UNIVERSITÉ DES SCIENCES ET TECHNOLOGIES DE LILLE (FRANCE)
ÉCOLE DOCTORALE DE BIOLOGIE-SANTÉ

UNIVERSITÉ DE SHERBROOKE (CANADA)
DÉPARTEMENT DE BIOCHIMIE

THÈSE DE DOCTORAT

En vue de l'obtention du grade de Docteur en Science
de l'Université des Sciences et Technologies de Lille et de l'Université de Sherbrooke

Présentée par

JUSAL QUANICO

Development of On-Tissue Mass Spectrometric Strategies for Protein Identification, Quantification and Mapping

Thèse sous la cotutelle internationale
du Pr Isabelle FOURNIER et du Pr Robert DAY

Soutenue à Lille le 11 Juillet 2014

Devant le jury composé de :

Dr Philippe BULET	**Examinateur**
Pr Xavier ROUCOU	**Examinateur**
Pr Michel SALZET	**Rapporteur**
Dr Virginie REDEKER	**Rapporteur**
Pr Morad ROUDBARAKI	**Rapporteur**
Isabelle FOURNIER	**Co-Directrice de Thèse**
Robert DAY	**Co-Directeur de Thèse**

This manuscript is dedicated to all graduate students, who indefatigably burned their candles all through the night just to"keep the science going".

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

-Robert Frost, 1916

ACKNOWLEDGEMENT

This work would not have been possible without the contributions of the following:

Prof. Isabelle Fournier, my thesis adviser in University Lille 1, whose tutelage and guidance permitted me to explore the world of Mass Spectrometry and its infinite possibilities;

Prof. Robert Day, my thesis adviser in University of Sherbrooke, for his continued support in this co-supervised thesis despite the complications it entailed. For giving me the opportunity to study in Canada and learn from his expertise;

Prof. Michel Salzet, head of our laboratory in Lille, who made this co-supervised thesis possible in the first place. For dealing with all of the administrative issues and logistics, and especially for handling the difficulties brought about by my sojourn as an international student in France;

My thesis reporters, Prof. Xavier Roucou, Prof. Klaus Klarskov and Dr. Philippe Bulet, for accepting to evaluate my work and taking time out of their busy schedules to peruse through the manuscript;

University Lille 1 for granting me the Bourse President and Fonds de recherche en santé du Québec (FRSQ), my sources of funding in this research, and in extension, the École Doctorale de Biologie Santé de Lille and the Biochemistry Department of University of Sherbrooke;

Dr. Julien Franck, my immediate supervisor in Lille, for dealing with my day to day struggles and being there during experiments that last through the night. For coming up with out-of-the-box ideas when I had none, thus allowing me to start anew in otherwise dead-end directions. For training me not only with physical manipulations necessary for work with MS, but also for helping me to develop critical thinking and question the status quo;

Dr. Maxence Wisztorski, my first immediate supervisor when I started working in Lille, who is always there when his assistance is needed and never fails to accommodate you in spite of his tons of work;

Members of the laboratory in Lille, namely, Benoit, Francesco, Jean Pascal, Annie, Stephanie, Pierre Eric, Jacopo, Dounia, Marie, Françoise, Christophe, Krystelle, Franck and Lucie for keeping a convivial atmosphere that stimulates productive work;

Members of Prof. Day's lab, namely, Frédéric, Kevin, Ania, Christine, Hugo, Frédérik, Roxane, Sandra, Xue Wen, Evelyne and Prof. Witold Neugebauer, for the warm welcome in

Sherbrooke. For their constructive critcisms and aid which were crucial in fulfilling the requirements of the department during my stay in Sherbrooke;

My nanay and tatay, and sisters Lyn and Lisa, for their unwavering support even though I'm far away from home. As well as my friends from all over the world who were one in heart and in spirit with me;

And finally to Marina, for the incessant motivation that is the driving force behind my struggle in this endeavor.

With all my heart.

RÉSUMÉ

L'imagerie par spectrométrie de masse est une technique sans marquage permettant la détection et la localisation de protéines à partir de coupes de tissus. Afin de repondre à des problematiques biologiques, le nombre de protéines identifiées doit etre amelioré. Une stratégie consiste à realiser une micro-jonction liquide sur des regions particulieres des coupes de tissus afin d'extraire les peptides issus de la digestion *in situ* des protéines. Plus de 1500 protéines ont identifié sur une zone de 650µm, correspondant à environ 1900 cellules. Une corrélation entre ces données avec celles générées par MSI a augmenté le nombre de protéines localisées. Afin d'obtenir dans le même temps, la localisation et l'identification de protéines, une méthode consiste à realiser la microdissection de l'ensemble de la coupe après l'avoir déposée sur une lame recouverte de prafilm. Parafilm-Assisted Microdissection (PAM) a également été appliquée à l'etude de l'expression différentielle de protéines dans des tumeurs de prostate. Les résultats identifiés glutamate oxaloacétate transférase 2 (GOT2) en tant que biomarqueur de protéine candidate impliquée dans le métabolisme du glucose, en plus de celles qui ont déjà été indiqué précédemment. Réunis ensemble, ces méthodes MS d'analyses directes fournissent un moyen robuste d'étude de protéines dans leur état natif afin de fournir des indications sur leur rôle dans des systèmes biologiques.

Mots clés : spectrométrie de masse, imagerie, protéomique, micro jonction liquide, microdissection assistée par parafilm, cancer de la prostate

SUMMARY

Mass spectrometry-based methods for direct tissue analysis, such as MS imaging, are label-free techniques that permit the detection and localization of proteins on tissue sections. There is a need to improve the number of protein identifications in these techniques for them to comprehensively address biological questions. One strategy to obtain high protein IDs is to realize liquid microjunction on localized regions of tissue sections to extract peptides from the *in situ* digestion of proteins. More than 1500 proteins were identified in a 650µm spot, corresponding to about 1900 cells. Matching these IDs with those from MSI increased the number of localized proteins. In order to achieve simultaneous identification and localization of proteins, a method consisting of microdissecting entire tissue sections mounted on parafilm-covered slides was developed. Spectral counting was then used to quantify identified proteins, and the values were used to generate images. Parafilm-Assisted Microdissection (PAM) was also used to examine the differential expression of proteins on prostate tumors. Results identified glutamate oxaloacetate transferase 2 (GOT2) as a candidate protein biomarker involved in glucose metabolism, in addition to those that have already been reported previously. Taken together, these direct MS analysis methods provide a robust means of analyzing proteins in their native state and are expected to provide insights to their role in biological systems.

Keywords: mass spectrometry, imaging, proteomics, liquid microjunction, parafilm-assisted microdissection, prostate cancer

TABLE OF CONTENTS

PUBLICATIONS

Accepted Publications

- Mériaux C, Franck J, Park DB, **Quanico J**, Kim YH, Chung CK, Park YM, Steinbusch H, Salzet M, Fournier I (2014) Human Temporal Lobe Epilepsy Analyses by Tissue Proteomics. *Hippocampus* 24(6): 628-42.

- Franck J, **Quanico J**, Wisztorski M,Day R, Salzet M, Fournier I (2013) Quantification-Based Mass Spectrometry Imaging of Proteins by Parafilm Assisted Microdissection. *Analytical Chemistry* 85(17): 8127-34.

- **Quanico J***, Franck J*, Dauly C, Strupat K, Dupuy J, Day R, Salzet M, Fournier I, Wisztorski M (2013) Development of Liquid Microjunction Extraction Strategy for Improving Protein Identification from Tissue Sections. *Journal of Proteomics* 79:200-18. *:co-first authors

Manuscripts in Progress

- **Quanico J**, Franck J, Salzet M, Day R, Fournier I. Differential Protein Expression of Candidate Prostate Cancer Biomarkers by Parafilm-Assisted Microdissection Coupled with Label-Free Mass Spectrometric Quantitation.

- **Quanico J**, Franck J, Wisztorski M, Fatou B, Salzet M, Day R, Fournier I. On-Line nanoLC-Coupled and Miniaturized Liquid Microjunction for Localized Extraction of Proteins Directly on Tissue Sections.

- **Quanico J**, Franck J, Day R, Salzet M, Fournier I. Global Hydrogen/Deuterium Exchange Mass Spectrometry Directly on Tissue.

Oral Communications

- **Quanico J**, Franck J, Wisztorski M, Dauly C, Day R, Salzet M, Fournier I. On-Tissue Micro-Extraction: The Key to Success for Identification of Less Abundant Proteins in MALDI MSI. 61st ASMS Conference on Mass Spectrometry an Allied Topics, Minneapolis, USA, June 9-13, 2013.

- **Quanico J**, Franck J, Dauly C, Day R, Salzet M, Fournier I, Wisztorski M. Liquid Microjunction Microextraction for In-Depth On-Tissue Protein Profiling and MS Imaging Back-Correlation. XVIIIèmes Rencontres et École de Printemps de Club Jeunes de la Société Française de Spectrométrie de Masse, Izeste, France April 8-12, 2013.

