampling point for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **F.** Hierarchical clustering of the three urinary proteomes of Patient IV. The protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

Patient V followed our standard procedure. As he presented for a potential BCa, one urine was taken, and a cystoscopy was carried out. Because T1 stage BCa was diagnosed, a TURBT was performed. The comprehensive timeline analysis of patient V is shown in Fig.7-9A. The evolution of the urine proteome reveals that this patient is not responding to medical surgery and treatment. As a matter of fact, after one month, the levels of the hallmark pathways were virtually the same. Then, a new cystoscopy was done at our request, and as a recurrence was detected, a new TURBT was done. Seven months later, the dPPI remains almost the same, the MAPK pathway presents higher when compared with the onset, and the single biomarkers APOA4 and APOE shows an increment. Taking it all together, we concluded that a new cystoscopy was needed. Afterwards, the patient was diagnosed with a Nonneoplastic bladder mass that was removed.

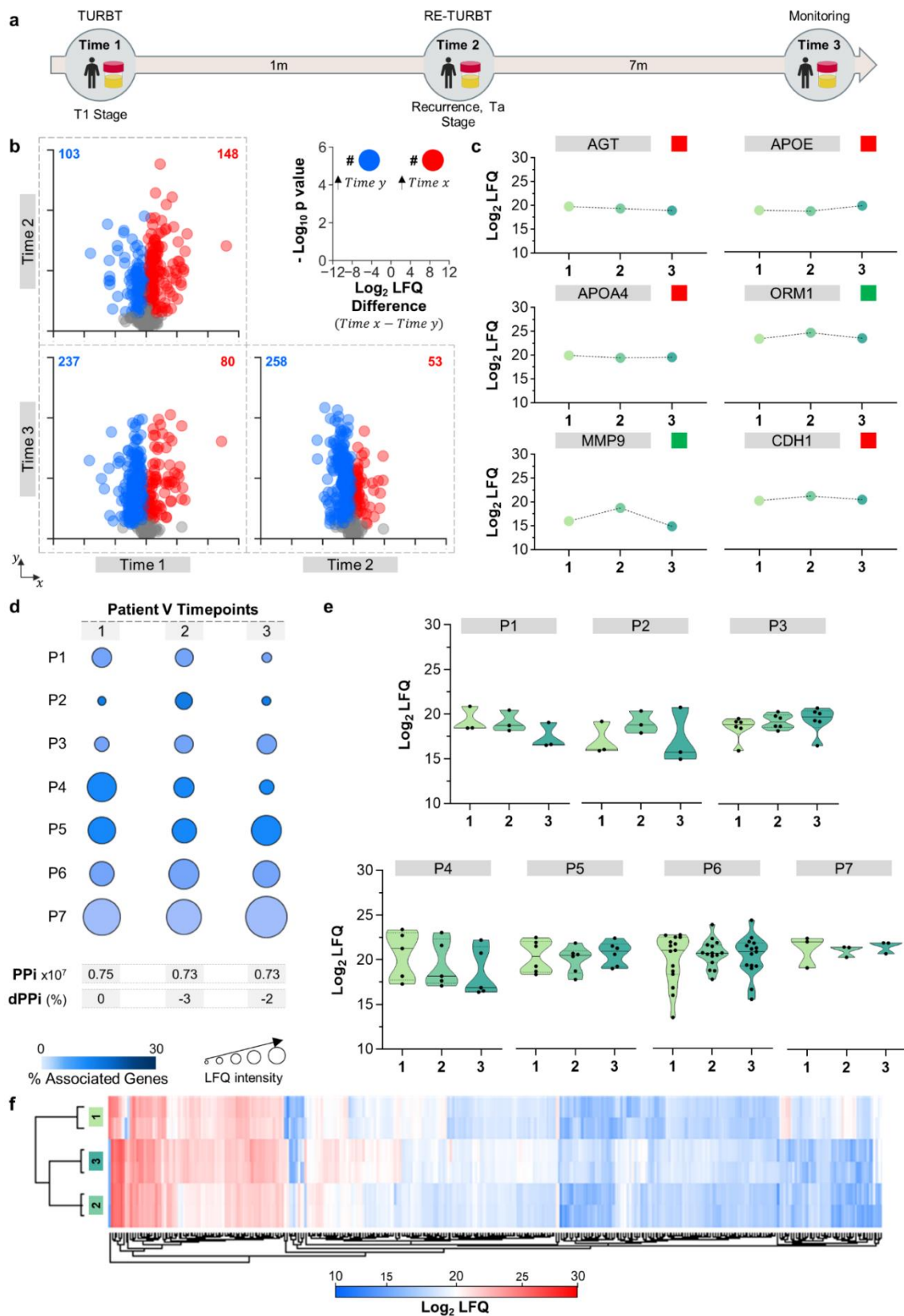


Figure 7-9. Longitudinal analysis of the urinary proteome of patient V. **A.** Timeline of disease course in Patient V showing the intervals in months (m) between medical interventions and urine sampling (time 1 to time 4). TURBT: transurethral resection of bladder tumour. **B.** Volcano plot matrix showing proteome changes in the urine collected

at timepoints 1-3. The dots represent the proteins that showed statistically significant increases (red), decreases (blue) or nonstatistically significant changes (grey) according to the Student's *t* test (FDR 0.05 and *S*₀ of 0.1). **C.** Variation in the known protein biomarkers for bladder cancer: angiotensinogen (AGT) [29], apolipoprotein E (APOE) [30], matrix metalloproteinase-9 (MMP9) [31], apolipoprotein A-IV (APOA4) [32], alpha-1-acid glycoprotein 1 (ORM1) [33] and cadherin-1 (CDH1) [34,35]. The green and red squares indicate whether the biomarker variation in the urine of Patient V matched or not, respectively, with trends reported in the literature (considering only the first two sampling points). Dots represent the average of two biological samples with two technical replicates each. **D.** Personal pathway index (PPI) at each time point and the estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **E.** Distribution and density variation protein LFQ values (including two biological replicates) at each sampling point for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **F.** Hierarchical clustering of the three urinary proteomes of Patient V. The protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

Patient VI was diagnosed as T1 BCa after cystoscopy and resection. 25 months later a urinary analysis revealed a dPPI level 41% lower than the one in the onset, with all the pathway hallmarks presenting lower values. In addition, the single markers MMP9 and APOA4 showed lower values. Our diagnosis was a good evolution of the patient, which the physicians confirmed via cystoscopy. However, eight months later, routine urine surveillance revealed an increment of the dPPI from -41% to -7%, with a general increment in all the pathway hallmarks. An analysis done using the cells present in the urine revealed no recurrence. However, as the single markers MMP9, APOA4 and APOE were found augmented, we have requested new urine to follow up the course of this patient before recommending a cystoscopy.

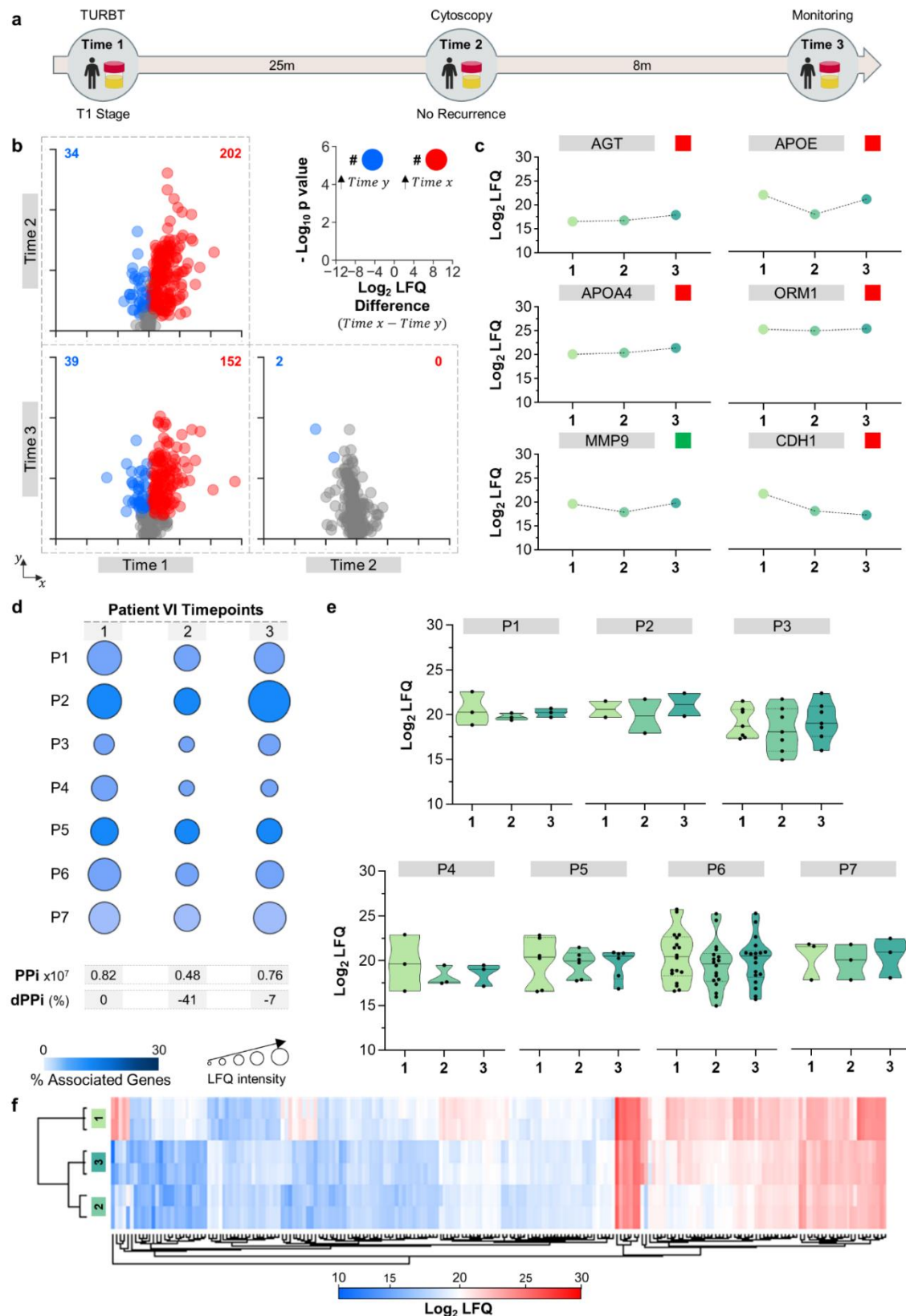


Figure 7-10. Longitudinal analysis of the urinary proteome of patient VI. A. Timeline of disease course in Patient VI showing the intervals in months (m) between medical interventions and urine sampling (time 1 to time 3). TURBT:

transurethral resection of bladder tumour. **B.** Volcano plot matrix showing proteome changes in urine collected at timepoints 1-3. The dots represent the proteins that showed statistically significant increases (red), decreases (blue) or nonstatistically significant changes (grey) according to Student's t test (FDR 0.05 and S0 of 0.1). **C.** Variation in the known protein biomarkers for bladder cancer: angiotensinogen (AGT) [29], apolipoprotein E (APOE) [30], matrix metalloproteinase-9 (MMP9) [31], apolipoprotein A-IV (APOA4) [32], alpha-1-acid glycoprotein 1 (ORM1) [33] and cadherin-1 (CDH1) [34,35]. The green and red squares indicate whether the biomarker variation in the urine of Patient V matched or not, respectively, with trends reported in the literature (considering only the first two sampling points). Dots represent the average of two biological samples with two technical replicates each. **D.** Personal pathway index (PPI) at each time point and the estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **E.** Distribution and density variation protein LFQ values (including two biological replicates) at each sampling point for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **F.** Hierarchical clustering of the three urinary proteomes of Patient VI. The protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

7.7 Contributions & Acknowledgements

7.7.1 Author Contributions

J.L.C., C.L., L.C.P., and H.M.S. designed the experimental work and provide financial support. L.B.C. and G.M. performed the laboratory work and data analysis under the supervision of H.M.S. and J.L.C. H.M.S and J.L.C. drafted the manuscript. L.B.C., J.L.C., C.L., L.C.P, R.D., M.M. and H.M.S. revisited the drafted version, corrected it, and made valuable suggestions. H.L.F. and F.D. performed the statistical analysis. L.C.P. and M.M. managed patient interventions and provided samples and medical data.

7.7.2 Acknowledgments

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV which is financed by national funds from FCT/MCTES (UIDB/50006/2020 and UIDP/50006/2020). H. M. S. acknowledges the Associate Laboratory for Green Chemistry-LAQV (LA/P/0008/2020) funded by Fundação para a Ciência e a Tecnologia, I.P for his research contract. L. B. C. is funded by the FCT PhD grant 2019 (SFRH/BD/144222/2019). G.M. is funded by the FCT PhD grant 2018 (SFRH/BD/139384/2018). H. López-Fernández is supported by a "María Zambrano" post-doctoral contract from Ministerio de Universidades (Gobierno de España).

7.8 References

- [1] A.M. Kamat, N.M. Hahn, J.A. Efsthathiou, S.P. Lerner, P.U. Malmström, W. Choi, C.C. Guo, Y. Lotan, W. Kassouf, Bladder cancer, *The Lancet*. 388 (2016) 2796–2810. [https://doi.org/10.1016/S0140-6736\(16\)30512-8](https://doi.org/10.1016/S0140-6736(16)30512-8).
- [2] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA: A Cancer Journal for Clinicians*. 71 (2021) 209–249. <https://doi.org/10.3322/caac.21660>.
- [3] J.A. Witjes, H.M. Bruins, R. Cathomas, E.M. Compérat, N.C. Cowan, G. Gakis, V. Hernández, E. Linares Espinós, A. Lorch, Y. Neuzillet, M. Rouanne, G.N. Thalmann, E. Veskimäe, M.J. Ribal, A.G. van der Heijden, European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 Guidelines, *European Urology*. 79 (2021) 82–104. <https://doi.org/10.1016/j.eururo.2020.03.055>.
- [4] B. Planz, E. Jochims, T. Deix, H.P. Caspers, G. Jakse, A. Boecking, The role of urinary cytology for detection of bladder cancer, *European Journal of Surgical Oncology*. 31 (2005) 304–308. <https://doi.org/10.1016/j.ejso.2004.12.008>.
- [5] L.B. Carvalho, J.L. Capelo-Martínez, C. Lodeiro, J.R. Wiśniewski, H.M. Santos, Ultrasonic-Based Filter Aided Sample Preparation as the General Method to Sample Preparation in Proteomics, *Analytical Chemistry*. 92 (2020) 9164–9171. <https://doi.org/10.1021/acs.anal-chem.0c01470>.
- [6] S. Ahadi, W. Zhou, S.M. Schüssler-Fiorenza Rose, M.R. Sailani, K. Contrepois, M. Avina, M. Ashland, A. Brunet, M. Snyder, Personal aging markers and ageotypes revealed by deep longitudinal profiling, *Nature Medicine*. 26 (2020) 83–90. <https://doi.org/10.1038/s41591-019-0719-5>.
- [7] B.D. Piening, W. Zhou, K. Contrepois, H. Röst, G.J. Gu Urban, T. Mishra, B.M. Hanson, E.J. Bautista, S. Leopold, C.Y. Yeh, D. Spakowicz, I. Banerjee, C. Chen, K. Kukurba, D. Perelman, C. Craig, E. Colbert, D. Salins, S. Rego, S. Lee, C. Zhang, J. Wheeler, M.R. Sailani, L. Liang, C. Abbott, M. Gerstein, A. Mardinoglu, U. Smith, D.L. Rubin, S. Pitteri, E. Sodergren, T.L. McLaughlin, G.M. Weinstock, M.P. Snyder, Integrative Personal Omics Profiles during Periods of Weight Gain and Loss, *Cell Systems*. 6 (2018) 157–170.e8. <https://doi.org/10.1016/J.CELS.2017.12.013/ATTACHMENT/408B8854-BB80-4961-B6B3-0284A0211754/MMC3.XLSX>.
- [8] S.M. Schüssler-Fiorenza Rose, K. Contrepois, K.J. Moneghetti, W. Zhou, T. Mishra, S. Mataraso, O. Dagan-Rosenfeld, A.B. Ganz, J. Dunn, D. Hornburg, S. Rego, D. Perelman, S. Ahadi,

- M.R. Sailani, Y. Zhou, S.R. Leopold, J. Chen, M. Ashland, J.W. Christle, M. Avina, P. Limcaoco, C. Ruiz, M. Tan, A.J. Butte, G.M. Weinstock, G.M. Slavich, E. Sodergren, T.L. McLaughlin, F. Haddad, M.P. Snyder, A longitudinal big data approach for precision health, *Nature Medicine* 2019 25:5. 25 (2019) 792–804. <https://doi.org/10.1038/s41591-019-0414-6>.
- [9] V.A. Nguyen, N. Riddell, S.G. Crewther, P. Faou, H. Rajapaksha, D.W. Howells, G.J. Hankey, T. Wijeratne, H. Ma, S. Davis, G.A. Donnan, L.M. Carey, Longitudinal Stroke Recovery Associated With Dysregulation of Complement System—A Proteomics Pathway Analysis, *Frontiers in Neurology*. 11 (2020) 692. <https://doi.org/10.3389/FNEUR.2020.00692/BIBTEX>.
- [10] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, *Cell*. 100 (2000) 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- [11] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: The next generation, *Cell*. 144 (2011) 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>.
- [12] V. Thongboonkerd, S. Putita, Bacterial overgrowth affects urinary proteome analysis: recommendation for centrifugation, temperature, duration, and the use of preservatives during sample collection, *J. Proteome Res.* 6 (2007) 4173–4181. <https://doi.org/10.1021/PR070311+>.
- [13] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat Methods*. 6 (2009) 359–362. <https://doi.org/10.1038/nmeth.1322>.
- [14] J.R. Wiśniewski, Filter Aided Sample Preparation – A tutorial, *Anal Chim Acta*. 1090 (2019) 23–30. <https://doi.org/10.1016/j.aca.2019.08.032>.
- [15] L.B. Carvalho, J.L. Capelo-Martínez, C. Lodeiro, J.R. Wiśniewski, H.M. Santos, Snap-heated freeze-free preservation and processing of the urine proteome using the combination of stabilizer-based technology and filter aided sample preparation, *Anal Chim Acta*. 1076 (2019) 82–90. <https://doi.org/10.1016/j.aca.2019.05.051>.
- [16] E.W. Deutsch, N. Bandeira, V. Sharma, Y. Perez-Riverol, J.J. Carver, D.J. Kundu, D. García-Seisdedos, A.F. Jarnuczak, S. Hewapathirana, B.S. Pullman, J. Wertz, Z. Sun, S. Kawano, S. Okuda, Y. Watanabe, H. Hermjakob, B. Maclean, M.J. Maccoss, Y. Zhu, Y. Ishihama, J.A. Vizcaíno, The ProteomeXchange consortium in 2020: Enabling “big data” approaches in proteomics, *Nucleic Acids Res.* 48 (2020) D1145–D1152. <https://doi.org/10.1093/nar/gkz984>.
- [17] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: Improving support for quantification data, *Nucleic Acids Res.* 47 (2019) D442–D450. <https://doi.org/10.1093/nar/gky1106>.

- [18] J. Cox, M.Y. Hein, C.A. Lubner, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Molecular and Cellular Proteomics*. 13 (2014) 2513–2526. <https://doi.org/10.1074/MCP.M113.031591/ATTACHMENT/B414BAF4-20AE-46D2-BB85-A06FBF3A11C3/MMC1.ZIP>.
- [19] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat Biotechnol*. 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.
- [20] S. Tyanova, T. Temu, A. Carlson, P. Sinitcyn, M. Mann, J. Cox, Visualization of LC-MS/MS proteomics data in MaxQuant, *Proteomics*. 15 (2015) 1453–1456. <https://doi.org/10.1002/pmic.201400449>.
- [21] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J. V. Olsen, M. Mann, Andromeda: A peptide search engine integrated into the MaxQuant environment, *Journal of Proteome Research*. 10 (2011) 1794–1805. <https://doi.org/10.1021/pr101065j>.
- [22] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data, *Nature Methods*. 13 (2016) 731–740. <https://doi.org/10.1038/nmeth.3901>.
- [23] Sho Minami, Yuichi Sato, Toshihide Matsumoto, Taihei Kageyama, Yusuke Kawashima, Koderia Yoshio, Jun-Ichiro Ishii, Kazumasa Matsumoto, Ryo Nagashio, Isao Okayasu, Proteomic study of sera from patients with bladder cancer: usefulness of S100A8 and S100A9 proteins, *Cancer Genomics & Proteomics*. 7 (2010) 181–189.
- [24] P.F. Liu, Y.H. Wang, Y.W. Cao, H.P. Jiang, X.C. Yang, X.S. Wang, H.T. Niu, Far from resolved: stromal cell-based iTRAQ research of muscle-invasive bladder cancer regarding heterogeneity, *Oncology Reports*. 32 (2014) 1489–1496. <https://doi.org/10.3892/OR.2014.3340/HTML>.
- [25] Y.T. Chen, C.L. Chen, H.W. Chen, T. Chung, C.C. Wu, C. De Chen, C.W. Hsu, M.C. Chen, K.H. Tsui, P.L. Chang, Y.S. Chang, J.S. Yu, Discovery of novel bladder cancer biomarkers by comparative urine proteomics using iTRAQ technology, *Journal of Proteome Research*. 9 (2010) 5803–5815. https://doi.org/10.1021/PR100576X/SUPPL_FILE/PR100576X_SI_012.PDF.
- [26] M. Lindén, S.B. Lind, C. Mayrhofer, U. Segersten, K. Wester, Y. Lyutvinskiy, R. Zubarev, P.U. Malmström, U. Pettersson, Proteomic analysis of urinary biomarker candidates for non-muscle invasive bladder cancer, *Proteomics*. 12 (2012) 135–144. <https://doi.org/10.1002/PMIC.201000810>.

- [27] S.J. London, T.A. Lehman, J.A. Taylor, Myeloperoxidase Genetic Polymorphism and Lung Cancer Risk, *Cancer Research*. 57 (1997).
- [28] D.C. Castillo-Tong, D. Pils, G. Heinze, I. Braicu, J. Sehouli, A. Reinthaller, E. Schuster, A. Wolf, R. Watrowski, R.A. Maki, R. Zeillinger, W.F. Reynolds, Association of myeloperoxidase with ovarian cancer, *Tumor Biology*. 35 (2014) 141–148. <https://doi.org/10.1007/s13277-013-1017-3>.
- [29] V. Soukup, O. Capoun, M. Pesl, L. Vavrova, R. Sobotka, K. Levova, T. Hanus, T. Zima, M. Kalousova, The significance of calprotectin, CD147, APOA4 and DJ-1 in non-invasive detection of urinary bladder carcinoma, *Neoplasma*. 66 (2019) 1019–1023. https://doi.org/10.4149/neo_2019_190124N74.
- [30] S. Hatipoglu, E. Sevetoglu, A. Gedikbasi, A. Yilmaz, A. Kiyak, M. Mulazimoglu, G. Aydogan, T. Ozpacaci, Urinary MMP-9/NGAL complex in children with acute cystitis, *Pediatric Nephrology*. 26 (2011) 1263–1268. <https://doi.org/10.1007/S00467-011-1856-3/TABLES/4>.
- [31] B. Kumar, J. Sinclair, L. Khandrika, S. Koul, S. Wilson, H.K. Koul, Differential effects of MAPKs signaling on the growth of invasive bladder cancer cells, *International Journal of Oncology*. 34 (2009) 1557–1564. https://doi.org/10.3892/ijo_00000285.
- [32] L. Stramucci, A. Pranteda, G. Bossi, Insights of crosstalk between p53 protein and the MKK3/MKK6/p38 MAPK Signaling pathway in cancer, *Cancers*. 10 (2018) 131. <https://doi.org/10.3390/cancers10050131>.
- [33] T. Yoshida, H. Kinoshita, K. Fukui, T. Matsuzaki, K. Yoshida, T. Mishima, M. Yanishi, Y. Komai, M. Sugi, T. Inoue, T. Murota, T. Matsuda, Prognostic impact of renin-angiotensin inhibitors in patients with bladder cancer undergoing radical cystectomy, *Ann Surg Oncol*. 24 (2017) 823–831. <https://doi.org/10.1245/s10434-016-5534-3>.
- [34] L. Ren, J. Yi, W. Li, X. Zheng, J. Liu, J. Wang, G. Du, Apolipoproteins and cancer, *Cancer Med*. 8 (2019) 7032–7043. <https://doi.org/10.1002/cam4.2587>.
- [35] F.C. Zeng, S. Cen, Z.Y. Tang, X.L. Kang, Elevated matrix metalloproteinase-9 expression may contribute to the pathogenesis of bladder cancer, *Oncol Lett*. 11 (2016) 2213–2222. <https://doi.org/10.3892/ol.2016.4187>.
- [36] F. Li, Z. Yu, P. Chen, G. Lin, T. Li, L. Hou, Y. Du, W. Tan, The increased excretion of urinary orosomucoid 1 as a useful biomarker for bladder cancer, *Am J Cancer Res*. 6 (2016) 331–40. <http://www.ncbi.nlm.nih.gov/pubmed/27186407> (accessed February 2, 2021).
- [37] M. Gurhan Balci, M. Tayfur, Loss of E-cadherin expression in recurrent non-invasive urothelial carcinoma of the bladder, 2018. www.ijcep.com/ (accessed February 2, 2021).

[38] N. Pećina-Šlaus, Tumor suppressor gene E-cadherin and its role in normal and malignant cells, *Cancer Cell Int.* 3 (2003) 17. <https://doi.org/10.1186/1475-2867-3-17>.

8.1 Abstract

This study investigates the proteomic landscapes of chromophobe renal cell carcinoma (chRCC) and renal oncocytomas (RO), two subtypes of renal cell carcinoma that together account for approximately 10% of all renal tumors. Despite their histological similarities and shared origins, chRCC is a malignant tumor necessitating aggressive intervention, while RO, a benign growth, is often subject to overtreatment due to difficulties in accurate differentiation. We conducted a label-free quantitative proteomic analysis on solid biopsies of chRCC (n = 5), RO (n = 5), and normal adjacent tissue (NAT, n = 5). The quantitative analysis was carried out by comparing protein abundances between tumor and NAT specimens. Our analysis identified a total of 1610 proteins across all samples, with 1379 (85.7%) of these proteins quantified in at least seven out of ten LC-MS/MS runs for one renal tissue type (chRCC, RO, or NAT). Our findings revealed significant similarities in the dysregulation of key metabolic pathways, including carbohydrate, lipid, and amino acid metabolism, in both chRCC and RO. Compared to NAT, both chRCC and RO showed a marked downregulation in gluconeogenesis proteins, but a significant upregulation of proteins integral to the citrate cycle. Interestingly, we observed a distinct divergence in the oxidative phosphorylation pathway, with RO showing a significant increase in the number and degree of alterations in proteins, surpassing that observed in chRCC. This study underscores the value of integrating high-resolution mass spectrometry protein quantification to effectively characterize and differentiate the proteomic landscapes of solid tumor biopsies diagnosed as chRCC and RO. The insights gained from this research offer valuable information for enhancing our understanding of these conditions and may aid in the development of improved diagnostic and therapeutic strategies.

8.2 Introduction

Renal cell carcinoma (RCC), a disease characterized by abnormal cell proliferation in the epithelial cells of the kidney, accounts for 400,000 new cases of adult kidney cancer globally each year [1]. There are several subtypes within the RCC classification, each displaying unique histological and cytological phenotypes [2]. Chromophobe renal cell carcinoma (chRCC) and renal oncocytomas (RO) are two RCC subtypes that represent approximately 10% of all renal tumors [3]. RO, a benign growth often found in the kidney collecting ducts, typically lends itself to conservative treatment approaches. Conversely, chRCC is a malignant tumor that, despite sharing a similar origin in distal kidney nephrons, requires more aggressive intervention to

prevent dire consequences to the patient[3]. Given the strikingly similar anatomical origins and histological features of these two tumor types, their accurate differentiation poses a significant challenge [4]. Therefore, most patients undergo surgery, often involving a partial or total nephrectomy. This can lead to overtreatment in RO, as benign lesions may not require such invasive interventions if correctly identified beforehand. Thus, accurately distinguishing between these two conditions is essential not only to guide optimal treatment strategies but also to reduce the potential burden on patients and the healthcare system, thereby improving overall patient care [4,5].

Building on our recent development of a highly efficient, ultrasonic-based methodology for extracting, identifying, and quantifying the proteome of solid renal biopsies embedded in optimal cutting temperature (OCT) compound [6], we sought further to leverage these technological advancements in our current study. In this context, we employed the total protein approach (TPA), a computational tool adept at converting spectral intensities into protein concentrations. The efficacy of the TPA methodology has been proven in various scenarios, such as evaluating metabolic pathways in slow and fast skeletal muscle [7], exploring biological processes associated with colorectal cancer[8], and assessing the impact on mitochondrial bioenergetics in disease-induced conditions[9]. Additionally, we previously investigated differentially expressed proteins in renal cancers using high-resolution MS combined with the TPA method [10]. In this work, we focused on the metabolic features associated with chRCC and RO versus NAT using the TPA method to identify undisclosed new features characterizing each tumor [11]. We compared our findings with the literature addressing the metabolic characteristics of these renal neoplasms [12,13].

8.3 Materials and Methods

8.3.1 Study design and sampling

In this manuscript, we have conscientiously reused a portion of the raw data from our previous by Jorge *et al*¹. The raw data in question forms the basis of a comprehensive dataset that was collected with significant effort and resources. This dataset is rich and multifaceted, allowing for various analyses and interpretations. While our previous publication focused on the identification of novel histochemical markers in renal neoplasm, the current manuscript focuses on the discovery of shared metabolic features between chromophobe renal cell carcinoma and renal oncocytoma. We believe that the use of this data, in this manner, is a

responsible use of resources, avoiding unnecessary duplication of data collection efforts. These data comprises the use of high-resolution mass spectrometry to examine the proteomes of 15 human renal tissue biopsies that were flash-frozen and embedded in OCT. These included chromophobe renal cell carcinoma (chRCC, $n = 5$), renal oncocytoma (RO, $n = 5$) and normal adjacent renal tissue (NAT, $n = 5$). The University of Pittsburgh Biospecimen Core provided the human kidney tissue samples used in this study, which received approval from the University of Pittsburgh's Institutional Review Board (IRB # 02-077). All neoplasms in the study consisted of at least 85% tumor cells. Table 1 summarizes the patient data involved in this research.

8.3.2 Ethics approval and consent to participate

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).

8.3.3 Proteomic analysis

The processing of biopsies followed the methodology outlined in Jorge *et al.*[6] Briefly, tissues were first cleaned of OCT using an ultrasonic bath model TI-H-5 from Elma (Singen, Germany). Then, proteins were extracted in 8 M urea prepared in 25 mM ammonium bicarbonate using an ultrasonic probe (UP50H from Hielscher Ultrasonics, Teltow, Germany). Next, protein digestion was carried out with trypsin for four minutes using an ultrasonic microplate horn assembly device (QSonica, Newtown, CT, USA). The digested proteome extracts were subsequently analyzed by a label-free nanoLC-HR-MS/MS approach (UHR-QqTOF IMPACT HD from Bruker Daltonics, Bremen, Germany) as described before by Jorge *et al.*¹¹ A step-by-step protocol, including OCT Cleaning, proteome extraction, proteome digestion, and LC-MS/MS data acquisition is available in Supplementary Material 1.

8.3.4 Data analysis and statistics

The MaxQuant software V1.6.0.16 was utilized for relative label-free quantification. All raw files were processed in a single run using default settings [14,15]. The Andromeda search engine was employed for database searches, using the UniProt-SwissProt Human database as a reference, along with a database of common contaminants.

For absolute protein quantification, the Total Protein Approach (TPA) was used to examine raw spectral intensities from the MaxQuant output[10]. The calculation of protein concentration was performed as follows:

$$c(i) = \frac{MS\ signal(i)}{total\ MS\ signal \times MW(i)} \left[\frac{mol}{g\ total\ protein} \right]$$

Data processing was performed using Perseus (version 1.6.15.0) with default settings [16,17]. In brief, reverse hits and proteins only identified by site were removed from the protein list, and TPA-based concentrations were log₂-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 values. Missing values were imputed from the total matrix with width = 0.5 and downshift = 1.8. PCA was performed on the filtered and imputed data. Log ratios were calculated as the difference in average log₂ values between the two conditions tested in volcano plots (two-tailed Student's t test, FDR = 0.01 and S0 = 0.1). Protein network analysis was integrated and visualized using the software platform Cytoscape V3.8.2 with the application StringApp V1.6.0 [18]. KEGG [19], Reactome [20] and GO terms [21] were used as ontology databases.

8.4 Results

8.4.1 Renal tumor profiling

The tissue biopsies were analyzed in duplicate using liquid chromatography coupled to tandem mass spectrometry (nanoLC-ESI-HR-MS/MS). Raw spectral intensity values were converted into absolute concentrations using the TPA method [10]. A total of 1610 proteins were identified (1% FDR) across all samples, with 1379 (85.7%) of these proteins being quantified in at least seven out of the ten LC-MS/MS runs for one renal tissue type (chRCC, RO or NAT). Fig.8-2A displays the principal component analysis (PCA) of the quantified proteins, showing that each specimen class segregates into distinct clusters.

The reproducibility of the MS biological replicates was verified by comparing the Pearson correlation between all pairs of samples as well as visualizing pairwise scatter plots in a matrix (Fig.8-2B). The calculated Pearson correlation coefficients spanned from 0.79 to 0.97 for chRCC, from 0.76 to 0.95 for RO, and from 0.83 to 0.95 for NAT. The Pearson coefficients were higher when tumors were compared to each other (0.68 - 0.86) than when compared to NAT (0.56 - 0.70). These results underline a marked deviation of both chRCC and RO from NAT.

Furthermore, the data also demonstrate that there are discernible differences between chRCC and RO, highlighting the potential to differentiate these two neoplasms using this profiling approach.

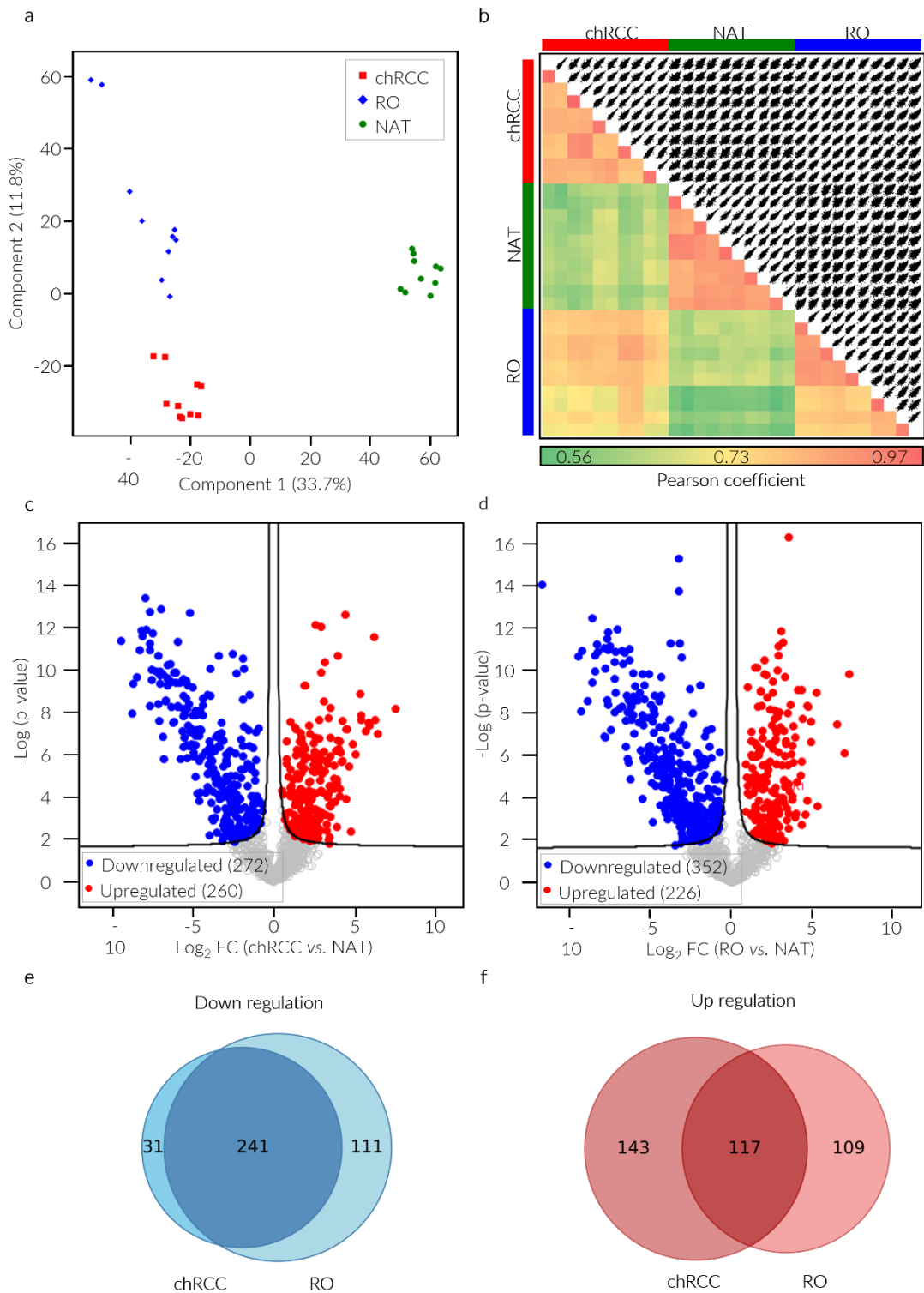


Figure 8-2. Classification of human renal tissue proteomes and protein expression profiles in tumors relative to NATs. Two instrumental replicates were run for each biopsy, resulting in 30 chromatograms that identified a total

of 1610 proteins. Of these proteins, a total of 1379 were obtained from at least one tissue group (chRCC, RO, or NAT), with a reproducibility ranging from 70% to 100%. **A.** Principal component analysis (PCA) of chRCC, RO, and NAT group samples. **B.** Heatmap representation of Pearson correlation of biological ($n = 5$) and technical ($n = 2$, each sample) replicates (lower left) in combination with a scatterplot matrix protein of TPA-based quantification (upper right) demonstrating the reproducibility of the MS data. Volcano plots illustrate the significantly different ($FDR = 0.01$, $S0 = 0.1$) protein expression levels between **C.** chRCC and NAT samples and **D.** RO and NAT samples. Venn diagrams comparing proteomes of chRCC and RO in terms of the **E.** downregulated proteins and **F.** upregulated proteins.

8.4.2 Dysregulated proteins between tumors and NATs

A two-tailed Student's t test ($FDR = 0.01$ and $S0 = 0.1$) was used to distinguish the abundance of proteins between each tumor and NAT. In chRCC, 532 proteins exhibited significant differences compared to NAT, with 260 upregulated and 272 downregulated proteins (Fig.8-2C). Similarly, in the case of RO versus NAT, 578 proteins showed significant differences, with 226 upregulated and 352 downregulated proteins (Fig.8-2D). Comparing the upregulated and downregulated proteins between each tumor and NAT, 241 downregulated proteins (63%) and 117 upregulated proteins (32%) were found to be common (Fig. 8-2E and Fig.8-2F).

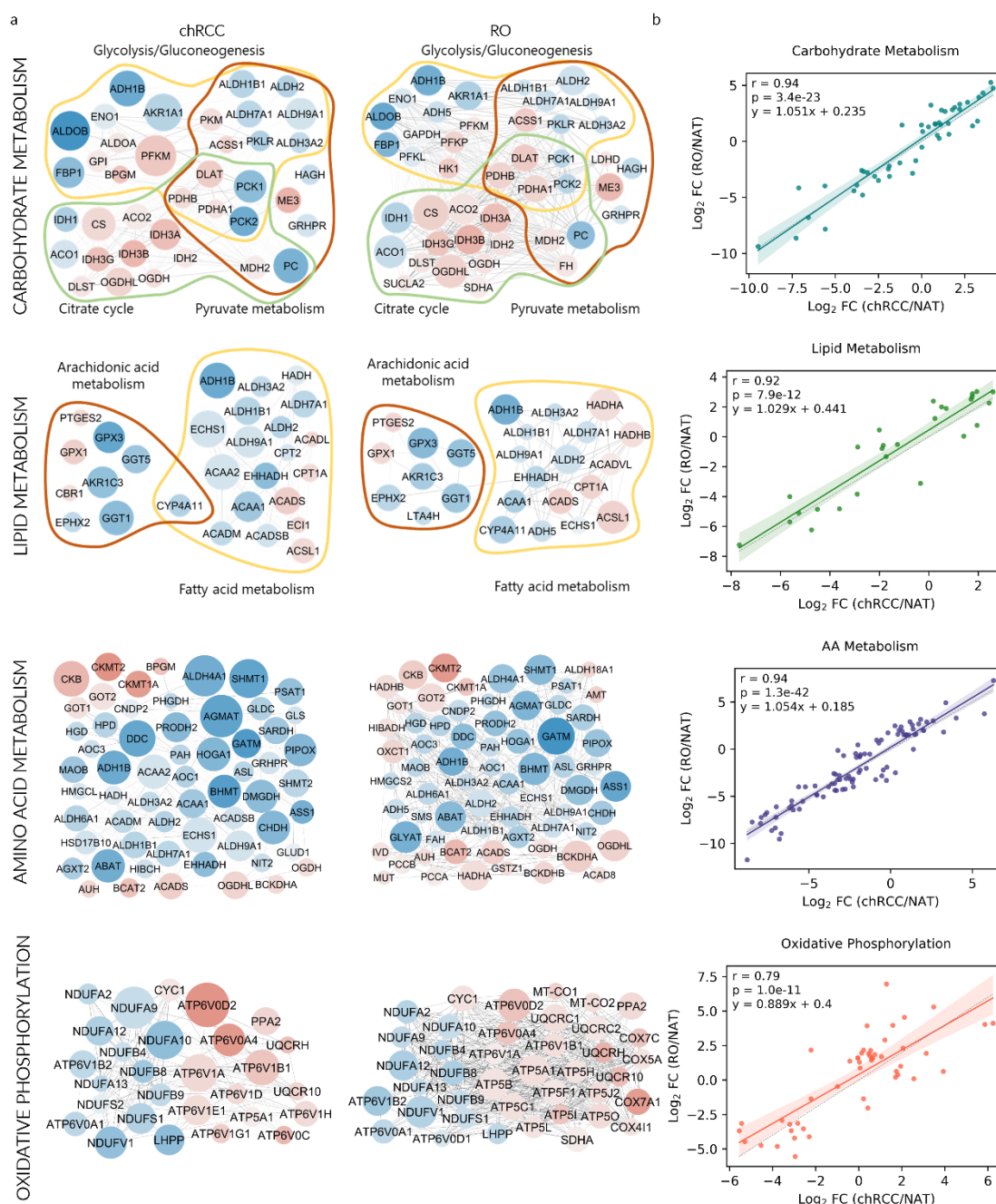


Figure 8-3. Metabolic pathway proteomes similarly affected in chRCC and RO tumors. A. Network representation of a subset of pathways in four areas of metabolism in which proteins were differentially regulated in tumors. Significantly ($FDR < 0.01$) upregulated and downregulated proteins are shown in red and blue, respectively. Protein pathway analysis was performed by searching the KEGG database [19] using differentially expressed proteins between tumor and NAT biopsies. **B.** Correlation between levels of dysregulation of individual proteins (log2 fold change (FC) in abundance in tumor vs NAT) in chRCC and RO for proteins in carbohydrate ($n = 47$), lipid ($n = 27$) and amino acid ($n = 91$) metabolism and oxidative phosphorylation ($n = 47$). The correlation coefficient r was calculated using the Pearson test. Shading areas represent the confidence of the interval, and p represents the p value of the test.

To elucidate the biochemical processes driving the phenotypes of chRCC and RO, the differentially expressed proteins for each tissue type were used to interrogate comprehensive, functional proteomic databases. We found a common signature of metabolic dysregulation for chRCC and RO compared to NAT (Fig.8-3A), affecting carbohydrate metabolism, lipid metabolism, amino acid metabolism and oxidative phosphorylation. Among KEGG protein pathways linked to carbohydrate metabolism, the most dysregulated were glycolysis/gluconeogenesis, pyruvate metabolism and the citric acid cycle ($\text{FDR} < 8.94 \times 10^{-14}$). In our analysis, we identified up to 23 proteins implicated in these pathways that exhibited dysregulation across both tumor subtypes.

The KEGG glycolysis/gluconeogenesis pathway includes processes involved in the degradation of glucose into pyruvate and the generation of glucose from noncarbohydrates. Based on the differential protein expression data, proteins involved in gluconeogenesis, such as fructose-1,6-bisphosphatase 1 (FBP1), pyruvate carboxylase (PC), and phosphoenolpyruvate carboxykinase 1 and 2 (PCK1 and PCK2), were significantly downregulated in tumors. In contrast, glycolytic proteins remained unchanged or upregulated, e.g., ATP-dependent 6-phosphofructokinase (PFKM). These data indicate that gluconeogenesis was downregulated in tumors compared to NATs. For instance, the results of FBP1 match well with its well-known inhibitory effects on glycolysis and tumor growth [22]. Furthermore, the downregulation of PC, which catalyzes the conversion of pyruvate to oxaloacetate in the first step of gluconeogenesis, suggests a shift in cellular metabolism toward the conversion of pyruvate into acetyl-CoA, as evidenced by the observed upregulation of proteins such as pyruvate dehydrogenase E1 component subunit alpha and beta (PDHA1 and PDHB) and the dihydrolipoyl lysine-residue acetyltransferase component of the pyruvate dehydrogenase complex (DLAT). Several TCA cycle proteins were upregulated, including citrate synthase (CS), aconitate hydratase (ACO2), isocitrate dehydrogenase [NADP] and [NAD] subunits alpha and beta (IDH2, IDH3A and IDH3B), dihydrolipoyl lysine-residue succinyltransferase component of the 2-oxoglutarate dehydrogenase complex (DLST), 2-oxoglutarate dehydrogenase (OGDH) and malate dehydrogenase (MDH2). The expression of proteins active in lipid metabolism was modified in both tumors, with most being downregulated relative to NAT controls. In the chRCC and RO samples, 91 downregulated proteins were linked to amino acid (AA) metabolism.

Pearson correlations were calculated for chRCC and RO based on the fold change in abundance of differentially expressed proteins in carbohydrate, lipid and AA metabolism pathways (Fig.8-3B). Positive correlations were found for the four pathways compared, with the

oxidative phosphorylation pathway showing the highest p and lowest Pearson correlation values.

8.4.3 Main chRCC and RO features

The functional protein pathway analysis also revealed divergent features between chRCC and RO that may underlie differences in the tumor subtypes. Proteins involved in energy metabolism of chRCC were less dysregulated than RO. For instance, in chRCC, 28 proteins involved in the oxidative phosphorylation pathway were expressed at different levels compared to NAT, while in RO, 40 proteins belonging to this pathway were dysregulated (Fig.8-3A). Comparing the variation between the proteins belonging to the oxidative phosphorylation pathway in both tumors versus NAT, a Pearson coefficient of 0.79 was obtained. The same comparison for carbohydrate, lipid, and AA metabolism resulted in Pearson coefficients of 0.94, 0.92 and 0.94, respectively (Fig.8-3B).

Mitochondrial KEGG pathways were dysregulated more extensively in RO biopsies, with 256 differentially expressed proteins (Fig.8-4A) compared to 175 in chRCC.

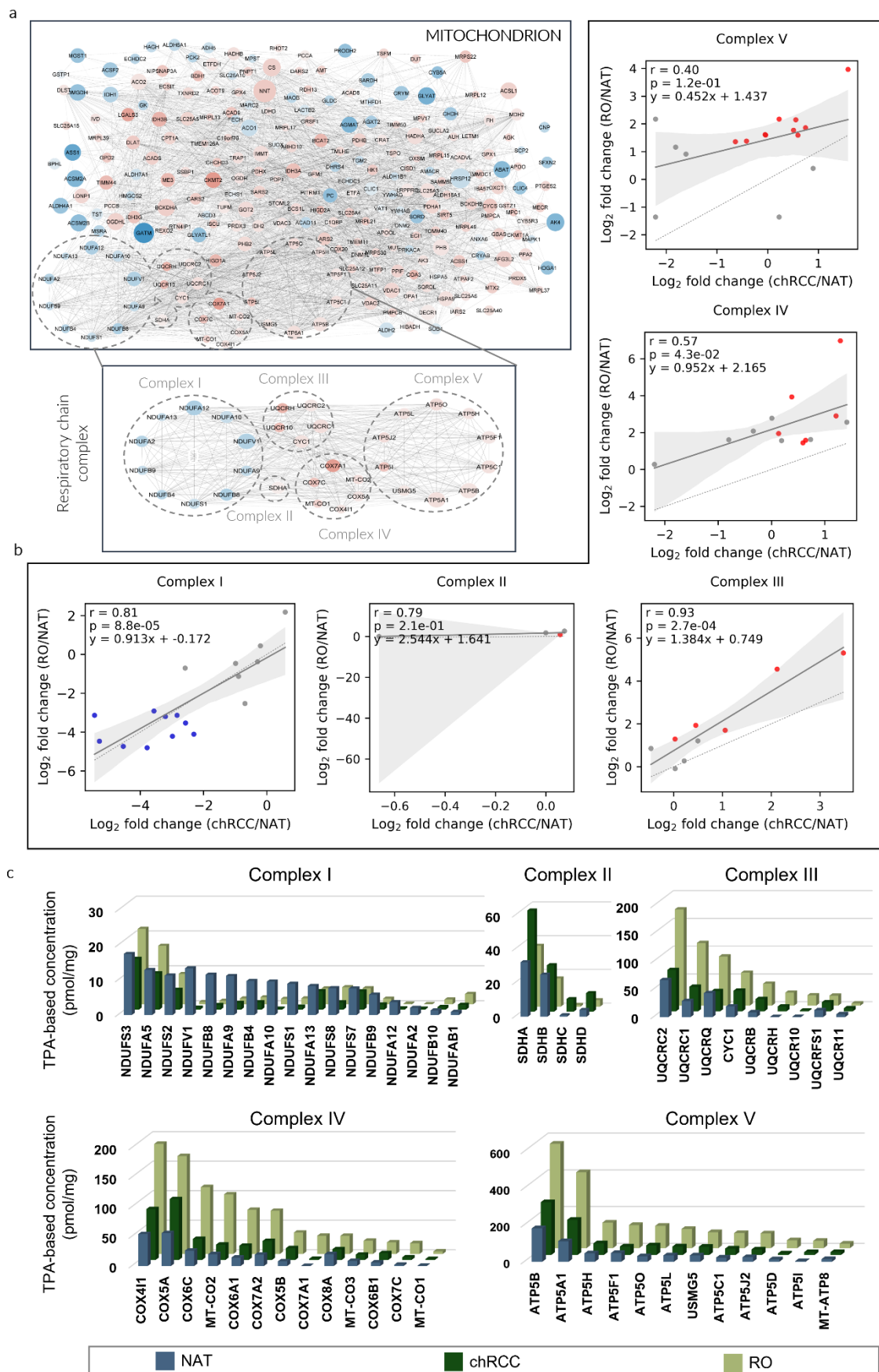


Figure 8-4. Differences in protein pathways underlying each tumor sub-type. A. Specific dysregulation of mitochondrial protein pathways in ROs, including those of respiratory chain complexes. Protein pathway analysis was performed by using the differentially expressed proteins in each tumor biopsy (relative to NAT) in searches against KEGG [19] and GO [21] databases. Significantly (FDR < 0.01) upregulated and downregulated proteins are shown in red and blue, respectively. **B.** Comparison of the protein deregulation (log2 fold change (FC) in abundance in tumor relative to NAT) between chRCC and RO for the five respiratory chain complexes [Complex I, n = 44; Complex II, n = 4; Complex III, n = 10; Complex IV, n = 19; Complex V, n = 20]. Blue dots, downregulated proteins (RO vs NAT); red dots, upregulated proteins (RO vs NAT). The correlation coefficient r was calculated using the Pearson test. Shading areas represent the confidence of the interval, and p represents the p value of the test. **C.** Absolute protein amount, expressed as pmol/mg of tissue, calculated through the TPA methodology.

The common chRCC versus NAT and RO versus NAT dysregulations of proteins related to respiratory chain complexes were evaluated and compared. The results depicted in Fig.8-4B present Pearson correlation coefficients of 0.81, 0.79, 0.93, 0.57, and 0.40 for complexes I, II, III, IV and V, respectively. The TPA-based concentration values for the respiratory chain complex proteins are presented in Fig.8-4C, Fig.8-5 and Fig.8-6.

8.5 Discussion

In our research, we utilized the total protein approach to quantify the absolute protein amounts in chRCC, RO, and NAT tissues. These data enabled us to examine the unique protein expression patterns in chRCC and RO compared to NAT and to delve into potential biological implications. In this context, we believed that the most informative approach was to assess and compare the deviations from the normal state in both chRCC and RO. By contrasting these differences, we aimed to elucidate the unique dysregulation patterns associated with each malignancy, shedding light on their molecular characteristics.

PCA of the variance in protein abundance showed that the proteomes of the two tumor subtypes exhibited greater similarity to each other than to the proteomes of NAT samples. This similarity suggests a shared tumor phenotype between the two subtypes. However, PCA also highlighted discernible differences in the quantified proteomes of the two tumor subtypes, indicating unique characteristics associated with each subtype (Fig.8-2A).

Comparing dysregulated pathways between chRCC and NAT with those between RO and NAT, as depicted in Fig.8-3, reveals a similarity in dysregulation levels. This is evidenced by a Pearson coefficient of 0.92 or higher for carbohydrate metabolism, lipid metabolism, and amino acid metabolism, which strongly suggests parallel patterns of dysregulation in both chRCC and RO subtypes when compared to normal tissue (Fig.8-3B). This observation aligns with the well-established hallmark of cancer, namely, the reprogramming of central metabolic

pathways. It also corroborates the findings reported by Xiao *et al.* [13] who similarly identified decreased levels of proteins involved in gluconeogenesis and fatty acid metabolism in chRCC compared to NAT. Additionally, our data suggest that gluconeogenesis and lipid metabolism are also reduced in RO compared to NAT, which aligns with the findings reported by Kürschner *et al.* [23].

This collective evidence sheds light on the shared metabolic traits between chRCC and RO, thus confirming common signatures of metabolic dysregulation for chRCC and RO compared to NAT. The evident suppression of amino acid synthesis in both cancer types further highlights the possibility of amino acid auxotrophy as a defining feature of these conditions. The downregulation of argininosuccinate synthetase (ASS1), phosphoglycerate dehydrogenase (PHGDH), and glutamic-oxaloacetic transaminase 1 and 2 (GOT1 and GOT2) in both chRCC and RO tumors compared with NAT confirms this result [24]. Inhibition of this protein leads to decreased production of important amino acids such as arginine, serine and glycine, forcing cancer cells to depend on healthy tissue surrounding the tumor as a source of nutrients [24].

Additionally, the analysis of the oxidative phosphorylation pathway revealed a Pearson coefficient of 0.74 (Fig.8-3B), suggesting the presence of distinct mechanisms of oxidative phosphorylation between the chRCC and RO subtypes. Notably, our findings align with previous studies, demonstrating significant alterations in mitochondrial pathways in both RO and chRCC when compared to NAT [25,26]. It is interesting to note that the levels of dysregulated mitochondrial-related proteins were higher in RO than in chRCC (Fig.8-4). Our findings agree with those reported by Joshi *et al.* [25], who suggest that mitochondrial activity acts as a barrier to tumorigenesis in RO, which means that normal mitochondrial activity can prevent or slow down the development of tumors. Reduced mitochondrial activity may contribute to the development of tumors by promoting a protumorigenic environment in cells. This is because normal mitochondrial activity is necessary to maintain healthy cellular metabolism, and altered mitochondrial activity in cancer cells can affect cellular metabolism and promote tumor growth. This suggests that targeting mitochondrial metabolism may be explored as a coadjutant therapeutic strategy for treating chRCC.

In addition to mitochondrial abnormalities, mutations in respiratory chain complex I genes are frequently observed in RO [23]. Our RO versus NAT proteomic data revealed downregulation of complex I-related proteins and predominant upregulation of proteins associated with complexes III, IV and V, as shown in Fig.8-4B and Fig.8-4C, 8-5 and Fig.8-6. Our data agree with the work of Xiao *et al.* [12] who also reported reduced expression of complex I-related proteins and an increase in all other proteins related to respiratory chain complexes. In this

way, the direct difference in the absolute levels of expression of these mitochondrial proteins demonstrates the utility of the TPA in discriminating between these two tumor subtypes.

8.6 Conclusions

In summary, this work presents an in-depth exploration of the contrasting protein expression profiles found in chRCC and RO compared to their normal adjacent tissue. The study used the TPA-based method to determine the absolute amounts of proteins in each tissue type. The findings suggest that both chRCC and RO show dysregulation of carbohydrate, lipid, and amino acid metabolism, as well as mitochondrial abnormalities, when compared to NAT. The study found that the proteomes of the two tumor subtypes were more similar to each other than to NAT, reflecting their shared tumor phenotype. The observed decrease in argininosuccinate synthetase (ASS1) expression in both chRCC and RO tumors relative to adjacent healthy tissue indicates shared metabolic characteristics between these cancers. One such shared trait appears to be auxotrophy, which is prominently featured in both conditions. Targeting mitochondrial metabolism may be explored as a coadjutant therapeutic strategy for treating both chRCC and RO. The reduced number of samples per group used in this study must be considered in future studies.

8.7 Supplementary material

8.7.1 Supplementary data

The mass spectrometry proteomics raw data have been deposited in the ProteomeXchange Consortium via the PRIDE [27] partner repository with the dataset identifier PXD022018.

8.7.2 Supplementary figures & tables

Table 8-1. 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
O15	55	Female	RCC	RO

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

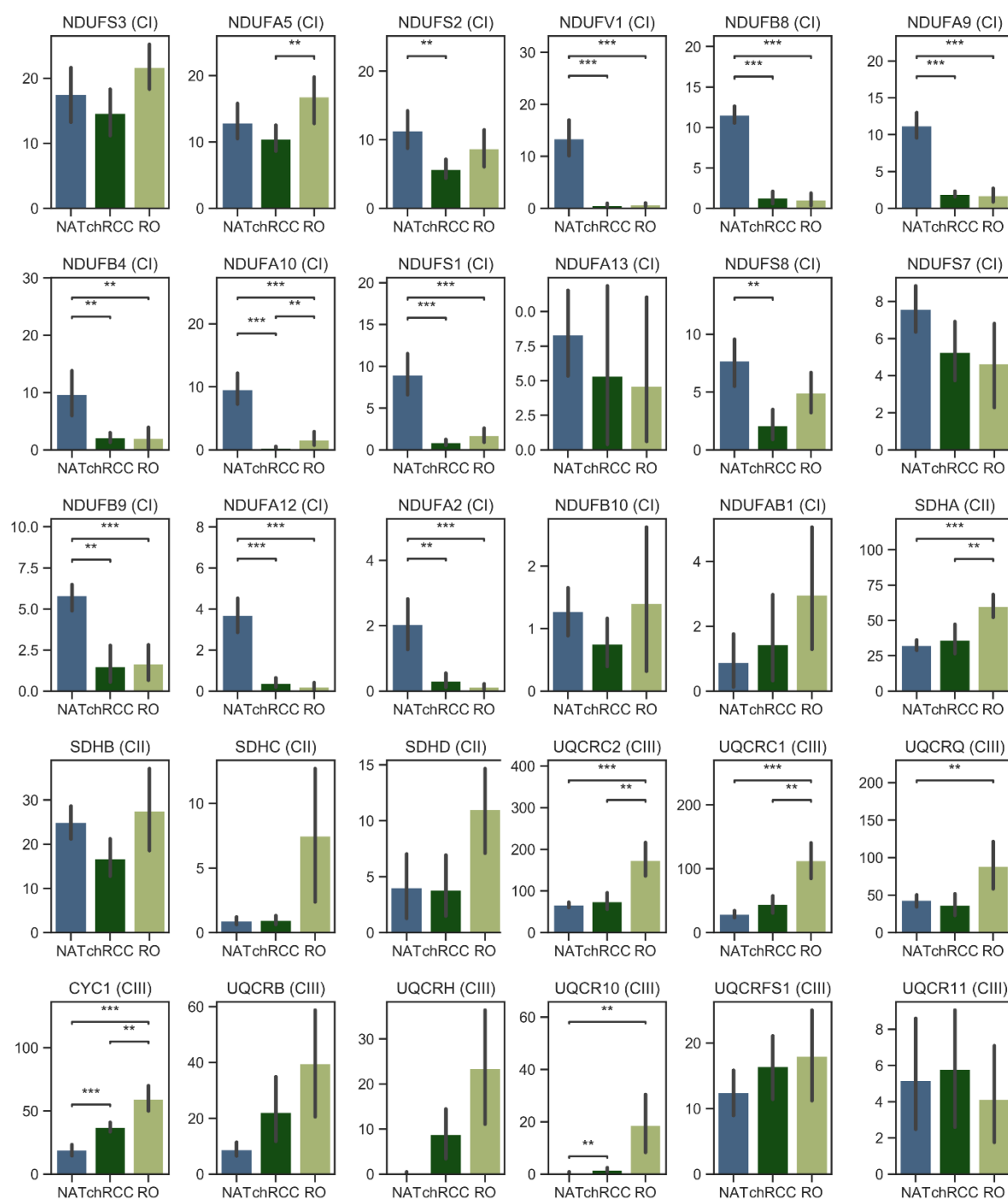


Figure 8-5. TPA concentration values (pmol/mg of tissue) of respiratory chain complex I (CI), II (CII) and III (CIII) proteins. Statistical analysis was performed using pairwise Mann Whitney test (** $p \leq 0.001$; *** $p \leq 0.0001$; **** $p \leq 0.00001$).

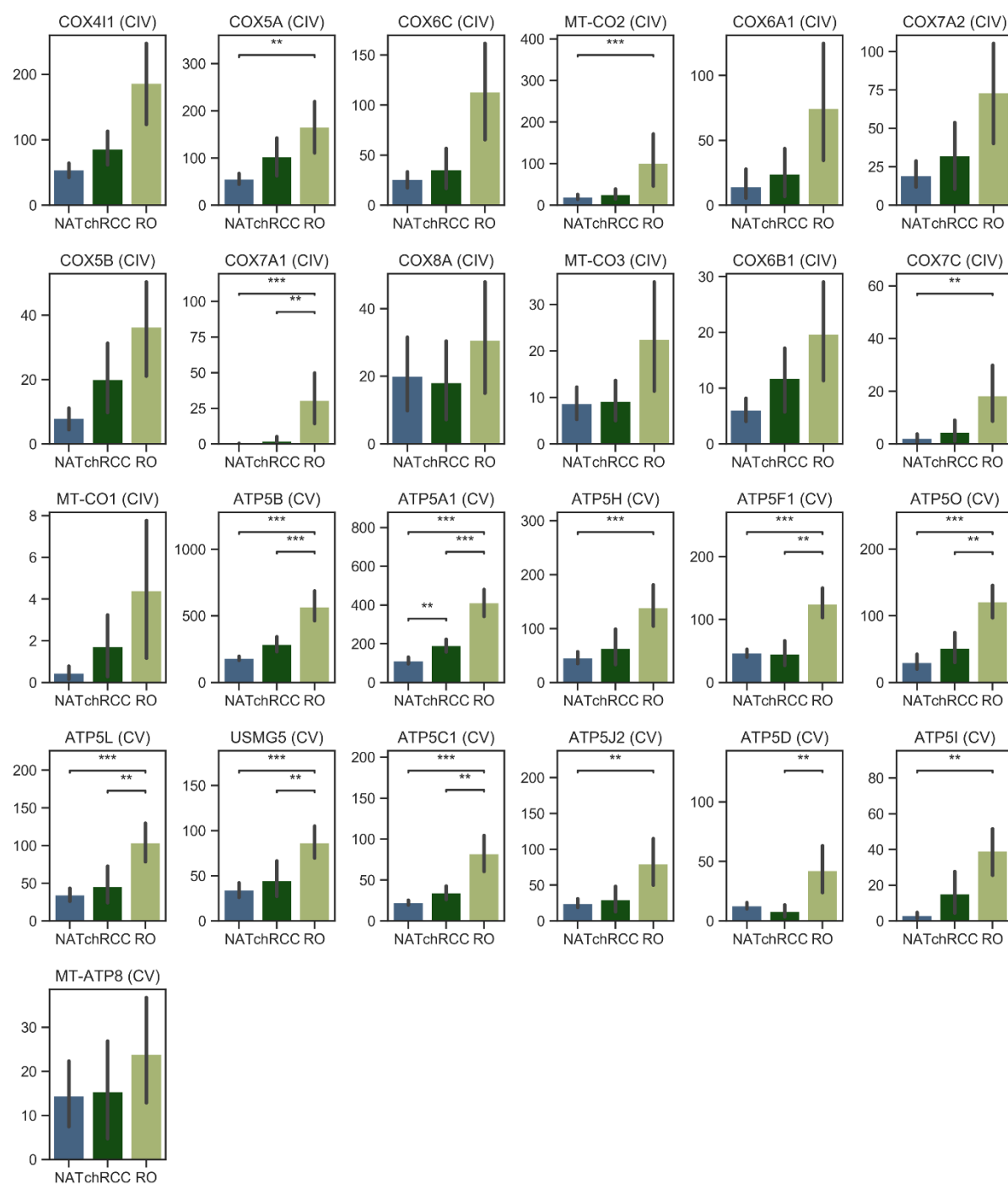


Figure 8-6. TPA concentration values (pmol/mg of tissue) of respiratory chain complex IV (CIV) and V (CV) proteins. Statistical analysis was performed using pairwise Mann Whitney test (**p ≤ 0.001; ***p ≤ 0.0001; ****p ≤ 0.00001).

8.8 Contributions & Acknowledgements

8.8.1 Author Contributions

RD, SS, DK, SB and GQG collected and organized all human tissue biopsies. JLC, HMS, CLE, and RD designed the experimental work and provided financial support. SJ performed the laboratory work under the supervision of HMS and JLC. LBC finished the data handling and contributed to the figure making and contributed significantly to the final manuscript version. SJ, LBC and HLF performed the bioinformatic analysis. SJ, LBC and JLC drafted the manuscript, and HMS, CLE, and RD revised and corrected the draft. All authors have approved the final version of the manuscript.

8.8.2 Acknowledgments

The PROTEOMASS Scientific Society (Portugal) is acknowledged for the funding provided through the General Funding Grant 2022-2023 and by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). The authors also acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV, which is financed by national funds from *Fundação para a Ciência e Tecnologia and Ministério da Ciência, Tecnologia e Ensino Superior* (FCT/MCTES) through the projects UIDB/50006/2020 and UIDP/50006/2020. HMS acknowledges the Associate Laboratory for Green Chemistry LAQV (LA/P/0008/2020) funded by FCT/MCTES for his research contract. LBC and SJ thank FCT/MCTES for their PhD studentships SFRH/BD/144222/2019 and SFRH/BD/120537/2016, respectively. HLF is supported by a 'María Zambrano' postdoctoral contract from *Ministerio de Universidades (Gobierno de España)*. This project benefited from the University of Pittsburgh Hillman Cancer Center shared resource facility (Cancer Genomics Facility) supported in part by award P30CA047904.

8.9 References

- [1] Cancer Today, (2020). 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] H. Moch, A.L. Cubilla, P.A. Humphrey, V.E. Reuter, T.M. Ulbright, The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours, *Eur Urol.* 70 (2016) 93–105. <https://doi.org/10.1016/j.eururo.2016.02.029>.
- [3] P. Cairns, Renal cell carcinoma, *Cancer Biomarkers.* 9 (2011) 461–473. <https://doi.org/10.3233/CBM-2011-0176>.
- [4] B. Ljungberg, L. Albiges, Y. Abu-Ghanem, K. Bensalah, S. Dabestani, S. Fernández-Pello, R.H. Giles, F. Hofmann, M. Hora, M.A. Kuczyk, T. Kuusk, T.B. Lam, L. Marconi, A.S. Merseburger, T. Powles, M. Staehler, R. Tahbaz, A. Volpe, A. Bex, European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update, *Eur Urol.* 75 (2019) 799–810. <https://doi.org/10.1016/j.eururo.2019.02.011>.
- [5] S.E. Wobker, S.R. Williamson, Modern Pathologic Diagnosis of Renal Oncocytoma, *J Kidney Cancer VHL.* 4 (2017) 1–12. <https://doi.org/10.15586/jkcvhl.2017.96>.
- [6] S. Jorge, J.L. Capelo, W. LaFramboise, R. Dhir, C. Lodeiro, H.M. Santos, Development of a Robust Ultrasonic-Based Sample Treatment To Unravel the Proteome of OCT-Embedded Solid Tumor Biopsies, *J Proteome Res.* 18 (2019) 2979–2986. <https://doi.org/10.1021/acs.jproteome.9b00248>.
- [7] D. Rakus, A. Gizak, A. Deshmukh, J.R. Winiewski, Absolute quantitative profiling of the key metabolic pathways in slow and fast skeletal muscle, *J Proteome Res.* 14 (2015) 1400–1411. <https://doi.org/10.1021/pr5010357>.
- [8] J.R. Wiśniewski, K. Duś-Szachniewicz, P. Ostasiewicz, P. Ziółkowski, D. Rakus, M. Mann, Absolute Proteome Analysis of Colorectal Mucosa, Adenoma, and Cancer Reveals Drastic Changes in Fatty Acid Metabolism and Plasma Membrane Transporters, *J Proteome Res.* 14 (2015) 4005–4018. <https://doi.org/10.1021/acs.jproteome.5b00523>.
- [9] E. Walheim, J.R. Wiśniewski, M. Jastroch, Respiromics – An integrative analysis linking mitochondrial bioenergetics to molecular signatures, *Mol Metab.* 9 (2018) 4–14. <https://doi.org/10.1016/j.molmet.2018.01.002>.
- [10] J.R. Wiśniewski, Label-Free and Standard-Free Absolute Quantitative Proteomics Using the “Total Protein” and “Proteomic Ruler” Approaches, in: *Methods Enzymol*, Academic Press Inc., 2017: pp. 49–60. <https://doi.org/10.1016/bs.mie.2016.10.002>.
- [11] S. Jorge, J.L. Capelo, W. LaFramboise, S. Satturwar, D. Korentzelos, S. Bastacky, G. Quiroga-Garza, R. Dhir, J.R. Wiśniewski, C. Lodeiro, H.M. Santos, Absolute quantitative proteomics using the total protein approach to identify novel clinical immunohistochemical markers in renal neoplasms, *BMC Medicine* 2021 19:1. 19 (2021) 1–13. <https://doi.org/10.1186/S12916-021-02071-9>.
-

- [12] Y. Xiao, R. Clima, J. Busch, A. Rabien, E. Kilic, S.L. Villegas, B. Timmermann, M. Attimonelli, K. Jung, D. Meierhofer, Decreased mitochondrial DNA content drives OXPHOS dysregulation in chromophobe renal cell carcinoma, *Cancer Res.* 80 (2020) 3830–3840. <https://doi.org/10.1158/0008-5472.CAN-20-0754>.
- [13] Y. Xiao, A. Rabien, R. Buschow, V. Amtislavskiy, J. Busch, E. Kilic, S.L. Villegas, B. Timmermann, M. Schutte, T. Mielke, M.L. Yaspo, K. Jung, D. Meierhofer, Endocytosis-mediated replenishment of amino acids favors cancer cell proliferation and survival in chromophobe renal cell carcinoma, *Cancer Res.* 80 (2020) 5491–5501. <https://doi.org/10.1158/0008-5472.CAN-20-1998>.
- [14] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat Biotechnol.* 26 (2008) 1367–1372. <https://doi.org/10.1038/nbt.1511>.
- [15] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, *Nat Protoc.* 11 (2016) 2301–2319. <https://doi.org/10.1038/nprot.2016.136>.
- [16] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data, *Nat Methods.* 13 (2016) 731–740. <https://doi.org/10.1038/nmeth.3901>.
- [17] S. Tyanova, J. Cox, Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research, in: *Methods in Molecular Biology*, 2018: pp. 133–148. https://doi.org/10.1007/978-1-4939-7493-1_7.
- [18] N.T. Doncheva, J.H. Morris, J. Gorodkin, L.J. Jensen, Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data, *J Proteome Res.* 18 (2019) 623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>.
- [19] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets, *Nucleic Acids Res.* 40 (2012) D109–D114. <https://doi.org/10.1093/nar/gkr988>.
- [20] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C.D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D’Eustachio, The reactome pathway knowledgebase, *Nucleic Acids Res.* 46 (2018) D649–D655. <https://doi.org/10.1093/nar/gkx1132>.
- [21] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C.

Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology, *Nat Genet.* 25 (2000) 25–29. <https://doi.org/10.1038/75556>.

[22] G. Grasmann, E. Smolle, H. Olschewski, K. Leithner, Gluconeogenesis in cancer cells – Repurposing of a starvation-induced metabolic pathway?, *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer.* 1872 (2019) 24–36. <https://doi.org/10.1016/j.bbcan.2019.05.006>.

[23] G. Kürschner, Q. Zhang, R. Clima, Y. Xiao, J.F. Busch, E. Kilic, K. Jung, N. Berndt, S. Bulik, H.-G. Holzhütter, G. Gasparre, M. Attimonelli, M. Babu, D. Meierhofer, Renal oncocytoma characterized by the defective complex I of the respiratory chain boosts the synthesis of the ROS scavenger glutathione, *Oncotarget.* 8 (2017) 105882–105904. <https://doi.org/10.18632/oncotarget.22413>.

[24] L. Vettore, R.L. Westbrook, D.A. Tennant, New aspects of amino acid metabolism in cancer, *Br J Cancer.* 122 (2020) 150–156. <https://doi.org/10.1038/s41416-019-0620-5>.

[25] S. Joshi, D. Tolkunov, H. Aviv, A.A. Hakimi, M. Yao, J.J. Hsieh, S. Ganesan, C.S. Chan, E. White, The Genomic Landscape of Renal Oncocytoma Identifies a Metabolic Barrier to Tumorigenesis, *Cell Rep.* 13 (2015) 1895–1908. <https://doi.org/10.1016/j.celrep.2015.10.059>.

[26] D. Lindgren, J. Sjölund, H. Axelson, Tracing Renal Cell Carcinomas back to the Nephron, *Trends Cancer.* 4 (2018) 472–484. <https://doi.org/10.1016/j.trecan.2018.05.003>.

[27] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Pérez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, Ş. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaíno, The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* 47 (2019) D442–D450. <https://doi.org/10.1093/nar/gky1106>.

GENERAL CONCLUSIONS

The present works show different types of contributions to the field of bladder cancer (BCa) research, offering innovative proteomics answers to the current challenges in BCa diagnosis, prognosis, and treatment. Additionally, the expanded analysis of a different urological cancer, renal carcinoma, further highlights the versatility and potential of such answers in different types of cancers and/or conditions.

By addressing the different objectives outlined throughout the thesis, significant advancements were made, yielding different contributions to both proteomics research and clinical practice:

- i. Urine sample treatment and storage by using an integration of heat stabilization with the Filter-Aided Sample Preparation (FASP) approach to preserve and transport the urine proteome at room temperature. This novel approach avoids the need for the standard expensive freeze storage while keeping the integrity of samples over extended periods at room temperature. With no risk of degradation facilitates the sample transport and collection for large-scale or longitudinal studies. Furthermore, this approach can be further expanded to a wide range of liquid biopsy samples, showing a promising transformative implication for diagnostic and prognostic.
- ii. Urine proteome analysis and preparation by the development of an ultrasonic-assisted FASP (US-FASP) protocol that increases sample processing efficiency. By using ultrasound energy, it substantially reduces the processing time and increases the scalability and compatibility with mass spectrometry platforms. This method holds the potential to be used as the general method in clinical proteomics workflows. Its simplicity of execution makes it suitable for implementation in hospital settings.

- iii. LC-MS/MS data processing and normalization, through the evaluation of different normalization methods. Different normalization approaches, particularly the Z-score normalization method, show the importance of data treatment to uncover clinically relevant protein alterations. Evaluating and finding the optimal data treatment method further enhances the robustness of biomarker discovery efforts and strengthens the development of personalization of targeted therapeutic interventions.
- iv. Analysis of bladder cancer protein networks, which uncovers the identification of inflammation dysregulation as a key pathway for BCa recurrence and progression. Such finding highlights the therapeutic potential of anti-inflammatory and immune checkpoint modalities. Additionally, the use of the Myeloperoxidase to Cubilin ratio as a prognostic tool underscores the clinical utility of urine-based biomarkers in complementing and guiding therapeutic decisions.
- v. Personalized medicine in bladder cancer, focused on the longitudinal monitoring of BCa patients. Using the differential personal pathway index (dPPi), allows for the follow-up of the patient's response to treatment. Focusing on the pathway analysis and the shift of different pathways regulation using the dPPi score as an indication represents a shift to the current way of follow-up BCa patients. Furthermore, due to the non-invasive nature of the method it can offer insights into disease trajectory and treatment response while minimizing the reliance on invasive methods.
- vi. Lastly, the application of such methods to a different oncoprotein study, focusing on renal cancer pathway analysis to comprehensively characterize protein expression profiles in chromophobe renal cell carcinoma (chRCC) and renal oncocytomas (RO). Different shared metabolic signatures and potential therapeutic targets such as mitochondrial metabolism offer a promising target for therapeutic target. Additionally, this approach further shows the versatility of proteomics to different types of cancer and/or condition.

This doctoral work not only advances the understanding of BCa pathophysiology but also aims to decrease the patient disease burden and improve the treatment outcome. Furthermore, more new perspectives arrive from this work, allowing for continuous work in the field of proteomics and personalized medicine, signalling a new era for precision oncology.

FUTURE PERSPECTIVES

Based on the considerable advances highlighted in this work, various next directions might be taken to improve bladder cancer (BCa) and other types of cancer research and proteomics applications. Such directions attempt to expand the influence of existing works and turn it into actual clinical solutions.

i. Expanding to other liquid biopsies. Extend the heat stabilization and Filter-Aided Sample Preparation (FASP) method to different types of liquid biopsies and assess the preservation of additional biological information such as DNA/RNA and metabolites. This can help with the early identification and monitoring of a variety of disorders, not just bladder cancer.

ii. Developing a portable urine processing device. Currently being developed in partnership with the University of Bologna, a portable device ready to incorporate the fast ultrasonic-assisted FASP (US-FASP) technology. These devices are design to be employed in remote or resource-constrained places, making sample processing and preparation for mass spectrometry easier.

iii. Integrating artificial intelligence. Use AI and machine learning to improve data processing and standardization procedures, as well as to easily construct an automated biomarker discovery pipeline. These enhanced models' goals are to detect small trends in proteomic data, increasing the accuracy of biomarker discovery and resulting in more tailored treatment recommendations.

iv. Conducting large-scale longitudinal studies with a focus on patient-centred approaches. Currently underway in partnership with Medical Harvard University, experimenting with new types of cancer, such as breast cancer, and conducting large-scale longitudinal

research to evaluate the stability and reproducibility of urine proteome analysis methodologies over time.

v. Create Real-Time Monitoring Platforms. Ongoing collaboration with Dr. Rajiv Dhir of the University of Pittsburgh and Dr. Nataly Kravchenko-Balasha of the Hebrew University of Jerusalem to build systems for real-time monitoring of patient response to treatment using the differential personal pathway index (dPPi) approach and prescriptomics. These approaches aim to provide clinicians with quick feedback, allowing them to make timely therapy adjustments.

vi. Establishing collaborative research networks. Different networks are being formed between research institutes and hospitals to accelerate the translation of discoveries into clinical practice. The vast projects described in the document are currently being implemented through collaboration with various institutions, including Harvard Medical School, University of Pittsburgh Medical Center, Washington University in St. Louis, University of Bologna, Max Planck Institute, Hospital São Jose, and Boston Children's Hospital.

By addressing these future ideas with some of them already ongoing, the work can continue to develop the area of proteomics and tailored medicine, providing up new paths for precision oncology and improved patient outcomes across a range of tumors.

| **A**

**PUBLISHED PAPER IN ANALYTICA CHIMICA
ACTA - CORRESPONDING TO CHAPTER 3**



Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Snap-heated freeze-free preservation and processing of the urine proteome using the combination of stabilizer-based technology and filter aided sample preparation

Luís B. Carvalho^{a, b}, José Luis Capelo-Martínez^{a, b}, Carlos Lodeiro^{a, b},
Jacek R. Wiśniewski^c, Hugo M. Santos^{a, b, *}

^a BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, Faculty of Science and Technology, Universidade NOVA de Lisboa, 2829-516, Campus de Caparica, Portugal

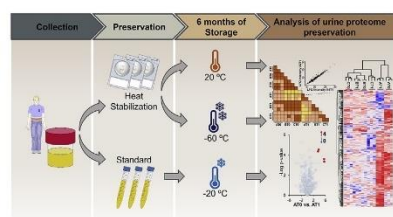
^b PROTEOMASS Scientific Society, Madan Parque, Rua dos Inventores, 2825-182, Caparica, Portugal

^c Biochemical Proteomics Group, Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152, Martinsried, Germany

HIGHLIGHTS

- Urine proteome preservation at room temperature in a membrane that is ready for processing using FASP digestion method.
- Stabilizer T1 as a method to preserve urine proteome by a heat treatment.
- Urine storage at room temperature for 6 months with no significant changes.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 6 April 2019

Received in revised form

19 May 2019

Accepted 22 May 2019

Available online 24 May 2019

Keywords:

Urine proteomics

Room temperature urine preservation

Heat stabilization

Filter-aided sample preparation

Label-free quantitative mass spectrometry

ABSTRACT

We present a method to preserve and process urine proteins for proteomic analysis in a filter aided sample preparation (FASP) format. The method combines concentration of urine proteins in ultrafiltration devices, their thermal stabilization, allowing long term storage of the samples, and filter aided sample preparation. Proteomes of four different urines were preserved during 48 h and 6 months using (i) the classic freeze preservation at -20°C , (ii) snap-heated freeze-free preservation at laboratory temperature (20°C) and (iii) snap-heated preservation at -60°C . The three storage methods can effectively preserve the urine proteome for at least 6 months without significant alterations. Abundances of more than 500 proteins and specially 24 selected -cleared or -approved protein assayed in serum or plasma were found similar within the three preservation methods assessed. The new method here proposed dramatically simplifies the conditions for preserving the urine proteome for biobanks in terms of space and storage, including lowering the risks of sample degradation caused by malfunction of the freezer. Furthermore, the shipping of large number of samples can be made without the need of freezing. The application of the FASP format to isolate and preserve the proteins facilitates long-term storage and processing of proteome of urine samples.

© 2019 Elsevier B.V. All rights reserved.

* Corresponding author. Bioscope Research Group, Laqv-Requimte, Department of Chemistry, Faculty of Science and Technology, Universidade NOVA de Lisboa, 2829-516, Campus de Caparica, Portugal.

E-mail address: hmsantos@fct.unl.pt (H.M. Santos).

<https://doi.org/10.1016/j.jaca.2019.05.051>

0003-2670/© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Human body fluids such as urine, serum, plasma, cerebrospinal fluid, saliva, and tears have characteristic protein profiles that can be used to unravel diseases [1,2]. Of the fluids mentioned above, serum and plasma are still the most used for biomarker discovery. However, urine is gaining momentum due to a number of unique advantages, such as that large sample volumes can be collected in a non-invasive and therefore painless manner, it can be collected at any time and at any place [3,4]. Urine contains proteins of which circa 70% are from the urinary tract, while the remaining 30% corresponds to blood proteins [5]. Urinary proteins can originate from (i) glomerular filtration of blood proteins, (ii) secretion of proteins by the renal tubular epithelial cells, (iii) shedding of cells along the urinary tract and (iv) exosome secretion [6]. Some reliable biomarkers present in urine has been identified as candidates to diagnostic and prognostic of some diseases [7,8]. Such biomarkers are mainly related to diseases of the urinary system, and some have been approved and also adopted by the FDA qualification program [9]. By now, the proteomics community is intensively working on finding new biomarkers for a number of diseases.