- **Quanico J**, Franck J, Wisztorski M, Salzet M, Fournier I, Dauly C, Day R. On-tissue proteome profiling by liquid extraction surface analysis (LESA): Method development

and application on prostate cancer. Desorption 14th International Conference, Schloss Rauischholzhausen, Ebsdorfergrund, Germany, June 3-7, 2012.

- **Quanico J**, Franck J, Salzet M, Fournier I, Wisztorski M. Liquid extraction surface analysis: Application on proteins. Advion Biosciences User Meeting, Spectrométrie de Masse et d'Analyse Ptotéomique (SMAP) 2011, Avignon, France. 19-22 September 2011.

Posters

- **Quanico J**, Franck J, Wisztorski M, Salzet M, Day R, Fournier I. Combining Identification and Quantification of Proteins to their Localization Using Parafilm-Assisted Manual Microdissection. EuPa 7th Annual Conference and Spectrométrie de Masse et d'Analyse Ptotéomique (SMAP) 2013, St. Malo, France. October 14-17, 2013.

- **Quanico J**, Franck J, Wisztorski M, Dauly C, Dupuy J, Strupat K, Day R, Salzet M, Fournier I. Probing the rat brain proteome using microextraction, nanoLC-MS/MS and MALDI imaging strategies. 16th EURON PhD student meeting, Maastricht, Netherlands, September 27-28, 2012.

- **Quanico J**, Franck J, Kuzminska M, Salzet M, Fournier I, Wisztorski M. Liquid extraction surface analysis (LESA) of proteins on tissue sections. Spectrométrie de Masse et d'Analyse Ptotéomique (SMAP), Avignon, France 19-22 September 2011.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1,5-DAN	1,5-Diaminonaphthalene
A2M	α-2-Macroglobulin
AAT	Aspartate Aminotransferase
ACN	Acetonitrile
ADME	Absorption, Distribution, Metabolism, and Excretion
Amu	Atomic Mass Unit
ANI	Aniline
AP MALDI	Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionization
APCI	Atmospheric Pressure Chemical Ionization
APEX	Absolute Protein Expression
APPI	Atmospheric Pressure Photoionization
APS	Ammonium persulfate
AR	Androgen Receptor
ASAP	Atmospheric Pressure Surface Analysis Probe
AUI	Advanced User Interface
BAC	Benzalkonium chloride
BSA	Bovine Serum Albumin
C1QBP	Complement 1q-Binding Brotein
CE	Capillary Electrophoresis
CGI	Cancer Gene Index
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
$CHCl_3$	Chloroform
CID	Collision-Induced Dissociation
CTAB	Cetyl trimethylammonium bromide
CTNNA1	Catenin (Cadherin-Associated Protein), α1
DAPPI	Desorption Atmospheric Pressure Photoionization
DART	Direct Analysis in Real Time
DBS	Dried Blood Spot
DDA	Data-Dependent Analysis
DESI	Desorption Electrospray Ionization

DeSSI	Desorption Sonic Spray Ionization
DHB	2,5-Dihydroxy benzoic acid
ΔM	Mass Difference
DMF	Dimethyl formamide
DMPK	Drug Metabolism and Pharmacokinetics
DNA	Deoxyribonucleic acid
DNA-PK or PRKDC	DNA-Dependent Protein Kinase catalytic subunit
DRE	Digital Rectal Exam
DTT	Dithiothreitol
ECD	Electron Capture Dissociation
EMT	Epithelial-to-Mesenchymal Transition
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
EtOH	Ethanol
FA	Formic acid
FASP	Filter-Aided Sample Preparation
FDR	False Discovery Rate
FFPE	Formalin-Fixed Paraffin-Embedded
FN1	Fibronectin 1
FOLDS	Fast On-Line Digestion System
FTICR	Fourier Transform Ion Cyclotron Resonance
FTIR	Fourier Transform Infrared Spectroscopy
FWHM	Full Width at Half Maximum
GDF15	Growth Differentiation Factor 15
GO	Gene Ontology
GOT2	Glutamate Oxaloacetate Transferase 2
GSTM3	Glutathione S-Transferase mu 3
GWAS	Genome-Wide Association Studies
HADHA	Hydroxyacyl-coenzyme A (CoA) Dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA Hydratase (trifunctional protein), Alpha subunit
HC	Hierarchical Clustering
HCCA	α–Cyano-4-hydroxycinnamic acid

HCD	High Collision-induced Dissociation
HeLA	Henrietta Lacks
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HES	Hematoxylin Erythrosine Saffron
HPLC	High Performance Liquid Chromatography
HR FTMS	High-Resolution Fourier Transform Mass Spectrometry
HYOU1	Hypoxia Upregulated Protein 1
IAA	2-Iodoacetamide
IACUC	Institutional Animal Care and Use Committee
ICP	Inductively-Coupled Plasma
ID	Identification
IFN-γ	γ-Interferon
IR	Infrared
ISD	In-Source Decay
ITO	Indium Tin Oxide
JeDI	Jet Desorption/Ionization
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAESI	Laser Ablation Electrospray Ionization
LCM	Laser Capture Microdissection
LDTD	Laser Diode Thermal Desorption
LESA	Liquid Extraction Surface Analysis
LMJ-SSP	Liquid Microjunction Surface Sampling Probe
MALDI	Matrix-Assisted Laser Desorption/Ionization
MDH2	Mitochondrial Malate Dehydrogenase 2
MeOH	Methanol
MHC	Major Histocompatibility Complex
MS	Mass Spectrometry
MSI	Mass Spectrometry Imaging
MSMB	β-microseminoprotein
nanoLC-MS	Nano Liquid Chromatography-Mass Spectrometry
NCBI	National Center for Biotechnology Information
NCI	National Cancer Institute

NCL	Nucleolin
Nd :YAG	Neodymium-Doped Yttrium Aluminium Garnet
NDEESI	Neutral Desorption Extractive Electrospray Ionization
NH_4HCO_3	Ammonium bicarbonate
NID2	Nidogen 2
NMR	Nuclear Magnetic Resonance
NSAF	Normalized Spectral Abundance Factor
OCT	Optimal Cutting Temperature
OGP	n-Octyl-β-d-glycopyranoside
PA28β or PSME2	Proteosome Activator complex 28 subunit beta
PADI	Plasma-Assited Desorption/Ionization
PAI	Protein Abundance Index
PAM	Parafilm-Assisted Microdissection
PARP1	Poly(ADP-ribose) polymerase 1
PDF	Prostate-derived factor
PEP	Peptide Identification Probability
PIKK	Phophatidyl Inositol 3-Kinase related Kinases
PMF	Peptide Mass Fingerprint
POSTN	Periostin
PPP2R1A	Serine/Threonine-Protein Phosphatase 2A 65kDa Regulatory subunit A, alpha isoform
PSA	Prostate-Specific Antigen
PSD	Post-Source Decay
PTM	Post-Translation Modification
ROI	Region Of Interest
ROs	Reactive Oxygen Species
RPLC	Reverse Phase Liquid Chromatography
RT	Retention Time
RT-PCR	Real Time-Polymerase Chain Reaction
S/N	Signal-to-Noise Ratio
SA	Sinapinic acid
SDS-PAGE	Sodium dodecyl sulfate-Polyacrylamide Gel Electrophoresis

SIMS	Secondary Ion Mass Spectrometry
SRM	Selected Reaction Monitoring
SSSP	Sealing Surface Sampling Probe
TCEP	Tris(2-carboxyethyl)phosphine
TD-APCI	Thermal Desorption Atmospheric Pressure Chemical Ionization
TEC	Tissue Extinction Coefficient
TEMED	Tetramethylethylenediamine
TGFβ	Transforming Growth Factor-β
TFA	Trofluoroacetic acid
TLC	Thin Layer Chromatography
TOF	Time Of Flight
TRIM28	Tripartite motif containing 28
TRUS	Transrectal Ultrasound
UV	Ultraviolet
VEGF	Vascular Endothelial Growth Factor
WBA	Whole Body Animal
XIC	Extracted Ion Chromatogram

INTRODUCTION

Biological processes are highly dynamic and involve millions of molecules of various physicochemical properties and functions. In order to have a relatively fine picture of their mechanisms we need to first be able to monitor the variation in terms of abundance of the molecules in time and space. This means that we need to at least first perform a large scale identification of these biomolecules, and quantify and localize them in the time course of the biological process. These first intention studies are generally performed to highlight the main features differentiating the studied physiological or physiopathological process from a normal situation. When studying physiopathological mechanisms, information about the changes in signaling pathways is expected. Omics is nowadays the related term used in general to refer to these large scale analyses which are classified according to the family of biomolecules to be studied due to the different methodologies and technologies they are relying on. Omics is thus a broad terminology including, genomics, transcriptomics, metabolomics, lipidomics and proteomics as main research fields. Omics techniques will give information on the modification of cell signalization and will help to determine markers of the physiological/physiopathological process as well as provide a better understanding of the mechanisms. These large scale analyses are generally followed by more focused studies in order to decipher precise signalization pathways which are required to obtain a more precise knowledge on the cascade of events, including specific modifications carried by the molecules, as well as their interactions and translocation into the cells requiring thus various methodologies including structural biology techniques such as NMR, X-Ray and a large panel of microscopy technologies. Various families of biomolecules act in synergy to maintain cell homeostasis and allow cell-to-cell communication in a normal function or in response to an external stimulus. Ideally, to obtain a complete picture we should need to perform systematic biology by correlating the data coming from the different Omics. Since this is a huge task, researches are generally focused on one family of biomolecules, although more and more researchers are trying to integrate several large analyses together (Weckwerth 2008; Griffiths and Wang 2009). Taking the problem from one side, proteins are one of the families of biomolecules that can be studied in order to understand physiopathological mechanisms. Large scale protein analyses are referred to as proteomics analyses as introduced in the mid 90's by Wilkins and coll (Wilkins, Pasquali et al. 1996).

Proteomics analyses rely on a panel of technologies but among them Mass Spectrometry (MS) plays a central role. Indeed, the development of soft ionization sources such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) in the mid 80's have paved the way for the establishment of MS as a tool for the analysis of molecules of biological importance (Yamashita and Fenn 1984; Karas, Bachmann et al. 1987). With these sources it is now easy to generate ions from high molecular weight compounds of high polarity (Karas and Hillenkamp 1988; Fenn, Mann et al. 1989), thus expanding the range of biological molecules that can be analyzed by MS from small metabolites, to most notably proteins. In a way, MS has revolutionized the field of proteomics by permitting large scale analysis (McDonald and Yates 2002; Yates, Ruse et al. 2009). Instrumental developments in the field of MS combined to improvements in separation methods such as liquid chromatography (LC (Zhao and Lin ; Ishihama 2005) and developments of bioinformatic tools (Swan, Mobasheri et al. ; Eng, McCormack et al. 1994; Fenyo and Beavis 2008) have greatly contributed to this achievement. Currently thousands of proteins can be identified from a single experiment using shotgun strategies (Geiger, Velic et al. 2013; Picotti, Clement-Ziza et al. 2013). In these strategies extracted proteins are altogether submitted to enzymatic digestion. After purification of the digestion peptides, they are conventionally separated using LC on-line coupled to the MS instrument. This strategy relies on the capacity of the instrument to record MS and MS/MS data from the solution eluted from the LC system. MS spectra provide access to the mass of the digestion peptides while MS/MS will give structural information on the primary sequence of these peptides. The measured masses of the proteolytic peptides and the mass collected from the structural analysis are then combined to be compared and matched with a protein or gene sequence database, and the protein identification is assigned based on the accuracy of the match. This identification process depends on the unequivocal assignment of ions, and parallel developments in instrumental design in terms of mass accuracy and resolving power led to improvements in the identification (Scigelova, Hornshaw et al. 2011). Increase in the duty cycle of instruments also greatly improves the number of identification by allowing to record more structural information (Liu, Huttenhain et al. 2013). Crucial to the development of MS-based proteomics is the parallel improvement in the classical analytical methods of sample preparation. For instance, the improvement of chromatographic separation allowed the refinement of protein purification protocols that are tailored to the demands of the protein of interest. Development of high-throughput and less cumbersome instruments permitted the interfacing of more and more methods with MS, and paved the way for incorporation of multi-dimensional preparative

procedures in an effort to decrease the complexity of samples and increase the identification potential. Currently, Shot-Gun strategies allow fast access to the identification and relative quantification through label free quantification methods to thousands of proteins (Wisniewski, Dus et al. 2013) at various levels (from whole body, organs, cells down to cell compartments or vesicles) in various biological complexes allowing now to study molecules of lower abundance with important biological functions such as chemokines or cytokines (Meissner and Mann 2014).

However, in spite of all these improvements in proteomics, one crucial dimension is lacking, which is the information about the evolution of protein abundances with respect to their localization. This information is not required when working on body fluids, but becomes crucial when studying tissues, such as solid tumors. In fact, most large scale proteomics analyses are performed on whole organs or regions of organs. This only provides access to averaged information and renders difficult the identification of proteins that are highly localized to a restricted area (even though they are in relatively high abundance) due to dilution of the total amount of proteins. Such approaches thus do not show the variation in abundance of the proteins in a specific micro-environment, which can be primordial, for example in the early development of pathologies. Various strategies were developed to circumvent these difficulties. One instance involves collection of a specific type of cells. Cell sorting can be carried out by various means including Flow Cytometry, but the most popular ones are based on Laser Capture Microdissection (LCM, (Emmert-Buck, Bonner et al. 1996)). In LCM, cells are sorted based on their morphologies, which were pre-determined by histological staining. In the beginning LCM was shown to be relatively laborious since it was difficult to collect a large number of cells sufficient for gel based strategies. The advent of shotgun strategies based on nanoflow LC systems and nanoESI ion sources, and requiring much less amount of material, has greatly helped to the improvement of these strategies (Thakur, Rejtar et al. 2011). However, even if LCM coupled to shotgun analysis is nicely working this strategy still presents certain limitations such as the number of cells to be collected (most of the time about 50,000 cells) and the morphological phenotyping-based cell selection criterion in place of molecular phenotyping. Moreover, obtaining numerous cells entails having to collect them from various tissue sections in which the micro-environment can greatly vary.

In contrast, MS Imaging (MSI) strategies provide direct access to the 2D-distribution of molecules without pre-fractionation or purification of samples (Jungmann and Heeren 2012).

This greatly simplifies the sample preparation and stresses on the importance of analyzing proteins in the context of their localization. By doing so, MSI allows for the ability to discriminate different types of tissues in the same organ, and permits the non-subjective identification of deseased areas in contrast to normal ones. This simple and straightforward approach permitted the identification of tumor regions in otherwise normal tissues, for example, and demonstrated the potential MSI as a diagnostic tool that can be developed for future use by clinicians in the hospital setting. In this regard, MALDI MSI has shown to be the more powerful and versatile of the various MSI modalities due to the wide range of molecules accessible to the MALDI ion source (Fournier, Wisztorski et al. 2008). MALDI MSI is applicable to a large variety of endogenous molecules including proteins as well as exogenous ones such as drugs, abiotics or xenobiotics. MSI has shown to be very powerful in obtaining the localization of known compounds; one of the best illustrations is the imaging of drugs and their metabolites in whole body animals. MSI can effectively replace whole body autoradiography (WBA), which uses radiolabeled drugs and does not allow the metabolites to be imaged (Solon, Schweitzer et al. 2010). Alone or combined to multivariate statistical analysis, MSI can also be used as a tool for cell phenotyping and finding markers of physiological or physiopathological processes. This approach is often used for clinical studies of pathologies such as cancer, where imaging has shown significant applications in tissue classification and biomarker discovery (Franck, Arafah et al. 2009). In spite of its success in recent years, there are still limitations that need to be addressed in MSI. Perhaps the biggest issue is the crucial need to improve its sensitivity especially in detecting less abundant molecules. This remains especially true for proteins that are more difficult to detect due to their physico-chemical properties. In particular, higher mass proteins (Mw >30,000 Da) are clearly not detected when using conventional sample preparation and only most soluble proteins of the lowest Mw are observed (Franck, Longuespee et al. 2010). Thus, for proteins the number of compounds that can be imaged is relatively limited compared to other molecule families such as lipids. Moreover, direct imaging often results to localization of the most abundant proteins. Although this is not an issue by itself, nonetheless, abundant proteins impose a limit to the number of biomarkers that can be discerned by MSI by hindering the identification of the less abundant ones. More detailed information on physiopathological mechanisms requires the identification of these low abundance proteins, especially if they are involved in signalization pathways implicated in the disease. Identification of imaged proteins is one of the other great challenges for MALDI MSI in these untargeted approaches. These two difficulties have led to the development of On Tissue Bottom-Up strategies (Lemaire, Desmons

et al. 2007; Franck, Arafah et al. 2009). Here, proteins are *in situ* digested in a controlled manner to generate peptides which should be easier to detect and identify while avoiding their delocalization. Direct protein identification is a challenge since most ion activation methods do not allow fragmentation of intact species due to internal redistribution on all the degrees of freedom of the energy transferred by means of ion activation. This is circumvented by on tissue digestion of proteins. Though resolving part of the problem, on tissue digestion can increase sample complexity by increasing the number of peptide species, and again due to ion suppression effects, only digestion peptides coming from most abundant proteins are predominantly observed. Many groups, including ours, have attempted to address this issue by developing peptide derivatization to increase their propensity for fragmentation (Franck, El Ayed et al. 2009), incorporating washing and abundant protein depletion strategies into the imaging protocol, designing novel matrices to improve protein ionization from single to multi-charged species, using different methods to digest the proteins *in situ*, and so on. Nonetheless, the results of these approaches are limited, and the challenge to be able to detect, analyze and identify proteins of low abundance in their native environment on tissue remains a daunting task.