In urine-based proteomics biomarker discovery, second-morning urine is becoming the standard sample to be collected for biomarker-based studies. Such urine is collected during mid-stream due to the low probability of protein contamination from overgrown bacteria and epithelial cells [10]. During collection, urine preservation is often done using bacterial growth inhibitors [11]. However, the use of protease inhibitors is no longer recommended [12].

The in-depth study of the urine proteome by mass spectrometry is time consuming due to the time required to complete the analysis, usually 90 min gradients, plus the subsequent bioinformatic data interrogation. The impossibility of analysing the proteome of many urine samples in short periods of time makes mandatory the need of urine preservation and storage. Thus, after collection, the conditions for the stabilization and storage of urine proteomes in the short, medium and long term are critical because urine can degrade. Currently, the gold standard methodology for urine storage for proteomics applications was defined by the Human Kidney and Urine Proteome Project (HUKPP) [13] and by the European Urine and Kidney Proteomics (EuroKUP) [14] as follows: after collection, it is recommended urine centrifugation to remove cell debris and then to storage it at -20°C or -80°C [15,16]. Such recommendations present some drawbacks, as there is the need for expensive freezers which, in addition to the price, must be connected to emergency energy supply systems to prevent the effects of electric power failures. Furthermore, repeated freezing and thawing cycles have a detrimental effect in the sample. Thus, it has been shown urine proteome alterations after just 5 freeze-thaw cycles [17,18]. Finally, to store a large amount of urine samples is space troubling.

Of the many different sample treatment methods available to deal with complex proteomes, the Filter-Aided Sample Preparation method, FASP [19], is one of the most versatile bottom-up proteomic approaches, outperforming another in-solution digestion procedures [20]. Thus, FASP consists in a protein extraction and subsequent digestion approach that uses an ultrafiltration weight cut-off molecular membrane inserted in a microtube. Such device facilitates removal of salts, detergents and other contaminants from the protein sample using centrifugation [21]. It was first designed to extract and digest proteins from tissues [22] and cell samples [23], and it has been further applied to a large variety of biological samples including urine [24]. In addition, large scale sample handling is possible as the application of FASP in 96 well plates has also been shown [25]. To date, the use of membranes to separate

and preserve the urine proteome has been cited in literature few times [26,27].

The Stabilizer™ T1 is a portable, benchtop instrument that uses rapid conductive heating to preserve biological samples as closely as possible to their original state. This technology has proven to be very useful in the preservation of the proteome complex samples such as tissues and blood [28,29].

In this work we describe a new approach to preserve the urinary proteome at constant laboratory temperature (20°C). The novel method combines urine proteome separation using the FASP membrane followed by proteome thermal snap using the Stabilizer™ T1. In brief, the urinary proteome was separated and concentrated using the filter aided sample preparation method, and then the proteome attached to the FASP membrane was snap-heated by a thermal shock treatment provided by a rapid conductive heating that results in irreversible denaturation of proteins, permanent elimination of all enzymatic activity and also in the inactivation of viral and bacterial pathogens.

We demonstrate in this work that by using a volume as low as 1.2 mL of urine, the proteome of urine samples can be preserved and stored in the FASP membrane during at least 6 months at a constant temperature of 20°C . Furthermore, the handling of the proteome for subsequent label-free quantification analysis by high resolution mass spectrometry is facilitated as the standard FASP-protocol can be followed. Therefore, sampling large volume of urine for proteomic analysis, and urine storage and preservation in large volume containers with the aid of freezers is no longer necessary, at least for a six months period. This work is not a biomarker discovery one but it presents a new analytical method to separate and preconcentrate the urine proteome, to preserve it and to further processing it for peptide-based mass spectrometry interrogation. The latter includes, but is not limited to, biomarker discovery.

2. Experimental section

2.1. Urine specimens and collection

Fig. 1 depicts the sample handling and sample treatment workflow. The urine samples were collected according the HUKPP guideline. Briefly, fresh second-morning mid-stream urines were collected from four healthy male volunteers in the same age group (24 ± 1 years old). All the volunteers provided a signed consent for the usage of their urine in this study. The urine samples were collected into 100 mL urine collection tube containing 37 mg of boric acid. The urine was divided into two 50 mL tubes and were then centrifuged at 1000 g for 10 min to remove cell debris. After centrifugation, the supernatants were extracted and divided into 10 mL aliquots.

2.2. Urinary proteome extraction

Urine sample concentration was processed in a Vivacon 500, 10,000 molecular weight cut-off hydrosart membrane. Before sample loading, 100 μL of 70% ethanol solution was added to the vivacon and then centrifugated at 14,000 g for 15 min. A urine volume of 400 μL was then loaded into the filter device and then was centrifuged during 15 min at 14,000 g. This step was repeated three times, and a total of 1.2 mL urine was passed through the membrane. In the last centrifugation, the time of centrifugation was 25 min to ensure that the membrane was dried and there was no urine on the top of the membrane (Supplementary material SM1a). The proteins remained on the membrane.

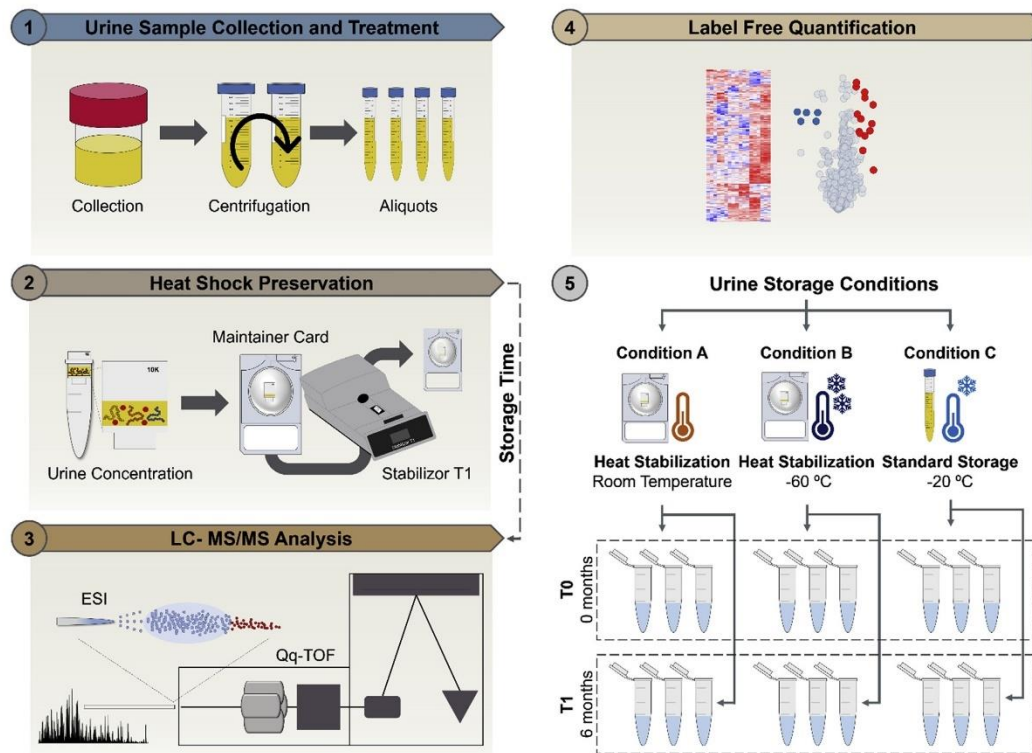


Fig. 1. Method Workflow chart. (1) First, urine is collected, then centrifuged to avoid cell debris and then aliquoted and (2) stored following one of these 3 conditions: (A) using the ultrafiltration membrane, the heat shock preservation and stored to laboratory temperature (20 °C); (B) using the membrane, the heat shock preservation and stored to -60°; (C) standard storage at -20 °C. (3 and 4). After 48 h or 6 months all stored samples are treated following the FASP protocol for protein cleavage and then samples are quantified by label-free high-resolution mass spectrometry. (5) Urine storage conditions and proteome comparisons done in this work. The proteomes were quantitatively compared to find out differences between the storage conditions within the same period of storage, 48 h or 6 months, and between periods of storage, 48 h versus 6 months.

2.3. Heat shock preservation

Urinary proteome present in the membrane was preserved by a heat treatment using the bench instrument stabilizer™ T1. The membrane containing the urine proteome was stored in a stabilizer™ T1 maintainer® tissue card that then undergoes inside the stabilizer™ T1 equipment where is applied a rapid conductive heating to preserve the biological samples (conditions A and B). The preservation is processed under a custom program with a controlled pressure and the following parameters: 30 s of sample treatment time, cavity vacuum pressure set to 5 mbar and heater position to 2.1 mm with a temperature of 95 °C (Supplementary material SM1 b–d). For control purposes, urine samples were stored at -20 °C without heating treatment (condition C).

2.4. Urine storage

Urine sample storage and preservation were processed in 3 different conditions, as it follows: (A) Urine proteome separation using ultrafiltration and heat stabilization followed by storage at

laboratory room temperature (RT, 20 °C), (B) Heat stabilization and storage at -60 °C and (C) and for control purposes, freezing at -20 °C with no heating stabilization. When urine samples in condition C were thawed a precipitate was observed. In order to resolubilize the precipitate, urine sample pH was adjusted to approximately 8 using 1 M tris-HCl. After storage, the total protein of each urine was quantified by a Bradford protein assay.

2.5. Protein digestion using filter aided sample preparation

The preserved and stored urinary proteins were analysed at two different time points: (i) T0, 48 h of storage and (ii) T1, 6 months of storage. After storage, the urinary proteins were digested using the FASP method with a few modifications [19]. Briefly, FASP, was used as the methodology to first separate, preserve and then digest the proteome of the urines. The protein present in the membrane were washed with 200 µL of 8 M urea 25 mM ammonium bicarbonate (AmBic) solution to remove low-weight material and then centrifuged for 20 min at 14,000 g. Next, the proteins were reduced by adding 200 µL of 50 mM dithiothreitol in 8 M urea and 25 mM

AmBic and incubated for 60 min at 37 °C. Then, a centrifugation of 20 min at 14,000 g was done, and then, the sample was alkylated during 45 min in the dark by the addition of 100 μ L 50 mM iodoacetamide in 8 M urea and 25 mM AmBic solution. Subsequently, a centrifugation of 20 min at 14,000 g was processed and then the sample was washed twice with 200 μ L 25 mM AmBic. Finally, protein digestion was initiated by the addition of 100 μ L 1:30 trypsin in 12.5 mM AmBic solution. Protein digestion proceeded overnight (approximately 14 h) at 37 °C and the peptides were collected with a centrifugation of 20 min at 14,000 g. The peptides were transferred to a 500 μ L microtube and were dried and stored at –20 °C.

2.6. Nano-LC-MS/MS

Peptides were resuspended in 100 μ L of 3% (v/v) acetonitrile containing 0.1% (v/v) aqueous formic acid (FA). Then, samples were homogenized by 5 min on vortex followed by 10 min on an ultrasonic bath at 100% ultrasonic amplitude. Afterwards, samples were quantified using a pierce™ quantitative colorimetric peptide assay. Following the previous step, 4 μ L containing 400 $\pm$ 6 ng of peptides were loaded onto a trap column (Acclaim PepMap100, 5 μ m, 100 Å, 300 μ m i.d. $\times$ 5 mm) and desalted for 5 min from 3% to 5% B (B: 90% acetonitrile 0.08% FA) at a flow rate of 15 μ L min^{–1}. Then the peptides were separated using an analytical column (Acclaim™ PepMap™ 100C18, 2 μ m, 0.075 mm i.d. $\times$ 150 mm) with a linear gradient at 300 nL min^{–1} (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) acetonitrile and 0.08% (v/v) FA) 5–90 min from 5% to 35% of mobile phase B, 90–100 min linear gradient from 35% to 95% of mobile phase B, 100–110 95% B. Chromatographic separation was carried out at 35 °C. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8–32 Hz), cycle time 3.0 s, 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.

2.7. Protein identification and quantification

LC-MS/MS data were analyzed using Data Analysis 4.2 software (Bruker). Relative label-free quantification was carried out using MaxQuant software V1.6.0.16. All raw files were processed in a single run with default parameters [30,31]. Database searches were performed using Andromeda search engine with the UniProt-SwissProt Human Uniprot Proteome database as a reference and a contaminants database of common contaminants [32]. Data processing was performed using Perseus (version 1.6.2.2) with default settings [33]. Protein group LFQ intensities were log2-transformed and the quantitative profiles were filtered for missing values with the following settings: min valid number 2 in at least one group and values greater than 0. To overcome the obstacle of missing LFQ values, missing values were imputed using default parameter, with = 0.3 and down shift = 1.8 [33]. Log ratios were calculated as the difference in average log2 LFQ intensity values between the tested conditions (two-tailed, Student's t-test, permutation-based FDR 0.05 and S0 of 1.5). S0 controls the relative importance of the resulted p-value and difference between means. At S0 = 0 only the p-value matters, while at nonzero S0 also the difference of means plays a role [34].

3. Results and discussion

3.1. Quantitative comparison between different storage conditions at T0 (48 h)

To assess the potential differences between the conditions of storage, the proteins levels of the urine samples were compared at T0 using label-free mass spectrometry-based protein quantification. The results presented in Supplementary material SM2 (volcano plots of the storage comparisons at T0) show that within 48 h the differences among the three conditions of preservation were minimal for the urines assessed. Red dots represent proteins with levels found statistically higher than expected between conditions (two-tailed, Student's t-test, permutation-based FDR 0.05, S0 of 1.5, n = 3). Blue dots represent proteins with levels found statistically lower than expected between conditions (two-tailed Student's t-test, permutation-based FDR 0.05, S0 of 1.5, n = 3). The number of proteins quantified was between 388 and 560, from which, less than 1% were found statistically significant (two-tailed, Student's t-test, permutation-based FDR 0.05, S0 of 1.5, n = 3), with the exception of condition A versus B in urine 1, where 20 proteins over a total of 560 were found statistically significant (two-tailed, Student's t-test, permutation-based FDR 0.05, S0 of 1.5, n = 3).

3.2. Quantitative comparison between different storage conditions at T1 (6 months)

The same comparison as for T0 was done to compare the storage conditions after 6 months of preservation. Volcano plots regarding the storage comparisons at T1 in Supplementary material SM2 show that less than 5% proteins were found statistically significant (up or down, two-tailed Student's t-test, permutation-based FDR 0.05, S0 of 1.5, n = 3) for most of the urine and conditions assessed, with the exception of condition A versus condition C in urine number 4. However, the number of proteins found different was lower than 15%.

3.3. Quantitative T0 versus T1 comparison for each storage condition

The variation of the levels of the common proteins with storage conditions and storage time was assessed using Pearson's correlation plots with all the technical replicates as shown in Fig. 2A. The Pearson's correlation coefficient was higher than 0.9 for the 87% of the comparisons done, thus showing the excellent preservation for a period of six months provided by each one of the different conditions tested in this work.

Once it was verified that the urine proteomes were preserved with similar performance between methods either at T0 or at T1 regardless the storage condition chosen, the variability in preservation for each method within the time period of 6 months was investigated next. Volcano plots from the urines 1, 2, 3 and 4 in Fig. 2B show that proteome preservation was successfully achieved within 6 months for the three storage conditions tested. Indeed, preservation conditions performed in a similar way, resulting in less than 2% statistically significant proteins (up or down, two-tailed Student's T-test, permutation-based FDR 0.05, S0 of 1.5, n = 3) for all urines and conditions with the unique exception of condition A in urine 1 which was lower than 6%. SM3, shows the analysis of the quantification coefficient of variation (CV%) for all the urines tested, as follows: the LFQ values obtained at T0 (three technical replicates) and the LFQ values at T1 (three technical replicates) were considered together, a medium and a standard deviation were calculated and then the CV% was also calculated. As SM3 shows the number of proteins with CV% below 5% for the LFQs

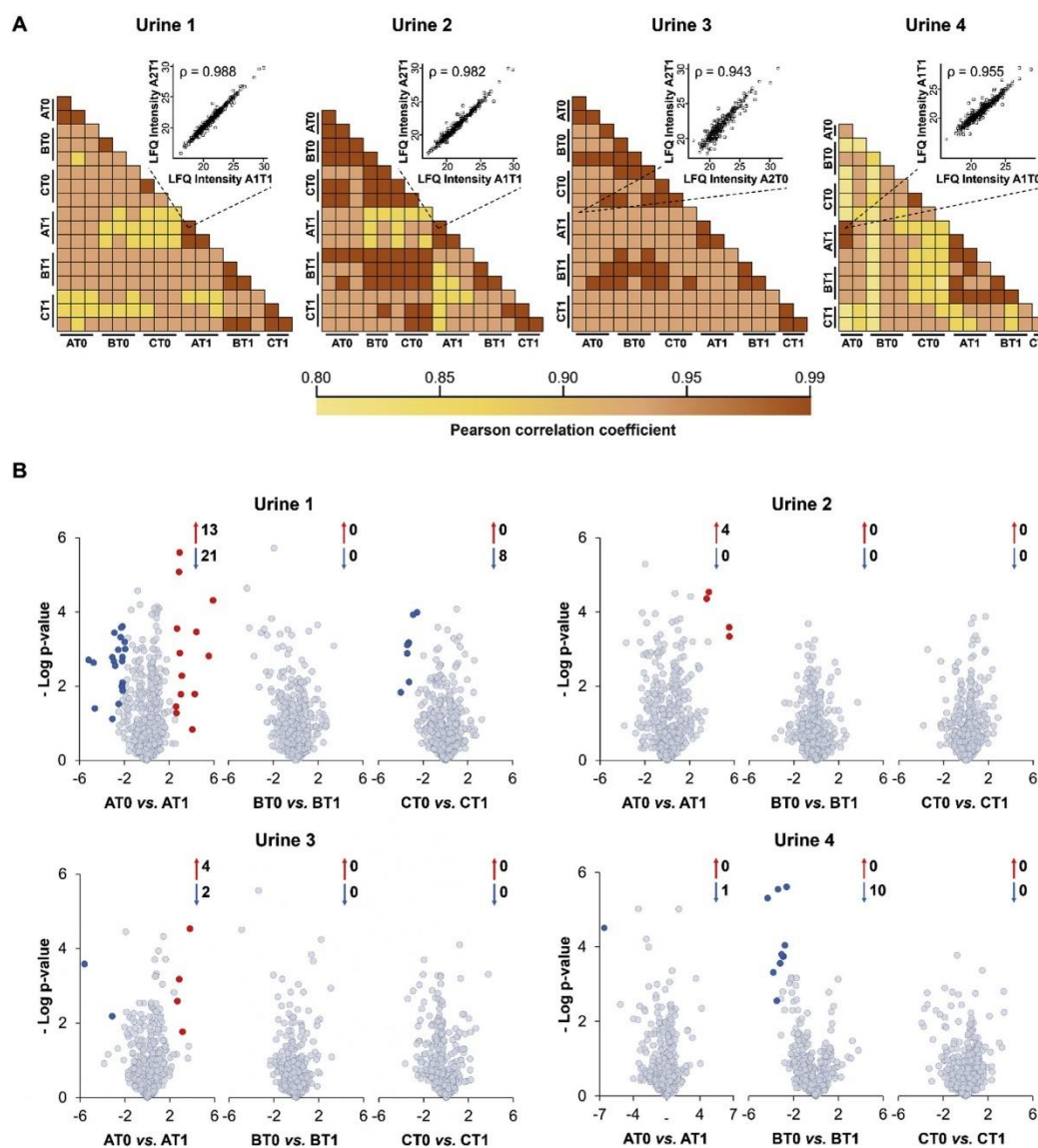


Fig. 2. (A) Color-coded Pearson's correlation values for the binary comparison of technical replicates and conditions tested showing the linearity of protein levels between conditions of storage and/or time of storage. Pearson's correlation values up to 0.99 demonstrate high reproducibility. The magnifications show the correlation between technical replicates (A1T1 vs. A2T1, for urine 1 and A1T1 vs. A2T1 for urine 2) and for urine 3 and 4 the magnifications show the correlation between different preservation times (urine 3 A2T0 vs. A2T1 and urine 4 A1T0 vs. A1T1). (B) Volcano plots showing the variation in protein levels with storage condition and storage time. T0: storage time 48 h. T1: storage time 6 months. Red and blue arrows indicate the statistically significant features (two-tailed Student's t-test, permutation-based FDR 0.05 and S0 of 1.5, $n = 3$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

obtained vary from a minimum of 62% to a maximum of 80% of the total proteins. This excellent result is even better if we consider the proteins with LFQs with a CV% below 10%, where the minimum was 95% and the maximum was 98%.

3.4. Hierarchical cluster analysis of each storage methods and time

Fig. 3 depicts the hierarchical clusters obtained when all the LFQ data series for each urine obtained at different storage conditions and storage time are compared. The LFQ data of each preservation method for T0 and T1 was compared using an ANOVA test (permutation-based FDR, FDR 0.05, S0 of 0 [34,35]) and the significant proteins were normalized by Z-scores. The Z-score offers information about how many standard deviations differ from the mean. Once this data for each technical replicate is selected as explained above, they are grouped by storage conditions and storage time. Then, all are compared by unsupervised clustering. The hierarchical clusters shown in Fig. 3 suggest that data obtained with condition of storage A (Stabilizer T1 and room temperature) tend to group together whilst data obtained with condition B (Stabilizer T1 and -60°C) and condition C (-20°C) tend to group together. This second group cannot be grouped correctly for urines 1, 2 and 3. In urine 4, conditions A and B are mixed. The aforementioned results obtained with the hierarchical cluster analysis are in agreement with the information retrieved from the volcano plots, which address that the three methods perform equally in terms of preservation.

3.5. Proof-of-concept: preservation of (i) acute kidney injury urinary biomarkers; (ii) kidney tubular injury; and (iii) FDA -cleared or -approved protein assayed in serum or plasma

To further investigate the preservation conditions studied in this work we decided to check the status of different urinary biomarkers of (i) Acute Kidney Injury (AKI), namely neutrophil gelatinase-associated lipocalin and cystatin-C; (ii) kidney tubular injury, namely clustering and osteopontin; and (iii) other 20 FDA -cleared or -approved protein assayed in serum or plasma [36]. Thus, a total of 24 biomarkers was assessed. Further details about the selected biomarkers, are shown in Fig. 4 and Supplementary Material SM4.

Over the last decade, different biomarkers have been studied and identified to diagnose AKI disease [37], and thus it can be found in literature the effect of urine storage at -70°C on the preservation of different AKI biomarkers. As an example, Liu et al. [38], have found in a study about biomarker preservation at -70°C that there was no significant impact of storage time over a median of 18 months on urine KIM-1, NGAL, IL-18 or L-FABP biomarkers in hospitalized samples; but a statistically significant effect towards a decrease over time was noted for KIM-1 in outpatient samples. In our case, the number of biomarkers with LFQ variation lower than 5% between the analysis done after 48 h and after 6 months (considering the three conditions evaluated) was around 20 ± 2 ($n = 24$, range 16–24) as depicted in Fig. 4. The number of biomarkers with a variation comprised between 5% and 10% was 3 ± 2 ($n = 24$ range 0–6) and the number of biomarkers with a variation between 10% and 20% was 1 ± 1 ($n = 24$, range 0–2). The aforementioned results are a good indicative that the three conditions assessed in this study works well. Interestingly, some of the biomarkers assessed in this proof of concept are typically used in studies done with plasma samples. Furthermore, we show here that such biomarkers can be found also in urine and that they are well preserved at least for six months. As an example, for the NGAL protein our result is in agreement with the one reported by Liu et al. [38] and by van de Vrie et al. [39] who claim that the levels of this

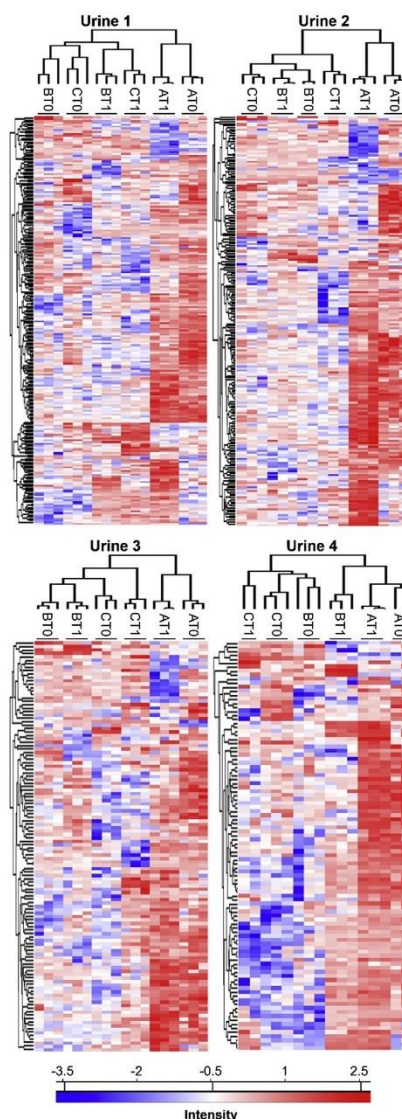


Fig. 3. Cluster analysis presenting the variation of proteins levels between the different conditions of storage and different time of storage. The profiles of preservation for conditions A present higher similarities among them and cluster together whilst the profiles for samples preserved at low temperature, condition B and condition C, cluster together.

biomarker remain substantially the same at -70°C and -80°C , respectively, for 6 months. The variation for this biomarker was lower than 5% in all the comparisons done by us as shown in Table SM1, thus confirming that preservation at low temperatures works well, as it also works the preservation using the snap-heat shock filter aided method. Further literature is in agreement with

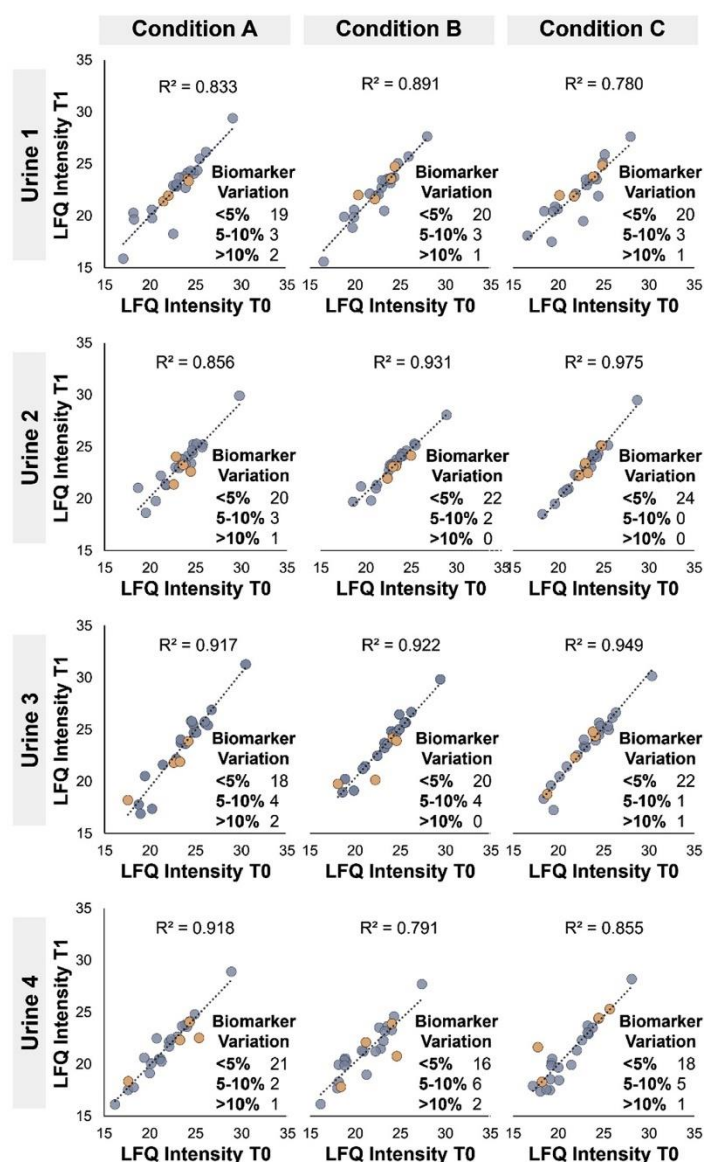


Fig. 4. Comparison of LFQ intensities for 24 selected -cleared or -approved protein assayed in serum or plasma between T0 (48 h) and T1 (6 months) for the 3 conditions tested in this study. It can be seen that most of the biomarker's present variations lower than 5% between T0 and T1. Serum and plasma biomarkers (blue): Albumin, α -1-acid glycoprotein, α -1-antitrypsin, α -2-HS-glycoprotein, α -2-macroglobulin, amylase, antithrombin III, apolipoprotein A1, β -2 microglobulin, biotinidase, ceruloplasmin, complement C3, fibrinogen, fibronectin, her-2/neu protein, lactoferrin, plasminogen, prostatic acid phosphatase, pyruvate kinase, thyroxine binding globulin. Urinary biomarkers (orange): clusterin, cystatin C, neutrophil gelatinase-associated lipocalin, osteopontin. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

us, as Collins et al. [40] reported albumin preservation during 6 months at -20°C with no significant changes; and Giensen et al. [41] reported albumin and NGAL preservation at room temperature (18°C) and at -20°C with the addition of boric acid with no changes for 1 week.

4. Conclusions

The work here presented shows clearly that the use of heat stabilization in combination with the FASP approach allows preservation and further processing of the entire urine proteome within six months without the need of storing the sample in a freezer. Such a novel approach is going to change the way urine proteome is preserved as space and conditions needed for storage are dramatically simplified. The risk of sample degradation caused by malfunction of freezers is also overcome, decreasing the cost of sample storage. Furthermore, the use of FASP approach to pre-concentrate the urinary proteins simplifies storage and transport of samples. The approach presented in this work opens new perspectives in sample preservation as it would be interesting to test for how long samples can be preserved at room temperature using heat stabilization. Finally, although this method has been applied to urine, it can be extended to any liquid biopsy such as serum and plasma. Therefore, it is expected that this study will open new avenues for other types of liquid biopsies.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Luís B. Carvalho: Methodology, Software, Writing - review & editing. **José Luis Capelo-Martínez:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Carlos Lodeiro:** Conceptualization, Funding acquisition, Project administration, Writing - review & editing. **Jacek R. Wiśniewski:** Writing - review & editing. **Hugo M. Santos:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

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). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [42] via the PRIDE partner repository with the data set PXD012880.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2019.05.051>.

References

- [1] D.M. Good, V. Thongboonkerd, J. Novak, J.-L. Bascands, J.P. Schanstra, J.J. Coon, A. Dominiczak, H. Mischak, Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future, *J. Proteome Res.* 6 (2007) 4549–4555, <https://doi.org/10.1021/pr070529w>.
- [2] J. Bergquist, M. Palmblad, M. Wetterhall, P. Håkansson, K.E. Markides, Peptide mapping of proteins in human body fluids using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Mass Spectrom. Rev.* 21 (2002) 2–15, <https://doi.org/10.1002/mas.10016>.
- [3] É. Csász, G. Kalló, B. Márkus, E. Deák, A. Csutak, J. Tózsér, Quantitative body fluid proteomics in medicine — a focus on minimal invasiveness, *J. Proteomics* 153 (2017) 30–43, <https://doi.org/10.1016/j.jprot.2016.08.009>.
- [4] S. Thomas, L. Hao, W.A. Rieke, L. Li, Biomarker discovery in mass spectrometry-based urinary proteomics, *Proteomics* 10 (2010) 358–370, <https://doi.org/10.1002/prca.201500102>.
- [5] R. Pieper, C.L. Gatlin, A.M. McGrath, A.J. Makusky, M. Mondal, M. Seonaraín, E. Field, C.R. Schatz, M.A. Estock, N. Ahmed, N.G. Anderson, S. Steiner, Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots, *Proteomics* 4 (2004) 1159–1174, <https://doi.org/10.1002/pmic.200300661>.
- [6] V. Thongboonkerd, P. Malasit, Renal and urinary proteomics: current applications and challenges, *Proteomics* 5 (2005) 1033–1042, <https://doi.org/10.1002/pmic.200401012>.
- [7] S. Decramer, A. Gonzalez de Peredo, B. Breuil, H. Mischak, B. Monsarrat, J.-L. Bascands, J.P. Schanstra, Urine in clinical proteomics, *Mol. Cell. Proteom.* 7 (2008) 1850–1862, <https://doi.org/10.1074/mcp.R800001-MCP200>.
- [8] M. Harpole, J. Davis, V. Espina, Current state of the art for enhancing urine biomarker discovery, *Expert Rev. Proteomics* 13 (2016) 609–626, <https://doi.org/10.1080/14789450.2016.1190651>.
- [9] FDA, List of Qualified Biomarkers, n.d. <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/ucm535383.htm>, (Accessed 10 January 2019).
- [10] V. Thongboonkerd, Practical points in urinary proteomics, *J. Proteome Res.* 6 (2007) 3881–3890, <https://doi.org/10.1021/pr0703285>.
- [11] V. Thongboonkerd, P. Saetun, Bacterial overgrowth affects urinary proteome analysis: recommendation for centrifugation, temperature, duration, and the use of preservatives during sample collection, *J. Proteome Res.* 6 (2007) 4173–4181, <https://doi.org/10.1021/pr070311>.
- [12] P. Havanapan, V. Thongboonkerd, Are protease inhibitors required for gel-based proteomics of kidney and urine? *J. Proteome Res.* 8 (2009) 3109–3117, <https://doi.org/10.1021/pr900015q>.
- [13] T. Yamamoto, R.G. Langham, P. Ronco, M.A. Knepper, V. Thongboonkerd, Towards standard protocols and guidelines for urine proteomics: a report on the human kidney and urine proteome project (HKUPP) symposium and workshop – 6 October 2007, Seoul, Korea and 1 November 2007, San Francisco, CA, USA, *Proteomics* 8 (2008) 2156–2159, <https://doi.org/10.1002/pmic.200800138>.
- [14] A. Vlahou, J. Schanstra, J. Frokiaer, M. El Nahas, G. Spasovski, H. Mischak, B. Dorn, G. Allmaier, E. Bongcam-Rudloff, T. Attwood, Establishment of a European network for urine and kidney proteomics, *J. Proteomics* 71 (2008) 490–492, <https://doi.org/10.1016/j.jprot.2008.06.009>.
- [15] HKUPP, Standard protocol for urine collection and storage, n.d. <http://www.hkupp.org/Urine%20collection.Documents.htm>, (Accessed 7 January 2019).
- [16] C.E. Thomas, W. Sexton, K. Benson, R. Sutphen, J. Koomen, Urine collection and processing for protein biomarker discovery and quantification, *Cancer Epidemiol. Biomark. Prev.* 19 (2010) 953–959, <https://doi.org/10.1158/1055-9965.EPI-10-0069>.
- [17] D.H. Jackson, R.E. Banks, Banking of clinical samples for proteomic biomarker studies: a consideration of logistical issues with a focus on pre-analytical variation, *Proteomics Clin. Appl.* 4 (2010) 250–270, <https://doi.org/10.1002/prca.200900220>.
- [18] Y. Zhang, Y. Luo, H. Lu, N. Wang, Y. Shen, R. Chen, P. Fang, H. Yu, C. Wang, W. Jia, Effect of freeze/thaw cycles on several biomarkers in urine from patients with kidney disease, *Biopreserv. Biobanking* 13 (2015) 144–146, <https://doi.org/10.1089/bio.2014.0033>.
- [19] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat. Methods* 6 (2009) 359–362, <https://doi.org/10.1038/nmeth.1322>.
- [20] J.R. Wiśniewski, Quantitative evaluation of filter aided sample preparation (FASP) and multienzyme digestion FASP protocols, *Anal. Chem.* 88 (2016) 5438–5443, <https://doi.org/10.1021/acs.analchem.6b00859>.
- [21] J.R. Wiśniewski, D.F. Zielinska, M. Mann, Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method, *Anal. Biochem.* 410 (2011) 307–309, <https://doi.org/10.1016/j.ab.2010.12.004>.
- [22] J.R. Wiśniewski, P. Ostasiewicz, M. Mann, High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers, *J. Proteome Res.* 10 (2011) 3040–3049, <https://doi.org/10.1021/pr200019m>.
- [23] A.H. Bryk, J.R. Wiśniewski, Quantitative analysis of human red blood cell proteome, *J. Proteome Res.* 16 (2017) 2752–2761, <https://doi.org/10.1021/acs.jproteome.7b00025>.
- [24] M. Zhao, M. Li, Y. Yang, Z. Guo, Y. Sun, C. Shao, M. Li, W. Sun, Y. Gao,

- A comprehensive analysis and annotation of human normal urinary proteome, *Sci. Rep.* 7 (2017) 3024, <https://doi.org/10.1038/s41598-017-03226-6>.
- [25] Y. Yu, M.-J. Suh, P. Sikorski, K. Kwon, K.E. Nelson, R. Pieper, Urine sample preparation in 96-well filter plates for quantitative clinical proteomics, *Anal. Chem.* 86 (2014) 5470–5477, <https://doi.org/10.1021/ac5008317>.
 - [26] L. Jia, X. Liu, L. Liu, M. Li, Y. Gao, Urinem, a membrane that can store urinary proteins simply and economically, makes the large-scale storage of clinical samples possible, *Sci. China Life Sci.* 57 (2014) 336–339, <https://doi.org/10.1007/s11427-013-4582-1>.
 - [27] J.R. Jesus, H.M. Santos, H. López-Fernández, C. Lodeiro, M.A.Z. Arruda, J.L. Capelo, Ultrasonic-based membrane aided sample preparation of urine proteomes, *Talanta* 178 (2018) 864–869, <https://doi.org/10.1016/j.talanta.2017.09.078>.
 - [28] C.B. Rountree, C.A. Van Kirk, H. You, W. Ding, H. Dang, H.D. VanGuilder, W.M. Freeman, Clinical application for the preservation of phospho-proteins through in-situ tissue stabilization, *Proteome Sci.* 8 (2010) 61, <https://doi.org/10.1186/1477-5956-8-61>.
 - [29] D. Blesborn, K. Sköld, D. Zeeberg, K. Kaewkhao, O. Sköld, M. Ahnoff, Heat stabilization of blood spot samples for determination of metabolically unstable drug compounds, *Bioanalysis* 5 (2013) 31–39, <https://doi.org/10.4155/bio.12.294>.
 - [30] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372, <https://doi.org/10.1038/nbt.1511>.
 - [31] S. Tyanova, T. Temu, A. Carlson, P. Sinitcyn, M. Mann, J. Cox, Visualization of LC-MS/MS proteomics data in MaxQuant, *Proteomics* 15 (2015) 1453–1456, <https://doi.org/10.1002/pmic.201400449>.
 - [32] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, *J. Proteome Res.* 10 (2011) 1794–1805, <https://doi.org/10.1021/pr101065j>.
 - [33] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (pro)teomics data, *Nat. Methods* 13 (2016) 731–740, <https://doi.org/10.1038/nmeth.3901>.
 - [34] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, *Proc. Natl. Acad. Sci. Unit. States Am.* 98 (2001) 5116–5121, <https://doi.org/10.1073/pnas.091062498>.
 - [35] S. Tyanova, J. Cox, Perseus, A bioinformatics platform for integrative analysis of proteomics data in cancer research, in: *Methods Mol. Biol.*, 1711th ed., Humana Press, New York, NY, 2018, pp. 133–148, https://doi.org/10.1007/978-1-4939-7493-1_7.
 - [36] N.L. Anderson, The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum, *Clin. Chem.* 56 (2010) 177–185, <https://doi.org/10.1373/clinchem.2009.126706>.
 - [37] J. Vanmassenhove, R. Vanholder, E. Nagler, W. Van Biesen, Urinary and serum biomarkers for the diagnosis of acute kidney injury: an in-depth review of the literature, *Nephrol. Dial. Transplant.* 28 (2013) 254–273, <https://doi.org/10.1093/ndt/gfs380>.
 - [38] K.D. Liu, E.D. Siew, W.B. Reeves, J. Himmelfarb, A.S. Go, C. Hsu, M.R. Bennett, P. Devarajan, T.A. Ikizler, J.S. Kaufman, P.L. Kimmel, V.M. Chinchilli, C.R. Parikh, Storage time and urine biomarker levels in the ASSESS-AKI study, *PLoS One* 11 (2016), e0164832, <https://doi.org/10.1371/journal.pone.0164832>.
 - [39] M. van de Vrie, J.K. Deegens, J. van der Vlag, L.B. Hilbrands, Effect of long-term storage of urine samples on measurement of kidney injury molecule 1 (KIM-1) and neutrophil gelatinase-associated lipocalin (NGAL), *Am. J. Kidney Dis.* 63 (2014) 573–576, <https://doi.org/10.1053/j.ajkd.2013.10.010>.
 - [40] A.C. Collins, M. Sethi, F.A. MacDonald, D. Brown, G.C. Viberti, Storage temperature and differing methods of sample preparation in the measurement of urinary albumin, *Diabetologia* 36 (1993) 993–997, <http://www.ncbi.nlm.nih.gov/pubmed/8243884>.
 - [41] C. Giesen, J.C. Lieske, The influence of processing and storage conditions on renal protein biomarkers, *Clin. J. Am. Soc. Nephrol.* 11 (2016) 1726–1728, <https://doi.org/10.2215/CJN.08800816>.
 - [42] J.A. Vizcaino, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Rios, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.-A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.-J. Kraus, J.P. Albar, S. Martinez-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination, *Nat. Biotechnol.* 32 (2014) 223–226, <https://doi.org/10.1038/nbt.2839>.

| **B**

**PUBLISHED PAPER IN ANALYTICAL CHEMIS-
TRY - CORRESPONDING TO CHAPTER 4**

Ultrasonic-Based Filter Aided Sample Preparation as the General Method to Sample Preparation in Proteomics

Luís B. Carvalho, José-Luis Capelo-Martínez, Carlos Lodeiro, Jacek R. Wiśniewski, and Hugo M. Santos*

Cite This: *Anal. Chem.* 2020, 92, 9164–9171

Read Online

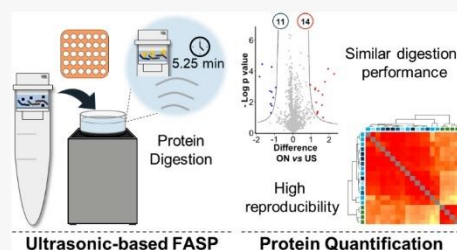
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: We propose a new high-throughput ultrafast method for large-scale proteomics approaches by speeding up the classic filter aided sample preparation protocol, FASP, from overnight to 2.5 h. Thirty-six samples can be treated in 2.5 h, and the method is scalable to 96-well plate-based pipelines. After a modification of the FASP-tube, the steps of protein reduction, protein alkylation, and protein digestion of complex proteomes are done in just 5.25 min, each one under the effects of an ultrasonic field (7 cycles: 30 s on and 15 s off). The new method was compared to the standard overnight digestion FASP protocol, and no statistical differences were found for more than 92.4%, 92%, and 93.3% of the proteins identified by studying the proteome of *E. coli*, mouse brain, and mouse liver tissue samples, respectively. Furthermore, the successful relative label-free quantification of four spiked proteins in *E. coli* samples, BSA, β -lactoglobulin, α -casein, and α -lactalbumin, was achieved, using either the ultrasonic-based FASP protocol or the classic overnight one. The new US-FASP method matches the analytical minimalism rules as time, cost, sample requirement, reagent consumption, energy requirements, and production of waste products are reduced to a minimum while maintaining high sample throughput in a robust manner as all of the advantages of the filter aided sample preparation protocol are maintained.



Mass spectrometry-based bottom-up protein quantification of complex proteomes relies on extensively handling and time-consuming approaches; all of them share several steps, including proteome extraction and solubilization, fractionation of the proteome (to reduce complexity), reduction and alkylation of protein disulfide bonds, and proteome digestion using enzymes.¹

The protein digestion step is the most time-consuming in the pipeline for proteome-wide quantification, as the standard protocols often rely on overnight incubations lasting up to 18 h.^{1,2} Therefore, attempts to reduce such time-consuming steps are reported in the literature. Thus, it has been demonstrated that the enzymatic digestion of proteins can be accelerated through an array of different tools, including microwave energy, infrared energy, heating, electrical fields, ultrasonic energy, and immobilized enzymes.^{2,3} In this context, ultrasonic energy as a way to speed up digestion of complex proteomes for protein identification from overnight to some minutes was first reported in 2005 by Ferrer et al.⁴ Subsequent developments of this methodology showed that ultrasonic energy can also accelerate from tens of minutes to some few the reduction and alkylation of protein's disulfide bonds.^{5–11} Furthermore, it has been shown recently that ultrasonic energy can also be successfully applied in the format of a 96-well plate, which thus makes it possible to digest 96 samples in just 4 min.¹² As the methodology was applied for protein identification, it remains

unknown if such an approach can also be applied for protein quantification by mass spectrometry.

Filter aided sample preparation (FASP) is a frequently used sample treatment method in proteomics.^{13,14} The method uses ultrafiltration membrane for protein handling and isolation of clean peptide fraction. In brief, in-solution complex protein mixtures are separated from other components using an ultrafiltration device where a membrane, with a certain cutoff, is used to retain the proteome content, while allowing the removal of smaller contaminant molecules and interfering buffers. With use of such membranes as a solid support, proteins are first reduced, then alkylated, and finally digested. Furthermore, there is an array of membranes types and cut-offs that make this method extremely versatile for proteomics. The FASP protocol offers a convenient way to handle the different steps of complicated workflows for proteome-wide quantification.

Received: April 6, 2020

Accepted: June 2, 2020

Published: June 2, 2020



ACS Publications

© 2020 American Chemical Society

9164

<https://dx.doi.org/10.1021/acs.analchem.0c01470>
Anal. Chem. 2020, 92, 9164–9171

In the present work, ultrasonic energy is introduced as a tool to speed up sample preparation in the standard FASP pipeline, and this new approach is named US-FASP. The ultrafiltration membrane in FASP acts as a proteomic reactor for detergent removal and buffer exchange, while ultrasonic energy is used to speed up chemical reactions and protein digestion. Overall, the treatment time is reduced from overnight to 2.5 h, which allows 36 samples to be processed simultaneously and facilitates high throughput sample preparation.

EXPERIMENTAL SECTION

***E. coli* Proteome.** *E. coli* samples were prepared using the ReadyPrep™ *E. coli* lysate sample from BioRad, following the manufacturers' instructions with minor modifications, as described in Supporting Information S11. In some experiments to access the different methods' reproducibility, five different known concentrations of bovine serum albumin (BSA), β -lactoglobulin, α -casein, and α -lactalbumin were spiked to the *E. coli* proteome.

Mouse Liver and Brain Proteomes. Fresh frozen mouse liver and brain were obtained from Patricell Limited. Tissue proteome extraction was performed as described in SI2. Total protein content was determined by the Bradford protein assay.

Filter Aided Sample Preparation (FASP). The standard overnight digestion FASP, ON-FASP, depicted in Figure 1A, was used to process the *E. coli*, mouse brain, and liver proteome with a few minor modifications.^{12–15} A total of 100 μ g of *E. coli* protein was loaded in a membrane with a 10 000 molecular weight cutoff (MWCO). The same amounts of mouse brain and liver protein were loaded in a membrane with a 30 000 molecular weight cutoff (MWCO). The proteins present in the membrane were washed with 200 μ L of 8 M urea, 25 mM AmBic solution and then centrifuged for 20 min at 14 000g. *E. coli* samples were reduced by adding 200 μ L of 50 mM dithiothreitol (DTT) in 8 M urea and 25 mM AmBic and incubated for 60 min at 37 °C. Mouse brain and liver proteomes skipped the step of reduction as the lysis buffer had this already incorporated. A centrifugation of 20 min at 14 000g was then done, and the sample was alkylated during 45 min in the dark by the addition of 100 μ L of 50 mM iodoacetamide (IAA) in 8 M urea and 25 mM AmBic solution. Subsequently, a centrifugation of 20 min at 14 000g was processed, and then the sample was washed twice with 200 μ L of 25 mM AmBic. Finally, protein digestion was initiated by the addition of 100 μ L of 1:30 trypsin in 12.5 mM AmBic solution. Protein digestion was proceeded overnight (approximately 16 h) at 37 °C, and then the peptides were collected by 20 min of centrifugation at 14 000g. To ensure that all of the peptides were extracted, 100 μ L of 3% (v/v) acetonitrile and 0.1% (v/v) formic acid were added followed by a centrifugation of 20 min at 14 000g. This previous step was repeated one more time, and then the peptides were transferred to a 500 μ L microtube, dried, and stored at –20 °C until further peptide quantification (SI3) and analysis by Nano-LC–MS/MS (SI4).

Ultrasonic-Based Filter Aided Sample Preparation (US-FASP). The US-FASP depicted in Figure 1A was used to process the protein *E. coli*, mouse brain, and mouse liver proteome and compared to the standard FASP described above. The same workflow, including cleaning steps, as described before, was used; however, ultrasound energy was applied to speed up the reduction, alkylation, and digestion steps. Briefly, the proteins were reduced by adding 200 μ L of

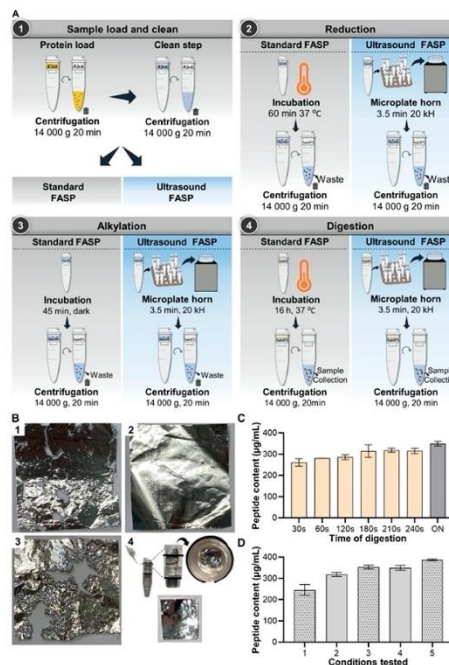


Figure 1. Ultrasonic-based filter aided sample preparation workflow and optimization. (A) Different methods, standard overnight digestion FASP (ON-FASP) and ultrasonic FASP (US-FASP), were used to process *E. coli*, mouse brain, and liver proteomes. Ultrasound energy was used to speed up the reduction, alkylation, and digestion FASP steps. (B) Effect of ultrasound energy in 5 cm × 5 cm aluminum foils. (1) 30 s ultrasonic time, UT, in an ultrasound bath at 100% ultrasonic amplitude, UA, and 35 kHz ultrasonic frequency, UF. (2) 30 s UT in an ultrasound bath at 100% UA and 130 kHz UF. (3) 30 s UT in an ultrasonic microplate horn assembly at 25% UA and 20 kHz UF. (4) 5.25 min UT in an ultrasonic microplate horn assembly at 25% UA and 20 kHz UF (0.5 cm × 0.5 cm aluminum foil inside the FASP tube). (C) Digestion time optimization. Different total elapsed ultrasound digestion times (30, 60, 180, 210, and 240 s) were accessed using the ultrasonic microplate horn assembly at 25% UA and 20 kHz UF and compared to the ON-FASP. (D) Ultrasound in the reduction, alkylation, and digestion steps. Bars with gray color with and without dots represent different batches of *E. coli*. Different conditions using the ultrasonic microplate horn assembly were done (1, no reduction and alkylation step, use of ultrasound energy in the digestion; 2, standard alkylation step and use of ultrasound energy in the reduction and digestion step; 3, use of ultrasound in the reduction, alkylation, and digestion step) and compared to the ON-FASP (4, 5).

50 mM dithiothreitol (DTT) prepared in 8 M urea and 25 mM AmBic. Tissue homogenization skipped the step of reduction as this was already done during the tissue homogenization using the lysis buffer. Ultrasound energy was then applied using the ultrasonic microplate horn assembly for 5.25 min (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). Afterward, centrifugation for 20 min at 14 000g was done, followed by protein alkylation by the addition of 100 μ L of 50 mM

iodoacetamide (IAA) prepared in 8 M urea and 25 mM AmBic solution. The alkylation step was sped up using the ultrasonic microplate horn assembly during 5.25 min (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). Finally, 100 μ L of 1:30 trypsin in 12.5 mM AmBic solution was added, and the protein digestion was processed using the ultrasonic microplate horn assembly for 5.25 min (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). Peptides were dried and stored at -20°C until further peptide quantification (SI3) and analysis by Nano-LC–MS/MS (SI4).

Nano-LC–MS/MS. Analysis was carried out using an Ultimate 3000 nano LC system coupled to an Impact HD (Bruker Daltonics) with a CaptiveSpray nanoBooster using acetonitrile as the dopant. Further details are provided in SI4. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium²⁷ via the PRIDE²⁸ partner repository with the data set identifier PXD018360.

Label-Free Protein Quantification and Statistical Analysis. Relative label-free quantification was carried out using MaxQuant software V1.6.10.43. All raw files were processed in a single run with default parameters.^{16,17} Database searches were performed using the Andromeda search engine with the UniProt - Proteome UP000000625, which contains proteins from *E. coli* and the sequences from the spiked proteins that were added to the fasta file.¹⁸ Data processing was performed using Perseus (version 1.6.5.0) with default settings.¹⁹ Protein group LFQ intensities were log 2-transformed, and the quantitative profiles were filtered for missing values with the following settings: min valid percentage of 70% in at least one group and values greater than 0. To overcome the obstacle of missing LFQ values, missing values were imputed using the parameters, with = 0.5 and down shift = 1.8.¹⁹ Log ratios were calculated as the difference in average log 2 LFQ intensity values between the tested conditions (two-tailed, Student's *t* test, permutation-based FDR 0.05 and S0 of 0.1).

■ RESULTS AND DISCUSSION

Initial Considerations about Variables Affecting Ultrasonic Performance. Previous research performed in our laboratory has recently shown that the ultrasonic microplate horn assembly is the ultimate ultrasonic tool to speed up complex proteomics pipelines as 96 samples can be digested in 4 min, and volumes as low as 10 μ L can be handled.¹² Thus, the analytical minimalism rules, as explained by Halls, are accomplished.^{12,20} We have also shown that high intensity focused ultrasonic energy has two effects over enzymatic kinetics, first speeding up the kinetics and then, after 5 min, inactivating the enzyme.²¹ Therefore, ultrasonic energy was not applied in this work for longer than 5 min. We have observed that ultrasonic amplitudes higher than 25% led to the rapid deterioration of the microplate horn assembly surface, but we demonstrated that for in-solution digestion of complex proteomes, such as *E. coli* lysates and tissue extracts, a 25% ultrasonic amplitude is adequate to speed the overnight proteome digestion process up to 4 min (intervals of 30 s on and 15 s off). Consequently, the ultrasonic amplitude chosen was 25%.¹²

Effect of Ultrasonic Frequency over the FASP Membrane. The implosion of ultrasonic nanobubbles generated by high-intensity ultrasound produced at frequencies between 20 and 40 kHz creates shock waves with disruptive effects.^{22–24} However, at higher frequencies, this disruptive

effect is less intense. Figure 1B1 and B2 shows two aluminum foils (5 $\times$ 5 cm) that were exposed to an ultrasonic field generated in an ultrasonic bath. Figure 1B1 clearly shows the disruptive effects over the aluminum foil of ultrasound generated at a frequency of 35 kHz, while Figure 1B2 shows that at a frequency of 130 kHz there is no disruption. The ultrasonic well plate works at a frequency of 20 kHz, and thus has a disruptive effect, which could potentially affect the membrane of the FASP tube. The difference in cavitation effects between the ultrasonic bath and the microplate horn is visualized in Figure 1B3, where it may be seen that the microplate horn's disruptive effect is more intense when the following conditions were used to deliver ultrasound: aluminum foil 5 $\times$ 5 cm, 30 s UT, 20 kHz UF, 25% UA. Therefore, the first task was to investigate the integrity of the FASP membrane. To this end, the effect of ultrasound energy over the FASP membrane was tested by introducing an aluminum foil in the FASP tube just over the membrane as depicted in Figure 1B4 (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). It was found that the microtube walls have a protective effect, which diminishes the disruptive effects of ultrasound, as no disruption was observed either in the aluminum foil or in the FASP membrane (Figure S1S). This fact was later confirmed by the results obtained for protein identification and quantification as shown below. Moreover, no membrane polymers were observed in the mass spectra. However, during the first experiments with ultrasound, water was observed inside the FASP tubes after some tens of seconds of ultrasonication. To avoid this effect, it was necessary to develop a small modification on the FASP tube design consisting of the addition of two O-rings, as shown in video S1. Finally, once it was verified that (i) no degradation of the FASP membranes was observed under the effects of an ultrasonic field and that (ii) no water entered inside the FASP tube during ultrasonication, the next step was to assess the acceleration of the FASP protocol using ultrasonic energy.

Effect of Time of Ultrasonication in the FASP-Based Digestion Protocol. The workflow of complex proteomes can be divided into three main steps: reduction, to break disulfide bonds; alkylation, to inactivate sulfur groups, which thus hampers disulfide bonds from forming again; and finally protein cleavage using trypsin, also known as the protein digestion step. The latter is by far the step that consumes more time as it is generally done overnight (18 h). Therefore, our first efforts were oriented to shorten this step. To this end, a series of *E. coli* samples were digested overnight following the standard FASP. Other series of the same *E. coli* were also digested under the effects of an ultrasonic field during different times, comprising the range 0.5–4 min (reduction and alkylation were done with no US applied for both sets of samples). If the success of the digestion is measured as the concentration of peptides obtained with the overnight digestion protocol, Figure 1C shows that after 3 min of digestion under the effects of an ultrasonic field, the peptide content of the ultrasonically digested sample is $91 \pm 8\%$ of the sample digested overnight. Therefore, an ultrasonication time of 5.25 min was settled as optimum for the digestion protocol (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF).

Effect of Ultrasonication in Reduction and Alkylation Steps. Once it was verified that the overnight digestion process in the FASP membrane can be accelerated to 5.25 min, it was investigated if the alkylation and digestion processes could be also shortened. To this end, a series of experiments

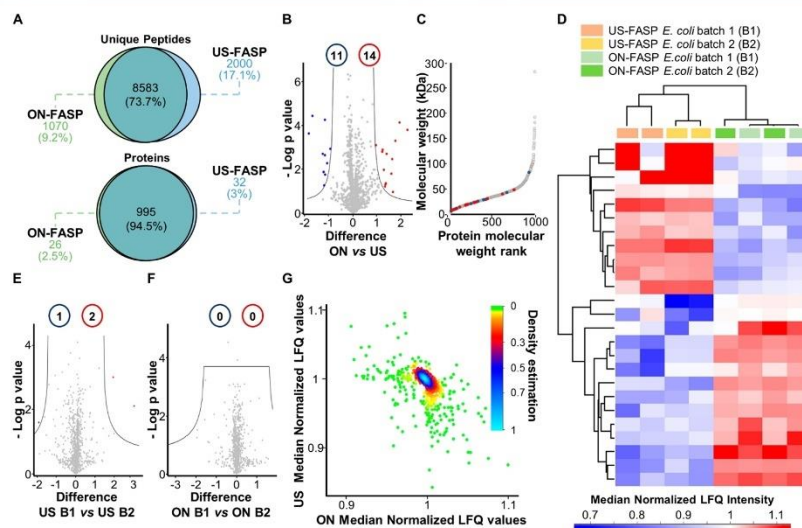


Figure 2. *E. coli* proteome digestion using the US-FASP and ON-FASP. (A) Venn diagram of the identified peptides and proteins in the US-FASP and ON-FASP. Different proteins identified in each method with a filter of 70% in at least one group. (B) Volcano plot showing the variation in protein levels in the comparison of ON-FASP versus US-FASP. Proteins were normalized by the median of each protein, and a two-tailed Student's *t* test (FDR 0.05 and S0 of 0.1, $n = 8$) was used for statistical analysis. Blue and red color dots represent the statistically significant proteins (numbers inside the colored circles), while the gray dots represent the nonstatistically significant proteins, respectively. (C) Protein molecular weight rank in the methods of digestion ON-FASP and US-FASP. (D) Hierarchical clustering with average linkage using the statically significant proteins (two-tailed Student's *t* test, FDR 0.05 and S0 of 0.1, $n = 8$). (E,F) Batch effect in each method. Protein levels were normalized by the median of each protein, and statistically significant changes were detected using a two-tailed Student's *t* test (FDR 0.05 and S0 of 0.1, $n = 4$). (G) Correlation plot and density from the comparison of ON-FASP versus US-FASP. Correlation plot of the proteins with median-based normalization shows a distribution centered around 1, which reflects a lower variation between the tested methods.

were designed in which *E. coli* extracts were submitted to reduction, alkylation, and digestion, with and without the effects of an ultrasonic field (7 cycles: 30 s on and 15 s off UT, 25% UA, 20 kHz UF). First, it was investigated whether the process could be done by applying ultrasound only in the digestion step, while skipping the alkylation and reduction steps. The results of this experiment are shown in Figure 1D1, and it may be seen that 60% of the expected peptide concentration is obtained. Next, we apply ultrasound only in the reduction and digestion steps, while skipping the application of ultrasound in the alkylation process. However, the results were not satisfactory as only 80% of the expected peptides were obtained (Figure 1D2). Finally, when all of the steps, reduction, alkylation, and digestion, were done under the effects of an ultrasonic field, the concentration of peptides was more than 90% of the expected amount (Figure 1D3). Thus, the set of experiments described above showed that ultrasonic energy conveniently speeds up each step of the proteomic workflow (Figure 1D4 and D5).

Finally, in further experiments, and as best conditions, for each step assessed (reduction, alkylation, and digestion), 5.25 min of ultrasonication time was applied in periods of 30 s followed by resting periods of 15 s seconds (total time of 5.25 min UT, with a 25% UA and at an UF of 20 kHz).

Mass Spectrometry Label Free-Based Protein Quantification: Ultrasonic-Based FASP versus Classic Overnight FASP. Figure 2 shows the comparison of ultrasonic-

based FASP protocol, US-FASP, versus classic overnight FASP, ON-FASP, protocol using *E. coli* samples and label free-based mass spectrometry ($n = 2$ different batches for each protocol run in two different days, with two technical replicates each batch, resulting in a total of four samples for each method). The two conditions tested resulted in 11 653 unique peptides and 1053 proteins, Figure 2A. Of the unique peptides identified, 8583 (73.7%) were found common for both methods, from which (i) 7022 peptides (81.8%) had no missed cleavages, (ii) 1432 peptides (16.7%) had one missed cleavage, and (iii) 125 unique peptides (1.5%) had two missed cleavages. Of the 1053 identified proteins, 995 (94.5%) were found common for both methods, Figure 2A. The second step was to quantify differences in digestion efficiency across the detected proteome. These differences are shown in Figure 2B. This volcano plot shows that only 25 proteins are differentially digested, which represents 2% of the detected proteome (two-tailed Student's *t* test, FDR 0.05 and S0 of 0.1, $n = 8$). This result shows the similarity of both proteomic pipelines (Figure S16). Furthermore, Figure 2C shows that the majority of the 2% of proteins found with different digestion efficiencies are of a molecular weight below 50 kDa. Interestingly, when these 25 proteins are used to generate a hierarchical clustering, Figure 2D, it becomes evident that such proteins are preferentially digested either by the US-FASP or by the ON-FASP. This observation is in agreement with the results reported previously.^{9,12}

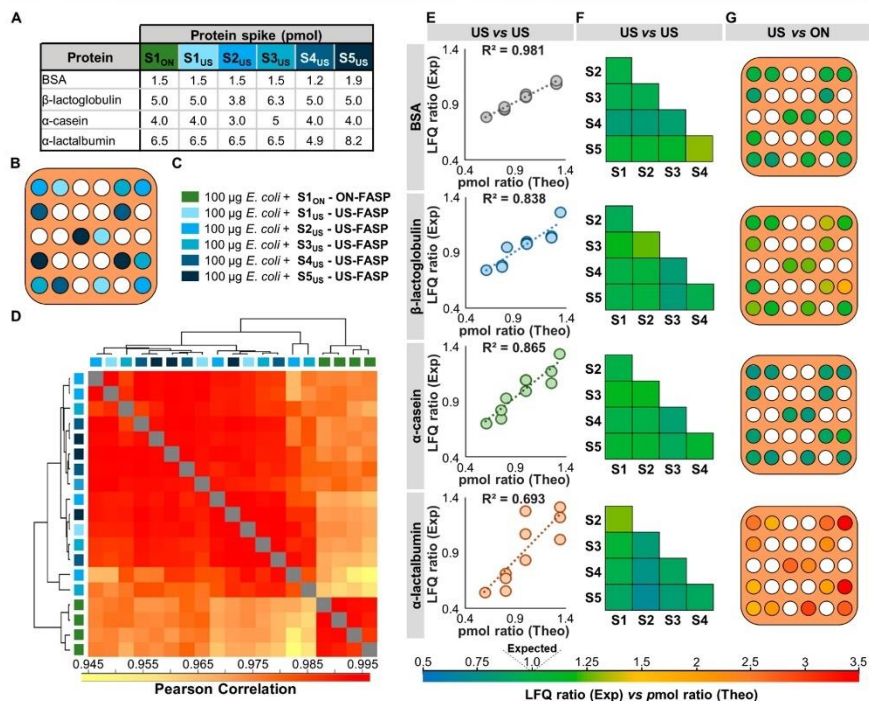


Figure 3. Ultrasonic space effectiveness in US-FASP. (A) Protein spikes and *E. coli* used in each condition. (B) Position of the different samples in the ultrasonication plate. (C) *E. coli* proteome and spiked proteins processed with ON-FASP and US-FASP. (D) Cluster by Pearson correlation of the *E. coli* samples with the different protein spikes using the ON-FASP and US-FASP. (E, F, and G) Comparison between the theoretical pmol ratio (theo) and the LFQ values obtained ratio (exp). The protein spikes were compared as follows: (E,F) US-FASP versus US-FASP and (G) US-FASP versus ON-FASP. The difference was obtained using formulas 1–5.

The reproducibility for each method was assessed as the difference between two batches of the same sample done in different days, and it is shown in Figure 2E and F. These volcano plots show that the interdaily variability of both methods is negligible (two-tailed Student's *t* test, FDR 0.05 and *S*0 of 0.1, *n* = 4 each method). When the 1053 protein LFQ values obtained after finishing each experimental pipeline are compared using the median-based normalization, the distribution is centered around 1, as shown in Figure 2G, which further reflects the few differences between both US- and ON-sample treatments.

Performance of Ultrasonic-Based FASP. To verify the homogeneity of the ultrasonic effects inside the microplate horn, *E. coli* samples were distributed randomly and were spiked at different levels with four proteins, as depicted in Figure 3A and B. The samples then were overnight or ultrasonically digested as depicted in Figure 3C. When the complete set of *E. coli* proteins of each sample is identified and quantified and then used to cluster the samples, the reproducibility as expressed by the Pearson correlation value is higher than 0.985 (Figure 3D), which shows that the digestion of the *E. coli* proteomes is done with the same efficiency in all of the positions and thus concludes that the effects of the ultrasound energy are homogeneously distributed

in the entire microplate horn. Moreover, when the ultrasonic treated samples are compared to the overnight treated ones (Figure 1D), the Pearson correlations are higher than 0.945, which again indicates that for a few proteins both methods perform slightly different but for the majority both perform equally.

Ultrasonic-Based FASP Label-Free Quantification of a Set of Spiked Proteins in *E. coli* Samples. To get a deeper insight into the potentialities of the proposed US-FASP method, *E. coli* samples were spiked with four known proteins, BSA, β-lactoglobulin, α-casein, and α-lactalbumin. The known protein amount ratios (Figure 3A) between samples were compared to their respective mass spectrometer signal ratios (LFQs ratio) using the formulas:

$$\text{pmol ratio (theo)} = \frac{\text{Spike}_{S_x}}{\text{Spike}_{S_y}} \quad (1)$$

$$\text{LFQ ratio (exp)} = \frac{\bar{X} \text{ LFQ}_{S_x}}{\bar{X} \text{ LFQ}_{S_y}} \quad (2)$$

where *x* and *y* can be samples S1, S2, S3, S4, or S5, Spike_{*S_x*} and Spike_{*S_y*} refer to the spikes added to the samples as they are presented in Figure 3A, and LFQ_{*S_x*} and LFQ_{*S_y*} refer to the

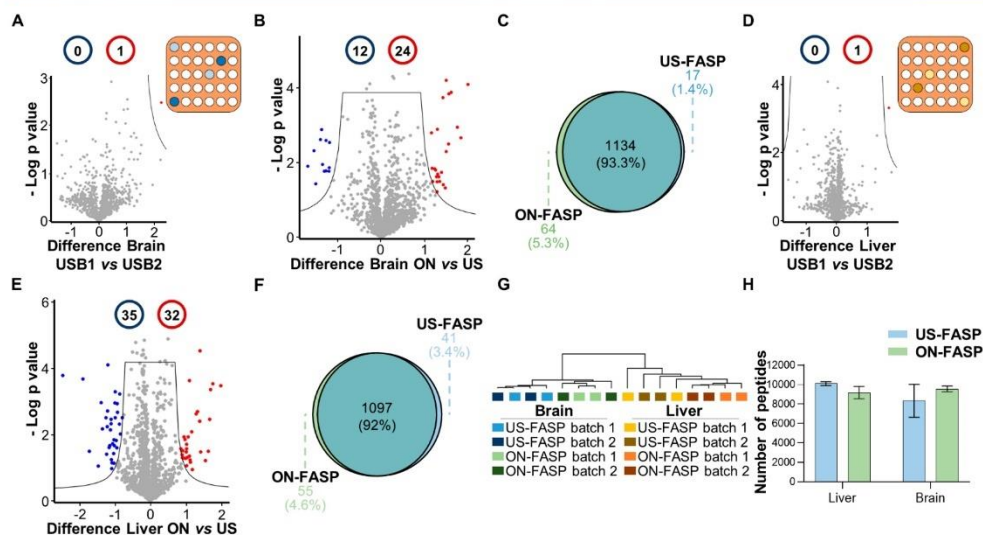


Figure 4. Mouse liver and brain proteome digestion using the US-FASP. (A–C) Quantitative proteomics analysis using the US-FASP and ON-FASP methods in brain mouse tissues. (D–F) Quantitative proteomics analysis using the US-FASP and ON-FASP methods in liver mouse tissues. (A and D) Batch and plate position effect in the US-FASP method. Volcano plot comparing each batch of tissue with random position in the ultrasonication plate. Protein levels were normalized by the median of each protein, and the different quantified proteins were obtained using a two-tailed Student's *t* test (FDR 0.05 and *S*₀ of 0.1, *n* = 4). Blue and red color dots represent the statistically significant proteins (numbers inside the colored circles), while the gray dots represent the nonstatistically significant proteins, respectively. (B and E) Protein quantitative levels comparison between the US-FASP and ON-FASP methods. Volcano plot comparing the protein levels in each method. Protein levels were normalized by the median of each protein, and the different quantified proteins were obtained using a two-tailed Student's *t* test (FDR 0.05 and *S*₀ of 0.1, *n* = 8). (C and F) Venn diagram with the proteins identified by US-FASP and ON-FASP method. Different proteins identified in each method with a filter of 70% presence in at least one group. (G) Cluster of the mouse brain and liver tissues processed by ON-FASP and US-FASP. Protein identified with a filter of 70% presence in at least one group was normalized by Z-score and a multisample test; ANOVA (FDR 0.05, *n* = 16) were plotted in a hierarchical clustering. (H) Number of peptides identified in each tissue type with the different FASP methods used. ON-FASP method is represented by the green color bars and US-FASP by the blue color bars.

respective mass spectrometry data of samples *x* and *y*. The quantification results are depicted in Figure 3E, where it may be seen that for proteins BSA, β -lactoglobulin, and α -casein the spiked ratios match close to the LFQ ratios, which is reflected in the regression coefficients close to 1. For the case of α -lactalbumin, the match is worst, with a regression coefficient of ca. 0.7.

In an ideal situation, where the proteins are efficiently digested and where the LFQs are proportional to the protein concentration, the ratio using the formula

$$\text{ratio} = \frac{\text{LFQ ratio (exp)}}{\text{pmol ratio (theo)}} \quad (3)$$

should be 1. When this ratio is calculated for the different positions inside the microplate horn, as shown in Figure 3F, it can be seen that the effects of ultrasonic energy are homogeneously distributed all over the surface, with the best results being obtained for BSA and the worst for α -lactalbumin. Thus, for BSA protein, the differences between the pmol ratios and the LFQs ratios were up to 6% (*n* = 10, range 1–13.5%), for the β -lactoglobulin they were 20% (*n* = 10, range 15–22%), and for α -casein they were 15% (*n* = 10, range 6–19%). For α -lactalbumin, the results were the worst, and the difference was 56% (*n* = 10, range 35–91%).

Ultrasonic-Based FASP versus Classic FASP Label-Free Quantification of a Set of Spiked Proteins in *E. coli* Samples. The final insight was done by comparing the values of the US-FASP versus ON-FASP for the four spiked proteins as follows:

$$\text{pmol ratio (theo)} = \frac{\text{Spike}_{\text{S1 ON}}}{\text{Spike}_{\text{Sx US}}} \quad (4)$$

$$\text{LFQ ratio (exp)} = \frac{\bar{X} \text{ LFQ}_{\text{S1 ON}}}{\text{LFQ}_{\text{Sx US}}} \quad (5)$$

where $\text{Spike}_{\text{Sx US}}$ can be ultrasonic treated samples S1, S2, S3, S4, or S5, $\text{Spike}_{\text{S1 ON}}$ refers to the sample treated overnight (see Figure 3C), $\text{LFQ}_{\text{Sx US}}$ refers to the mass spectrometry data of ultrasonic treated samples S1, S2, S3, S4, or S5, and $\text{LFQ}_{\text{S1 ON}}$ refers to the mass spectrometry data obtained for the sample treated overnight.

The ratio, using formula 3, was then calculated. This ratio should ideally be 1. It can be seen in Figure 3G that proteins BSA, β -lactoglobulin, and α -casein are digested at the same level using either the ultrasonic or the overnight method. For the case of the α -lactalbumin, the digestion was better for the overnight method, almost 1.5 times better.

Application to Brain and Liver Tissue Samples. The optimum sample treatment conditions were also applied to tissue samples. To this end, brain and liver tissue samples were submitted to the protein extraction explained in the Experimental Section, and, of each protein tissue extract, four subsamples were made. These subsamples were divided into two groups of two samples each. Samples were digested with the optimized conditions in different positions inside the microplate horn in different days as depicted in Figure 1. The results are shown in Figure 4. The first interesting fact is that no differences were found neither in the number of proteins nor in their quantification values when the analysis was done on different days (Figure 4A and D). Moreover, the same data also reflect that the position where the digestion is done inside the microplate horn does not affect the digestion efficiency (Figure 4B and E). Furthermore, the number of common proteins identified for each tissue type was almost the same regardless of the sample treatment used, 1134 (93.3%) for the brain tissue and 1097 (92%) for the liver tissue (Figure 4 C and F, respectively). In addition, the number of peptides identified was also similar for each method and each tissue type (Figure 4H). Finally, the data obtained also allow a nice classification of both tissues (Figure 4G).

CONCLUSION

A novel ultrasonic-assisted filter aided sample preparation protocol has been developed for large-scale proteomics approaches, which allows high sample throughput and reduces the treatment time from 18 to 2.5 h for a total of 36 samples. The method is scalable to 96 samples using 96-well plates and virtually applicable to similar digestion platforms such as S-Trap.²⁵ We foresee this method as the universal one for proteomics pipelines as it paves the way to clinical proteomics in hospitals because it is of easy handling as the operator needs no special skills. Also, this method matches perfectly the new high-resolution mass spectrometers that allow fast chromatographic runs in 15 min, which thus allows one to measure the proteome of about 100 samples a day per spectrometer. Further, shortening of the sample processing time can be achieved by skipping the cysteine alkylation step.²⁶

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c01470>.

Appendix A: Supplementary data containing the reagents and apparatus used, extended experimental section (S11–4), optical microscopy FASP membrane images (Figure S15), hierarchical clustering of the *E. coli* proteome processed with US-FASP and ON-FASP (Figure S16), and a custom-built wood plate vivacon support floater (Figure S17) (PDF)

Video S1: Video showing the digestion step in the US-FASP digestion method (MP4)

AUTHOR INFORMATION

Corresponding Author

Hugo M. Santos – BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, Faculty of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; PROTEOMASS Scientific Society, Madan Parque, Caparica 2825-182, Portugal; Department of

Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213, United States; orcid.org/0000-0002-6032-8679; Phone: +351 934 432 320; Email: hmsantos@fct.unl.pt

Authors

Luís B. Carvalho – BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, Faculty of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; PROTEOMASS Scientific Society, Madan Parque, Caparica 2825-182, Portugal; orcid.org/0000-0002-3761-8512

José-Luis Capelo-Martínez – BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, Faculty of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; PROTEOMASS Scientific Society, Madan Parque, Caparica 2825-182, Portugal; orcid.org/0000-0001-6276-8507

Carlos Lodeiro – BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, Faculty of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal; PROTEOMASS Scientific Society, Madan Parque, Caparica 2825-182, Portugal; orcid.org/0000-0001-5582-5446

Jacek R. Wiśniewski – Biochemical Proteomics Group, Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Martinsried D-82152, Germany; orcid.org/0000-0002-8452-5095

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.0c01470>

Author Contributions

J.-L.C.-M., C.L., and H.M.S. designed the experimental work and provided financial support. L.B.C. performed the laboratorial work under the supervision of H.M.S. J.-L.C.-M. and H.M.S. drafted the manuscript. J.-L.C.-M., C.L., J.R.W., and H.M.S. revisited the drafted version, corrected it, and made valuable suggestions.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

PROTEOMASS Scientific Society is acknowledged for the funding provided to the Laboratory for Biological Mass Spectrometry – Isabel Moura. We acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV, which is financed by national funds from FCT/MEC (UID/QUI/S0006/2019). H.M.S. and L.B.C. acknowledge the funding provided by FCT/MEC within the framework of the FCT Investigator Program (IF/00007/2015) and the Ph.D. grant (SFRH/BD/144222/2019), respectively. We acknowledge Eng. F. Silvestre from FCT FabLab (powered by USA Embassy) for his help in the manufacturing of the wood plate vivacon support floater, Figure S7.

REFERENCES

- (1) López-Ferrer, D.; Cañas, B.; Vázquez, J.; Lodeiro, C.; Rial-Otero, R.; Moura, I.; Capelo, J. L. *TrAC, Trends Anal. Chem.* **2006**, *25* (10), 996–1005.
- (2) Capelo, J. L.; Carreira, R.; Diniz, M.; Fernandes, L.; Galesio, M.; Lodeiro, C.; Santos, H. M.; Vale, G. *Anal. Chim. Acta* **2009**, *650* (2), 151–159.
- (3) Shen, X.; Sun, L. *Proteomics* **2018**, *18* (9), 1700432.

- (4) López-Ferrer, D.; Capelo, J. L.; Vázquez, J. J. *Proteome Res.* **2005**, *4* (5), 1569–1574.
- (5) Rial-Otero, R.; Carreira, R. J.; Cordeiro, F. M.; Moro, A. J.; Fernandes, L.; Moura, I.; Capelo, J. L. *J. Proteome Res.* **2007**, *6* (2), 909–912.
- (6) Santos, H. M.; Rial-Otero, R.; Fernandes, L.; Vale, G.; Rivas, M. G.; Moura, I.; Capelo, J. L. *J. Proteome Res.* **2007**, *6* (9), 3393–3399.
- (7) Galesio, M.; Vieira, D. V.; Rial-Otero, R.; Lodeiro, C.; Moura, I.; Capelo, J. L. *J. Proteome Res.* **2008**, *7* (5), 2097–2106.
- (8) Santos, H. M.; Mota, C.; Lodeiro, C.; Moura, I.; Isaac, I.; Capelo, J. L. *Talanta* **2008**, *77* (2), 870–875.
- (9) Jesus, J. R.; Santos, H. M.; López-Fernández, H.; Lodeiro, C.; Arruda, M. A. Z.; Capelo, J. L. *Talanta* **2018**, *178*, 864–869.
- (10) Martins, G.; Fernández-Lodeiro, J.; Djafari, J.; Lodeiro, C.; Capelo, J. L.; Santos, H. M. *Talanta* **2019**, *196*, 262–270.
- (11) Jorge, S.; Capelo, J. L.; LaFramboise, W.; Dhir, R.; Lodeiro, C.; Santos, H. M. *J. Proteome Res.* **2019**, *18* (7), 2979–2986.
- (12) Jorge, S.; Araújo, J. E.; Pimentel-Santos, F. M.; Branco, J. C.; Santos, H. M.; Lodeiro, C.; Capelo, J. L. *Talanta* **2018**, *178*, 1067–1076.
- (13) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat. Methods* **2009**, *6* (5), 359–362.
- (14) Wiśniewski, J. R. *Anal. Chim. Acta* **2019**, *1090*, 23–30.
- (15) Carvalho, L. B.; Capelo-Martínez, J. L.; Lodeiro, C.; Wiśniewski, J. R.; Santos, H. M. *Anal. Chim. Acta* **2019**, *1076*, 82–90.
- (16) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.
- (17) Tyanova, S.; Temu, T.; Carlson, A.; Sinitcyn, P.; Mann, M.; Cox, J. *Proteomics* **2015**, *15* (8), 1453–1456.
- (18) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.
- (19) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*; Nature Publishing Group: New York, 2016; pp 731–740.
- (20) Halls, D. J. *J. Anal. At. Spectrom.* **1995**, *10* (3), 169–175.
- (21) Galesio, M.; Loureno, J.; Madeira, D.; Diniz, M.; Capelo, J. L. *J. Mol. Catal. B: Enzym.* **2012**, *74* (1–2), 9–15.
- (22) Suslick, K. S.; Nyborg, W. L. *J. Acoust. Soc. Am.* **1990**, *87* (2), 919–920.
- (23) Mason, T. J. *Dev. Cardiovasc. Med.* **1996**, *178*, 25–54.
- (24) Capelo-Martínez, J.-L. *Ultrasound in Chemistry: Analytical Applications*; Wiley-VCH: New York, 2009.
- (25) Ludwig, K. R.; Schroll, M. M.; Hummon, A. B. *J. Proteome Res.* **2018**, *17*, 2480–2490.
- (26) Wiśniewski, J. R.; Zettl, K.; Pilch, M.; Rysiewicz, B.; Sadok, I. *Anal. Chim. Acta* **2020**, *1100*, 131–137.
- (27) Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; García-Seisdedos, D.; Jarnuczak, A. F.; Hewapathirana, S.; Pullman, B. S.; Wertz, J.; Sun, Z.; Kawano, S.; Okuda, S.; Watanabe, Y.; Hermjakob, H.; MacLean, B.; MacCoss, M. J.; Zhu, Y.; Ishihama, Y.; Vizcaino, J. A. *Nucleic Acids Res.* **2019**, *48* (D1), D1145–D1152.
- (28) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaino, J. A. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450.

| c

PUBLISHED PAPER IN TALANTA - CORRESPONDING TO CHAPTER 5

Talanta 266 (2024) 124953



Contents lists available at ScienceDirect

Talanta

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

Normalization methods in mass spectrometry-based analytical proteomics: A case study based on renal cell carcinoma datasets

Luis B. Carvalho^{a,b,1}, Pedro A.D. Teigas-Campos^{a,b,1}, Susana Jorge^{a,b}, Michele Protti^c, Laura Mercolini^c, Rajiv Dhir^d, Jacek R. Wiśniewski^e, Carlos Lodeiro^{a,b}, Hugo M. Santos^{a,b,d,**}, José L. Capelo^{a,b,*}

^a BIOSCOPE Group, LAQV-REQUIMTE, Chemistry Department, NOVA School of Science and Technology, FCT NOVA, Universidade NOVA de Lisboa, 2829-516, Caparica, Portugal

^b PROTEOMASS Scientific Society, Madan Park, 2829-516, Caparica, Portugal

^c Research Group of Pharmacotoxicological Analysis (PTA Lab), Department of Pharmacy and Biotechnology (FaBiT), Alma Mater Studiorum - University of Bologna, Via Belmeloro 6, 40126, Bologna, Italy

^d Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

^e Biochemical Proteomics Group, Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

ARTICLE INFO

Keywords:

Normalization methods
Mass spectrometry
Proteomics
Renal carcinoma

ABSTRACT

Normalization is a crucial step in proteomics data analysis as it enables data adjustment and enhances comparability between datasets by minimizing multiple sources of variability, such as sampling, sample handling, storage, treatment, and mass spectrometry measurements. In this study, we investigated different normalization methods, including Z-score normalization, median divide normalization, and quantile normalization, to evaluate their performance using a case study based on renal cell carcinoma datasets. Our results demonstrate that when comparing datasets by pairs, both the Z-score and quantile normalization methods consistently provide better results in terms of the number of proteins identified and quantified as well as in identifying statistically significant up or down-regulated proteins. However, when three or more datasets are compared at the same time the differences are found to be negligible.