It is thus mandatory for the understanding of physiopathological mechanisms to elucidate prognostic and diagnostic biomarkers and their targets, and to be able to identify and image a much larger number of proteins. Clearly there is currently a need to fill the gap between the potential of conventional proteomics strategies and the potential of MS imaging to bring the localization of the molecules. This was the first goal of my PhD project. In this work, I present different strategies to demonstrate alternative ways to address the limitation of protein detection on tissue, as well as other developments I have tried to address in the course of three years of study. The first part focuses on the development of a strategy using the liquid microjunction extraction to be able to extract proteins on a localized region on a tissue section. The liquid microjunction extraction is considered as one of the ambient surface sampling techniques. In this published work, as much as 1500 proteins were identified in a region as small as 650 µm, roughly corresponding to 1900 cells. We were able to discriminate two regions on a rat brain tissue section based on these protein identifications and managed to correlate these results on the regions identified by hierarchical clustering analysis following a MALDI MSI experiment. As such, we have started to incorporate the liquid microjunction extraction strategy into our MSI platform and used it in further studies in our group. This localized extraction strategy is presented as an alternative to on-tissue cell collection done using laser capture microdissection (LCM) and

related strategies, which could sample homogeneous cell populations but at the expense of losing localization due to limited number of cells in the sampling region, and which still also suffers from poor protein identification yield in spite of the cell population enrichment. Although the strategy presently cannot sample homogeneous cell populations given the current size of the microdigested and microextracted area attainable, nonetheless, it can be used to examine cell populations in the context of their microenvironment. The ideal situation, in retrospect, would be to sample proteins of individual cells in the context of their tissue localization, as we expect the protein expression to change in a similar manner as has been demonstrated for metabolites even for cells of the same population (Zenobi 2013). Concordance with this idea, a chapter on the recent results of the continuation of the work on this subject is presented, focusing on decreasing both the size of the microextracted and microdigested areas. This consists of the use of alternative means to create a smaller liquid microjunction and the direct coupling of this method to nanoLC-MS. Although a similar set-up has already been previously reported, the current one was designed specifically for a proteomic approach and aimed for the sampling of even smaller regions, and results of its comparison with the conventional off-line liquid microjunction nanoLC-MS approach is discussed.

In the second part, I present the parafilm-assisted microdissection method. Here, an alternative approach dealing with the limited protein identification problem is shown. This consists of mounting a tissue specimen on a parafilm-covered slide, excising regions from the tissue, and subjecting them individually to conventional shotgun proteomics MS analysis. Furthermore, the published work explores the possibility of generating molecular images from the protein identifications, using the quantitative values of the spectral counts of peptide assignments. With this approach, we were able to identify and localize high and low abundance proteins, and are currently working towards the refinement of this strategy in order to improve the resolution of the images. An application of the PAM method on prostate tumors is then presented. Here, we examined fresh frozen tumor sections, excised regions of interest, and subjected them to proteomics MS analysis. Here I demonstrate that using the PAM approach allows for the identification and quantification of protein expression levels. The importance of this in regard to prostate cancer is that the data can be used to detect changes in the regulation of proteins, and eventually biological pathways, as the prostate undergoes pathological changes.

In summary, this work is largely divided into two categories: development of MS-based methods for on-tissue protein identification, and linking these methods with quantification and use this to be able to map protein distribution across the tissue. In addition, application of on-tissue protein identification and quantification strategies in the context of prostate cancer for the discovery and development of candidate biomarkers is also presented.

Hypothesis

Current MS-based methods for the direct analysis of proteins on tissue, such as MS imaging, suffer from poor detection, identification and quantification of proteins thereby warranting the development of alternative approaches that could circumvent these inherent limitations. One way is to improve the number of protein identifications by localized microextraction on discreet regions of the tissue section. Restricting microextraction on discreet regions could improve the identification of proteins of low abundance by limiting the amount of abundant proteins co-extracted, compared for example when extraction was performed on the entire organ or tissue section. These protein IDs could then be used for back-correlation with MSI data to increase the number of protein assignments that have been localized. Another way is to perform simultaneous identification and localization of proteins in a single approach, without sacrificing the high number of protein identifications. This can be achieved by dividing a tissue section into millimeter-sized portions that can then be individually subjected to shotgun MS. The spectral counts of the proteins identified can then be used to reconstruct the image of the tissue section and map their spatial distribution.

Objectives

Develop alternative methods to ameliorate MSI protein identification strategies.

This can be achieved by developing a localized protein extraction method based on the liquid microjunction approach. The approach is similar in principle to the generation of protein ID databases that are then used for matching with peaks in MSI spectra. In contrast though, the protein database will be obtained from localized regions on the tissue section. This will be done by performing *in situ* tryptic digestion on those regions, followed by extracting the peptides using the liquid microjunction approach, then subjecting the extracts to shotgun nanoLC MS.

Develop quantification-based imaging strategies to directly correlate protein identification data with localization information.

Simultaneous identification, quantification and localization of proteins on tissue sections can be realized by obtaining millimeter-sized pieces of the tissue and subjecting each piece individually to shotgun nanoLC MS. Label-free quantification of the identified proteins can then be done by spectral counting, and the relative quantification data used to generate the image of the tissue by noting the positions where the tissue pieces have been excised from the section. The tissue pieces can be obtained from the tissue section by mounting the latter on a parafilm-covered slide and excising the pieces using a scalpel. Parafilm is a relatively inert material that can act as a support for the tissue during cutting of the pieces.

BIBLIOGRAPHY

Direct Analysis of Proteins on Tissue by Mass Spectrometry

Direct on-tissue protein analysis refers to methods that permit the sampling, fragmentation, sequencing and identification of proteins directly from tissue specimens with or without the assistance of intermediary fractionation or purification methods. This is in stark contrast to classical MS-based proteomics methods that employ extensive fractionation techniques to come up with semi-purified samples that are then injected to a mass spectrometer in tandem with a chromatograph, the latter serving as the final stage of the purification procedure. MS-based methods that are employed for this type of analysis use either ambient or vacuum-type ion sources. In general, little emphasis is given on the mechanism of fragmentation and the type of mass analyzer used. Instruments that are capable of providing rapid acquisition rates and good spectral resolution are generally prefered for proteomics-type work because of the need to decrease sample complexity, whereas portable and less complicated systems are generally prefered for clinical diagnostics or random testing type scenarios where only comparative spectral differences are needed to make inferences. On the other hand, extensive elucidation of the mechanisms of ionization is always highlighted as this plays the pivotal role on the type of sample that can be analyzed and the type of analyte that can be examined.

An overview of a typical MS analysis in **Figure 1** places direct analysis methods at the front line of the procedure, their impact thus determines the outcome of the experiment.

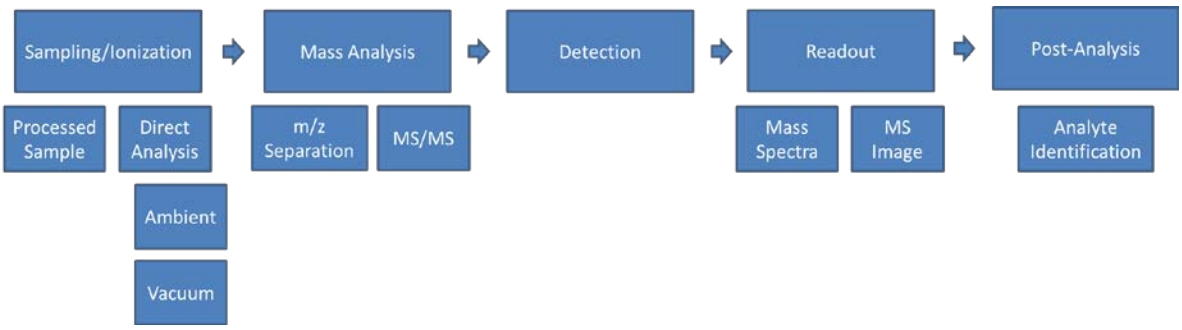


Figure 1. Schematic of a typical MS analysis.

A typical MS analysis can begin with the processing of biological samples tailored depending on the type of analyte to be examined, or with the placement of the material onto a mounting stage for ambient or vacuum-based surface sampling. The sample is then subjected to ionization, where

the sampling of charged analytes actually begins. The ionized species are then transported to the mass analyzer where their resolution in terms of m/z occurs. In tandem MS experiments, additional fragmentation of ions detected after m/z separation can be performed and the resultant MS/MS spectra can be used for analyte identification in the post-analysis step. After detection, the analysis can yield spectra for precursor or product ion scans, MS/MS spectra, as well as MS images constructed from the mapping of ion density on the selected or entire region of the sample section. Note that direct MS analysis methods are at the frontline of the scheme and thus are important in determining the outcome of the entire analysis.