1. Introduction

Some analytical chemists encounter significant challenges in analyzing large-scale proteomics data using bioinformatics tools, often because these topics are not sufficiently developed in undergraduate or post-graduate programs. As a result, analytical chemists may need to acquire additional skills and knowledge to effectively handle these types of data. From an analytical chemistry point of view, the so-called shot-gun-based proteomics workflows are characterized by multiple sources of variability, including sampling, sample handling, sample storage, sample treatment and mass spectrometry measurements [1,2]. For instance, analytical chemists know that large-scale and longitudinal studies often encounter a source of variation associated with the

difference in storage time and ageing of the samples [2,3]. Samples are typically collected and prepared at different times during the study, introducing an environmental factor that may affect the data. This factor may be compounded by other variables, such as the collection of samples in different centres or countries, which may be subjected to varying numbers of freeze-thaw cycles. In addition, the sample collection protocol may be applied with different levels of attention by different operators, which can lead to unwanted differences in the samples. Furthermore, transporting the samples to the laboratory where they will be stored or analysed can also introduce noise into the data, as the transportation conditions may vary, potentially altering the samples differently. Finally, in a typical shot-gun proteomics workflow, samples are collected and treated in the analytical laboratory before being stored

* Corresponding author. BIOSCOPE Group, LAQV-REQUIMTE, Chemistry Department, NOVA School of Science and Technology, FCT NOVA, Universidade NOVA de Lisboa, 2829-516, Caparica, Portugal.

** Corresponding author. BIOSCOPE Group, LAQV-REQUIMTE, Chemistry Department, NOVA School of Science and Technology, FCT NOVA, Universidade NOVA de Lisboa, 2829-516, Caparica, Portugal.

E-mail addresses: hmsantos@fct.unl.pt (H.M. Santos), jlem@fct.unl.pt (J.L. Capelo).

¹ These two authors have equally worked for this manuscript.

<https://doi.org/10.1016/j.talanta.2023.124953>

Received 17 May 2023; Received in revised form 7 July 2023; Accepted 14 July 2023

Available online 17 July 2023

0039-9140/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

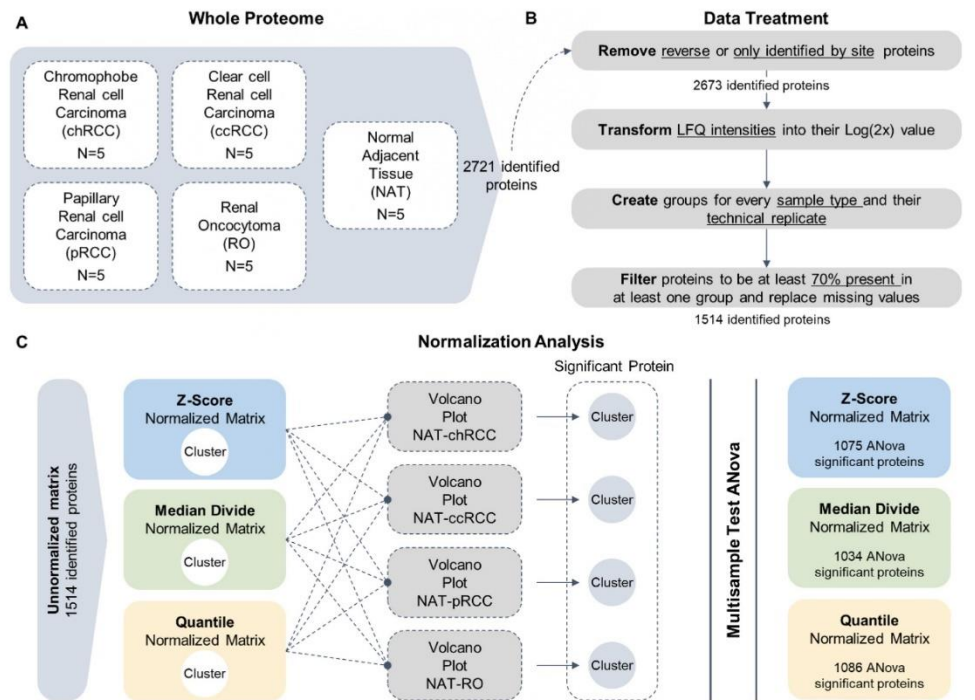


Fig. 1. Schematic representation of the data process applied to the proteomic data when comparing Neat Adjacent Tissue, NAT (control group) against multiple renal tumour types. **A:** Type and number of tumours (description of patients is given in Table SM1). **B:** Data treatment for the proteins quantified using the five datasets. **C:** Types of normalization made: Z-Score, median divide, Quantile and types of comparisons done: Volcano plots, Cluster analysis and Multisample Test ANOVA.

or investigated using mass spectrometry. The samples are then cleaved with various enzymes depending on the experimental strategy, and the resulting cleavage proteome composed of thousands of peptides is subjected to chromatography-mass spectrometry analysis. The noise generated during the pipeline of work, from sampling to mass spectrometry, makes shot-gun proteomics challenging to compare samples, extract meaningful information and derive accurate conclusions. The ultimate tools to overcome this challenge of variability in shotgun proteomics workflows are the so-called normalization methods [4–6].

Normalization is a method used to adjust the data obtained from different samples, platforms, or batches to remove unwanted variability and enhance comparability between the datasets. It involves applying a mathematical correction to the data to reduce the effects of systematic technical variation and other sources of bias. The choice of normalization method depends on the nature of the data and the research question at hand [7–9]. However, this choice is not as simple as it seems at first glance. The importance of selecting an appropriate normalization method is critical to obtain well-grounded results. Therefore, selecting the optimal way to normalize data has important implications, as it significantly impacts downstream analyses, delivering different results depending on the normalization method selected [1,10,11]. Although several studies have evaluated different proteome normalization approaches, primarily focusing on their ability to enhance data repeatability, there is no consensus about which normalization method must be used, or whether the use is conditioned by the number of groups or the number of samples in each group [10,12–15].

There are several normalization methods, but the most popular ones include Z-score normalization, median divide normalization, and

quantile normalization.

The z-score normalization method creates an induced normal matrix by subtracting to each protein in each sample their respective median across all samples in the dataset and then dividing each protein value by the standard deviation of the respective protein across all samples. This process forces the dataset to follow a normal distribution, which can improve the accuracy of downstream statistical analyses. By normalizing the data in this way, the z-score normalization method can address variations in the data, making it a valuable tool for data normalization in proteomic studies [10].

The median divide normalization approach involves dividing all results by the median, which reduces the impact of bias caused by equipment variations and makes the data more comparable to other assays. This method is particularly useful when dealing with large datasets that are collected over time and across different batches, as it reduces the variation in the data introduced by the technical aspects of the assay. The median divide normalization method is a reliable approach for comparing protein expression levels between different samples and provides an accurate representation of the overall distribution of the data. However, this method may not be appropriate for datasets with extreme values or when most values are close to zero [7].

The quantile normalization involves substituting the value of each data point in a sample with the mean of the corresponding quantile, thereby modifying the data distribution of each sample to be identical. This normalization method is particularly useful for comparing samples with different distributions. By standardizing the distribution of each sample, quantile normalization can improve the accuracy and reproducibility of downstream analyses, such as differential expression

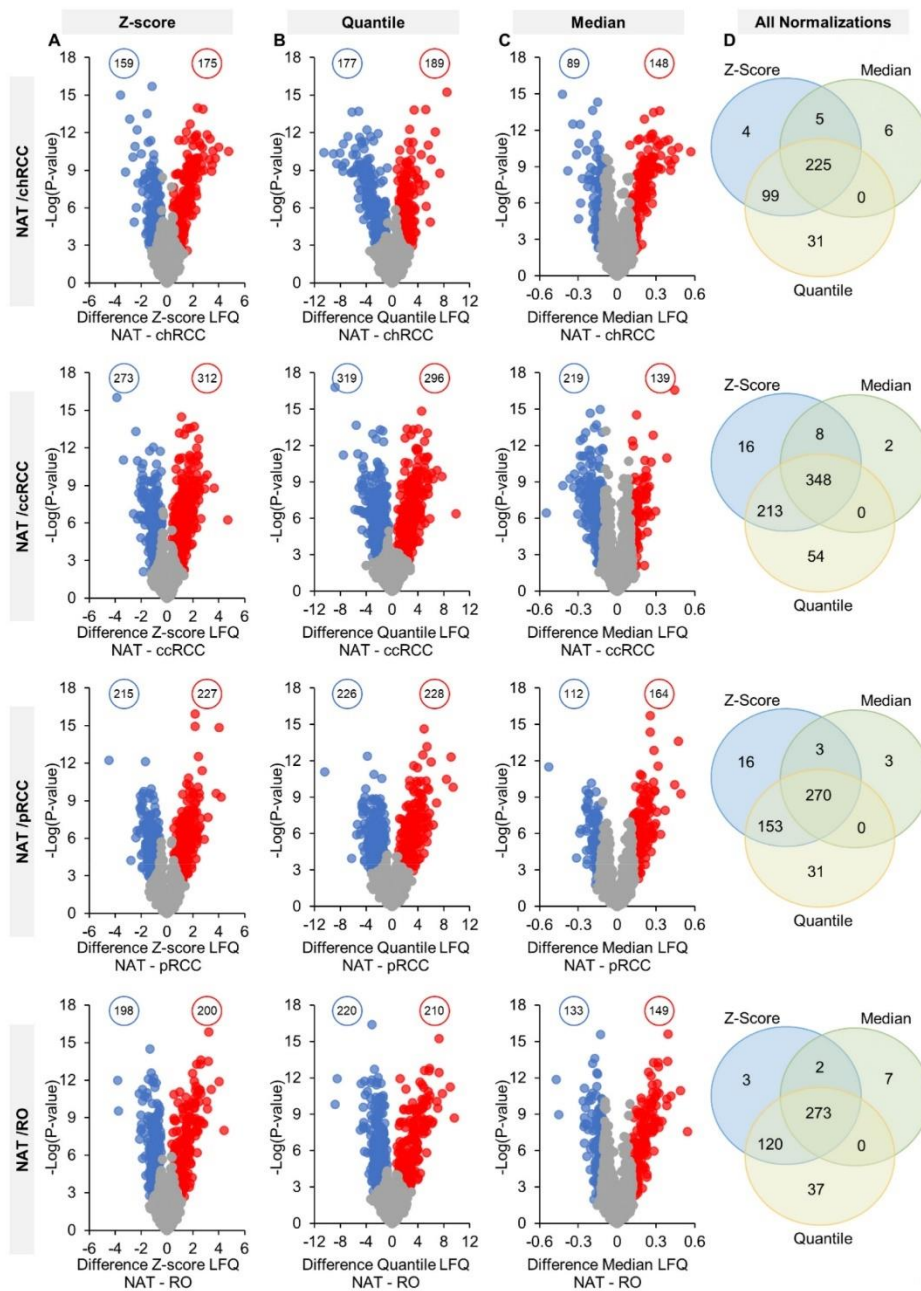


Fig. 2. Comparison of the numbers of proteins found statistically differentially expressed using each normalization method when comparing the near adjacent tissue (NAT) against each renal tumour type. Each column represents the volcano plots obtained using one normalization method: A: Z-score; B: Quantile; C: Median. Each row shows the comparison by pairs (Volcano plot, according to Student's *t*-test, FDR 0.05 and S_0 of 0.1, red dots: overexpressed proteins in NAT; blue dots over-expressed proteins in tumours). chRCC: chromophobe renal cell carcinoma; ccRCC: clear cell renal cell carcinoma; pRCC: papillary renal cell carcinoma; RO: Renal oncocytoma. D: column summarizes the common and uncommon proteins found among pair comparisons for each normalization method.

analysis and pathway analysis [6,16,17].

In a previous work [18], we applied Z-Score normalization and the Total Protein Approach (TPA) to proteomic data derived from renal carcinoma solid biopsies. Thus, near adjacent tissue (NAT), Papillary Renal Cell Carcinoma (pRCC), Clear Cell Renal Carcinoma (ccRCC), Chromophobe Renal Cell Carcinoma (chRCC) and Renal Oncocytoma (RO), samples were interrogated with the main aim of obtaining new immunohistochemical markers for diagnosis and prognosis. The best biomarkers found via mass spectrometry, namely perilipin-2 (PLIN2), beta-tubulin III (TUBB3), lysosomal-associated membrane protein-1 (LAMP1) and hexokinase-1 (HK1) were later successfully validated using tissue microarrays.

In the present work, the complete mass spectrometry datasets obtained for the different types of renal carcinomas and NAT are interrogated using bioinformatic pipelines differing only in the selected normalization methods, namely Z-score, median divide and quantile normalization. Selecting these specific tumour types allows for a comprehensive analytical assessment of the normalization methods in a clinically relevant context. Through this evaluation, we aim to provide insights into the selection process of appropriate normalization techniques that can enhance the accuracy and reproducibility of proteomic analyses in tumour research. The results obtained were compared with those previously validated using tissue microarrays.

2. Methods

2.1. Patients and sample collection

The experimental workflow developed to interrogate the renal carcinoma and the NAT tissue samples, along with the extended study design and patient sampling collection is described elsewhere [19], furthermore the validation of the biomarkers found using tissue microarrays is also described elsewhere [18]. A brief patient summary is provided in Supplementary material Table SM1.

2.2. Mass spectrometry raw data

This study used frozen tissue biopsies of human renal tissues from four different tumour types (ccRCC, chRCC, pRCC, and RO) and a control group (NAT). The samples were properly treated and analysed by mass spectrometry to identify and quantify the protein expression levels [19]. Mass spectrometry raw data was accessed in ProteomeXchange Consortium [20] using the Proteomics Identifications Database (PRIDE) [21] with the unique identifier PXD023296.

2.3. Identification and quantification of proteins

Protein identification and relative label-free quantification (LFQ) was newly performed using the previously described raw data using the precursor signal intensity method and delayed normalization, MaxLFQ, on MaxQuant software V2.0.3.0 [12]. The newly generated proteins group table containing protein identification and quantification can be found in Supplementary material 2. All raw files were processed in a single run with default parameters [22,23]. Database searches were performed using peptide search engine Andromeda against the human UniProt UP000005640.9606 and UniProt UP000005640.9606 additional database [24]. Searches were configured with cysteine carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidation as variable modifications. We set the false discovery rate (FDR) to 0.01 for protein and peptide levels with a minimum length of seven amino acids. The FDR was determined by searching a reverse database. Trypsin enzyme specificity was set as C-terminal to arginine and lysine with a maximum allowance of two missed cleavages.

2.4. Bioinformatics data processing

Bioinformatic data processing was conducted using Perseus V1.6.0 [25,26] as explained in Fig. 1B. Briefly, the data processing can be roughly divided into stages: (i) Elimination of protein groups that were reversely identified or only identified by site from the list of proteins; (ii) Annotation of sample condition and technical replicates; (iii) Transformation of LFQ intensities into Log (2x) value to lessen the impact of outliers and to understand the protein changes proportional across the conditions; and finally (iv) Proteins were filtered so that they were at least present in 70% of the samples in at least one of the created groups and the remaining missing values were imputed according with the respective sample normal distribution with a width of 0.3 and down shift of 1.8.

2.5. Bioinformatic data analysis and normalization comparisons

Bioinformatic data analysis was carried out in Perseus V1.6.0 [25,26] as explained in Fig. 1C in which the analytical comparison of three different types of commonly used normalizations was conducted (Z-Score normalization, Median divide normalization and Quantile normalization). Z-Score and Median divide normalization were obtained using the standard tools in Perseus while Quantile normalization was acquired by using the R package limma add-on for Perseus [27,28]. To access the impact of normalization, the NAT (control group) data set of protein expression was compared with the renal tumours using the three different types of normalization. The number of proteins whose levels were found to be statistically different between the group comparison were accessed by using a volcano plot analysis (Student's *t*-test, FDR 0.05 and S_0 of 0.1). Finally, all the datasets were compared together for each type of normalization using a Multisample Test ANOVA with a FDR of 0.05. All comparison resulting in statistically different proteins were used respectively in a hierarchical clustering analysis with an average linkage, no constraint, pre-processing with k-means, and Euclidean distance. The proteomic data was also processed (i) by comparing the NAT group of samples against two, three or four tumour types at the same time, as shown in Fig. 1 and (ii) by comparing each tumour condition and the control group at a time, as shown in Supplementary material Fig. SM1. This approach was adopted to ensure that the results would be consistent across different subtypes of renal tumour and would be useful for researchers focusing on precision medicine.

These analytical approaches allowed for a comprehensive evaluation of the normalization methods in terms of their ability to accurately identify significant proteins and their impact on downstream analyses, such as differential expression analysis and pathway analysis. Enrichment pathways analysis was carried out using the software platform Cytoscape v3.9.1 [29] and the application StringApp [30] v2.0.0 for protein-protein analysis and biochemical pathway analysis. Different types of databases for enrichment pathway analysis (GO biological process [31], The Kyoto Encyclopedia of Genes and Genomes - KEGG [32], Reactome pathways [33]) were considered to access the impact of normalization on downstream analysis.

3. Results

The data pre-processing steps were standardized across all normalization methods. A detailed workflow of the pre-processing steps is illustrated in Fig. 1 and Supplementary material Fig. SM1.

3.1. Comparing each type of renal carcinoma against NAT

After comparing the datasets in pairs, we tested three normalization methods. Our findings revealed that both z-score normalization and quantile normalization produced similar results in terms of the number of statistically significant differentially expressed proteins throughout all comparisons, as shown in Fig. 2A–D. However, quantile

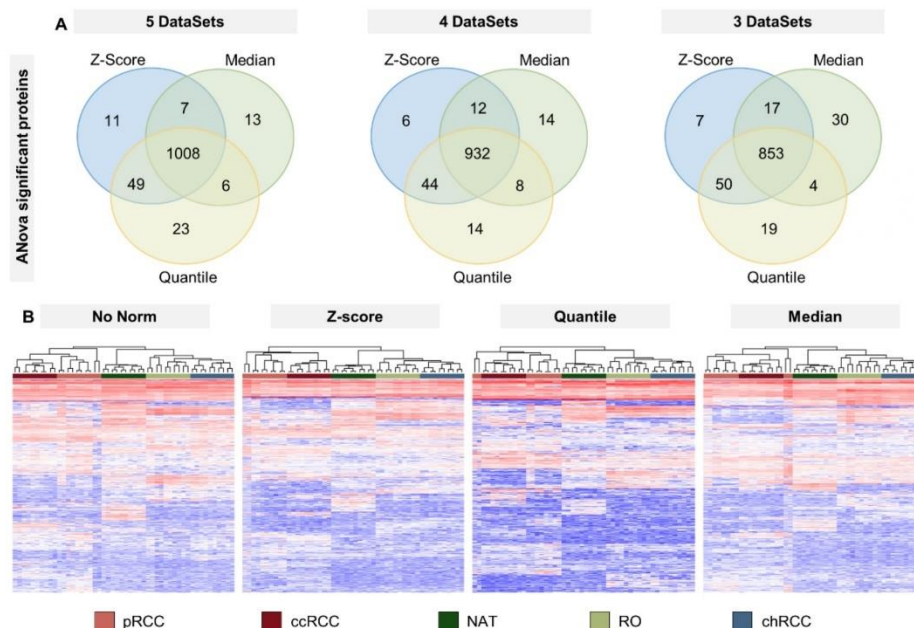


Fig. 3. Proteins found statistically differentially expressed using the comparison of: **A:** 5 datasets: All tumours against NAT; 4 datasets: chRCC, ccRCC and RO tumours against NAT and 3 datasets: ccRCC and RO against NAT. **B:** Hierarchical clusters were performed by using proteins that were found to be statistically different using each normalization method when comparing all tumours against NAT (average linkage, no constraint, pre-processing with k-means, and Euclidean distance, $n = 25$).

normalization consistently identified a slightly higher number of proteins, depending on the carcinoma compared against the near adjacent tissue, NAT. On the other hand, the median divide method consistently produced about 37% fewer differentially expressed proteins for all the pairs compared, as depicted in Fig. 2. This trend was also observed for the down- and upregulated proteins as shown in Fig. SM2.

In a previous study utilizing the same raw data, our team employed mass spectrometry to identify four potential biomarkers: PLIN2, TUBB3, LAMP1, and HK1. The Z-score and Total Protein Approach (TPA) were utilized for this purpose, and these markers are depicted in Fig. SM2. TPA is a label-free mass spectrometry method developed by Wiśniewski et al. [34,35] that enables the measurement of absolute expression levels of numerous proteins without the need for standards.

Using TPA, we were able to highlight these potential biomarkers, and their validity was subsequently confirmed through immunohistochemical tissue arrays [18]. Interestingly, in our current study, we observed statistically significant differential expression for the same proteins across all three normalization methods employed, with similar p-values. This suggests that when a protein exhibits high significance, it will be consistently identified as significant regardless of the specific normalization method used.

These findings emphasize the robustness and reliability of the identified biomarkers and support the notion that they are biologically relevant. It further reinforces the notion that the choice of normalization method does not significantly impact the detection of highly significant proteins.

Furthermore, we compared the top 100 proteins with the highest log P -values across the three normalization methods in all two-group comparisons. We found that the z-score and quantile normalization methods had 94 proteins in common, while the median method shared 86 and 84 proteins with the z-score and the quantile methods, respectively. These

results suggest that the z-score and quantile normalization methods may be more consistent with each other in identifying highly differentially expressed proteins, while the median method may identify a slightly different set of proteins.

3.2. Comparison of the two, three and four renal carcinomas at a time against NAT

Our next step was to compare the tumours datasets in groups of 2, 3 and 4 against NAT. Proteins for further processing were selected if they were found in any dataset in at least 70% of the replicates.

Focusing on the results obtained comparing the four neoplasia's datasets against NAT (Fig. 3A) it can be concluded that when the comparison is not made in pairs, the numbers of proteins found to be statistically different is almost the same (i.e., 1075, 1086 and 1034 for the Z-score, the quantile and the median divide, respectively). In addition, 1008 of all the proteins were identified with any normalization method. The same patterns were repeated comparing three or two tumours against NAT, Fig. 3A respectively.

These results suggest that the choice of normalization method did not have a substantial impact on the number of differentially expressed proteins identified when datasets are not compared in pairs, as explained in the previous section. This is due to the fact that the Max-Quant statistics varied differently, from comparing pairs to comparing three or more datasets, so the differences were not on the same scale as those observed when comparing datasets in pairs, as shown in the previous section.

To cluster the samples, we exploited the differentially expressed proteins identified by each normalization method using the five datasets (Supplementary Material 3). As noted in Fig. 3B, we found that three distinct clusters were formed regardless of the normalization method.

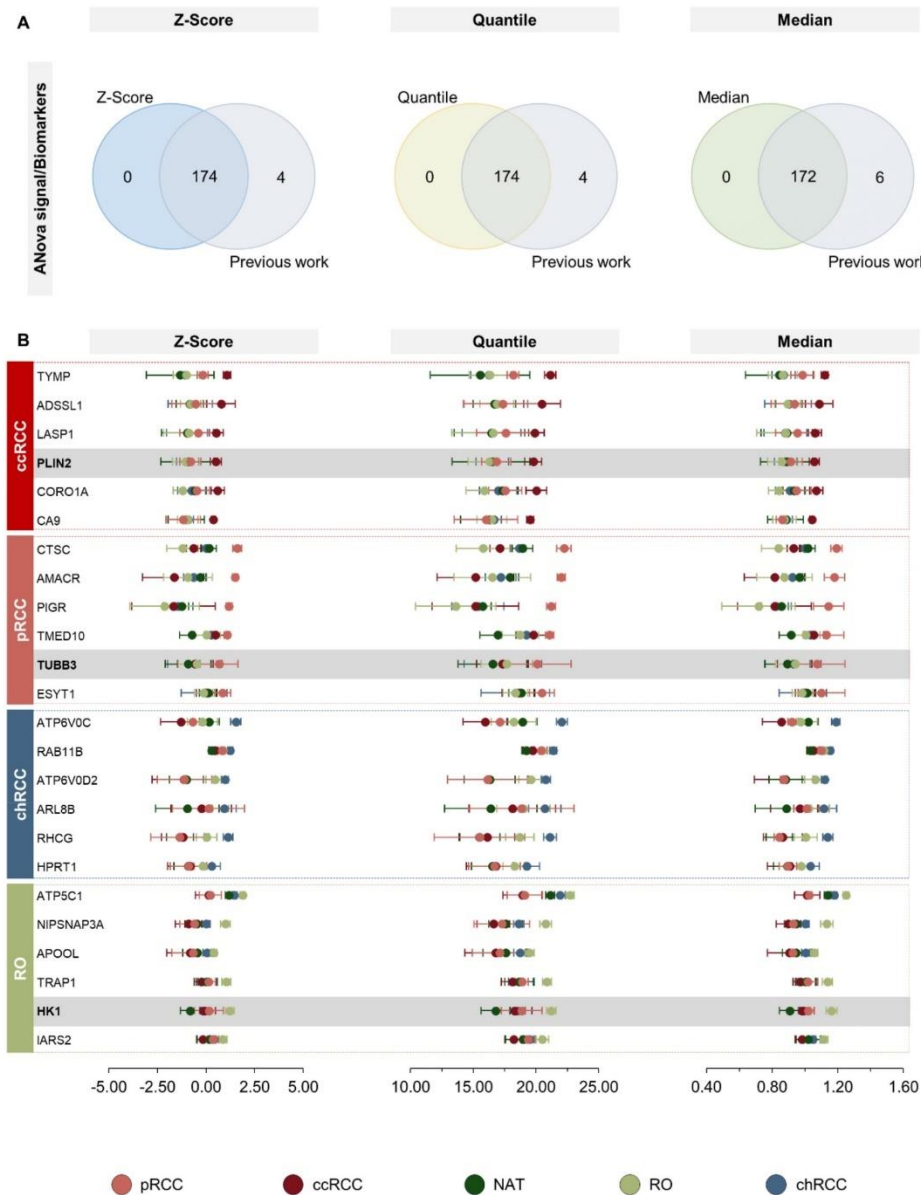


Fig. 4. Comparison of candidate biomarkers for each renal tumour subtype across all different type of data normalization. **A:** Biomarkers found in this work for each normalization method versus those found in our previous work [18]. **B:** Biomarkers proposed in our previous work (Fig. SM2) found with each normalization method in the present work. Note that for all of them the standard deviations of the normalized LFQ values overlap.

The tumours pCRR and ccRCC were grouped together in the first cluster, while RO and chRCC formed a distinct cluster in the second. The third one was NAT tissue. Notably, we observed that the proteomic expressions of RO and chRCC tumours were more similar to NAT than to pRCC and ccRCC.

When the differentially expressed proteins were analysed to extract those whose levels make them unique for just one type of carcinoma, the number of proteins was 178, regardless of the normalization process, as depicted in Fig. 4A. Furthermore, we also found that 174 out of these 178 biomarkers were present in our previous work [18], including the



B. Carvalho | NOVA University

L.B. Carvalho et al.

Talanta 266 (2024) 124953

four biomarkers validated by us using immunohistochemistry. These results suggest that the choice of normalization method has a negligible impact on the identification of specific biomarkers.

3.3. Comparison of the best biomarkers for immunohistochemistry

In a previous work we recommended several proteins as new potential biomarkers for immunohistochemistry (Supplementary material Fig. SM3). These proteins were first identified using mass spectrometry (LFQs) and then with TPA. Of these proteins, those represented in grey in Fig. 4B were selected in the previous work because the standard deviations did not overlap. These proteins were later validated against tissue arrays [18].

When we used the levels obtained through each normalization method for each protein to identify statistically significant proteins, we obtained two interesting results. Firstly, the three normalization methods delivered almost the same results, but they differed substantially from those obtained using TPA as described above. As shown in Fig. 4B, proteins PLIN2, TUBB3, and HK1 would have never been identified as the best potential biomarkers because the standard deviations overlap each other (Fig. 4). This unexpected result highlights the need to re-evaluate how data is treated and the importance of TPA in biomarker discovery.

3.4. Exploring the impact of normalization methods on identifying main pathways affected in renal tumours

To determine whether the identified dysregulated pathways changed as a function of the selected normalization method, we used the set of proteins found to be dysregulated when the five different datasets were compared. Then, through this data set, using an ANOVA test, we compared the proteins differentially expressed between NAT and chRCC.

Fig. 5 shows the 10 most important (as per p-value) dysregulated pathways for (i) GO biological process, (ii) KEGG, and (iii) Reactome pathways. The results show that for the GO biological processes the first 7 pathways affected are the same, with no variations nor in the p-value, neither in the number of proteins involved, thus showing no differences in the normalization methods. For the KEGG, the Z-score and the quantile normalizations were found more consistent between them than when compared to the median, as they have the same top-10 pathways. For the Reactome pathways, the three normalization methods were found to provide similar results.

4. Conclusions

When comparing two groups of datasets, such as healthy versus diseased samples, the Z-score normalization and the quantile methods tend to yield a higher number of differentially expressed proteins compared to the median normalization method. Thus, in these cases, we recommend using the Z-score normalization method. However, when comparing a larger number of datasets groups (3, 4, or 5), the differences in terms of proteins and biochemical pathways statistically different among the three normalization methods are negligible, and any can be selected. Interestingly, our findings show that four biomarkers, previously identified using the z-score normalization (LFQs) in combination with the TPA approach, and later validated with immunohistochemistry [18], could not have been easily highlighted using any of the normalization methods assessed in this study. This discovery opens new avenues of research in biomarker discovery as it emphasises the TPA approach as a powerful tool in biomarker discovery.

Author contributions

J.L.C., H.M.S., and L.B.C. designed the bioinformatics pipelines. L.B.C. and P.A.D.T.C. performed the bioinformatics analysis. J.L.C. drafted

the first manuscript. P.A.D.T.C., H. M. S., L.B.C., R.D., J.W., C.L., L.M. and M.P. read the final draft, corrected it and made valuable suggestions. R.D. provided the solid biopsies, and J.W. helped with the total protein approach. S.J. made the proteomics sample treatment and H.M. S. acquire the raw LC-MS/MS data.

Funding

PROTEOMASS Scientific Society, #PM001/2019, #PM003/2016. Fundacao para a Ciencia e a Tecnologia (FCT/MCTES), UIDB/50006/2020, UIDP/50006/2020, LA/P/0008/2020, SFRH/BD/144222/2019, SFRH/BD/120537/2016. University of Pittsburgh Hillman Cancer Center shared resource facilities (Cancer Genomics Facility and The Health Science Tissue Bank) supported in part by award P30CA047904.

Availability of data and materials

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [21] partner repository with the dataset identifier PXD023296.

Declaration of competing interest

All authors declare nor financial or personal relationships that may be perceived as influencing this work.

Data availability

Data is submitted n Pride. We provide the link in the text.

Acknowledgements

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry - LAQV which is financed by national funds from FCT/ MCTES, *Fundação para a Ciência e a Tecnologia* and *Ministério da Ciência, Tecnologia e Ensino Superior*, through the projects UIDB/50006/2020 and UIDP/50006/2020). H. M. S. acknowledges LAQV (LA/P/0008/2020) funded by FCT/ MCTES for his research contract. L. B. C. thanks the FCT/MEC for the FCT PhD grant 2019 (SFRH/BD/144222/2019). S. J. thanks FCT/MEC (Portugal) for her PhD grant reference SFRH/BD/120537/2016. This project utilized the University of Pittsburgh Hillman Cancer Center shared resource facilities (Cancer Genomics Facility and The Health Science Tissue Bank) supported in part by award P30CA047904 (Dr.R. Dhir).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.talanta.2023.124953>.

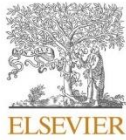
References

- [1] E. Dubois, A.N. Galindo, L. Dayon, O. Cominetti, Assessing normalization methods in mass spectrometry based proteome profiling of clinical samples, *Biosystems* 215–216 (2022), 104661, <https://doi.org/10.1016/j.biosystems.2022.104661>.
- [2] B.J.A. Mertens, Transformation, normalization, and batch effect in the analysis of mass spectrometry data for omics studies, in: *Statistical Analysis of Proteomics, Metabolomics, and Lipidomics Data Using Mass Spectrometry*, Springer International Publishing, Cham, 2017, pp. 1–21, https://doi.org/10.1007/978-3-319-45809-0_1.
- [3] A.M. De Livera, M. Sysi-Aho, L. Jacob, J.A. Gagnon-Bartsch, S. Castillo, J. A. Simpson, T.P. Speed, Statistical methods for handling unwanted variation in metabolomics data, *Anal. Chem.* 87 (2015) 3606–3615, <https://doi.org/10.1021/ac502439y>.

- [4] A. Chawade, E. Alexandersson, F. Levander, Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets, *J. Proteome Res.* 13 (2014) 3114–3120, <https://doi.org/10.1021/pr401264n>.
- [5] Y. V. Karpievitch, A.R. Dabney, R.D. Smith, Normalization and missing value imputation for label-free LC-MS analysis, *BMC Bioinf.* 13 (2012) S5, <https://doi.org/10.1186/1471-2105-13-S16-S5>.
- [6] B.M. Bolstad, R.A. Irizarry, M. Åstrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, *Bioinformatics* 19 (2003) 185–193, <https://doi.org/10.1093/bioinformatics/19.2.185>.
- [7] S.J. Callister, R.C. Barry, J.N. Adkins, E.T. Johnson, W. Qian, B.-J.M. Webb-Robertson, R.D. Smith, M.S. Lipton, Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics, *J. Proteome Res.* 5 (2006) 277–286, <https://doi.org/10.1021/pr050300l>.
- [8] K. Kultina, A. Nilsson, B. Scholz, U.L. Rossbach, M. Fälth, P.E. Andrén, Development and evaluation of normalization methods for label free relative quantification of endogenous peptides, *Mol. Cell. Proteomics* 8 (2009) 2285–2295, <https://doi.org/10.1074/mcp.M800514.MCP200>.
- [9] T. Välikangas, T. Suominen, L.L. Elo, A Systematic Evaluation of Normalization Methods in Quantitative Label-free Proteomics, *Brief Bioinform.* 2016, p. bbw095, <https://doi.org/10.1093/bib/bbw095>.
- [10] B.-J.M. Webb-Robertson, M.M. Matzke, J.M. Jacobs, J.G. Pounds, K.M. Waters, A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors, *Proteomics* 11 (2011) 4736–4741, <https://doi.org/10.1002/pmic.201100078>.
- [11] A. Chawade, M. Sandini, J. Telenius, J. Mahström, F. Levander, Data processing has major impact on the outcome of quantitative label-free LC-MS analysis, *J. Proteome Res.* 14 (2015) 676–687, <https://doi.org/10.1021/pr500665j>.
- [12] J. Cox, M.Y. Hein, C.A. Luber, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Mol. Cell. Proteomics* 13 (2014) 2513–2526, <https://doi.org/10.1074/mcp.O115.055475>.
- [13] B. Zhang, L. Käll, R.A. Zubarev, DeMix-Q: quantification-centered data processing workflow, *Mol. Cell. Proteomics* 15 (2016) 1467–1478, <https://doi.org/10.1074/mcp.O115.055475>.
- [14] A. Pursiainen, A.P. Vehmas, S. Afzal, T. Suominen, T. Chand, L. Strauss, M. Poutanen, A. Rokka, G.L. Cortiada, L.L. Elo, Optimization of statistical methods impact on quantitative proteomics data, *J. Proteome Res.* 14 (2015) 4118–4126, <https://doi.org/10.1021/acs.jproteome.5b00183>.
- [15] D.L. Tabb, L. Vega-Montoto, P.A. Rudnick, A.M. Variyath, A.-J.L. Ham, D.M. Bunk, L.E. Kilpatrick, D.D. Billheimer, R.K. Blackburn, H.L. Cardasis, S.A. Carr, K. R. Clauser, J.D. Jaffe, K.A. Kowalski, T.A. Neubert, F.E. Regnier, B. Schilling, T. J. Tegeler, M. Wang, P. Wang, J.R. Whiteaker, L.J. Zimmerman, S.J. Fisher, B. W. Gibson, C.R. Kinsinger, M. Mesri, H. Rodriguez, S.E. Stein, P. Tempst, A. G. Paulovich, D.C. Liebler, C. Spiegelman, Repeatability and reproducibility in proteomic identifications by liquid chromatography–tandem mass spectrometry, *J. Proteome Res.* 9 (2010) 761–776, <https://doi.org/10.1021/pr9006365>.
- [16] B. Bolstad, preprocessCore: A collection of pre-processing functions. R package version 1.62.1., 2023, in: <https://github.com/bnbolstad/preprocessCore>. (Accessed 21 July 2023).
- [17] Y. Zhao, L. Wong, W.W. Bin Goh, How to do quantile normalization correctly for gene expression data analyses, *Sci. Rep.* 10 (2020), 15534, <https://doi.org/10.1038/s41598-020-72664-6>.
- [18] S. Jorge, J.L. Capelo, W. LaFramboise, S. Satturwar, D. Korentzelos, S. Bastacky, G. Quiroga-Garza, R. Dhir, J.R. Wisniewski, C. Lodeiro, H.M. Santos, Absolute quantitative proteomics using the total protein approach to identify novel clinical immunohistochemical markers in renal neoplasms, *BMC Med.* 19 (2021) 196, <https://doi.org/10.1186/s12916-021-02071-9>.
- [19] S. Jorge, J.L. Capelo, W. LaFramboise, R. Dhir, C. Lodeiro, H.M. Santos, Development of a robust ultrasonic-based sample treatment to unravel the proteome of OCT embedded solid tumor biopsies, *J. Proteome Res.* 18 (2019) 2979–2986, <https://doi.org/10.1021/acs.jproteome.9b00248>.
- [20] E.W. Deutsch, A. Csordas, Z. Sun, A. Januszczak, Y. Perez-Riverol, T. Tertent, D. S. Campbell, M. Bernal-Linares, S. Okuda, S. Kawano, R.L. Moritz, J.J. Carver, M. Wang, Y. Ishihama, N. Bandeira, H. Hermjakob, J.A. Vizcaino, The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition, *Nucleic Acids Res.* 45 (2017), <https://doi.org/10.1093/nar/gkx936>, D1100–D1106.
- [21] Y. Perez-Riverol, J. Bai, C. Bandla, D. García-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D.J. Kimdu, A. Prakash, A. Frericks Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, J.A. Vizcaino, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, *Nucleic Acids Res.* 50 (2022), <https://doi.org/10.1093/nar/gkab1038>, D543–D552.
- [22] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat. Biotechnol.* 26 (2008) 1367–1372, <https://doi.org/10.1038/nbt.1511>.
- [23] S. Tyanova, T. Tenu, A. Carlson, P. Sinitcyn, M. Mann, J. Cox, Visualization of LC-MS/MS proteomics data in MaxQuant, *Proteomics* 15 (2015) 1453–1456, <https://doi.org/10.1002/pmic.201400449>.
- [24] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, *J. Proteome Res.* 10 (2011) 1794–1805, <https://doi.org/10.1021/pr101065j>.
- [25] S. Tyanova, T. Tenu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote) omics data, *Nat. Methods* 13 (2016) 731–740, <https://doi.org/10.1038/nmeth.3901>.
- [26] S. Tyanova, J. Cox, Perseus, A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research, 2018, pp. 133–148, https://doi.org/10.1007/978-1-4939-7493-1_7.
- [27] S. Yu, D. Ferretti, J.P. Schlessner, J.D. Rudolph, G.H.H. Börner, J. Cox, Expanding the perseus software for omics data analysis with custom plugins, *Curr. Protoc. Bioinformatics* 71 (2020), <https://doi.org/10.1002/cpbi.105>.
- [28] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-seq and microarray studies, *Nucleic Acids Res.* 43 (2015) e47, <https://doi.org/10.1093/nar/gkv007>, e47.
- [29] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504, <https://doi.org/10.1101/gr.1239303>.
- [30] N.T. Doncheva, J.H. Morris, J. Gorodkin, L.J. Jensen, Cytoscape StringApp: network analysis and visualization of proteomics data, *J. Proteome Res.* 18 (2019) 623–632, <https://doi.org/10.1021/acs.jproteome.8b00702>.
- [31] The gene ontology resource: 20 years and still going strong, *Nucleic Acids Res.* 47 (2019) D330, <https://doi.org/10.1093/nar/gky1055>, –D338.
- [32] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic Acids Res.* 45 (2017) D353–D361, <https://doi.org/10.1093/nar/gkx1092>.
- [33] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C.D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusi, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase, *Nucleic Acids Res.* 46 (2018) D649–D655, <https://doi.org/10.1093/nar/gkx1132>.
- [34] J.R. Wisniewski, M.Y. Hein, J. Cox, M. Mann, A “proteomic ruler” for protein copy number and concentration estimation without spike in standards, *Mol. Cell. Proteomics* 13 (2014) 3497–3506, <https://doi.org/10.1074/mcp.M113.037309>.
- [35] J.R. Wisniewski, Label free and standard free absolute quantitative proteomics using the “total protein” and “proteomic ruler” approaches, *Methods Enzymol.* 585 (2017) 49–60, <https://doi.org/10.1016/bs.mie.2016.10.002>.

| D

PUBLISHED PAPER IN JOURNAL OF PROTE- OMICS - CORRESPONDING TO CHAPTER 6



Contents lists available at ScienceDirect

Journal of Proteomics

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

Biochemical network analysis of protein-protein interactions to follow-up T1 bladder cancer patients

Luís B. Carvalho^{a,b}, José Luis Capelo Martínez^{a,b}, Carlos Lodeiro^{a,b}, Rafael Bento^{a,b}, Rajiv Dhir^c, Jeremiah J. Morrissey^d, Luis Campos Pinheiro^{e,f}, Mariana Medeiros^{e,f}, Hugo M. Santos^{a,b,c,*}

^a BIOSCOPE Research Group, LAQV REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, NOVA University of Lisbon, Caparica 2829-516, Portugal

^b PROTEOMASS Scientific Society, Madan Parque, Rua dos Inventores, Caparica 2825-182, Portugal

^c Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

^d Department of Anesthesiology, Division of Clinical and Translational Research, School of Medicine and Siteman Cancer Center, Washington University, St. Louis, MO, USA

^e Urology Department, Central Lisbon Hospital Center, Lisbon, Portugal

^f NOVA Medical School, NOVA University of Lisbon, Lisbon, Portugal

ARTICLE INFO

Keywords:

Bladder cancer
Urine proteomics
Mass spectrometry
Neutrophil extracellular traps,
myeloperoxidase (MPO)
Cubilin (CUBN)

ABSTRACT

Bladder cancer (BCa) is a prevalent disease with a high risk of aggressive recurrence in T1-stage patients. Despite the efforts to anticipate recurrence, a reliable method has yet to be developed. In this work, we employed high-resolution mass spectrometry to compare the urinary proteome of T1-stage BCa patients with recurring versus non-recurring disease to uncover actionable clinical information predicting recurrence. All patients were diagnosed with T1-stage bladder cancer between the ages of 51 and 91, and urine samples were collected before medical intervention. Our results suggest that the urinary myeloperoxidase to cubilin ratio could be used as a new tool for predicting recurrence and that dysregulation of the inflammatory and immune systems may be a key driver of disease worsening. Furthermore, we identified neutrophil degranulation and neutrophil extracellular traps (NETs) as key pathways in the progression of T1-stage BCa. We propose that proteomics follow-up of the inflammatory and immune systems may be useful for monitoring the effectiveness of therapy.

Significance: This article describes how proteomics can be used to characterize tumor aggressiveness in patients with the same diagnosis of bladder cancer (BCa). LC-MS/MS in combination with label free quantification (LFQ) were used to explore potential protein and pathway level changes related to the aggressiveness of the disease in 13 and 17 recurring and non-recurring T1 stage BCa patients. We have shown that the MPO/CUBN protein ratio is a candidate for a urine prognosis tool in BCa. Furthermore, we identify dysregulation of inflammation process as a driver for BCa recurrence and progression. Moreover, we propose using proteomics to track the effectiveness of therapy in the inflammatory and immune systems.

1. Introduction

Precision medicine, which tailors treatment to the individual patient, is becoming increasingly important. One way to achieve this is through the use of personalized proteogenomics, which looks at the unique proteins and genetic information of each patient rather than grouping and pooling data from multiple patients [1]. This approach allows for more accurate diagnosis and treatment, as it considers each patient's

unique characteristics. Precision medicine can retrieve unprecedented levels of biological information by combining robust fast-sample treatment protocols, high-resolution mass spectrometry, and bioinformatics [2]. Furthermore, the use of mass spectrometry-based multiple reaction monitoring has made possible the large-scale validation of candidates as it has never been seen before, thus ensuring the correctness and applicability of the clinical information retrieved from each patient [3].

BCa is the ninth most common neoplasm, second in incidence and

* Corresponding author at: BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, NOVA University Lisbon, Caparica 2829-516, Portugal.

E-mail address: hmsantos@fct.unl.pt (H.M. Santos).

<https://doi.org/10.1016/j.jprot.2023.104865>

Received 30 November 2022; Received in revised form 3 February 2023; Accepted 27 February 2023

Available online 2 March 2023

1874-3919/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

mortality of the genitourinary system [4]. Literature shows that BCa has a high incidence, progression, and recurrence rates. Therefore, it is extremely important to monitor BCa patients, especially those with high-risk tumors, which require frequent and invasive examination procedures [5]. Urinary cytology and cystoscopy are the current gold standard tools to diagnose BCa. However, cystoscopy is an invasive examination with a high burden for the patient due to pain, bleeding, urinary tract infections and other complications. On the other hand, Cytology is a non-invasive test that can identify tumor cells shed in the urine. It is simple to use and inexpensive, but it has low sensitivity and low diagnostic efficiency [6]. Much effort has been made to diagnose and prognose BCa patients using DNA, RNA, or proteins, focusing on solid or liquid biopsies (blood-derived samples) [7]. Despite the valuable information the urinary proteome carries, this liquid biopsy remains to be further explored to disclose new features about BCa.

The clinical outcome of individuals with T1-stage BCa with a high incidence of recurrence and progression is highly variable. The aetiology behind T1-stage BCa aggressiveness remains unclear [8]. Early and accurate diagnosis of T1-stage BCa that will recur or progress is now a significant challenge in treating patients with BCa.

In this work, the urinary proteome of T1-stage BCa with and without recurrence was interrogated via high-resolution mass spectrometry and bioinformatics to unveil prognosis signatures. Herein, we present the myeloperoxidase to cubilin ratio as a potential indicator of recurrence, and we propose to flag the course of BCa patients following the changes in the information related to the immune and inflammatory systems encoded in the urinary proteome.

2. Methods

2.1. Patients

Mid-stream second void urine specimens were collected from 30 patients, all male, diagnosed with T1-stage BCa, ranging between 51 and 91 years. Urine samples were collected before the medical intervention, and T1-stage BCa diagnosis was assessed by pathological examination. Exclusion criteria: (1) unclear BCa diagnosis, (2) urinary cancer history, (3) HIV affected, and (4) organ transplant receivers or in recent chemo/radiotherapy (last six months). Further information is given in supplementary material 1.

2.2. Urine collection and preservation

Mid-stream second void morning urine samples (30 mL) were

collected in centrifuge tubes containing 38 mg of boric acid (Sigma-Aldrich) to prevent bacterial growth. Urine samples were placed on ice and were processed within the first two hours after collection. Samples were centrifuged for 20 min at 5000g to remove cell debris and divided into 10 mL aliquots, and these aliquots were stored at -80°C until further processing. The urinary proteome was concentrated by ultrafiltration from 10 mL to a final volume of c.a 300 μL using VivaSpin 15R (cut-off 10 kDa, Sartorius) at 5000 g for 20 min. The concentrated urinary proteomes were quantified for total protein using the Bradford protein assay method with Bradford reagent (Sigma-Aldrich) and Bovine Serum Albumin (BSA, Sigma-Aldrich) to perform the standard curve. Each concentrated urine sample and BSA standard were measured in duplicate. (See Fig. 1 flow chart for further information).

2.3. Filter-aided sample preparation

Filter Aided Sample Preparation (FASP) was used for urinary proteome digestion using a modified protocol [9]. Two technical replicates of each urine sample were prepared. An extended FASP protocol description can be found in supplementary data 1.

2.4. LC-MS/MS

LC-MS/MS analysis was performed using an Ultimate 3000 nano LC system in coupled to an Impact HD (Bruker Daltonics) with a Captive-Spray nanoBooster using acetonitrile as dopant. Peptides were resuspended in 100 μL of 3% (v/v) acetonitrile containing 0.1% (v/v) aqueous formic acid (FA). The samples were homogenized by 5 min on vortex followed by 10 min on an ultrasonic bath at 100% ultrasonic amplitude, 35 kHz. Afterward, samples were quantified using a Pierce™ quantitative colorimetric peptide assay (Thermo Fisher Scientific). Following the previous step, 3 μL containing 563 ng of peptides were loaded onto a trap column (Acclaim PepMap100, 5 μm , 100 $\text{\AA}$, 300 μm i.d. $\times$ 5 mm) and desalted for 5 min from 3% to 5% B (B: 90% acetonitrile 0.08% FA) at a flow rate of 15 $\mu\text{L}\cdot\text{min}^{-1}$. Then the peptides were separated using an analytical column (Acclaim™ PepMap™ 100C18, 2 μm , 0.075 mm i.d $\times$ 150 mm) with a linear gradient at 300 nL·min⁻¹ (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) acetonitrile and 0.08% (v/v) FA) 5–90 min from 5% to 35% of mobile phase B, 90–100 min linear gradient from 35% to 95% of mobile phase B, 100–110 95% B. Chromatographic separation was carried out at 35 $^{\circ}\text{C}$. The MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8–32 Hz), cycle time 3.0 s, with active exclusion (precursors were excluded from precursor selection for 0.5 min after the acquisition of 1 MS/MS

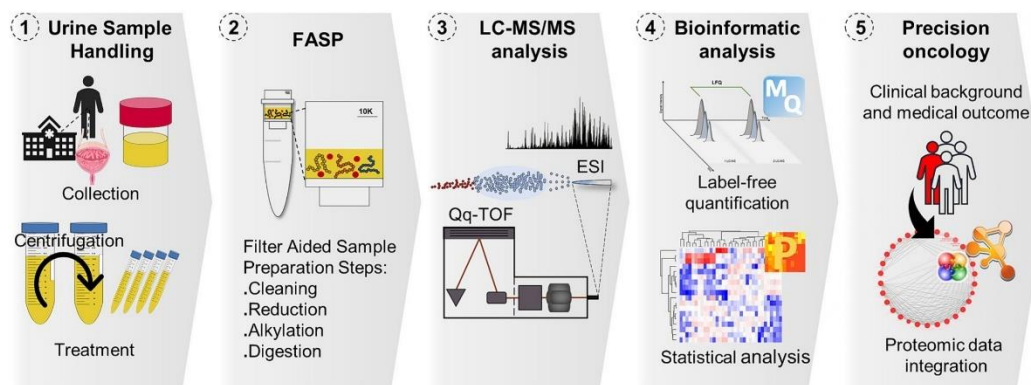


Fig. 1. Proteomics and bioinformatics workflow. 1. Urine sample collection pre-treatment and preservation. 2. Filter-aided sample preparation (FASP). 3. LC-MS/MS analysis. 4. Bioinformatic and statistical analysis. 5. Data interpretation and correlation with a clinical background.

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 was 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*.

2.5. Bioinformatics

Relative label-free quantification (LFQ) was carried out using MaxQuant software V1.6.0.16. All raw files were processed in a single run with default parameters [10]. Database searches were performed using the Andromeda search engine against the UniProt UP000005640.9606 database (20,600 sequences; 11,395,157 residues, downloaded on April 27, 2021). Data processing was performed using Perseus software V1.6.15.0 using default settings using the workflow depicted in Fig. 1. Protein group LFQ intensities were log2-transformed, and the quantitative profiles were filtered for missing values with the following settings: min valid percentage of 50% in at least one group and values >0. To overcome the obstacle of missing LFQ values, missing values were imputed from the total matrix with width = 0.5 and down-shift = 1.8 [11]. Log ratios were calculated as the difference in average log2 LFQ intensity values between the tested conditions (two-tailed, Student's *t*-test, permutation-based FDR 0.05, and SO of 0.1). Perseus was also used to obtain clusters, using average linkage, no constraint, pre-process with k-means, and Euclidean distance between column trees. Protein complex networks were integrated and visualized using Cytoscape V3.8.2 with the applications ClueGo V2.5.8, CluepediaV1.5.8, and StringApp V1.6.0. Reactome and biological process go terms were used as ontologies databases [12].

2.6. MPO in vitro enzyme-linked immunosorbent assay

Quantitative measurement of urinary MPO was performed using an in vitro ELISA kit from Abcam (ab119605) following manufacturer instructions. The plate was read for absorbance at 450 nm using a CLARIOstar® High-Performance Monochromator Multimode from BMG LABTECH within 30 min after adding the stop solution.

3. Results

The clinical information of the patient cohort used in the study is presented in Table SM1. The comparison of the urine proteome of T1-stage BCa patients that had experienced recurrence (RG) versus patients that had not (NRG) is explained in a comprehensive flow-chart in Fig. SM1. This comparison is presented in the format of a volcano plot in Fig. SM2. A set of 197 proteins differentially expressed between both groups was found and is listed in supplementary material 2. The bioinformatics analysis of the set of 197 proteins, focusing on both single proteins and biochemical networks, showed that (i) myeloperoxidase (MPO) was up-regulated (12.6 fold change and *p*-value = 0.0028) in the RG group and (ii) cubilin (CUBN) was up-regulated (3.6 fold change and *p*-value = 0.0289) in the NRG group. The variation in the levels of both proteins for each patient is presented in Fig. 2A, whilst the CUBN/MPO ratio is presented in Fig. 2B. The orthogonal validation of the MPO levels in urine was done using ELISA as shown in Fig. 2G. The set of 197 proteins dysregulated between both groups of patients was used to obtain the biochemical networks dysregulated between RG and NRG. Such networks are shown in Fig. 2H and I, displaying the protein-protein interaction and network analysis focusing on the inflammation process. Finally, the most interesting protein-protein interactions and biological functions addressed by 197 statistically significant proteins are integrated comprehensively in Fig. 3.

4. Discussion

4.1. CUBN to MPO ratio

As the bladder is in direct contact with the urine, we hypothesized that the urine proteome of BCa patients should contain critical and diagnostic/prognostic information about the disease. So far, we have studied the differences in the urinary proteome of BCa patients with and without a background of recurrence and/or progression, as this stratification is significant for medical treatment, and there is no current way to make it. The groups are called RG and NRG, respectively. The comparison is presented as a volcano plot in Fig. SM2. In this comparison, the myeloperoxidase (MPO) was found on the proteins most up-regulated in the RG group. On the other hand, the same data showed cubilin (CUBN), and lipopolysaccharide-binding protein 2 (LRP2) proteins down-regulated in the same group.

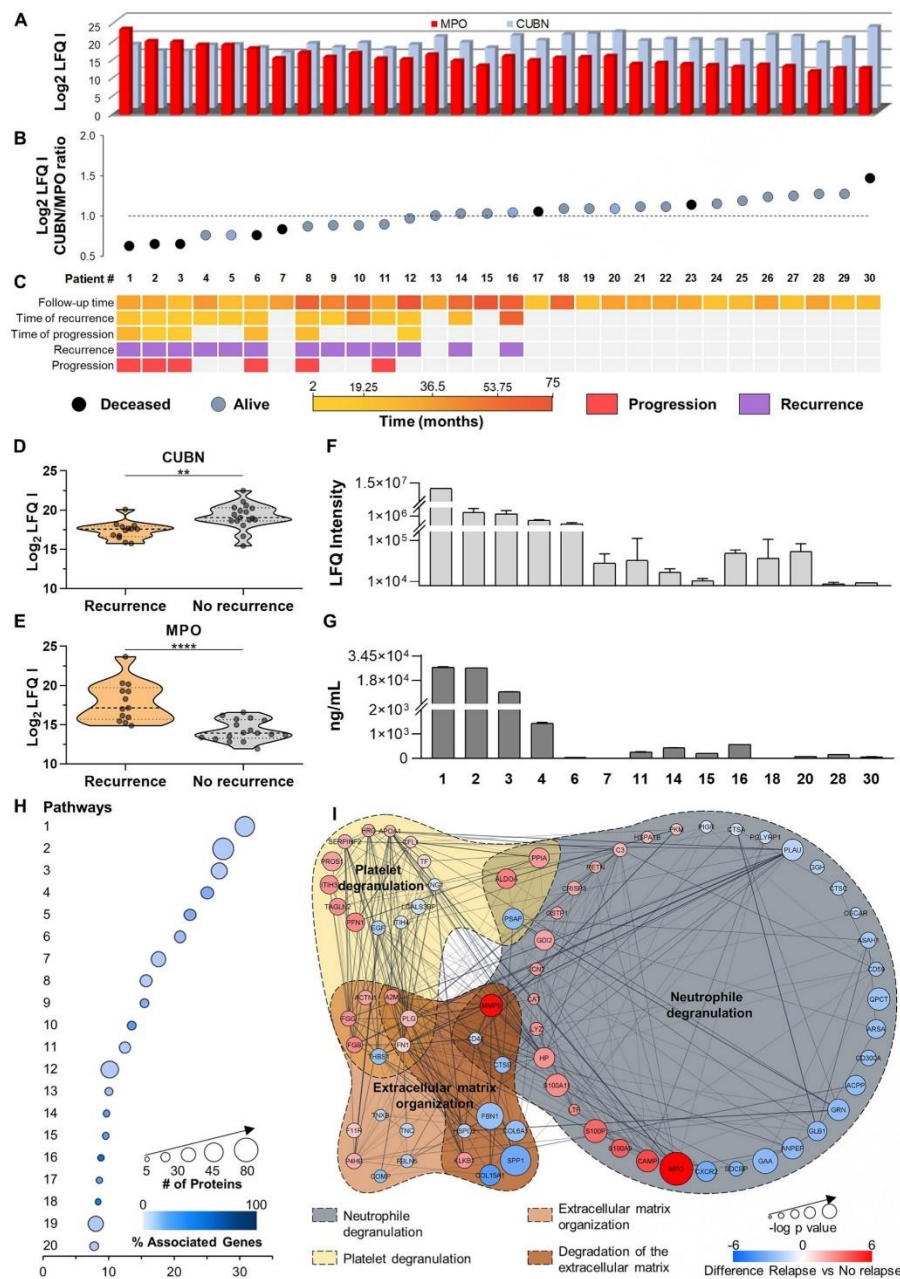
The MPO is a hemoprotein expressed by polymorphonuclear neutrophils, which is secreted during leucocyte degranulation in a cellular process that releases antimicrobial cytotoxic or mediator molecules as a defense response to invading microorganisms. More precisely, MPO catalyzes the conversion of H₂O₂ to hypochlorous acid, an efficient intracellular bacterial killer. In addition, it inactivates protease inhibitors, thus allowing lytic enzymes released from neutrophils to degrade tissue matrix and foreign materials facilitating tumor growth [13].

Furthermore, MPO has been proposed recently as a marker of poor prognosis in the lung [14] and ovarian carcinomas [15]. However, to the best of our knowledge, no studies report MPO levels as a prognosis biomarker for BCa carcinoma. Our data in Fig. 2A shows that those patients experiencing recurrence and progression presented the highest levels of MPO. In addition to MPO, the RG presented down-regulation of CUBN and LRP2. Remarkably, the down-regulation of CUBN has already been noted as a poor prognosis biomarker for renal cell carcinoma [16]. Thus, the lower the CUBN level, the worst the prognosis for renal carcinoma. Our data address the same trend also for BCa. It is worth noting that the group with recurrence and progression has the lowest levels of CUBN (see Fig. 2A, B, and C). Therefore, an index based on the MPO to CUBN ratio can be established such that patients with ratios equal to or lower than one present recurrence and/or progression.

4.2. Biochemical networks

The study of the differences between the biochemical networks of the RG and NRG groups using the 197 proteins differentially expressed between them were analyzed, and the first 20 pathways with the lowest *p*-value were selected (Fig. 2 H). Most of them are linked to immune and inflammatory responses. Therefore, we decided to assess the pathways that markedly differentiate RG and NRG groups, namely neutrophil degranulation and degradation of extracellular matrix pathways (Fig. 2I).

It is well known that the tumorigenesis process comprises the activation of several biochemical pathways, among which the ones related to invasion, proliferation, and inflammation are critical in tumor progression. Accordingly, our results address such pathways overexpressed in the recurrence group and are presented comprehensively in Fig. 3. Thus, our results suggest that the first line of defense, the neutrophils, are recruited and deployed near the tumor to destroy the tumoral cells. The neutrophils mature into the tumor-associated neutrophils, TAN-N1 and TAN-N2 [17]. The TAN-N1 Type has pro-inflammatory and anti-tumor effects, while the TAN-N2 type has anti-inflammatory and pro-tumor effects. The TAN-N2 type tries to destroy the tumoral cells by releasing their cellular content rich in proteases, histones, and DNA, forming the so-called neutrophil extracellular traps, NETs, that are thought to coat tumor cells, shielding them from immune attack [18]. Pointing this process to occur, we found that matrix metalloproteinase 9 (MMP9) was up-regulated in the RG group. This protein and MMP8 are



(caption on next page)

Fig. 2. T1 Bca patients MPO/CUBN levels and protein-protein pathway analysis. **A.** Relative label-free quantification of CUBN and MPO within the T1 patients. Log₂-transformed Label-free quantification intensity (Log₂ LFQ I). **B.** CUBN/MPO ratio correlation with patient survival. **C.** Patients with T1 clinical and medical follow-up information. **D and E.** Violin plot showing the relative expression of MPO and CUBN by the protein label-free quantification (LFQ) levels. Expression of those proteins is shown in the group of patients with recurrence and no recurrence. Student's *t*-test was used for statistical analysis. ** *p* ≤ 0.01; **** *p* ≤ 0.0001. **F.** Label-free quantification of myeloperoxidase and **G.** Validation of mass spectrometry data using myeloperoxidase (MPO) protein by ELISA (ng/mL). X-axes represent the patient number. **H.** The top 20 pathways were obtained with the differential protein expression between the RG and NRG groups (x-axis -Log₁₀ p-value). Pathways were obtained by the software platform cytoscope using StringApp with the Reactome pathway database. Pathways: (1) Innate immune system; (2) Immune system; (3) Neutrophil degranulation; (4) Regulation of insulin-like growth factor (IGF) transport and uptake by Insulin-like growth factor binding proteins (IGFBPs); (5) Post-translational protein phosphorylation; (6) Platelet degranulation; (7) Hemostasis; (8) Platelet activation, signaling and aggregation; (9) Regulation of complement cascade; (10) Formation of fibrin clot (clotting cascade); (11) Extracellular matrix organization; (12) Metabolism of proteins; (13) Integrin cell surface interactions; (14) Common pathway of fibrin clot formation; (15) Intrinsic pathway of fibrin clot formation; (16) Terminal pathway of complement; (17) Plasma lipoprotein assembly; (18) Chylomicron assembly; (19) Post-translational protein modification; (20) Degradation of the extracellular matrix. **I.** Protein-protein interaction and pathway analysis related to immune responses focus on the neutrophil action.

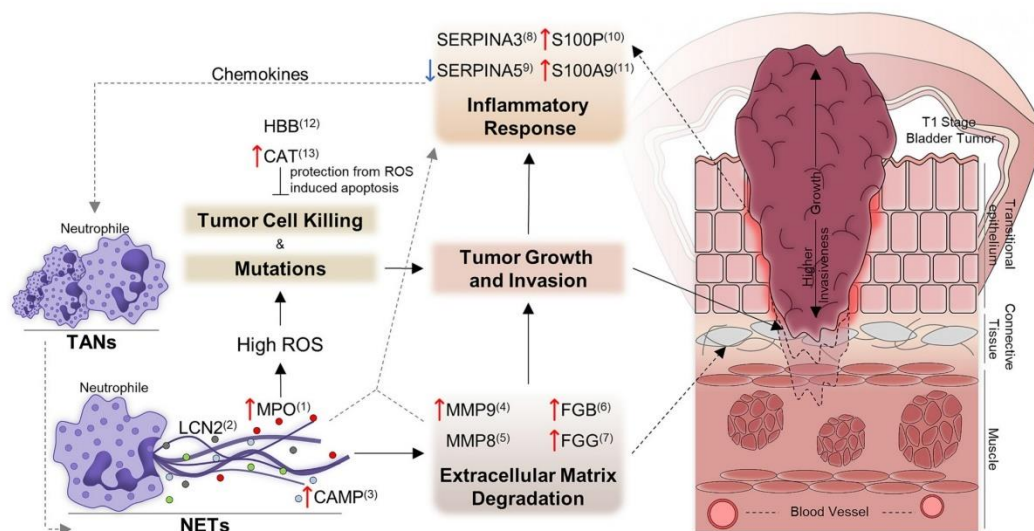


Fig. 3. Data integration for T1 Bca stage patients. Gene name, protein name and protein ID: (1) Myeloperoxidase (MPO, P05164–2); (2) Neutrophil gelatinase-associated lipocalin (LCN2, X6R8F3); (3) Cathelicidin antimicrobial peptide (CAMP, J3KNB4); (4) Matrix metalloproteinase-9 (MMP9, P14780); (5) Matrix metalloproteinase 8 (MMP8, P22894); (6) Fibrinogen beta chain (FGB, P02675); (7) Fibrinogen gamma chain (FGG, P02679–2); (8) Alpha-1-antichymotrypsin (SERPINA 3, P01011); (9) Plasma serine protease inhibitor (SERPINA 5, P05154); (10) Protein S100—P (S100P, P25815); (11) Protein S100-A9 (S100A9, P06702); (12) Hemoglobin subunit beta (HBB, P68871); (13) Catalase (P04040, CAT).

proteases that, in conjunction with other enzymes released during the NETs formation, are known to degrade the extracellular matrix, collagen, and gelatin [19]. This extracellular degradation also increases the concentration of fibrin-related proteins, such as fibrinogen beta and gamma (FGB and FGG), that we found in the RG group. These proteins further promote tumor invasion [20].

Furthermore, several proteins released during NETs formation have been associated with a poor prognosis in other cancer types [21,22]. Belonging to this group, we have found proteins Neutrophil Gelatinase-Associated Lipocalin (LNC2), Cathelicidin antimicrobial peptide CAMP, and MPO up-regulated in the RG group (see Fig. 3). The last one, MPO, is a notable finding because here we have found it markedly up-regulated in the patients with RG in agreement with results reported in other cancer types [14,15]. As mentioned above, MPO is involved in forming H₂O₂ to kill cancer cells via reactive oxygen species (ROS)-induced apoptosis. This also leads to an increment in the ROS and a higher number of mutations in the cancer cells. This may promote, to some extent, for the tumor to evolve and change expression patterns to survive by escaping the immune system [23]. The cancer cells respond to this attack by overexpressing proteins that promote protection against ROS,

such as catalase (CAT) and beta-globulin (HBB), which we also found in the RG group [24].

Finally, we have found the inflammatory response unbalanced for RG patients. Thus, whilst the serpin protein family involved in controlling the inflammatory response is down-regulated (SERPINA5), the S100 family, which promotes inflammation, is overexpressed (S100P, S100A9). Remarkably, S100A9 is a calcium- and zinc-binding protein that plays a prominent role in regulating inflammatory processes and the immune response [25]. It can induce neutrophil chemotaxis and adhesion and increase the bactericidal activity of neutrophils by promoting phagocytosis via activation of SYK, PI3K/AKT, and ERK1/2 pathways. Also, it can induce the degranulation of neutrophils by a MAPK-dependent mechanism [26].

In summary, our data show the presence of an inflammatory microenvironment that facilitates immune cell-matrix cross-talk, and promotes chronic inflammation. This state of inflammation further recruits more neutrophils and immune cells into the tumor microenvironment resulting in a cycle that the tissue cannot resolve. In agreement with the literature [27,28], our urinary proteome-based data suggest that the deregulation of inflammatory processes leads to an accelerated

BCa progression and a worse outcome for patients with BCa leading to recurrence and tumor growth. On the other hand, when the inflammatory process is balanced, the BCa seems to evolve slowly and with a better prognosis for the patient. In support of our conclusions, it has been recently reported that NET formation in chronic inflammation leads to radio-resistance in a mouse heterotopic model of invasive BCa [29].

5. Conclusions

We conclude that inflammation dysregulation is an essential driver of BCa recurrence and progression and that anti-inflammatory or immune checkpoint therapies should be further explored in treating patients with BCa. Furthermore, consider the MPO to CUBN ratio a promising candidate for a urine prognosis tool in BCa. A limitation of this study is its small patient sample, which may not fully capture the variability of T1-stage bladder cancer. Another issue is that some patients' prognoses were not evaluated within the time frame of this study. Further validation with larger patient groups is necessary.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2023.104865>.

Author Contributions

Conceptualization: JLCM, CL, RD, LCP, and HMS; Data curation: LBC and HMS; Formal analysis: LBC, JLCM, RB and HMS; Funding acquisition: JLCM, CL and HMS; Investigation: LBC, JLCM, LCP, MM and HMS; Methodology: LBC, JLCM, CL, RD, JJM, LCP, MM and HMS; Project administration: JLCM and HMS; Resources: JLCM, CL, LCP, MM and HMS; Supervision: JLCM and HMS; Visualization: LBC, JLCM, RB and HMS; Writing – original draft: LBC, JLCM and HMS; Writing – review & editing: LBC, JLCM, CL, RB, RD, JJM, LCP, MM and HMS.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set number PXD026784.

Acknowledgments

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). LBC, JLC, CL, RB, and HMS acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV which is financed by national funds from FCT/MCTES, *Fundação para a Ciência e Tecnologia e Ministério da Ciência, Tecnologia e Ensino Superior*, through the projects UIDB/50006/2020 and UIDP/50006/2020. HMS acknowledges the Associate Laboratory for Green Chemistry-LAQV (LA/P/0008/2020) funded by FCT/MCTES for his research contract. LBC thanks the FCT/MCTES for his Ph.D. grant (SFRH/BD/144222/2019). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [30] partner repository with the data set number PXD026784.

References

- [1] D. Ho, S.R. Quake, E.R.B. McCabe, W.J. Chng, E.K. Chow, X. Ding, et al., Enabling Technologies for Personalized and Precision Medicine, *Trends Biotechnol.* 38 (2020) 497–518.
- [2] H. Rodriguez, J.C. Zenklisen, L.M. Staudt, J.H. Doroshov, D.R. Lowy, The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment, *Cell* 184 (2021) 1661–1670, <https://doi.org/10.1016/j.cell.2021.02.055>.
- [3] G. Kontostathi, M. Makridakis, V. Bitsika, N. Tsolakos, A. Vlahou, J. Zoidakis, Development and validation of multiple reaction monitoring (MRM) assays for clinical applications, *Methods Mol. Biol.* 1959 (2019) 205–223, https://doi.org/10.1007/978-1-4939-9164-8_14. Humana Press Inc.
- [4] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* 71 (2021) 209–249, <https://doi.org/10.3322/caac.21660>.
- [5] C.Z. Zhu, H.N. Ting, K.H. Ng, T.A. Ong, A review on the accuracy of bladder cancer detection methods, *J. Cancer* 10 (17) (2019 Jul 8) 4038–4044, <https://doi.org/10.7150/jca.28989>.
- [6] J.A. Witjes, H.M. Bruins, R. Cathomas, E.M. Compérat, N.C. Cowan, G. Gakis, et al., European Association of Urology guidelines on muscle-invasive and metastatic bladder Cancer: summary of the 2020 guidelines, *Eur. Urol.* 79 (2021) 82–104, <https://doi.org/10.1016/j.eururo.2020.03.055>.
- [7] A. Kamoun, A. de Reyniès, Y. Allory, G. Sjödh, A.G. Robertson, R. Seiler, et al., A consensus molecular classification of muscle-invasive bladder Cancer, *Eur. Urol.* 77 (2020) 420–433, <https://doi.org/10.1016/j.eururo.2019.09.006>.
- [8] B. Jordan, J.J. Meeks, T1 bladder cancer: current considerations for diagnosis and management, *Nat. Rev. Urol.* 16 (2018) 23–34, <https://doi.org/10.1038/s41585-018-0105-y>.
- [9] L.B. Carvalho, J.L. Capelo-Martínez, C. Lodeiro, J.R. Wiśniewski, H.M. Santos, Snap-frozen freeze-free preservation and processing of the urine proteome using the combination of stabilizer-based technology and filter-aided sample preparation, *Anal. Chim. Acta* 1076 (2019) 82–90, <https://doi.org/10.1016/j.aca.2019.05.051>.
- [10] S. Tyanova, T. Temu, A. Carlson, P. Sinitcyn, M. Mann, J. Cox, Visualization of LC-MS/MS proteomics data in MaxQuant, *Proteomics* 15 (2015) 1453–1456, <https://doi.org/10.1002/pmic.201400449>.
- [11] S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, et al., The Perseus computational platform for proteomics data, *Nat. Methods* 13 (2016) 731–740, <https://doi.org/10.1038/nmeth.3901>.
- [12] N.T. Donecheva, J.H. Morris, J. Gorodkin, L.J. Jensen, Cytoscape StringApp: network analysis and visualization of proteomics data, *J. Proteome Res.* 18 (2019) 623–632, https://doi.org/10.1021/ACS.JPROTEOME.8B00702/SUPPL_FILE/PR8B00702_SI_002.XLSX.
- [13] S.J. Klebanoff, A.J. Kettle, H. Rosen, C.C. Winterbourn, W.M. Nauseef, Myeloperoxidase: a front line defender against phagocytosed microorganisms, *J. Leukoc. Biol.* 93 (2013) 185–198, <https://doi.org/10.1189/jlb.0712349>.
- [14] A.L. Rymaszewski, E. Tate, J.P. Yimbasah, A.E. Gelman, J.A. Jarzenbowski, H. Zhang, et al., The Role of Neutrophil Myeloperoxidase in Models of Lung Tumor Development, *Cancers* 6 (2014) 1111–1127, <https://doi.org/10.3390/CANCERS6021111>.
- [15] D.C. Castillo-Tong, D. Pils, G. Heinze, I. Braicu, J. Sehoul, A. Reinthaller, et al., Association of myeloperoxidase with ovarian cancer, *Tumor Biol.* 35 (2014) 141–148, <https://doi.org/10.1007/s13277-013-1017-3>.
- [16] G. Grenel, D. Djureinovic, M. Niniivirta, A. Laird, O. Ljungqvist, H. Johansson, et al., A systematic search strategy identifies cubilin as independent prognostic marker for renal cell carcinoma, *BMC Cancer* 17 (2017) 9, <https://doi.org/10.1186/s12885-016-3030-6>.
- [17] N. Mukaida, S. Sasaki, T. Baba, Two-faced roles of tumor-associated neutrophils in cancer development and progression, *Int. J. Mol. Sci.* 21 (2020) 3457, <https://doi.org/10.3390/ijms21103457>.
- [18] S. Jaillon, A. Ponzezza, D. Di Mitri, A. Santoni, R. Bonicchi, A. Mantovani, Neutrophil diversity and plasticity in tumour progression and therapy, *Nat. Rev. Cancer* 20 (2020) 485–503, <https://doi.org/10.1038/s41568-020-0281-y>.
- [19] M.T. Masucci, M. Minopoli, M.V. Carriero, Tumor associated neutrophils. Their role in tumorigenesis, metastasis, prognosis and therapy, *Front. Oncol.* 9 (2019) 1146, <https://doi.org/10.3389/fonc.2019.01146>.
- [20] M.W. Mosesson, Fibrinogen and fibrin structure and functions, *J. Thromb. Haemost.* 3 (2005) 1894–1904, <https://doi.org/10.1111/j.1538-7836.2005.01365.x>. John Wiley & Sons, Ltd.
- [21] S. Toluie, H.O. Yazdani, A.B. Al Khafaji, A.P. Chidi, O. Loughran, K. Mowen, et al., Neutrophil extracellular traps promote the development and progression of liver metastases after surgical stress, *Cancer Res.* 76 (2016) 1367–1380, <https://doi.org/10.1158/0008-5472.CAN.15.1591>.
- [22] E.B. Eruslanov, Phenotype and function of tumor associated neutrophils and their subsets in early-stage human lung cancer, *Cancer Immunol. Immunother.* 66 (2017) 997–1006, <https://doi.org/10.1007/s00262-017-1976-0>.
- [23] V. Aggarwal, H.S. Tuli, A. Varol, F. Thakral, M.B. Yerer, K. Sak, et al., Role of reactive oxygen species in cancer progression: molecular mechanisms and recent advancements, *Biomolecules* 9 (2019) 735, <https://doi.org/10.3390/biom9110735>.
- [24] P.S. Smith, W. Zhao, M. Robbins, Knocking down catalase leads to increased ROS production and proliferation in glioma cells, *Cancer Res.* 66 (2006).
- [25] S. Wang, R. Song, Z. Wang, Z. Jing, S. Wang, J. Ma, S100A8/A9 in inflammation, *Front. Immunol.* 9 (2018) 1298, <https://doi.org/10.3389/fimmu.2018.01298>.
- [26] J. C. Simard, M. M. Simon, P.A. Tessier, D. Girard, Damage associated molecular pattern S100A9 increases bactericidal activity of human neutrophils by enhancing phagocytosis, *J. Immunol.* 186 (2011) 3622–3631, <https://doi.org/10.4049/jimmunol.1002956>.
- [27] X. Sui, L. Lei, L. Chen, T. Xie, X. Li, X. Sui, et al., Inflammatory microenvironment in the initiation and progression of bladder cancer, *Oncotarget* 8 (2017) 93279–93294, <https://doi.org/10.18632/oncotarget.21565>.
- [28] R. Arantes Rodrigues, R. Pinto Leite, R. Ferreira, M.J. Neuparth, M.J. Pires, I. Gaivão, et al., Meloxicam in the treatment of in vitro and in vivo models of

L.B. Carvalho et al.

Journal of Proteomics 278 (2023) 104865

- urinary bladder cancer, *Biomed. Pharmacother.* 67 (2013) 277–284, <https://doi.org/10.1016/j.biopha.2013.01.010>.
- [29] S. Shinde-Jadhav, J.J. Mansure, R.F. Rayes, G. Marcq, M. Ayoub, R. Skowronski, et al., Role of neutrophil extracellular traps in radiation resistance of invasive bladder cancer, *Nat. Commun.* 12 (2021) 1–14, <https://doi.org/10.1038/s41467-021-23086-z>.
- [30] Y. Perez Riverol, A. Csordas, J. Bai, M. Bernal Llinares, S. Hewapathirana, D. J. Kundu, et al., The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* 47 (2019) D442–D450, <https://doi.org/10.1093/nar/gky1106>.

| E

**PUBLISHED PAPER IN NATURE COMMUNICA-
TIONS MEDICINE - CORRESPONDING TO
CHAPTER 7**

communications medicine

ARTICLE

<https://doi.org/10.1038/s43856-023-00238-4>

OPEN



Pathway-guided monitoring of the disease course in bladder cancer with longitudinal urine proteomics

Luís Botelho Carvalho^{1,2}, José Luis Capelo^{1,2}, Carlos Lodeiro^{1,2}, Rajiv Dhir³, Luis Campos Pinheiro^{4,5}, Hugo López-Fernández^{6,7}, Gonçalo Martins^{1,2}, Mariana Medeiros^{4,5}, Fernando Díaz⁸ & Hugo Miguel Santos^{1,2,3}✉

Abstract

Background Monitoring bladder cancer over time requires invasive and costly procedures. Less invasive approaches are required using readily available biological samples such as urine. In this study, we demonstrate a method for longitudinal analysis of the urine proteome to monitor the disease course in patients with bladder cancer.

Methods We compared the urine proteomes of patients who experienced recurrence and/or progression ($n = 13$) with those who did not ($n = 17$). We identified differentially expressed proteins within various pathways related to the hallmarks of cancer. The variation of such pathways during the disease course was determined using our differential personal pathway index (dPPI) calculation, which could indicate disease progression and the need for medical intervention.

Results Seven hallmark pathways are used to develop the dPPI. We demonstrate that we can successfully longitudinally monitor the disease course in bladder cancer patients through a combination of urine proteomic analysis and the dPPI calculation, over a period of 62 months.

Conclusions Using the information contained in the patient's urinary proteome, the dPPI reflects the individual's course of bladder cancer, and helps to optimise the use of more invasive procedures such as cystoscopy.

Plain language summary

Bladder cancer must be closely monitored for progression, but this requires expensive and invasive procedures such as cystoscopy. Less invasive procedures using readily available samples such as urine are needed. Here, we present an approach that measures the levels of various proteins in the urine. We compare protein levels at different points during the disease course in patients with bladder cancer, and show this helps to flag disease recurrence and the need for medical intervention. Our approach could help clinicians to determine which patients require more invasive testing and treatment.

¹BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal. ²PROTEOMASS Scientific Society, Madan Parque, Rua dos Inventores, 2825-182 Caparica, Portugal. ³Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA. ⁴Urology Department, Central Lisbon Hospital Center, Lisbon, Portugal. ⁵NOVA Medical School, NOVA University of Lisbon, Lisbon, Portugal. ⁶CINBIO, Universidade de Vigo, Department of Computer Science, ESEI-Escuela Superior de Ingeniería Informática, 32004 Ourense, Spain. ⁷SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, 36213 Vigo, Spain. ⁸Department of Computer Science, Universidad de Valladolid, Escuela de Ingeniería Informática, 40005 Segovia, Spain. ✉email: hmsantos@fct.unl.pt

ARTICLE

COMMUNICATIONS MEDICINE | <https://doi.org/10.1038/s43856-023-00238-4>

Bladder cancer (BC) is among the worst neoplasms because of its high incidence and mortality^{1,2} and because cystoscopy is necessary to diagnose BC. If BC is confirmed, the patient undergoes surgery, then must undergo a control cystoscopy approximately every 3 months^{3,4}. Cystoscopy is an expensive intervention that requires surgical facilities and a minimum of two physicians and one nurse. However, the procedure can also have a detrimental effect on the patient's state of mind, as cystoscopy is highly invasive and is performed through the urethra. Furthermore, some patients experience a rapid recurrence; thus, the control cystoscopy is performed too late, and transurethral resection of the bladder tumour is needed. Therefore, new approaches to diagnose and monitor BC are needed.

The biochemical changes that occur at the protein level in an individual after they develop a disease provides that can enable physicians to adjust therapy, provide follow up medical care and personalise their care^{5,6}. The concept of performing longitudinal profiling with individual proteomes has been described in the literature;^{7–9} however, to date, there are few easy ways to achieve this in a simple, straightforward and robust manner. To our knowledge, the possibility of monitoring the course of BC over time using the urinary proteome has not yet been described in the literature. To overcome this gap, we explored changes in the urinary proteome of BC patients using high-resolution mass spectrometry-based proteomics in conjunction with advanced bioinformatics tools. Thus, we introduced, a concept named the differential personal pathway index (dPPI), which uses the variation in the expression of selected biochemical pathways, which is calculated using a large number of urine proteins that are linked to Hanahan and Weinberg's biological hallmarks of cancer (HWhc^{10,11}). Thus, based on variation in the levels of tens of signature urine proteins, the dPPI reflects both the response of the individual's urine proteome to whatever course the disease takes and any medical care that ensues. The variation reflects the status of the biochemical pathways to which the proteins correspond, and these pathways are specifically chosen to combine multiple sources of diagnostic information. Thus, in this work, we showed how to monitor the progression of BC in a comprehensive yet straightforward and personalised manner using the patient's urine proteome. Essentially, a more positive dPPI indicates a potentially more severe clinical course of BC; whereas a more negative dPPI indicates a better clinical outlook. Therefore, the dPPI flags the need for medical intervention at an early stage. This work further demonstrates this concept using six patients diagnosed with T1-stage BC who were followed for up to 62 months.