The major advantage of direct on-tissue protein analysis in contrast to classical MS-based methods is the minimization or total elimination of complicated fractionation procedures that tend to introduce artifacts to the sample with each subsequent purification step (Huang, Cheng et al. 2011). By doing the analysis directly on tissue, it is possible to analyze proteins in their native state and in the presence of their interaction partners, thus providing a more realistic view of the binding of these proteins compared when their interactions are studied in solution. The point by point examination of tissue surfaces typical of direct tissue analysis methods allows the mapping of protein distribution across the whole tissue, permiting the localization of proteins.

Ambient MS

Ambient MS-based methods are those that employ direct tissue sampling under ambient conditions of temperature and pressure. Their development is mainly driven by the need for mass spectrometers that are capable of performing on the spot analysis of real-world samples thus giving access to this technology to everyday applications, such as for instance, in the determination of pesticide content in food products or, detection of substances of abuse or of stimulants in race animals and athletes, to name a few. This is possible by liberating the analysis from constraints presented by difficulties of introducing the sampling surface into the vacuum, including sample size and analyte volatility. In these methods, there is a great emphasis in the examination of samples that are preserved in their native state with minimal or no interference due to the analysis. Their use in proteomics is limited to the identification or quantification of highly abundant endogenous proteins, and proteome-wide discoveries involving these methods still have to be demonstrated.

A total of 42 ambient MS-based methods are currently reported in literature (**Table 1**, (Badu-Tawiah, Eberlin et al. 2013)). Several reviewers have classified them into various categories based on the similarities and differences in their mode of sampling the surface (**Figure 2**). It is not our current objective to provide a comprehensive review of these methods, but rather to

highlight major surface sampling principles in order to emphasize our choice of approaches used in this work.

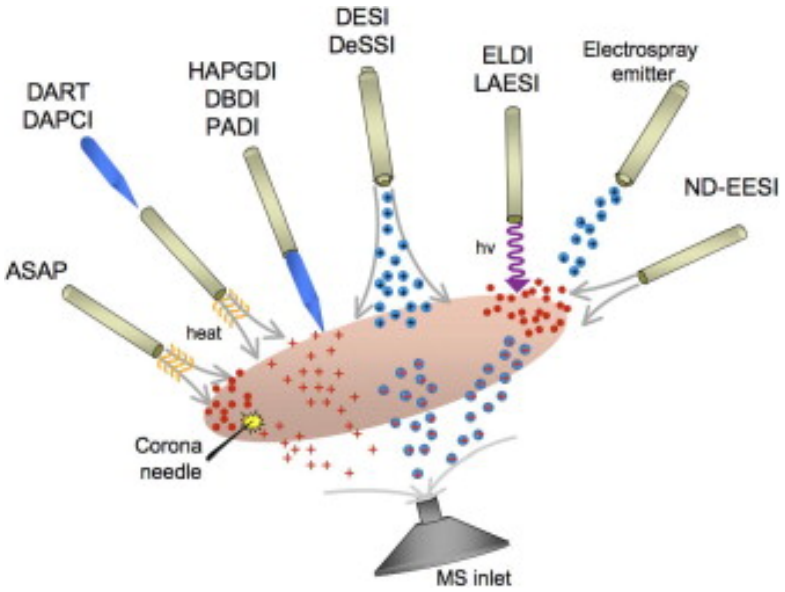


Figure 2. Overview of ambient MS-based methods.

Venter & Cooks' classification of ambient MS-based methods based on the mode of sample extraction and transport to the mass spectrometer inlet. Reprinted from (Venter, Nefliu et al. 2008), Copyright (2008) with permission from Elsevier.

Table 1. Ambient MS methods as recently reviewed by Badu-Tawiah and Cooks, 2013.

	Method name	Acronym	Agent	Characteristics
1	Desorption electrospray ionization	DESI	Sprays	Imaging capabilities, reactive mode, small (<500 Da) and large molecules
2	Easy ambient sonic-spray ionization	EASI	Sprays	Imaging capabilities, small (<500 Da) molecules
3	Desorption ionization by charge exchange	DICE	Sprays, chemical ionization	Small (<500 Da) molecules
4	Transmission mode desorption electrospray ionization	TM-DESI	Sprays	Reactive mode
5	Nanospray-desorption/electrospray ionization	Nano-DESI	Sprays	Imaging capabilities, small (<500 Da) and large molecules
6	Probe electrospray ionization	PESI	Sprays	Imaging capabilities, small (<500 Da) and large molecules
7	Liquid microjunction-surface sampling probe	LMJ-SSP	Sprays	Imaging capabilities, reactive mode, small (<500 Da) molecules
8	Paper spray	PS	Sprays	Reactive mode, small (<500 Da) and large molecules
9	Direct analysis in real time	DART	Plasmas	Small (<500 Da) and large molecules
10	Low-temperature plasma probe	LTP	Plasmas	Reactive mode, small (<500 Da) molecules
11	Flowing atmospheric pressure afterglow	FAPA	Plasmas	Small (<500 Da) molecules
12	Desorption atmospheric pressure chemical ionization	DAPCI	Plasmas	Small (<500 Da) molecules
13	Desorption corona beam ionization	DCBI	Plasmas	Small (<500 Da) molecules
14	Dielectric barrier discharge ionization	DBDI	Plasmas	Small (<500 Da) molecules
15	Electrospray-assisted laser desorption ionization	ELDI	Lasers, sprays	Imaging capabilities, small (<500 Da) and large molecules
16	Laser ablation electrospray ionization	LAESI	Lasers, sprays	Imaging capabilities, small (<500 Da) molecules
17	Laser-assisted desorption electrospray ionization	LADESI	Lasers, sprays	Imaging capabilities, small (<500 Da) and large molecules
18	Laser desorption electrospray ionization	LDESI	Lasers, sprays	Imaging capabilities, small (<500 Da) and large molecules
19	Laser-induced acoustic desorption-electrospray ionization	LIAD-ESI	Lasers, acoustic, sprays	Small (<500 Da) and large molecules
20	Neutral desorption extractive electrospray ionization	ND-EESI	Sprays	Reactive mode, small (<500 Da) molecules
21	Radiofrequency acoustic desorption and ionization	RADIO	Sprays, acoustic	Small (<500 Da) molecules
22	Atmospheric pressure solids analysis probe	ASAP	Heat, plasmas	Small (<500 Da) molecules
23	Infrared laser ablation metastable-induced chemical ionization	IR-LAMICI	Lasers, chemical ionization	Small (<500 Da) and large molecules
24	Rapid evaporative ionization mass spectrometry	REIMS	Chemical ionization	Heating, small (<500 Da) molecules
25	Desorption atmospheric pressure photoionization	DAPPI	Chemical ionization	Photons as proton source, small (<500 Da) molecules
26	Beta electron-assisted direct chemical ionization	BADCI	Chemical ionization	Beta particles as proton source, small (<500 Da) molecules
27	Extractive electrospray ionization	EESI	Sprays	Reactive mode, small (<500 Da) and large molecules
28	Remote analyte sampling transport and ionization relay	RASTIR	Sprays	Small (<500 Da) and large molecules
29	Laser ablation flowing atmospheric-pressure afterglow	LA-FAPA	Plasmas	Small (<500 Da) molecules
30	Surface activated chemical ionization	SACI	Plasmas	Small (<500 Da) molecules
31	Single-particle aerosol mass spectrometry	SPAMS	Lasers	Small (<500 Da) molecules
32	Laser diode thermal desorption	LDTD	Lasers, plasmas	Small (<500 Da) molecules
33	Helium atmospheric pressure glow discharge ionization	HAPGDI	Plasmas	Small (<500 Da) molecules
34	Surface acoustic wave nebulization	SAWN	Acoustic	Small (<500 Da) molecules
35	Ultrasonication-assisted spray ionization	UASI	Acoustic	Small (<500 Da) and large molecules
36	Atmospheric pressure-thermal desorption/electrospray ionization	AP-TD/ESI	Heat, sprays	Small (<500 Da) molecules
37	Microplasma discharge ionization	Microplasma	Plasmas	Small (<500 Da) molecules
38	Atmospheric pressure thermal desorption ionization	APTDI	Heat	Small (<500 Da) molecules
39	Desorption electrospray/metastable-induced ionization	DEMI	Sprays, plasmas	Small (<500 Da) molecules
40	Switched ferroelectric plasma ionizer	SwiFerr	Plasmas	Small (<500 Da) molecules
41	Laser electrospray mass spectrometry	LEMS	Lasers, sprays	Small (<500 Da) and large molecules
42	Plasma-assisted desorption ionization	PADI	Plasmas	Small (<500 Da) molecules

Reproduced with permission. For references, see (Badu-Tawiah, Eberlin et al. 2013).

Thermal Desorption/Ionization Methods

Ambient MS-based methods can be divided into several major categories based on the mechanism by which they extract samples from the surface and transport them to the inlet of the mass spectrometer (**Figure 2**). Thermal desorption/ionization-based methods are those that employ a heating mechanism to convert molecular species from the condensed phase in the sample surface into the vapor phase. This is traditionally followed by secondary ionization effected through the use of a corona discharge (atmospheric pressure chemical ionization, (APCI (Van Berkel, Pasilis et al. 2008)), or atmospheric pressure surface analysis probe, (ASAP (Ellis, Brown et al. 2013)). The mode by which heat is used to convert condensed phase material into gas phase, the manner of ionization, or both, can be varied thus leading to methods based on the TD-APCI principle. In laser diode thermal desorption (LDTD), the heated gas is replaced by an infrared (IR) laser that sequentially fires at the back of the sample surface, heating it until the gas phase molecules are ejected from the sample surface (Lohne, Andersen et al. 2012). In direct analysis in real-time (DART), heated He gas is converted into metastable atoms, and serves both to desorb material into gas phase and ionize them (Gross 2014). In plasma-assisted desorption/ionization (PADI), the He gas is initially made to pass between two electrodes, one of which is covered by a dielectric layer. As the gas flows, alternating voltage is applied to the two electrodes, generating plasma composed of metastable He, ions, radicals and electrons. The plasma jet is made into contact with the sample surface for desorption/ionization to occur. Molecular ionic species are generally formed using the method, although in some instances fragment ions can also be observed (Ratcliffe, Rutten et al. 2007). Lastly, in desorption atmospheric pressure photoionization (DAPPI), a microchip nebulizer delivers the heated solvent into the sample surface (Haapala, Pol et al. 2007; Luosujarvi, Arvola et al. 2008). This is then followed by ionization using an ultraviolet lamp in place of the corona discharge. The mechanism of ionization is similar with APPI using liquid sample introduction sources.

Laser Ablation Methods

Another type of ambient MS-based methods involves the use of a laser to generate an ablation plume of gas phase molecules which is then subjected to secondary ionization. Early examples of this approach used inductively-coupled plasma (ICP) as a means for secondary ionization (Van Berkel, Pasilis et al. 2008), making it useful in geology-type applications that demand elemental analysis of samples. Laser ablation has also been coupled with APCI-type ion sources. One

example uses a CO_2 laser to effect desorption of neutral molecules, followed by APCI ionization to analyze trypsinized proteins on SDS-PAGE gels (Coon, Steele et al. 2002). But perhaps the most widespread examples of this type of ambient MS-based method is when it is coupled with the ESI source. One method, termed laser ablation electrospray ionization (LAESI, (Nemes and Vertes 2007)), fundamentally accomplishes ESI-like ionization, and follows two phases. During the first phase, a strong 1μs pulse at 2940 nm excites the –OH vibration in water present in the sample, resulting to the creation of a hemispherical plume that initially expands and eventually collapses due to pressure exerted by the atmosphere during the second phase, lasting 300 μs. The explosion ejects secondary particles composed of a mixture of molecules, clusters and ablated particulate matter, which then coalesce with the charged droplets of the ESI source. A variant consists of using a 337-nm nitrogen laser operating at 20 μJ per pulse, the focused laser spot size of which is 100 μm x 150 μm with the laser beam incident angle at 45° (Shiea, Huang et al. 2005).

The main advantage of LAESI-type methods, in addition to high ionization yield, is the possibility of producing multiply-charged molecules from laser ablation modes of sampling. This limitation is inherent in laser desorption-type sampling where only singly-charged and rarely doubly-charged ions are generated. This capability stems from the use of the ESI to induce the charging of the ejected particles. As such, protein spectra having multiply-charged envelopes of ions similar to those observed in ESI-generated spectra are observed when using laser ablation-sampled spectra. Reports on top-down analysis of proteins using LAESI have recently appeared in literature (Kiss, Smith et al. 2014), and it is expected to find wider applications in proteomics, in addition to its established application in the in-depth profiling of small metabolites and in particular lipids (Shrestha, Nemes et al. 2010). Its main disadvantage though is the fact that ablation using a mid-IR range laser is dependent on the water content of the sample (Ye, Greer et al. 2011). Variations in the water content and tensile properties of the sample thus limit the replicability of this technique particularly for time-resolved imaging because they grossly affect the LAESI ion yield. The use of lower, mid-IR absorbances would also entail the use of significantly higher laser fluences in order to ablate biological samples.

Liquid and Gas Jet Desorption/Ionization Methods

Methods using liquid or gas jet, on the other hand, are those that rely on the collision of charged solvent droplets with the analyte on the surface for sample desorption, followed by charge

transfer upon impact. The mechanism of sample desorption in this case can be referred to as a liquid film-droplet sputtering process. Exemplified mainly by desorption electrospray ionization (DESI) (Takats, Wiseman et al. 2004), this method uses a pneumatically-driven electrospray that provides the charged droplets that wet the surface. The wet area is continuously impacted by the droplets, causing splashing, jetting, or creation of secondary droplets that may contribute to the ejection of the sample in addition to those that have initially been ejected upon impact of the spray (Takats, Wiseman et al. 2005). As such, no sample pre-treatments are used prior to its injection to the mass spectrometer, although it has also been demonstrated that the system can be coupled to an auxilliary chromatograhic set-up for use in more sensitive sample detection (Zhang, Yuan et al.). As in conventional ESI, a high voltage, typically 4 kV, is applied to create the charged droplets (Bereman and Muddiman 2007). By removing the voltage and increasing the nebulizing gas velocity to a sufficient level to permit ejection of analytes, a variation called desorption sonic spray ionization (DeSSI) was developed (Haddad, Sparrapan et al. 2006). Since no voltage is applied, the ESI source generates a distribution of positive and negative charged droplets. Further increasing the liquid flow up to 0.05-0.15 mL/min, among other modifications, leads to the another DESI variation called jet desorption/ionization (JeDI). In this case, the jet spray continuously erodes the sampling surface to generate gaseous analytes, making it ideal for depth profiling studies. Finally, in neutral desorption extractive electrospray ionization (NDEESI), both voltage and liquid were eliminated (Chen and Zenobi 2008). Only a moderate flow of gas at room temperature is applied on the surface to induce analyte desorption, followed by a secondary ESI ionization.

As surface sampling is done under ambient conditions, DESI can easily be applied to any type of sample surface, greatly expanding its applicability to diverse real-world problems. Its proteomics applications though are limited to the analysis of liquid (Miao and Chen 2009) or spotted protein standards (Rao, Celiz et al. 2013; Montowska, Rao et al. 2014) and imaging of the tryptic digests of these (Pasilis, Kertesz et al. 2008). It has been reported that the major limitation of DESI is its limited sensitivity; for instance, in the analysis of propanolol metabolites, it failed to detect the hydroxypropanolol glucoronide metabolite compared when using other ambient surface sampling MS methods or MALDI imaging (Kertesz, Van Berkel et al. 2008).

Liquid Extraction Surface Sampling/Ionization Methods

Surface liquid extraction methods are typically liquid-solid microextraction methods that can be coupled with any ambient ionization sources such as ESI or APCI (Van Berkel, Kertesz et al. 2008). The success of these methods is due to their greater extracting efficiency and larger surface sampling areas. It has been shown, for instance, that phase II metabolites of propranolol were detected using these methods (Van Berkel and Kertesz 2009; Kertesz and Van Berkel 2010; Kertesz and Van Berkel 2010), whereas they were unsuccessfully detected using DESI(Kertesz and Van Berkel 2010). In these techniques, the extraction of materials from the sample surface is independent of ionization by having a liquid solvent in contact with the sample surface so that extraction can occur. The solvent may then be introduced directly into the inlet of the mass spectrometer, as in continuous type liquid microjunction set-ups, or they can be collected and may be subjected to sample pre-treatment steps, as in discrete type ones. Since ionization is independent of extraction, the method can be used in conjunction with various types of ambient sources, as has been demonstrated in APCI and ESI applications. Examples of the continuous mode are the liquid microjunction surface sampling probe (LMJ-SSP, (Van Berkel, Kertesz et al. 2008)) and the sealing-surface sampling probe (SSSP, (Luftmann 2004)). This mode has been demonstrated to be useful for MS imaging, where an ESI source was used in place of the typical laser desorption/ionization source (Van Berkel, Kertesz et al. 2008). Alternatively, the solvent can be made in contact with the sample surface discretely as single droplets the formation and size of which is regulated by a syringe pump, as seen in the liquid extraction surface analysis probe (LESA, (Kertesz and Van Berkel 2010)). In this case, the solvent is carried using a conductive tip where the droplet is formed and where the subsequent extract is aspirated back to be introduced into the electrospray source. This mode is mostly applied in sample profiling, where discrete spots within the sample are extracted and the levels of analyte determined and compared for each sampling region.