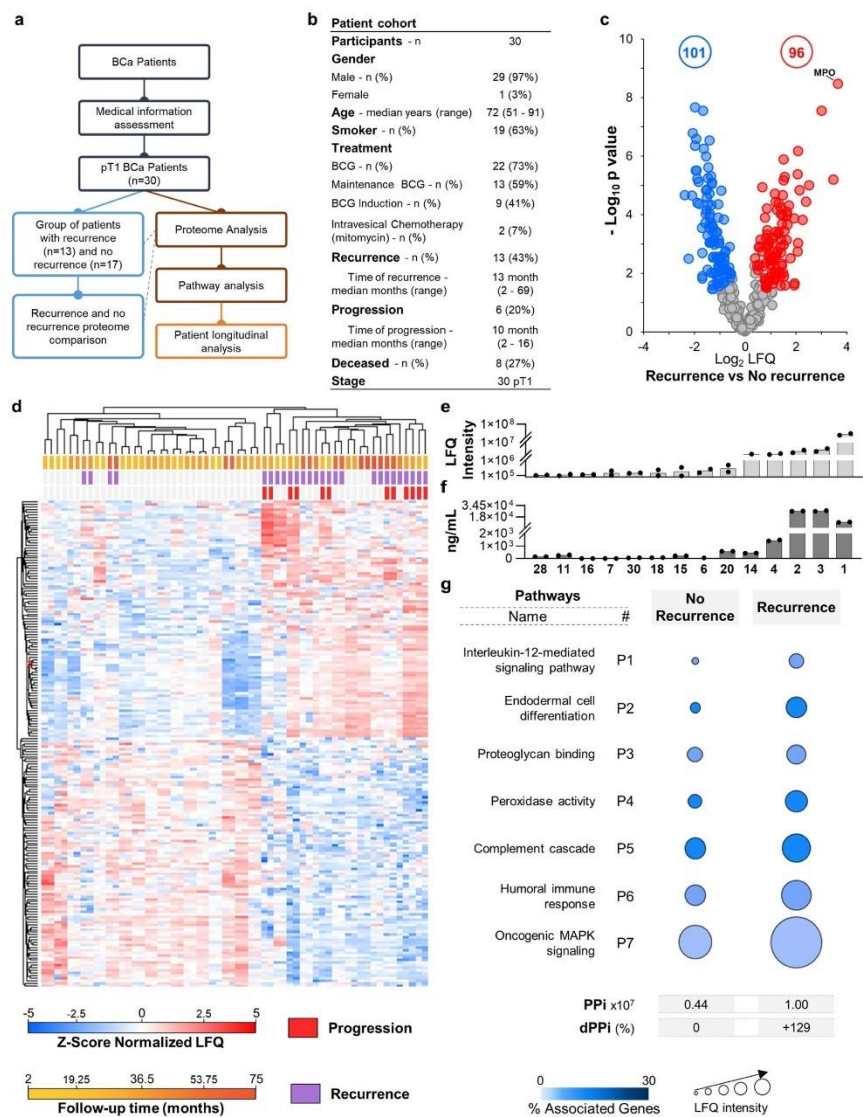
Methods

Patients. Human mid-stream second void morning urine specimens were collected from patients with a pT1 bladder cancer diagnosis that was confirmed by pathological examination. All patients were informed about the study and signed informed consent according to the policies of the Central Lisbon Hospital Center Ethics Committee, who approved our study. The ethics approval number is 669/2018. The patients that were enrolled in the study were selected based on the following criteria: (a) Inclusion—clear bladder cancer diagnosis, and (b) Exclusion—no records of urinary cancer history, no HIV, no organ transplant, and no recent chemo/radiotherapy. We compared the proteome of 30 patients from two groups—those who had experienced BC recurrence and/or progression (a total of 13) and those who had not (a total of 17). Fig. 1a depicts the steps from proteome analysis to pathway, and clinical data interpretation, while the medical information for the cohorts of patients is presented in Fig. 1b. Furthermore, we followed six patients for up to 62 months.

Sample treatment. Urine samples were collected in 50 mL centrifuge tubes (DB Falcon) that contained 38 mg of boric acid (Sigma-Aldrich) to prevent bacterial growth¹². Samples with haematuria were not included in our study. The urine samples were centrifuged at 5000 × g for 20 min to remove cell debris, and 10 mL aliquots of the resulting supernatants were stored at –60 °C until further use. An aliquot of urine (10 mL) was concentrated by centrifugal ultrafiltration using a VivaSpin 15R (10 kDa MWCO, Sartorius) at 5000 × g for 20 min. Finally, the concentrated urinary proteomes were quantified via a 96 well plate Bradford protein assay (Sigma-Aldrich) using bovine serum albumin to generate the calibration curve.

Filter-aided sample preparation (FASP). Filter Aided Sample Preparation (FASP) was used to perform urinary proteome digestion, following a modified FASP method^{5,13–15}. Briefly, different amounts of protein ranging from 50 and 100 µg were diluted with H₂O to obtain 200 µL and an additional 200 µL of 8 M urea (Sigma-Aldrich), 75 mM Tris-HCl (Sigma-Aldrich), 100 mM NaCl (Sigma-Aldrich), and 0.02% SDS (Sigma-Aldrich). Then this protein solution was loaded onto a VivaSpin 500 (10 kDa MWCO, Sartorius) and centrifuged for 20 min at 14,000 × g. Each sample was processed in duplicate. The proteins in the membrane were washed with 200 µL of 8 M urea 25 mM AmBic solution and then centrifuged for 20 min at 14,000 × g. Protein disulphide bonds were reduced by adding 200 µL of 50 mM dithiothreitol (DTT) in 8 M urea and 25 mM AmBic and incubated for 60 min at 37 °C. Then, centrifugation for 20 min at 14,000 × g was performed, and the sample was alkylated for 45 min in the dark by adding 100 µL 50 mM iodoacetamide in 8 M urea and 25 mM AmBic solution. Subsequently, centrifugation for 20 min at 14,000 × g was performed, and then the samples were washed twice with 200 µL of 25 mM AmBic. Finally, protein digestion was performed by adding 100 µL of trypsin solution (1:30 trypsin to protein ratio) prepared in 12.5 mM AmBic. Protein digestion was performed overnight (~16 h) at 37 °C, and then the peptides were collected by 20 min centrifugation at 14,000 × g, followed by two additional membrane washing steps with 100 µL of 3% (v/v) acetonitrile that contained 0.1% (v/v) aqueous formic acid and centrifugation of 20 min at 14,000 × g. Finally, peptides were transferred to a 500 µL microtube, dried and stored at –20 °C until further analysis by Nano-LC-MS/MS.

LC-MS/MS. LC-MS /MS analysis was carried out using an Ultimate 3000 nano-LC system coupled to an Impact HD (Bruker Daltonics) with a CaptiveSpray nanoBooster that used acetonitrile as a dopant. The protein digests were resuspended in 100 µL of 3% (v/v) acetonitrile containing 0.1% (v/v) aqueous formic acid (FA) and were sonicated for 10 min using an ultrasonic bath at 100% ultrasonic amplitude, and 35 kHz ultrasonic frequency. Afterwards, the samples were quantified using a Pierce™ quantitative colorimetric peptide assay. Then, 3 µL of protein digest containing 563 ng of peptides was loaded onto a trap column (Acclaim PepMap100, 5 µm, 100 Å, 300 µm i.d. × 5 mm) and desalted for 5 min from 3 to 5% B (B: 90% acetonitrile 0.08% FA) at a flow rate of 15 µL min^{–1}. The peptides were separated using an analytical column (Acclaim™ PepMap™ 100C18, 2 µm, 0.075 mm i.d. × 150 mm) with a linear gradient at 300 nL min^{–1} (mobile phase A: aqueous FA 0.1% (v/v); mobile phase B 90% (v/v) acetonitrile and 0.08% (v/v) FA) 5–90 min from 5 to 35% mobile phase B, 90–100 min linear gradient from 35 to 95% of mobile phase B, 100–110 min 95% B. Chromatographic separation was carried out at 35 °C. MS acquisition was set to cycles of MS (2 Hz), followed by MS/MS (8–32 Hz) with a cycle time of 3.0 s and active exclusion (precursors were excluded from the precursor selection for 0.5 min after the



acquisition of the 1 MS/MS spectrum; the intensity threshold for fragmentation was 2500 counts). As, the active exclusion was set to one, we reconsidered the precursor and whether the intensity of a precursor increased by a factor of three; this mass was taken from the 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. The mass spectrometry proteomics data were deposited in the ProteomeXchange Consortium¹⁶ via the PRIDE¹⁷ partner repository with the data set identifier PXD025139. To confirm myeloperoxidase label-free quantification levels, validation was performed using a commercially available ELISA kit (ab136943 - Myeloperoxidase (MPO)

Human ELISA kit from Abcam). The method was carried out according to the standard protocol described by the manufacturer.

Bioinformatics and statistical analyses. Relative label-free quantification was performed using the precursor signal intensity method and delayed normalisation, MaxLFQ, on MaxQuant software V2.0.3.0¹⁸. All raw files were processed in a single run with default parameters^{19,20}. Database searches were performed using Andromeda against the human UniProt UP000005640_9606 database (20,600 sequences; 11,395,157 residues, downloaded on April 27 2021)²¹. Searches were configured with cysteine

ARTICLE

COMMUNICATIONS MEDICINE | <https://doi.org/10.1038/s43856-023-00238-4>

Fig. 1 Proteome analysis of 30 pT1 bladder cancer patients. **a** Workflow of the experimental design. **b** Clinical information of the patient cohort used in the study. **c** Volcano plots showing statistically significant changes in protein levels in T1-BC patients with recurrences and/or progression ($n = 13$ patients, two technical replicates per patient) compared to the group with no recurrence ($n = 17$ patients, two technical replicates per patient), according to Student's t test (FDR 0.05 and S_0 of 0.1). Dots represent proteins that show statistically significant increases (red), decreases (blue) nonstatistically significant changes (grey). The total numbers of proteins that exhibited statistically significant differential expression are shown in the corresponding-coloured circles. **d** Hierarchical clustering of the T1-BC patients into two groups, as expected, recurrence and no recurrence (average linkage, no constraint, preprocessing with k-means, and Euclidean distance, $n = 30$ patients, two technical replicates per patient). **e, f** Validation of mass spectrometry data using myeloperoxidase (MPO) protein label-free quantification (LFQ) and ELISA (ng/mL). Bars represent the average of two technical replicates. The MPO protein was recently proposed to be a marker of poor prognosis in lung²⁷ and ovarian²⁸ carcinomas. To the best of our knowledge, no studies report MPO overexpression as a prognostic tool for BC carcinoma. **g** Pathways selected for calculating the Personal Pathway Index, after applying the guidelines provided by Hanahan and Weinberg^{10, 11} to the proteins showing statistically significant differential expression derived from the volcano plot in **c** (Supplementary Data 1). Differentially expressed proteins were grouped according to their functions, and networks were analysed in ClueGo (v 2.5.8), and Cytoscape (v 3.8.2). The LFQ average for all proteins involved in the pathway is expressed as a circle area plot. The maximum number of proteins identified in one single urine of our set was 494. The number of unique proteins considering all the samples of our set was about 754. For our studies, we have chosen the proteins that are found common in at least 50% of the urines of each group. This way, the number of proteins used in Fig. 1 (recurrence versus no recurrence) was 380.

Table 1 Pathways and proteins used on the longitudinal follow up of bladder cancer patients.

Pathway Number	GOID	Name	Ontology Source	Associated Proteins Found
P1	GO:0035722	interleukin-12-mediated signalling pathway	GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_08.05.2020_00h00	[LCP1, PPIA, S100A9 ^a]
P2	GO:0035987	endodermal cell differentiation	GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_08.05.2020_00h00	[FN1, MMP8, MMP9 ^a]
P3	GO:0043394	proteoglycan binding	GO_MolecularFunction-EBI-UniProt-GOA-ACAP-ARAP_08.05.2020_00h00	[AZU1, CFH, CTSB, FN1, HRG, LCP1, NID1]
P4	GO:0004601	peroxidase activity	GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_08.05.2020_00h00	[CAT, HBA1, HBB, HBD, MPO]
P5	R-HSA:166658	Complement cascade	REACTOME_Pathways_08.05.2020	[C3, C4A, C9, CD55, CFH, F2]
P6	GO:0006959	humoral immune response	GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_08.05.2020_00h00	[A2M, APCs, AZU1, C3, C4A, C9, CAMP, CD55, CFH, F2, FGB ^a , HPX, HRG, IGHG1, IGKV2D-30, KRT1, LCN2, LTF, LYZ, PRTN3, S100A9 ^a]
P7	R-HSA:6802957	Oncogenic MAPK signalling	REACTOME_Pathways_08.05.2020	[FGB ^a , FGG ^a , FN1]

^aIndicates a protein previously identified as a BC biomarker

carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidation as variable modifications. We set the false discovery rate (FDR) to 0.01 for protein and peptide levels with a minimum length of seven amino acids for peptides. The FDR was determined by searching a reverse database. Enzyme specificity was set as C-terminal to arginine and lysine as expected using trypsin. A maximum of two missed cleavages were allowed. Data processing was performed using Perseus (version 1.6.15.0) with default settings²². Protein group LFQ intensities were log2-transformed, and the quantitative profiles were filtered for missing values with the following settings: min valid percentage of 50% in at least one group and values greater than 0. To overcome the obstacle introduced by missing LFQ values, the missing values were imputed using the parameters, with = 0.5 and downshift = 1.8. The list of differentially expressed proteins was obtained with a two-tailed Student's t test between the two groups (permutation-based FDR 0.05 and S_0 of 0.1), and then the log ratios were calculated as the difference in average log₂ LFQ intensity values between the tested conditions. Perseus was also used to obtain clusters, using average linkage, no constraint, preprocess with k-means and Euclidean distance between column trees. GraphPad Prism 8.2.1. was employed for graphic construction and statistical analysis. Cytoscape V3.8.2 and the application ClueGo V2.5.8 was used for

pathway enrichment analysis. Reactome and biological process GO terms were used as ontology databases.

dPPI calculation. The dPPI is calculated from the numerical output of proteomics analysis of urine at different time points, as follows:

$$dPPI = \frac{\sum_{nx=1}^{nx=i} Prot_{nx} LFQ_{t_y} - \sum_{nx=1}^{nx=i} Prot_{nx} LFQ_{t_1}}{\sum_{nx=1}^{nx=i} Prot_{nx} LFQ_{t_1}} \quad (1)$$

where $\sum_{nx=1}^{nx=i} Prot_{nx} LFQ_{t_y}$ accounts for the sum of all label-free quantitation (LFQ) values of proteins of interest at a given time t_y , while $\sum_{nx=1}^{nx=i} Prot_{nx} LFQ_{t_1}$ accounts for all LFQ measurements of the same proteins at the time of disease onset t_1 .

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results and discussion

Selection of pathway hallmarks of cancer. The differentially expressed proteins between the two groups ($n = 197$, Fig. 1c and

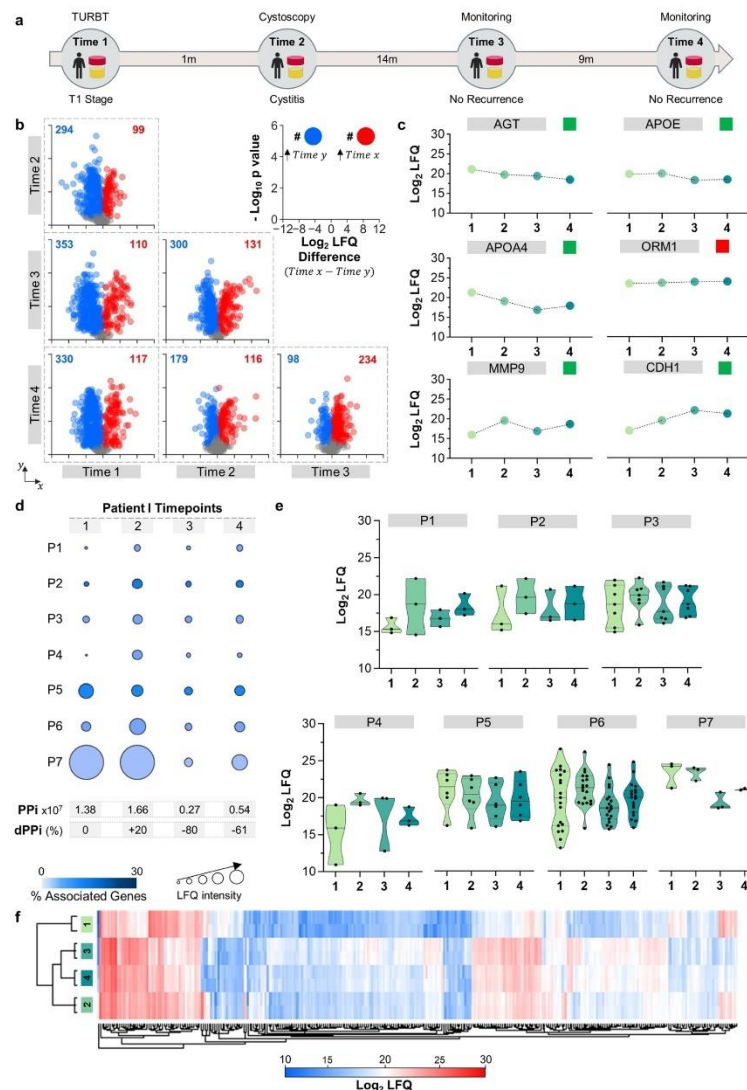
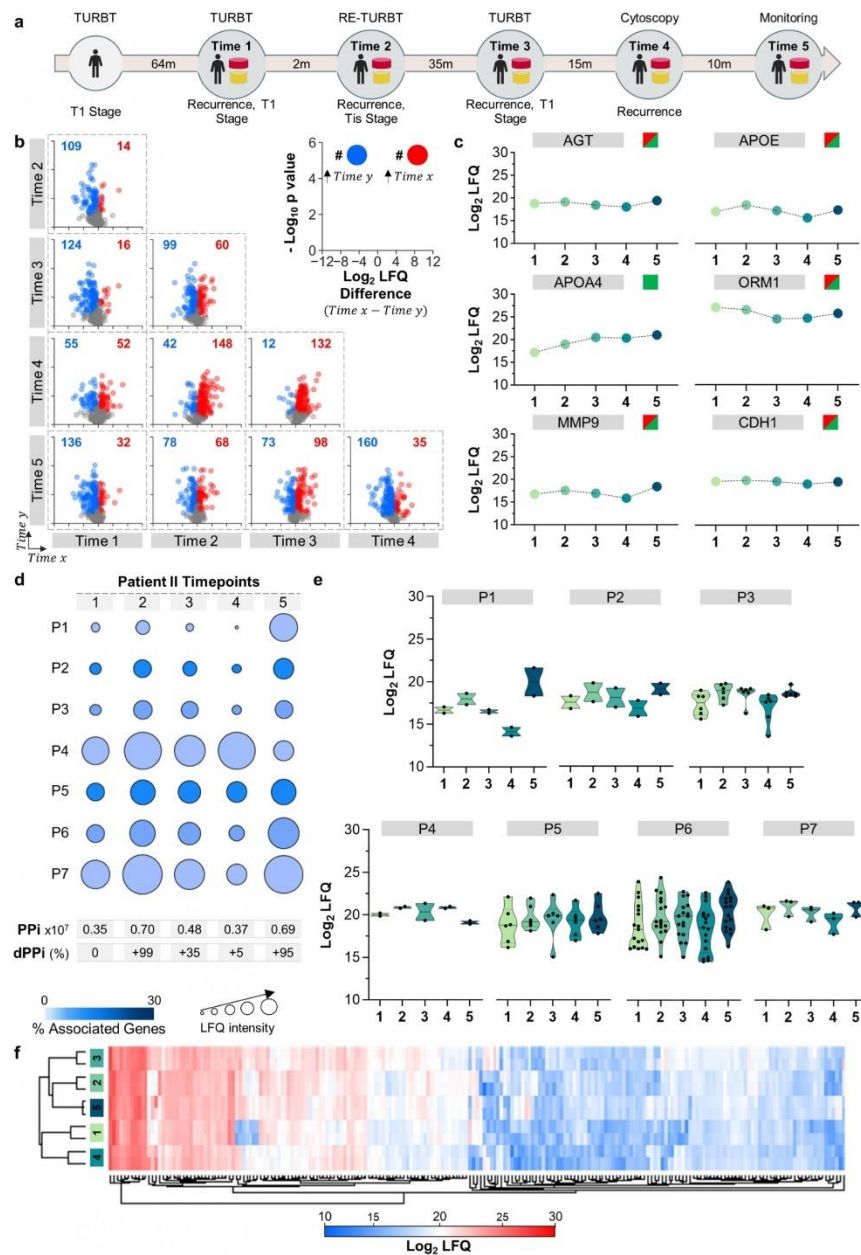


Fig. 2 Longitudinal analysis of the urinary proteome of patient I. **a** Timeline of the disease course in Patient I shows the intervals in months (m) between medical interventions and urine timepoint 1–4. TURBT: transurethral resection of bladder tumour; Re-TURBT: repeated transurethral resection of bladder tumour. **b** Volcano plot matrix showing proteome changes in the urine samples at time points 1–4. The dots represent the proteins that showed a statistically significant increase (red), decrease (blue) or nonstatistically significant changes (grey) according to the Student's *t* test (FDR 0.05 and SO of 0.1). **c** Variation in the following known protein biomarkers for bladder cancer: angiotensinogen (AGT)³⁵, apolipoprotein E (APOE)³⁶, matrix metalloproteinase-9 (MMP9)³⁷, apolipoprotein A-IV (APOA4)²⁹, alpha-1-acid glycoprotein 1 (ORM1)³⁸ and cadherin-1 (CDH1)^{30, 31}. The green and red squares indicate whether the variation in the urine of Patient I matched or not, respectively, with trends reported for each marker in the literature. Dots represent the average of two biological samples with two technical replicates each. **d** Personal pathway index (PPI) at each time point and estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3: proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **e** Distribution and density variation protein LFQ values (including two biological replicates) at each timepoint point for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **f** Hierarchical clustering of the three urinary proteomes of Patient I. Protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with *k*-means and Euclidean distance).

ARTICLE

COMMUNICATIONS MEDICINE | <https://doi.org/10.1038/s43856-023-00238-4>

Supplementary Data 1) were submitted to clue-go analysis (cystoscopy) to select the biochemical pathways involved in recurrence and/or progression. The 197 differentially expressed proteins distinguished 93.5% of patients with recurrence and 100% of patients with progression, as shown in Fig. 1d. The statistically significant pathways were cross-checked against the

HWHC^{10,11}, and seven pathways were selected. These pathways involve 35 related proteins (Fig. 1g. and Table 1). These 35 proteins were used to calculate the dPPI as defined in Eq. (1). Four of these 35 proteins have been described in literature as potential biomarkers for bladder cancer (indicated with ^a in Table 1)^{23–25}. Two of these four proteins namely fibrinogen β chain^{25,26} and

Fig. 3 Longitudinal analysis of the urinary proteome of patient II. **a** Timeline of disease course in Patient II showing the intervals in months (m) between medical interventions and urine timepoint 1 to 5. TURBT: transurethral resection of bladder tumour; Re-TURBT: repeated transurethral resection of bladder tumour. **b** Volcano plot matrix showing proteome changes in urine collected at time points 1–5. The dots represent proteins showing statistically significant increases (red), decreases (blue) or nonstatistically significant changes (grey) according to the Student's *t* test (FDR 0.05 and SO of 0.1). **c** Variation in the known protein biomarkers for bladder cancer, including angiotensinogen (AGT)³⁵, apolipoprotein E (APOE)³⁶, matrix metalloproteinase-9 (MMP9)³⁷, apolipoprotein A-IV (APOA4)²⁹, alpha-1-acid glycoprotein 1 (ORM1)³⁸ and cadherin-1 (CDH1)^{30, 31}. The green and red squares indicate whether the variation in the urine of Patient II matched or not, respectively, with trends reported for each marker in the literature. Dots represent the average of two biological samples with two technical replicates each. **d** Personal pathway index (PPI) at each time point and the estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **e** Distribution and density variation protein LFQ values (including two biological replicates) at each timepoint for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **f** Hierarchical clustering of the four urinary proteomes for Patient II. The protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

fibrinogen γ chain²⁵ have been described as having increased expression in urine of BC patients. In addition, we found the protein S100A9 which has been described as a biomarker candidate in plasma of BC patients²³. Protein myeloperoxidase (MPO) was found to exhibit the greatest increases in levels in patients with recurrence (Fig. 1c). MPO has been recently proposed to be a marker of poor prognosis in lung²⁷ and ovarian²⁸ carcinomas. Therefore, MPO was used for quality control to validate the mass spectrometry data using ELISA as an orthogonal method (Fig. 1e, f). Next, we investigated how the status of the seven pathways changed over time by following six patients with BC for up to 40 months. A comprehensive flow chart of the process is provided in Supplementary Fig. 1.

Longitudinal follow-up of BC patients. We followed six patients longitudinally over the course of their diagnosis and treatment. In this section, we will discuss three of the six patients in detail, and the other three are displayed in Supplementary Figs. 3, 4 and 5.

Patient I presented with symptoms that matched BC and was subsequently diagnosed. One reference urine sample was taken before urinary and bladder exploration was performed (see Fig. 2a for a comprehensive analysis timeline). Then, the urine proteome was extracted and analysed via mass spectrometry, and the PPI was calculated as explained in Eq. (1) and is presented in Fig. 2d. Physicians diagnosed this patient with T1-stage BC and performed a transurethral resection of the bladder tumour, TURBT. Then, one month later, a second analysis of the urinary proteome revealed a difference of +35% between the PPI at time point 1 and at time point 2 (Fig. 2d, PPI2 – PPI1), suggesting that the patient had not improved. In fact, all the hallmark pathways were obtained with higher values. This prompted us to call the patient and perform a medical inspection of the bladder, and the patient was diagnosed with cystitis; thus, no surgical intervention was needed. Another advantage of this type of analysis is that the mass spec LFQ values of individual protein biomarkers can also be used to obtain additional information about the course of the disease (Fig. 2c). For instance, APOA4, a urinary candidate biomarker for the presence of BC, exhibited lower value at timepoint 2 which might be a good indicator in terms of cancer progression. Increased levels of APOA4 in urine have been linked to bladder cancer²⁹. The same applies to the biomarker CDH1, in which an increment might also be a good signal^{30,31}. On the other hand, MMP9 is a biomarker linked to inflammation, and MMP9 exhibited an augmented value at timepoint 2, which is a negative indicator that is not necessarily related to cancer³².

In summary, the changes in these biomarkers are consistent with cystitis rather than a recurrence or progression. Further information supporting this is that the complex MMP9/NGAL has been reported as increased in the urine of children with acute cystitis³². At timepoint 2, both proteins are augmented. This result shows that

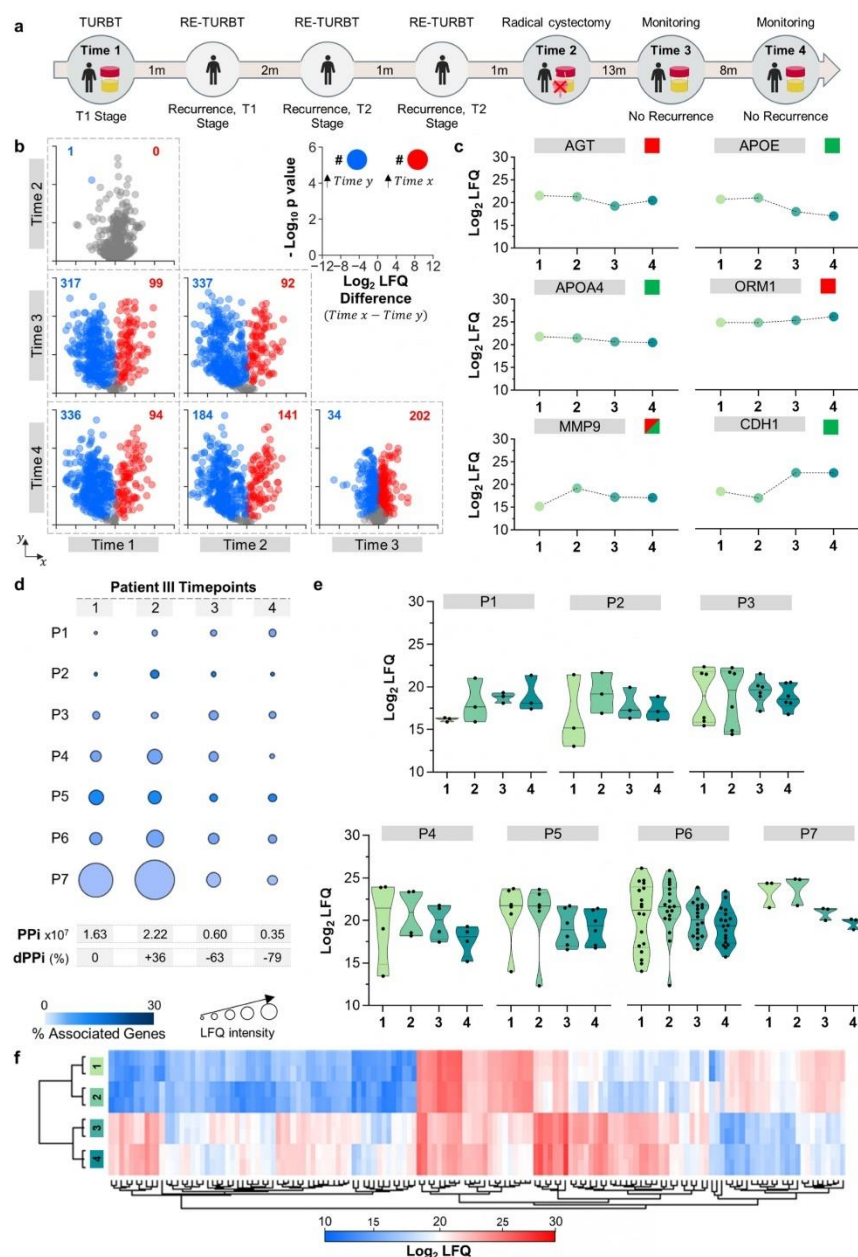
dPPI is useful as an alert tool to follow individual variations in the urine proteome caused by disease. Another cystoscopy three months later was performed and revealed no evidence of recurrence. Fourteen months later, a third urine sample was monitored again, showing a dPPI of –80% relative to time 1 (Fig. 2d timepoint 3), indicating potentially important immunological and inflammatory responses were downregulated (Fig. 2e), reflecting an improvement in the patient's condition, as confirmed by medical analysis. Nine months later, a fourth urine analysis revealed an increase in the dPPI index from –80% to –61%, so we concluded that no intervention was necessary. This was confirmed by the physicians. However, there is a clear increase in the MAPK oncogenic signalling pathway. This pathway has been reported to be involved in BC development³³, and overexpression of MAPK signalling proteins has been linked to a mutation in the P53 gene and poor prognosis³⁴. Furthermore, the variations in APOA4 and CDH1 indicated that the cancer progression had potentially worsened. Additionally, a change in MMP9 indicated a potential inflammatory response. Altogether, it was concluded that the patient was in a possible initial stage of recurrence. We recommended following this patient more frequently to check the dPPI levels in urine.

Patient II is an interesting case due to their frequent recurrences (Fig. 3). The first proteome analysis of the urine was performed 64 months after the onset of BC. Then, the PPI for this urine, which is presented in Fig. 3d timepoint 1, was calculated. At this point, the patient had a TURBT intervention. Two months later, a new urine analysis revealed a dPPI variation of +99%, (Fig. 3d timepoint 2), indicating that the patient's course was worsening. It is worth noting the increase in the expression of the MAPK pathway (P7, Fig. 3d timepoint 2). Based on these data, the patient underwent a urological assessment that revealed another tumour recurrence, confirming that the alert from the dPPI approach is valuable. Another TURBT was performed, and BC was documented.

After 35 months, a routine surveillance assessment was performed, and the urine proteome analysis revealed a dPPI of +35% relative to the time one reference value. The patient thus underwent another cystoscopy evaluation, which agreed with the dPPI alert and demonstrated the recurrence of BC. Consequently, a new TURBT was performed. A fourth urine proteomic analysis was completed 15 months later (Fig. 3d timepoint 4), which showed a change in the dPPI value of +5%, suggesting potential disease progression. Our prediction was again confirmed by cystoscopy as recurrence was recorded. Finally, ten months later an increase of +95% in the dPPI suggested that the disease was worsening. According to our medical intervention advice, the patient underwent one cystoscopy that confirmed our result. The patient was submitted to a TURBT.

Patient III (Fig. 4) presented with symptoms of BC, and a TURBT was performed because the cystoscopy results were positive. Histopathological analysis documented stage T1 BC. The patient's

ARTICLE

COMMUNICATIONS MEDICINE | <https://doi.org/10.1038/s43856-023-00238-4>

urine sample was collected before medical intervention, and the dPPI was calculated, as shown in Fig. 4d timepoint 1. Afterwards, the patient underwent three more TURBT procedures within the following four months. A second urine sample was taken one month after the fourth TURBT was performed (time point 2). The dPPI augmented a +36% (Fig. 4d timepoint 2), presenting seven

pathways that were increased. Again, the value of the dPPI approach was illustrated by the variation observed in the MAPK signalling pathway (P7 in Fig. 4d).

Our proteomics-based alert was confirmed as the medical exploration indicated that the disease had progressed to the T2-stage, and a radical cystectomy was performed. The urine

Fig. 4 Longitudinal analysis of the urinary proteome of patient III. **a** Timeline of the disease course in Patient III showing intervals in months (m) between medical interventions and urine timepoints 1 to 4. TURBT: transurethral resection of bladder tumour; Re-TURBT: repeated transurethral resection of bladder tumour. **b** Volcano plot matrix showing proteome changes in urine collected at time points 1–4. The dots represent proteins showing statistically significant increases (red), decreases (blue) or nonstatistically significant changes (grey) at the different time points according to Student's *t* test (FDR 0.05 and SO of 0.1). **c** Variation in the following known protein biomarkers for bladder cancer: angiotensinogen (AGT)³⁵, apolipoprotein E (APOE)³⁶, matrix metalloproteinase-9 (MMP9)³⁷, apolipoprotein A-IV (APOA4)²⁹, alpha-1-acid glycoprotein 1 (ORM1)³⁸ and cadherin-1 (CDH1)^{30, 31}. The green and red squares indicate whether the variation in the urine of Patient III matched or not, respectively, with trends reported for each marker in the literature. Dots represent the average of two biological samples with two technical replicates each. **d** Personal pathway index (PPI) for each time point and the estimated differential PPI (dPPI) were calculated as explained in the text. P1: interleukin-12-mediated signalling pathway; P2: endodermal cell differentiation; P3, proteoglycan binding; P4: peroxidase activity; P5: complement cascade; P6: humoral immune response; P7: oncogenic MAPK signalling. **e** Distribution and density variation protein LFQ values (including two biological replicates) at each timepoint for each pathway assessed. Continuous bar in the middle represents the median. The thin discontinuous line represents the quartile lines. **f** Hierarchical clustering of the four urinary proteomes of Patient III. The protein LFQ values were used to perform the cluster analysis (with average linkage, no constraint, preprocessing with k-means and Euclidean distance).

Table 2 The dPPI index column shows the index calculated as explained in the methods section.

Patient	Time point	dPPI index	dPPI Intervention?	Medical procedure	Outcome	dPPI intervention true?
1	2	20	Yes	Cystoscopy	Cystitis	Yes
1	3	−80	No	Medical Monitoring*	No tumour	Yes
1	4	−61	No	Medical Monitoring*	No tumour	Yes
2	2	99	Yes	Cystoscopy/RE-TURBT	T1 stage recurrence	Yes
2	3	35	Yes	Cystoscopy/TURBT	T1 stage recurrence	Yes
2	4	5	Yes	Cystoscopy	Recurrence (no AP exam)	Yes
2	5	95	Yes	Medical Monitoring*	Recurrence (no AP exam)	Yes
3	2	36	Yes	Cystoscopy/RE-TURBT	T2 Stage recurrence	Yes
3	3	−63	No	Medical Monitoring*	No recurrence	Yes
3	4	−79	No	Medical Monitoring*	No Recurrence	Yes
4	2	−80	No	Medical Monitoring*	No Recurrence	Yes
4	3	−56	No	Medical Monitoring*	No Recurrence	Yes
5	2	−3	Yes	Cystoscopy/RE-TURBT	Ta stage recurrence	Yes
5	3	−2	Yes	Medical Monitoring*	Nonneoplastic Bladder Mass	Yes
6	2	−41	No	Cystoscopy	No recurrence	Yes
6	3	−7	No	Medical Monitoring*	No Recurrence	Yes

The dPPI intervention column shows whether medical intervention is necessary or not (Yes/No). The medical procedure column indicates the medical intervention performed for the patient, while the outcome column highlights the diagnosis given by the physicians. The column named 'dPPI intervention true?' confirms whether the intervention that was recommended by the dPPI index was necessary or not (Yes/No). Note that patient numbers 4, 5 and 6 are described in the supplementary Fig. 3, 4, and 5.

proteome was analysed again eight and eleven months after cystectomy, as shown in Fig. 4d timepoints 3 and 4, respectively. The expression levels of all the pathways were found to be lower after resection of the bladder (Fig. 4d). The dPPI decreased 63 and 79% between onset and time points three and four, respectively. The latter is consistent with the recovery experienced by this patient. Currently, this patient is in good health. The profile of the four proteomes since the onset of disease were compared (Fig. 4f) and clearly show that the urines were clustered into two groups (before and after radical cystectomy).

Three further patient examples are presented in the Supplementary Fig. 3, 4 and 5.

Table 2 shows the dPPI index for each time point for each patient in a comprehensive manner. First, the urine was monitored, the dPPI was calculated, and an alert for intervention was set as yes or no. Next, the patient was examined by the physician's team through a medical procedure as described in the table. The diagnostic provided was then compared with the alert provided by the dPPI. As shown in Table 2, 100% of the alerts (yes or no) were accurate.

In this work, we demonstrated how using the dPPI algorithm to assess the urine proteome is a robust tool to (i) monitor the course of BC disease and (ii) guide clinical decisions regarding intervention. As the disease progresses, the levels of selected

cancer hallmark pathways change. This variation is used to monitor the course of BC disease in a simple, straightforward, and robust manner. This study would benefit from a large cohort of patients and will need to be replicated on a larger scale to validate the reported conclusions.

Data availability

The mass spectrometry proteomics data that support the findings of this study have been deposited in ProteomeXchange Consortium¹⁶ via the PRIDE¹⁷ partner with the PXD025139 accession codes. Source data for the Figs. 1–4 and Supplementary Fig. 2–5 presented in the manuscript are available in the Supplementary Data 2.

Received: 23 April 2021; Accepted: 6 January 2023;

Published online: 16 January 2023

References

- Kamat, A. M. et al. Bladder cancer. *Lancet* **388**, 2796–2810 (2016).
- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).

ARTICLE

COMMUNICATIONS MEDICINE | <https://doi.org/10.1038/s43856-023-00238-4>

3. Witjes, J. A. et al. European Association of Urology guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2020 guidelines. *Eur. Urol.* **79**, 82–104 (2021).
4. Planz, B. et al. The role of urinary cytology for detection of bladder cancer. *Eur. J. Surg. Oncol.* **31**, 304–308 (2005).
5. Carvalho, L. B., Capelo-Martínez, J. L., Lodeiro, C., Wiśniewski, J. R. & Santos, H. M. Ultrasonic-based filter aided sample preparation as the general method to sample preparation in proteomics. *Anal. Chem.* **92**, 9164–9171 (2020).
6. Ahadi, S. et al. Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nat. Med.* **26**, 83–90 (2020).
7. Piening, B. D. et al. Integrative personal omics profiles during periods of weight gain and loss. *Cell Syst.* **6**, 157–170.e8 (2018).
8. Schüssler-Fiorenza Rose, S. M. et al. A longitudinal big data approach for precision health. *Nat. Med.* **25**, 792–804 (2019). 2019.255.
9. Nguyen, V. A. et al. Longitudinal stroke recovery associated with dysregulation of complement system—a proteomics pathway analysis. *Front. Neurol.* **11**, 692 (2020).
10. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
11. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
12. Thongboonkerd, V. & Putita, S. Bacterial overgrowth affects urinary proteome analysis: recommendation for centrifugation, temperature, duration, and the use of preservatives during sample collection. *J. Proteome Res.* **6**, 4173–4181 (2007).
13. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
14. Wiśniewski, J. R. Filter aided sample preparation—a tutorial. *Anal. Chim. Acta* **1090**, 23–30 (2019).
15. Carvalho, L. B., Capelo-Martínez, J. L., Lodeiro, C., Wiśniewski, J. R. & Santos, H. M. Snap-heated freeze-free preservation and processing of the urine proteome using the combination of stabilizer-based technology and filter aided sample preparation. *Anal. Chim. Acta* **1076**, 82–90 (2019).
16. Deutsch, E. W. et al. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **48**, D1145–D1152 (2020).
17. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).
18. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* **13**, 2513–2526 (2014).
19. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
20. Tyanova, S. et al. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453–1456 (2015).
21. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
22. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (pro)teomic data. *Nat. Methods* **13**, 731–740 (2016).
23. Minami, S. et al. Proteomic study of sera from patients with bladder cancer: usefulness of S100A8 and S100A9 proteins. *Cancer Genom. Proteom.* **7**, 181–189 (2010).
24. Liu, P. F. et al. Far from resolved: stromal cell-based iTRAQ research of muscle-invasive bladder cancer regarding heterogeneity. *Oncol. Rep.* **32**, 1489–1496 (2014).
25. Chen, Y. T. et al. Discovery of novel bladder cancer biomarkers by comparative urine proteomics using iTRAQ technology. *J. Proteome Res.* **9**, 5803–5815 (2010).
26. Lindén, M. et al. Proteomic analysis of urinary biomarker candidates for nonmuscle invasive bladder cancer. *Proteomics* **12**, 135–144 (2012).
27. London, S. J., Lehman, T. A. & Taylor, J. A. Myeloperoxidase genetic polymorphism and lung cancer risk. *Cancer Res.* **57**, 5001–5003 (1997).
28. Castillo-Tong, D. C. et al. Association of myeloperoxidase with ovarian cancer. *Tumor Biol.* **35**, 141–148 (2014).
29. Soukup, V. et al. The significance of calprotectin, CD147, APOA4 and DJ-1 in non-invasive detection of urinary bladder carcinoma. *Neoplasma* **66**, 1019–1023 (2019).
30. Gurhan Balci, M. & Tayfur, M. Loss of E-cadherin expression in recurrent non-invasive urothelial carcinoma of the bladder. *Int. J. Clin. Exp. Pathol.* **11**, 4163–4168 (2018).
31. Pečina-Slaus, N. Tumor suppressor gene E-cadherin and its role in normal and malignant cells. *Cancer Cell Int.* **3**, 17 (2003).
32. Hatipoglu, S. et al. Urinary MMP-9/NGAL complex in children with acute cystitis. *Pediatr. Nephrol.* **26**, 1263–1268 (2011).
33. Kumar, B. et al. Differential effects of MAPKs signaling on the growth of invasive bladder cancer cells. *Int. J. Oncol.* **34**, 1557–1564 (2009).
34. Stramucci, L., Pranteda, A. & Bossi, G. Insights of crosstalk between p53 protein and the MKK3/MKK6/p38 MAPK Signaling pathway in cancer. *Cancers* **10**, 131 (2018).
35. Yoshida, T. et al. Prognostic impact of renin-angiotensin inhibitors in patients with bladder cancer undergoing radical cystectomy. *Ann. Surg. Oncol.* **24**, 823–831 (2017).
36. Ren, L. et al. Apolipoproteins and cancer. *Cancer Med.* **8**, 7032–7043 (2019).
37. Zeng, F. C., Cen, S., Tang, Z. Y. & Kang, X. L. Elevated matrix metalloproteinase-9 expression may contribute to the pathogenesis of bladder cancer. *Oncol. Lett.* **11**, 2213–2222 (2016).
38. Li, F. et al. The increased excretion of urinary orosomucoid 1 as a useful biomarker for bladder cancer. *Am. J. Cancer Res.* **6**, 331–340 (2016).