Because of the independence of sample extraction from ionization, liquid extraction surface sampling techniques permit the further treatment of samples with coupled devices prior to introduction to the mass analyzer. In essence, this does not belong to the realm of ambient surface sampling anymore, however, the use of these samplers as microextraction devices is very tempting especially if the whole system is fully automated (Van Berkel, Sanchez et al. 2002; Kertesz, Ford et al. 2005). This has been shown, for instance, by coupling the LMJ-SSP device with a high performance liquid chromatography (HPLC) system (Kertesz and Van Berkel 2010).

Here, ambient sampling of a whole body tissue section and monitoring of the distribution of total drug related compounds were performed without employing the typical methods used for these techniques (MALDI-MS and whole-body autoradiography, respectively). The advantages in this case are the detection of more metabolites compared when using MALDI-MS and the avoidance of using potentially harmful radioactive tracers. Similar results were obtained when compared with the typical methods, but with less tedious and less time-consuming means.

Because of their flexibility and adaptability, this type of ambient surface extraction technique is expected to find widespread applications in the future, particular in the proteomics field. At present their use has been demonstrated in the analysis of thin layer chromatography spots (Van Berkel, Sanchez et al. 2002; Luftmann 2004 20), dried blood spots on paper (Kertesz and Van Berkel 2010; Edwards, Creese et al. 2011), thin tissue sections (Kertesz and Van Berkel 2010; Marshall, Toteu-Djomte et al. 2010), and spotted sample arrays on a MALDI plate (Kertesz and Van Berkel 2010) as is or with hydrophobic treatment (Walworth, Stankovich et al. 2011). **Figure 3** shows a schematic of the three currently available liquid surface sampling techniques. The main difference in the design of the probes is the mode of recovering the extract, either continuously using the LMJ-SSP and SSSP designs or discretely using a single droplet as in LESA.

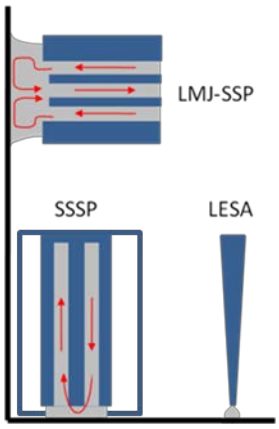


Figure 3. Schematic representation of the three currently available liquid surface sampling techniques.

The major variation in the design of the sampling probes is the manner of extract recovery either continuously using the LMJ-SSP or SSSP confirgurations, or discreetly using a single droplet as in LESA.

Sealing-Surface Sampling Probe

Much of the development of the current commercially available SSSP type configuration stems from the efforts of Luftmann (Luftmann 2004 20) to design a sampling probe for the automated extraction and analysis of TLC plates. SSSP works by enclosing the sample surface using a probe with a circular cutting edge that goes through the surface deep into the TLC backing aided by a frame press. This seals the surface of interest and somehow creates a partial vacuum that aids in the extraction of the sample. Solvent flows into the probe through an inlet capillary connected to an HPLC pump that regulates the flow, and comes into contact with the surface so that extraction can take place. The solvent then exits the probe thru an outlet capillary connected to the ion source for introduction to the mass analyzer. The typical diameter of the sampling head is 2 and 4 mm. In another design, the sampling head is elliptical (2 x 4 mm) so that it adapts more to the shape of the TLC spots (Morlock and Jautz 2008). The pump can be switched from standby mode, where solvent flows directly to the ion source, to sampling mode, where the solvent enters the probe and sampling occurs, or to docking mode, where the probe can be cleaned automatically using pressurized air (Aranda and Morlock 2007; Luftmann, Aranda et al. 2007).

As previously mentioned the SSSP system is currently designed for analysis of samples spotted on TLC plates, although it is imminent that this can also be applied to analogous surfaces especially those that are porous. Thus it may also be applied to systems utilizing polymer or metallic supports. The only limitations are that it is only restricted to samples where the pressure-aided sealing mechanism is possible, and that the pressure to be used for sealing must be optimized for each surface being used to avoid leakage of the extracting solvent, although modifications have been implemented to avoid this problem (Alpmann and Morlock 2006).

Liquid Microjunction Surface Sampling Probe

The LMJ-SSP system is mainly developed by Van Berkel and co-workers (Van Berkel, Sanchez et al. 2002), although prototypes of this device have been previously reported (Modestov, Srebnik et al. 2001; Wachs and Henion 2001). This probe consists of a capillary inserted inside a bigger one so that the space between them is used as an inlet connected to a solvent reservoir. Solvent enters and fills this space until the excess goes out of the capillary and comes into contact with the sampling surface. In this so-called wall-less ‘microjunction’ formed between the solid sample surface and the liquid, extraction occurs as the analytes with preferential affinity to the solvent polarity migrate to the liquid interphase and leave the sample by concentration gradient-driven

diffusion. The hollow portion of the inner capillary of the probe serves as the outlet line and is connected to the ion source. The resulting extract enters this line because of a self-aspirating mechanism driven by the pneumatic flow of the nebulizing gas towards the source.

The LMJ-SSP device has found applications in extraction of spots from hydrophobic TLC plates (Van Berkel, Sanchez et al. 2002; Ford, Deibel et al. 2005), drugs obtained by solid phase extraction cards (Wachs and Henion 2003), affinity-captured proteins (Van Berkel, Ford et al. 2006), and it has also been used for the imaging of inks on paper (Kertesz, Ford et al. 2005). In the latter case, resolution was achieved at 0.5 mm, the size of the probe. By modifying and automating the adjustment of the microjunction formation, LMJ-SSP in discrete sampling mode was used for high-throughput analyses of drugs (Van Berkel and Kertesz 2009). This also lessened the need for technical expertise in adjusting the probe-to-sample distance which is crucial to microjunction formation, and thus enables the automatic probing of a sample surface with varying thicknesses as in whole body tissue section for example. Problems in sample carry-over was observed, however, and it was suggested to perform a blank run in between samples to ensure complete cleaning of the probe prior to the next analysis. Coupling with other post-treatment devices, such as an HPLC set-up, was also recently accomplished using the LMJ-SSP (Kertesz and Van Berkel 2010). This can be very important with regard to future advances in ambient sampling because this allows clean-up of complex sample matrices that may hinder efficient sample detection when they are introduced directly to MS.

Liquid Extraction Surface Analysis

Further exploration of the discrete ambient surface sampling as an alternative to the continuous type led to the design of LESA (Kertesz and Van Berkel 2010). The principle is that the formation of the wall-less microjunction does not have to be restricted with the use of probes as in the case of LMJ-SSP. In fact, the microjunction can be formed simply by putting into contact a solvent droplet with the sampling surface, a function that can also be achieved using disposable pipette tips. After extraction, the extract is aspirated by means of a plunger mechanism using the same tip and then delivered into the electrospray head for ionization. With this system, flexibility is offered on the modulation of the size of the microjunction formed and variation of extraction time to suit a particular analysis. Currently, LESA has been demonstrated on the quantitation of drugs on dried blood spots (DBS), profiling of total drug distribution on whole body tissue sections, and extraction and quantitation on MALDI spots (Kertesz and Van Berkel 2010). It has

also recently been used for the direct and rapid identification of intact hemoglobin variants from DBS samples used for the neonatal screening of sickle cell and other hemoglobin-related diseases (Edwards, Creese et al. 2011). LESA has also been used to correlate results obtained from MS imaging experiments (Marshall, Toteu-Djomte et al. 2010). Although imaging investigations using LESA have yet to be reported, the possibility of using this technique for imaging has been implied.

Much of the developments in LESA have been obtained through the use of the TriVersa NanoMate system marketed by Advion Biosciences. Using the Advanced User Interface (AUI), Van Berkel and Kertesz (Kertesz and Van Berkel 2010) were able to use the NanoMate platform to perform LESA procedures. The TriVersa NanoMate is actually a coupling device used for automated delivery of infused samples for ESI MS analysis. It uses a robotic arm to pick up specially designed conductive tips and uses them to aspirate the sample by attaching the open end into a disposable nano-electrospray chip. The chip contains 384 electrospray emitters created using a deep reactive ion etching chemical process (Schultz, Corso et al. 2000). These nozzles provide long and stable sprays and eliminate sample-to-sample carry-over by the use of new nozzles for individual analyses. In addition, the TriVersa NanoMate platform offers portability and adaptability with a variety of compatible mass spectrometers. As such, prior to the development of the LESA mode, this device has already seen various applications in the analysis of lipids (Stahlman, Ejsing et al. 2009), metabolites (Giavalisco, Hummel et al. 2008), proteins and protein-related fields like post-translational modifications, non-covalent interactions, and top-down proteomics (Zhang, Van Pelt et al. 2003), antibodies and glycosylation (Ge, Rybakova et al. 2009), and others (Stumpo and Reinhold 2010).