Acknowledgements

PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). Authors acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV which is financed by national funds from FCT/MCTES, *Fundação para a Ciência e a Tecnologia* and *Ministério da Ciência, Tecnologia e Ensino Superior*, through the projects UIDB/50006/2020 and UIDP/50006/2020. H.M.S. acknowledges the Associate Laboratory for Green Chemistry-LAQV (LA/P/0008/2020) funded by FCT/MCTES for his research contract. L.B.C. is funded by the FCT/MCTES PhD grant 2019 (SFRH/BD/144222/2019). G.M. is funded by the FCT/MCTES PhD grant 2018 (SFRH/BD/139384/2018). H.López-Fernández is supported by a 'Maria Zambrano' post-doctoral contract from Ministerio de Universidades (Gobierno de España).

Author contributions

J.L.C., C.L., L.C.P., and H.M.S. designed the experimental work and provide financial support. L.B.C. and G.M. performed the laboratory work and data analysis under the supervision of H.M.S. and J.L.C. H.M.S. and J.L.C. drafted the manuscript. L.B.C., J.L.C., C.L., L.C.P., R.D., M.M. and H.M.S. revisited the drafted version, corrected it, and made valuable suggestions. H.L.F. and F.D. performed the statistical analysis. L.C.P. and M.M. managed patient interventions and provided samples and medical data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-023-00238-4>.

Correspondence and requests for materials should be addressed to Hugo Miguel Santos.

Peer review information *Communications Medicine* thanks Christiana Kartsonaki, Hideki Furuya, Joshua Meeks and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

| F

**PUBLISHED PAPER IN CLINICAL PROTEOMICS
MEDICINE - CORRESPONDING TO CHAPTER 8**

RESEARCH

Open Access



Proteomic analysis of chromophobe renal cell carcinoma and benign renal oncocytoma biopsies reveals shared metabolic dysregulation

Luis B. Carvalho^{1,2†}, Susana Jorge^{1,2†}, Hugo López-Fernández^{3,4}, Carlos Lodeiro^{1,2}, Rajiv Dhir⁵, Luis Campos Pinheiro^{6,7}, Mariana Medeiros^{6,7}, Hugo M. Santos^{1,2,5} and José L. Capelo^{1,2*}

Abstract

Background This study investigates the proteomic landscapes of chromophobe renal cell carcinoma (chRCC) and renal oncocytomas (RO), two subtypes of renal cell carcinoma that together account for approximately 10% of all renal tumors. Despite their histological similarities and shared origins, chRCC is a malignant tumor necessitating aggressive intervention, while RO, a benign growth, is often subject to overtreatment due to difficulties in accurate differentiation.

Methods We conducted a label-free quantitative proteomic analysis on solid biopsies of chRCC (n=5), RO (n=5), and normal adjacent tissue (NAT, n=5). The quantitative analysis was carried out by comparing protein abundances between tumor and NAT specimens. Our analysis identified a total of 1610 proteins across all samples, with 1379 (85.7%) of these proteins quantified in at least seven out of ten LC–MS/MS runs for one renal tissue type (chRCC, RO, or NAT).

Results Our findings revealed significant similarities in the dysregulation of key metabolic pathways, including carbohydrate, lipid, and amino acid metabolism, in both chRCC and RO. Compared to NAT, both chRCC and RO showed a marked downregulation in gluconeogenesis proteins, but a significant upregulation of proteins integral to the citrate cycle. Interestingly, we observed a distinct divergence in the oxidative phosphorylation pathway, with RO showing a significant increase in the number and degree of alterations in proteins, surpassing that observed in chRCC.

Conclusions This study underscores the value of integrating high-resolution mass spectrometry protein quantification to effectively characterize and differentiate the proteomic landscapes of solid tumor biopsies diagnosed as chRCC and RO. The insights gained from this research offer valuable information for enhancing our understanding of these conditions and may aid in the development of improved diagnostic and therapeutic strategies.

Keywords OCT-embedded tissues, Total Protein Approach, Label-free Quantification, Mass Spectrometry, Chromophobe Renal Cell Carcinoma, Renal Oncocytoma

[†]Luis B. Carvalho and Susana Jorge have equally worked on this manuscript.

*Correspondence:

José L. Capelo

jlc@fct.unl.pt

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Renal cell carcinoma (RCC), a disease characterized by abnormal cell proliferation in the epithelial cells of the kidney, accounts for 400,000 new cases of adult kidney cancer globally each year [1]. There are several subtypes within the RCC classification, each displaying unique histological and cytological phenotypes [2]. Chromophobe renal cell carcinoma (chRCC) and renal oncocytomas (RO) are two RCC subtypes that represent approximately 10% of all renal tumors [3]. RO, a benign growth often found in the kidney collecting ducts, typically lends itself to conservative treatment approaches. Conversely, chRCC is a malignant tumor that, despite sharing a similar origin in distal kidney nephrons, requires more aggressive intervention to prevent dire consequences to the patient [3]. Given the strikingly similar anatomical origins and histological features of these two tumor types, their accurate differentiation poses a significant challenge [4]. Therefore, most patients undergo surgery, often involving a partial or total nephrectomy. This can lead to overtreatment in RO, as benign lesions may not require such invasive interventions if correctly identified beforehand. Thus, accurately distinguishing between these two conditions is essential not only to guide optimal treatment strategies but also to reduce the potential burden on patients and the healthcare system, thereby improving overall patient care [4, 5].

Building on our recent development of a highly efficient, ultrasonic-based methodology for extracting, identifying, and quantifying the proteome of solid renal biopsies embedded in optimal cutting temperature (OCT) compound [6], we sought further to leverage these technological advancements in our current study. In this context, we employed the total protein approach (TPA), a computational tool adept at converting spectral intensities into protein concentrations. The efficacy of the TPA methodology has been proven in various scenarios, such as evaluating metabolic pathways in slow and fast skeletal muscle [7], exploring biological processes associated with colorectal cancer [8], and assessing the impact on mitochondrial bioenergetics in disease-induced conditions [9]. Additionally, we previously investigated differentially expressed proteins in renal cancers using high-resolution MS combined with the TPA method [10]. In this work, we focused on the metabolic features associated with chRCC and RO versus NAT using the TPA method to identify undisclosed new features characterizing each tumor [11]. We compared our findings with the literature addressing the metabolic characteristics of these renal neoplasms [12, 13].

Materials and methods

Study design and sampling

In this manuscript, we have conscientiously reused a portion of the raw data from our previous by Jorge et al. [11]. The raw data in question forms the basis of a comprehensive dataset that was collected with significant effort and resources. This dataset is rich and multifaceted, allowing for various analyses and interpretations. While our previous publication focused on the identification of novel histochemical markers in renal neoplasm, the current manuscript focuses on the discovery of shared metabolic features between chromophobe renal cell carcinoma and renal oncocytoma. We believe that the use of this data, in this manner, is a responsible use of resources, avoiding unnecessary duplication of data collection efforts. These data comprises the use of high-resolution mass spectrometry to examine the proteomes of 15 human renal tissue biopsies that were flash-frozen and embedded in OCT. These included chromophobe renal cell carcinoma (chRCC, $n=5$), renal oncocytoma (RO, $n=5$) and normal adjacent renal tissue (NAT, $n=5$). The University of Pittsburgh Biospecimen Core provided the human kidney tissue samples used in this study, which received approval from the University of Pittsburgh's Institutional Review Board (IRB # 02-077). All neoplasms in the study consisted of at least 85% tumor cells. Additional file 1: Table S1 summarizes the patient data involved in this research.

Proteomic analysis

The processing of biopsies followed the methodology outlined in Jorge et al. [6]. Briefly, tissues were first cleaned of OCT using an ultrasonic bath model TI-H-5 from Elma (Singen, Germany). Then, proteins were extracted in 8 M urea prepared in 25 mM ammonium bicarbonate using an ultrasonic probe (UP50H from Hielscher Ultrasonics, Teltow, Germany). Next, protein digestion was carried out with trypsin for four minutes using an ultrasonic microplate horn assembly device (QSonica, Newtown, CT, USA). The digested proteome extracts were subsequently analyzed by a label-free nanoLC-HR-MS/MS approach (UHR-QqTOF IMPACT HD from Bruker Daltonics, Bremen, Germany) as described before by Jorge et al. [11]. A step-by-step protocol, including OCT Cleaning, proteome extraction, proteome digestion, and LC-MS/MS data acquisition is available in Additional file 2.

Data analysis and statistics

The MaxQuant software V1.6.0.16 was utilized for relative label-free quantification. All raw files were processed in a single run using default settings [14, 15]. The

Andromeda search engine was employed for database searches, using the UniProt-SwissProt Human database as a reference, along with a database of common contaminants.

For absolute protein quantification, the Total Protein Approach (TPA) was used to examine raw spectral intensities from the MaxQuant output [10, 11]. The calculation of protein concentration was performed as follows:

$$c(i) = \frac{MS\ signal\ (i)}{total\ MS\ signal \times MW\ (i)} \left[\frac{mol}{g\ total\ protein} \right]$$

Data processing was performed using Perseus (version 1.6.15.0) with default settings [16, 17]. In brief, reverse hits and proteins only identified by site were removed from the protein list, and TPA-based concentrations were log₂-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 values. Missing values were imputed from the total matrix with width=0.5 and downshift=1.8. PCA was performed on the filtered and imputed data. Log ratios were calculated as the difference in average log₂ values between the two conditions tested in volcano plots (two-tailed Student's t test, FDR=0.01 and S0=0.1). Protein network analysis was integrated and visualized using the software platform Cytoscape V3.8.2 with the application StringApp V1.6.0 [18]. KEGG [19], Reactome [20] and GO terms [21] were used as ontology databases.

Results

Renal tumor profiling

The tissue biopsies were analyzed in duplicate using liquid chromatography coupled to tandem mass spectrometry (nanoLC-ESI-HR-MS/MS). Raw spectral intensity values were converted into absolute concentrations using the TPA method [10]. A total of 1610 proteins were identified (1% FDR) across all samples, with 1379 (85.7%) of these proteins being quantified in at least seven out of the ten LC-MS/MS runs for one renal tissue type (chRCC, RO or NAT). Figure 1a displays the principal component

analysis (PCA) of the quantified proteins, showing that each specimen class segregates into distinct clusters.

The reproducibility of the MS biological replicates was verified by comparing the Pearson correlation between all pairs of samples as well as visualizing pairwise scatter plots in a matrix (Fig. 1b). The calculated Pearson correlation coefficients spanned from 0.79 to 0.97 for chRCC, from 0.76 to 0.95 for RO, and from 0.83 to 0.95 for NAT. The Pearson coefficients were higher when tumors were compared to each other (0.68–0.86) than when compared to NAT (0.56–0.70). These results underline a marked deviation of both chRCC and RO from NAT. Furthermore, the data also demonstrate that there are discernible differences between chRCC and RO, highlighting the potential to differentiate these two neoplasms using this profiling approach.

Dysregulated proteins between tumors and NATs

A two-tailed Student's t test (FDR=0.01 and S0=0.1) was used to distinguish the abundance of proteins between each tumor and NAT. In chRCC, 532 proteins exhibited significant differences compared to NAT, with 260 upregulated and 272 downregulated proteins (Fig. 1c). Similarly, in the case of RO versus NAT, 578 proteins showed significant differences, with 226 upregulated and 352 downregulated proteins (Fig. 1d). Comparing the upregulated and downregulated proteins between each tumor and NAT, 241 downregulated proteins (63%) and 117 upregulated proteins (32%) were found to be common (Fig. 1e, f).

To elucidate the biochemical processes driving the phenotypes of chRCC and RO, the differentially expressed proteins for each tissue type were used to interrogate comprehensive, functional proteomic databases. We found a common signature of metabolic dysregulation for chRCC and RO compared to NAT (Fig. 2a), affecting carbohydrate metabolism, lipid metabolism, amino acid metabolism and oxidative phosphorylation. Among KEGG protein pathways linked to carbohydrate metabolism, the most dysregulated were glycolysis/gluconeogenesis, pyruvate metabolism and the citric acid cycle (FDR<8.94×10⁻¹⁴). In our analysis, we identified up to

(See figure on next page.)

Fig. 1 Classification of human renal tissue proteomes and protein expression profiles in tumors relative to NATs. 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 1379 were obtained from at least one tissue group (chRCC, RO, or NAT), with a reproducibility ranging from 70 to 100%. **a** Principal component analysis (PCA) of chRCC, RO, and NAT group samples. **b** Heatmap representation of Pearson correlation of biological (n=5) and technical (n=2, each sample) replicates (lower left) in combination with a scatterplot matrix protein of TPA-based quantification (upper right) demonstrating the reproducibility of the MS data. Volcano plots illustrate the significantly different (FDR=0.01, S0=0.1) protein expression levels between **(c)** chRCC and NAT samples and **(d)** RO and NAT samples. Venn diagrams comparing proteomes of chRCC and RO in terms of the **(e)** downregulated proteins and **(f)** upregulated proteins

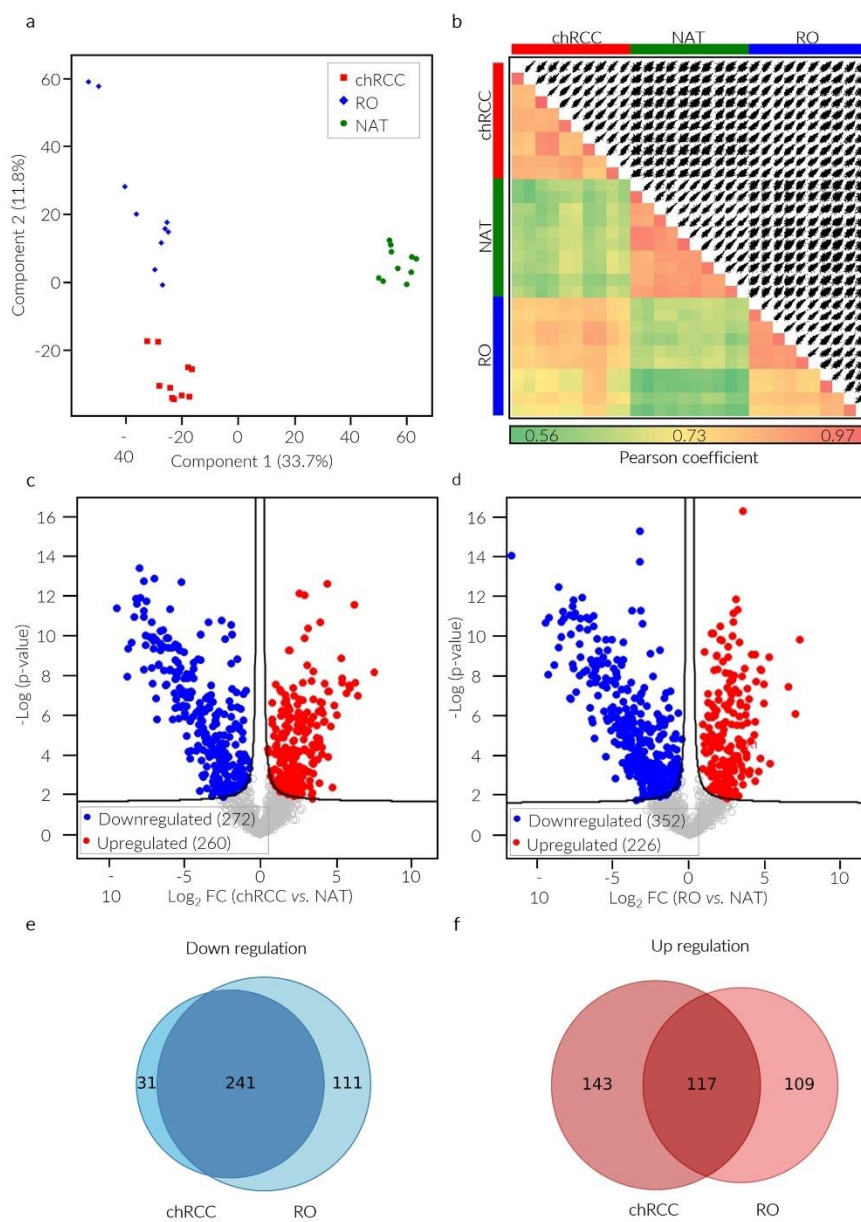


Fig. 1 (See legend on previous page.)

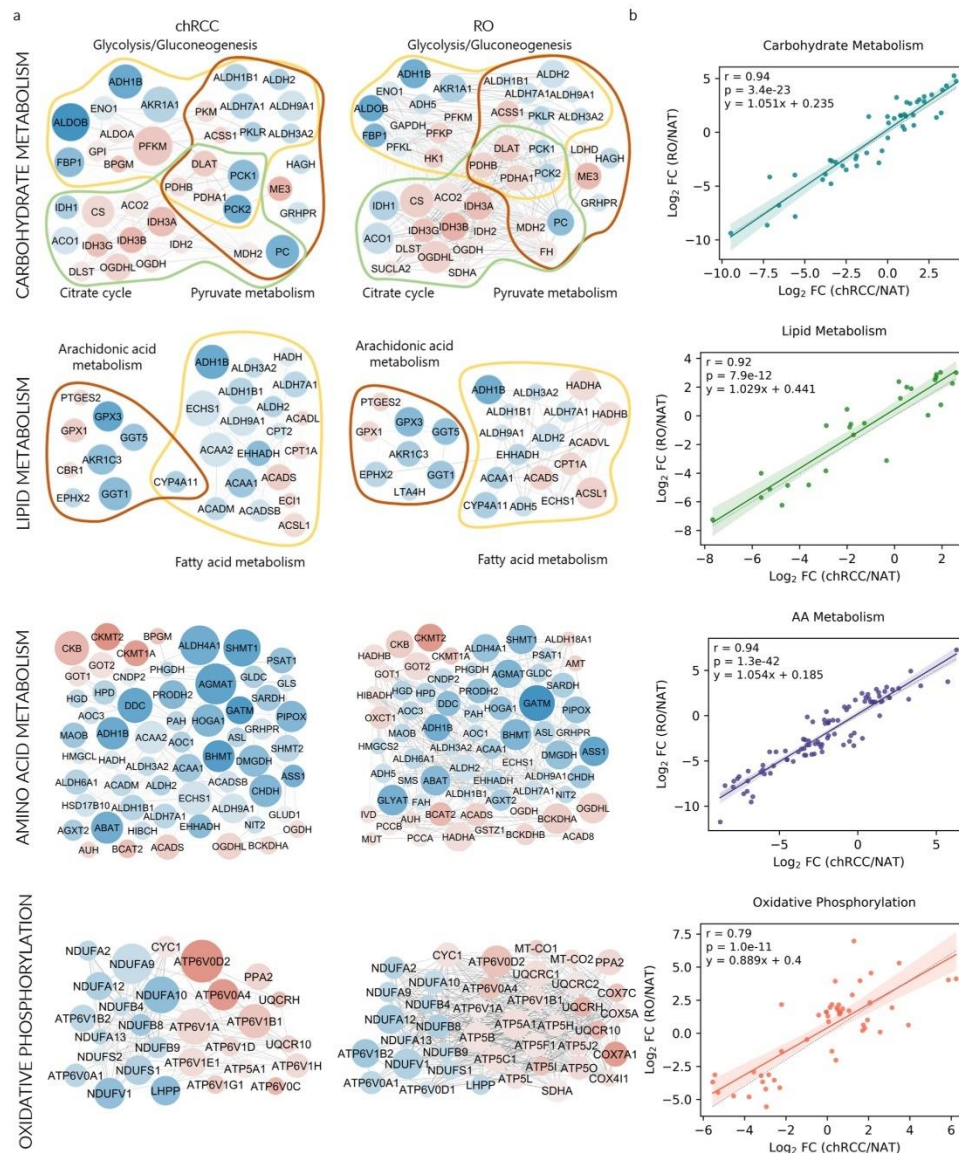


Fig. 2 Metabolic pathway proteomes similarly affected in chRCC and RO tumors. **a** Network representation of a subset of pathways in four areas of metabolism in which proteins were differentially regulated in tumors. Significantly ($FDR < 0.01$) upregulated and downregulated proteins are shown in red and blue, respectively. Protein pathway analysis was performed by searching the KEGG database [19] using differentially expressed proteins between tumor and NAT biopsies. **b** Correlation between levels of dysregulation of individual proteins ($\log_2$ fold change (FC) in abundance in tumor vs NAT) in chRCC and RO for proteins in carbohydrate ($n = 47$), lipid ($n = 27$) and amino acid ($n = 91$) metabolism and oxidative phosphorylation ($n = 47$). The correlation coefficient r was calculated using the Pearson test. Shading areas represent the confidence of the interval, and p represents the p value of the test

23 proteins implicated in these pathways that exhibited dysregulation across both tumor subtypes.

The KEGG glycolysis/gluconeogenesis pathway includes processes involved in the degradation of glucose into pyruvate and the generation of glucose from non-carbohydrates. Based on the differential protein expression data, proteins involved in gluconeogenesis, such as fructose-1,6-bisphosphatase 1 (FBP1), pyruvate carboxylase (PC), and phosphoenolpyruvate carboxykinase 1 and 2 (PCK1 and PCK2), were significantly downregulated in tumors. In contrast, glycolytic proteins remained unchanged or upregulated, e.g., ATP-dependent 6-phosphofructokinase (PFKM). These data indicate that gluconeogenesis was downregulated in tumors compared to NATs. For instance, the results of FBP1 match well with its well-known inhibitory effects on glycolysis and tumor growth [22]. Furthermore, the downregulation of PC, which catalyzes the conversion of pyruvate to oxaloacetate in the first step of gluconeogenesis, suggests a shift in cellular metabolism toward the conversion of pyruvate into acetyl-CoA, as evidenced by the observed upregulation of proteins such as pyruvate dehydrogenase E1 component subunit alpha and beta (PDHA1 and PDHB) and the dihydrolipoyl lysine-residue acetyltransferase component of the pyruvate dehydrogenase complex (DLAT). Several TCA cycle proteins were upregulated, including citrate synthase (CS), aconitate hydratase (ACO2), isocitrate dehydrogenase [NADP] and [NAD] subunits alpha and beta (IDH2, IDH3A and IDH3B), dihydrolipoyl lysine-residue succinyltransferase component of the 2-oxoglutarate dehydrogenase complex (DLST), 2-oxoglutarate dehydrogenase (OGDH) and malate dehydrogenase (MDH2). The expression of proteins active in lipid metabolism was modified in both tumors, with most being downregulated relative to NAT controls. In the chRCC and RO samples, 91 downregulated proteins were linked to amino acid (AA) metabolism.

Pearson correlations were calculated for chRCC and RO based on the fold change in abundance of differentially expressed proteins in carbohydrate, lipid and AA metabolism pathways (Fig. 2b). Positive correlations were found for the four pathways compared, with the oxidative

phosphorylation pathway showing the highest p and lowest Pearson correlation values.

Main chRCC and RO features

The functional protein pathway analysis also revealed divergent features between chRCC and RO that may underlie differences in the tumor subtypes. Proteins involved in energy metabolism of chRCC were less dysregulated than RO. For instance, in chRCC, 28 proteins involved in the oxidative phosphorylation pathway were expressed at different levels compared to NAT, while in RO, 40 proteins belonging to this pathway were dysregulated (Fig. 2a). Comparing the variation between the proteins belonging to the oxidative phosphorylation pathway in both tumors versus NAT, a Pearson coefficient of 0.79 was obtained. The same comparison for carbohydrate, lipid, and AA metabolism resulted in Pearson coefficients of 0.94, 0.92 and 0.94, respectively (Fig. 2b).

Mitochondrial KEGG pathways were dysregulated more extensively in RO biopsies, with 256 differentially expressed proteins (Fig. 3a) compared to 175 in chRCC (Additional file 3: Table S2).

The common chRCC versus NAT and RO versus NAT dysregulations of proteins related to respiratory chain complexes were evaluated and compared. The results depicted in Fig. 3b present Pearson correlation coefficients of 0.81, 0.79, 0.93, 0.57, and 0.40 for complexes I, II, III, IV and V, respectively. The TPA-based concentration values for the respiratory chain complex proteins are presented in Fig. 3c, Additional file 4: Figs. S1, 2.

Discussion

In our research, we utilized the total protein approach to quantify the absolute protein amounts in chRCC, RO, and NAT tissues. These data enabled us to examine the unique protein expression patterns in chRCC and RO compared to NAT and to delve into potential biological implications. In this context, we believed that the most informative approach was to assess and compare the deviations from the normal state in both chRCC and RO. By contrasting these differences, we aimed to elucidate the unique dysregulation patterns associated with

(See figure on next page.)

Fig. 3 Differences in protein pathways underlying each tumor subtype. **a** Specific dysregulation of mitochondrial protein pathways in ROs, including those of respiratory chain complexes. Protein pathway analysis was performed by using the differentially expressed proteins in each tumor biopsy (relative to NAT) in searches against KEGG [19] and GO [21] databases. Significantly (FDR < 0.01) upregulated and downregulated proteins are shown in red and blue, respectively. **b** Comparison of the protein deregulation ($\log_2$ fold change (FC) in abundance in tumor relative to NAT) between chRCC and RO for the five respiratory chain complexes [Complex I, n = 44; Complex II, n = 4; Complex III, n = 10; Complex IV, n = 19; Complex V, n = 20]. Blue dots, downregulated proteins (RO vs NAT); red dots, upregulated proteins (RO vs NAT). The correlation coefficient r was calculated using the Pearson test. Shading areas represent the confidence of the interval, and p represents the p value of the test. **c** Absolute protein amount, expressed as pmol/mg of tissue, calculated through the TPA methodology

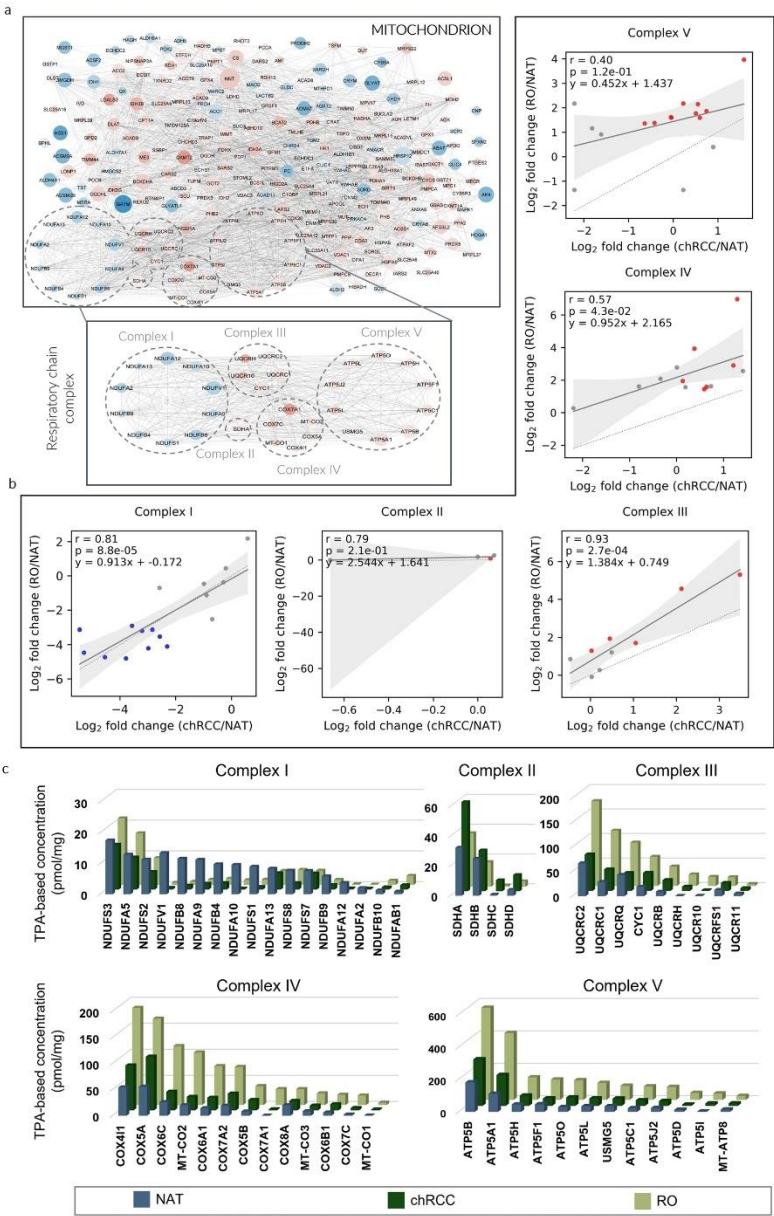


Fig. 3 (See legend on previous page.)

each malignancy, shedding light on their molecular characteristics.

PCA of the variance in protein abundance showed that the proteomes of the two tumor subtypes exhibited greater similarity to each other than to the proteomes of NAT samples. This similarity suggests a shared tumor phenotype between the two subtypes. However, PCA also highlighted discernible differences in the quantified proteomes of the two tumor subtypes, indicating unique characteristics associated with each subtype (Fig. 1a).

Comparing dysregulated pathways between chRCC and NAT with those between RO and NAT, as depicted in Fig. 2, reveals a similarity in dysregulation levels. This is evidenced by a Pearson coefficient of 0.92 or higher for carbohydrate metabolism, lipid metabolism, and amino acid metabolism, which strongly suggests parallel patterns of dysregulation in both chRCC and RO subtypes when compared to normal tissue (Fig. 2b). This observation aligns with the well-established hallmark of cancer, namely, the reprogramming of central metabolic pathways. It also corroborates the findings reported by Xiao et al. [13] who similarly identified decreased levels of proteins involved in gluconeogenesis and fatty acid metabolism in chRCC compared to NAT. Additionally, our data suggest that gluconeogenesis and lipid metabolism are also reduced in RO compared to NAT, which aligns with the findings reported by Kürschner et al. [23].

This collective evidence sheds light on the shared metabolic traits between chRCC and RO, thus confirming common signatures of metabolic dysregulation for chRCC and RO compared to NAT. The evident suppression of amino acid synthesis in both cancer types further highlights the possibility of amino acid auxotrophy as a defining feature of these conditions. The downregulation of argininosuccinate synthetase (ASS1), phosphoglycerate dehydrogenase (PHGDH), and glutamic-oxaloacetic transaminase 1 and 2 (GOT1 and GOT2) in both chRCC and RO tumors compared with NAT confirms this result [24]. Inhibition of this protein leads to decreased production of important amino acids such as arginine, serine and glycine, forcing cancer cells to depend on healthy tissue surrounding the tumor as a source of nutrients [24].

Additionally, the analysis of the oxidative phosphorylation pathway revealed a Pearson coefficient of 0.74 (Fig. 2b), suggesting the presence of distinct mechanisms of oxidative phosphorylation between the chRCC and RO subtypes. Notably, our findings align with previous studies, demonstrating significant alterations in mitochondrial pathways in both RO and chRCC when compared to NAT [25, 26]. It is interesting to note that the levels of dysregulated mitochondrial-related proteins were higher in RO than in chRCC (Fig. 3). Our findings agree with those reported by Joshi et al. [25], who suggest that mitochondrial activity acts as

a barrier to tumorigenesis in RO, which means that normal mitochondrial activity can prevent or slow down the development of tumors. Reduced mitochondrial activity may contribute to the development of tumors by promoting a protumorigenic environment in cells. This is because normal mitochondrial activity is necessary to maintain healthy cellular metabolism, and altered mitochondrial activity in cancer cells can affect cellular metabolism and promote tumor growth. This suggests that targeting mitochondrial metabolism may be explored as a coadjutant therapeutic strategy for treating chRCC.

In addition to mitochondrial abnormalities, mutations in respiratory chain complex I genes are frequently observed in RO [23]. Our RO versus NAT proteomic data revealed downregulation of complex I-related proteins and predominant upregulation of proteins associated with complexes III, IV and V, as shown in Fig. 3b, c, Additional file 4: Figs. S1, 2. Our data agree with the work of Xiao et al. [12], who also reported reduced expression of complex I-related proteins and an increase in all other proteins related to respiratory chain complexes. In this way, the direct difference in the absolute levels of expression of these mitochondrial proteins demonstrates the utility of the TPA in discriminating between these two tumor subtypes.

Conclusions

In summary, this work presents an in-depth exploration of the contrasting protein expression profiles found in chRCC and RO compared to their normal adjacent tissue. The study used the TPA-based method to determine the absolute amounts of proteins in each tissue type. The findings suggest that both chRCC and RO show dysregulation of carbohydrate, lipid, and amino acid metabolism, as well as mitochondrial abnormalities, when compared to NAT. The study found that the proteomes of the two tumor subtypes were more similar to each other than to NAT, reflecting their shared tumor phenotype. The observed decrease in argininosuccinate synthetase (ASS1) expression in both chRCC and RO tumors relative to adjacent healthy tissue indicates shared metabolic characteristics between these cancers. One such shared trait appears to be auxotrophy, which is prominently featured in both conditions. Targeting mitochondrial metabolism may be explored as a coadjutant therapeutic strategy for treating both chRCC and RO. The reduced number of samples per group used in this study must be considered in future studies.

Abbreviations

chRCC
LFQ

Chromophobe renal cell carcinoma
Label-free quantification

nanoLC-HR-MS/MS	Nanoliquid chromatography-high resolution-tandem mass spectrometry
NAT	Normal adjacent tissue
OCT	Optimum cutting temperature
PCA	Principal component analysis
RCC	Renal cell carcinoma
RO	Benign renal oncocytoma

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12014-023-09443-8>.

Additional file 1: Table S1. Description of human kidney biopsies used in the study.

Additional file 2: Step-by-step protocol, that includes OCT cleaning, proteome extraction, proteome digestion, and LC-MS/MS data acquisition.

Additional file 3: Table S2. Protein network analysis.

Additional file 4: Figure S1. TPA concentration values (pmol/mg of tissue) of respiratory chain complex I (CI), II (CII) and III (CIII) proteins. Statistical analysis was performed using pairwise Mann Whitney test (** $p \leq 0.001$; *** $p \leq 0.0001$; **** $p \leq 0.00001$). **Figure S2.** TPA concentration values (pmol/mg of tissue) of respiratory chain complex IV (CIV) and V (CV) proteins. Statistical analysis was performed using pairwise Mann Whitney test (** $p \leq 0.001$; *** $p \leq 0.0001$; **** $p \leq 0.00001$).

Acknowledgements

The PROTEOMASS Scientific Society (Portugal) is acknowledged for the funding provided through the General Funding Grant 2022–2023 and by the funding provided to the Laboratory for Biological Mass Spectrometry Isabel Moura (#PM001/2019 and #PM003/2016). The authors also acknowledge the funding provided by the Associate Laboratory for Green Chemistry LAQV, which is financed by national funds from *Fundação para a Ciência e Tecnologia and Ministério da Ciência, Tecnologia e Ensino Superior* (FCT/MCTES) through the projects UIDB/50006/2020 and UIDP/50006/2020. HMS acknowledges the Associate Laboratory for Green Chemistry LAQV (LA/P/0008/2020) funded by FCT/MCTES for his research contract. LBC and SJ thank FCT/MCTES for their PhD studentships SFRH/BD/144222/2019 and SFRH/BD/120537/2016, respectively. HLF is supported by a 'María Zambrano' postdoctoral contract from *Ministerio de Universidades (Gobierno de España)*. This project benefited from the University of Pittsburgh Hillman Cancer Center shared resource facility (Cancer Genomics Facility) supported in part by award P30CA047904.

Author contributions

RD, SS, DK, SB and GQG collected and organized all human tissue biopsies. JLC, HMS, CLE, and RD designed the experimental work and provided financial support. SJ performed the laboratory work under the supervision of HMS and JLC. LBC finished the data handling and contributed to the figure making and contributed significantly to the final manuscript version. SJ, LBC and HLF performed the bioinformatic analysis. SJ, LBC and JLC drafted the manuscript, and HMS, CLE, and RD revised and corrected the draft. All authors have approved the final version of the manuscript.

Funding

This work received financial support from PT national funds (FCT/MCTES) through the projects LA/P/0008/2020, UIDB/50006/2020 and UIDP/50006/2020. This work received support from the PROTEOMASS Scientific Society through the General Funding Grant 2022–2023 and the projects #PM001/2019 and #PM003/2016.

Availability of data and materials

The mass spectrometry proteomics raw data have been deposited in the ProteomeXchange Consortium via the PRIDE [27] partner repository with the dataset identifier PXD022018.

Declarations

Ethics approval and consent to participate

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).

Competing interests

The authors declare no potential conflicts of interest.

Author details

¹BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, FCT NOVA, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal. ²PROTEOMASS Scientific Society, Departmental Building, FCT-NOVA, Caparica Campus, 2829-516 Caparica, Portugal. ³CINBIO, Department of Computer Science, ESEI-Escuela Superior de Ingeniería Informática, Universidade de Vigo, 32004 Ourense, Spain. ⁴SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, 36213 Vigo, Spain. ⁵Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA. ⁶Urology Department, Central Lisbon Hospital Center, Lisbon, Portugal. ⁷NOVA Medical School, NOVA University of Lisbon, Lisbon, Portugal.

Received: 10 August 2023 Accepted: 7 November 2023

Published online: 28 November 2023

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 (2020).
2. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO classification of tumors of the urinary system and male genital organs—part a: renal, penile, and testicular tumors. *Eur Urol*. 2016;70:93–105.
3. Cairns P. Renal cell carcinoma. *Cancer Biomark*. 2011;9:461–73.
4. Jungberg B, et al. European association of urology guidelines on renal cell carcinoma: the 2019 update. *Eur Urol*. 2019;75:799–810.
5. Wobker SE, Williamson SR. Modern pathologic diagnosis of renal oncocytoma. *J Kidney Cancer VHL*. 2017;4:1–12.
6. 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;18:2979–86.
7. Rakus D, Gizak A, Deshmukh A, Winiewski JR. Absolute quantitative profiling of the key metabolic pathways in slow and fast skeletal muscle. *J Proteome Res*. 2015;14:1400–11.
8. 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;14:4005–18.
9. Walheim E, Wiśniewski JR, Jastroch M. Respiromics – An integrative analysis linking mitochondrial bioenergetics to molecular signatures. *Mol Metab*. 2018;9:4–14.
10. Wiśniewski JR. Label-free and standard-free absolute quantitative proteomics using the “total protein” and “proteomic ruler” approaches. In *Methods in enzymology* 2017 Jan 1 (Vol. 585, pp. 49–60). Academic Press.
11. Jorge S, et al. Absolute quantitative proteomics using the total protein approach to identify novel clinical immunohistochemical markers in renal neoplasms. *BMC Med*. 2021;19:1–13.
12. Xiao Y, et al. Decreased mitochondrial DNA content drives OXPHOS dysregulation in chromophobe renal cell carcinoma. *Cancer Res*. 2020;80:3830–40.
13. Xiao Y, et al. Endocytosis-mediated replenishment of amino acids favors cancer cell proliferation and survival in chromophobe renal cell carcinoma. *Cancer Res*. 2020;80:5491–501.
14. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26:1367–72.

15. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc.* 2016;11:2301–19.
16. Tyanova S, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods.* 2016;13:731–40.
17. Tyanova S, Cox J. Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. *Methods Mol Biol.* 2018;1711:133–48.
18. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. *J Proteome Res.* 2019;18:623–32.
19. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40:D109–14.
20. Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2018;46:D649–55.
21. Ashburner M, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–9.
22. Grasmann G, Smolle E, Olschewski H, Leithner K. Gluconeogenesis in cancer cells—repurposing of a starvation-induced metabolic pathway. *Biochimica et Biophysica Acta Rev Cancer.* 2019;1872:24–36.
23. 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;8:105882–904.
24. Vettore L, Westbrook RL, Tennant DA. New aspects of amino acid metabolism in cancer. *Br J Cancer.* 2020;122:150–6.
25. Joshi S, et al. The genomic landscape of renal oncocytoma identifies a metabolic barrier to tumorigenesis. *Cell Rep.* 2015;13:1895–908.
26. Lindgren D, Sjölund J, Axelsson H. Tracing renal cell carcinomas back to the nephron. *Trends Cancer.* 2018;4:472–84.
27. 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:D442–50.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions





LUÍS ANDRÉ BOTELHO
DE CARVALHO

2024

EMPOWERING PERSONAL HEALTH IN BLADDER CANCER PATIENTS BY
DISCLOSING MEDICAL INFORMATION HIDDEN ON COMPLEX URI-NARY
PROTEOMES: DIAGNOSTIC, PROGNOSTIC AND RESPONSE TO THERAPY