The NanoMate system has three modes of operation: 1) infusion mode, 2) fraction collection and 3) LESA. In the infusion mode, the device allows for the direct introduction of samples into the ESI emitter, as previously described. The fraction collection mode is used to interface the device (and the coupled mass spectrometer) to a liquid chromatography (LC) set-up. In this mode the eluant coming from the LC system is transported to the NanoMate through a microcapillary tubing connected to a tee. The tee is connected to the mandrel and to the nanospray adapter, so that simultaneous collection of the eluant at constant time segments (through the mandrel) is possible while a portion is being electrosprayed. The resulting MS spectra can thus be correlated

with the collected eluants and, depending on the interfaced LC set-up, with other detection systems such as a UV detector for example.

In LESA mode, the NanoMate system works as follows. The robotic arm picks up the conductive tip, aspirates the solvent from a reservoir and dispenses it on a specific location on the sample surface. It then aspirates back the extract and attaches the tip to an ESI chip nozzle for electrospraying. In fact, there is very little difference in the mode of operation as in the usual infusion mode, except that the solvent is allowed to flow out of the conductive tip and be in contact with the surface first before introducing to the electrospray emitter.

Crucial to the optimization of LESA using the NanoMate platform are the distance of the tip relative to the sampling surface, the kind and amount of solvent used for microjunction formation, the hydrophobicity and matrix of the surface to be analyzed, the contact time between the solvent and the surface during the extraction process, the number of extraction repeats used for sampling, as well as the parameters used for the generation of the electrospray. The first three conditions determine the quality of the microjunction created, ie., its size, shape and recovery. Depending on the application, the size is of particular importance because it determines the spatial resolution that can be achieved when using LESA. The contact time and number of extractions, on the other hand determine the extent of the extraction and thus has particular importance in quantitation applications (sample recovery, repeatability of measurements, and method feasibility). Whereas the electrospray parameters such as the capillary voltage and gas pressure, as well as the MS parameters, determine the overall quality of the spectra being generated.

MS Imaging

MS imaging is a direct surface sampling method that allows the mapping of analyte distribution based on their peak intensities in the mass spectra collected throughout the entire sample surface. This involves the use of a sampling source, typically a laser or secondary ions, that rasters on the surface to desorb/ionize material. For each point on the surface, mass spectra and coordinates are simultaneously obtained, and then processed so that each point will correspond to a pixel in the molecular image. As such, a single pixel can contain a multititude of information from the peaks corresponding to analytes contained within the sample, spanning a wide range of biomolecule families like small molecule metabolites, lipids, peptides and proteins. The major strength of this

approach is the simultaneous analysis of these components in a one-shot analysis, with minimal sample preparation steps. The generation of molecular maps allows for their spatial localization approaching subcellular resolution. Fundamentally a non-targeted, multiplexed approach, MS imaging allows for the detection of changes across the sample due to its high data acquisition capability. In contrast to traditional approaches such as whole body autoradiography, analytes are detected from the image based on their characteristic mass-to-charge (m/z) ratios and without the use of radioactive labeled compounds.

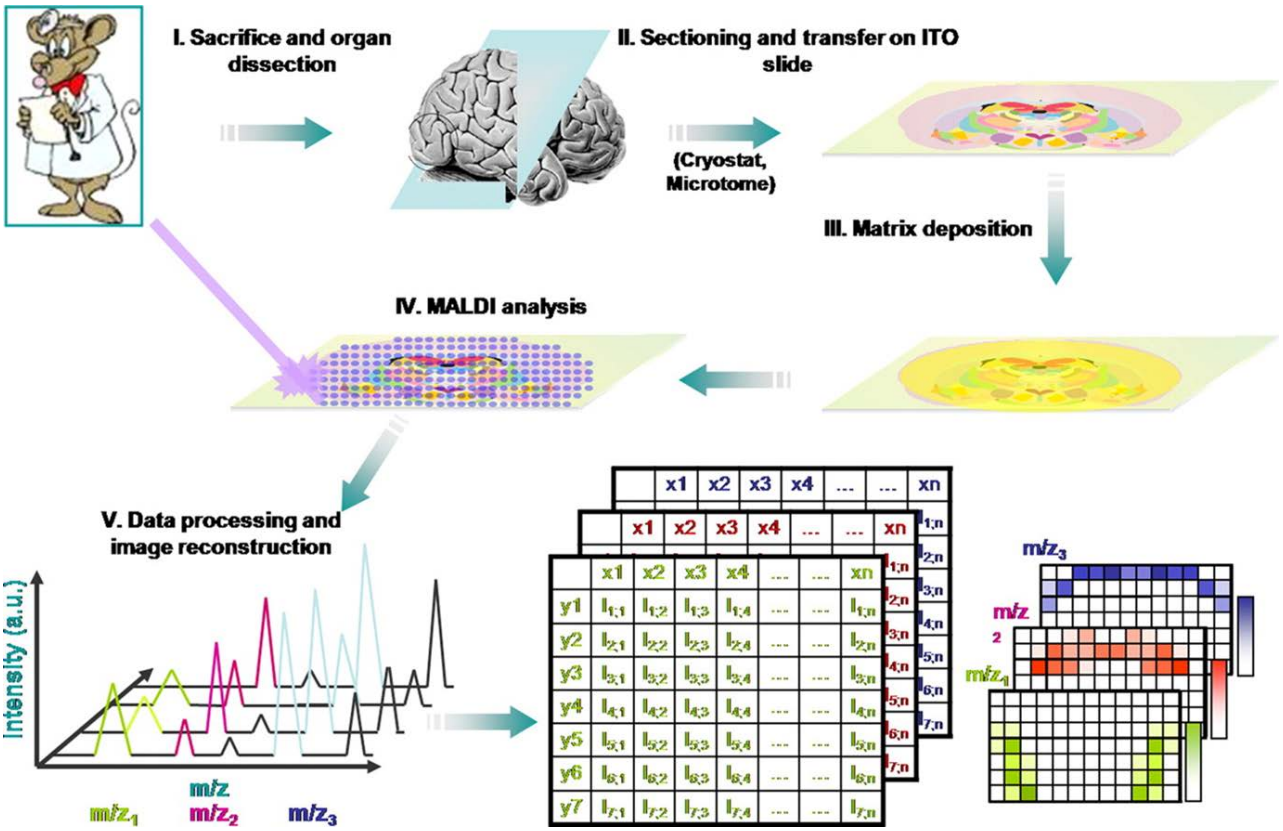


Figure 4. MSI workflow.

Tissues are obtained after sacrificing the organism and dissecting the organs; they are then kept viable by storage at cryogenic temperature or fixation. Micrometer-thin sections can be obtained from the tissue by cryosectioning, and these are mounted on conductive glass slides. To prevent analyte delocalization, the tissue is dried under vacuum. Matrix is then deposited and once dry, the sample is irradiated point by point while the laser is rastering across the entire section. For each point, spectra are recorded, and the ion distribution for each m/z peak is mapped using the point coordinates (pixels) during data processing and image reconstruction. (Reprinted from (Franck, Arafah et al. 2009) with permission from the American Society for Biochemistry and Molecular Biology).

A typical MSI experiment is outlined in **Figure 4**. By means of cryosectioning, micrometer-thin sections are obtained from tissue specimens that have either been frozen, formalin fixed and

paraffin-embedded, or preserved by other means and mounted on a slide. Point by point collection of spectra is then performed by rastering across the section, generating a grid of points that are then converted into spatial coordinates once the spectral acquisition step is done. Finally, the image is generated by displaying the intensity of an ion at the selected m/z value. The quality of the generated image can be assessed in terms of three figures of merit, two of which are coming from MS (Pol, Strohalm et al. 2010). Mass accuracy refers to the agreement of an ion's measured mass (or m/z) with its theoretical mass (expressed in parts per million units) and is dependent on the type of mass analyzer used (Rompp and Spengler 2013). This parameter is crucial particularly for peptide/protein identification in proteomics experiments because of the multitude of isobaric peptides that can exist in a complex sample that could lead to false positive identifications. Mass resolution or resolving power, on the other hand, refers to the minimum difference by which two ions can be discerned from each other (Smith, Kilgour et al. 2013). This is usually obtained by dividing the m/z value with the full width at half maximum (FWHM) of the ion peak (Marshall and Hendrickson 2008). This parameter is important in spatial localization because low-resolution images can generate false distributions especially if the individual isobaric ions produce distinct distributions themselves. The third figure of merit, spatial resolution, is a microscopy imaging parameter defined as the minimum distance between two objects that can be used to discern them from each other. In MSI, spatial resolution can be determined as the center-to-center distance between two sampling regions assuming that each sampling region corresponds to a single pixel in the image (Rubakhin, Jurchen et al. 2005).

The mass accuracy, resolutive power and spatial resolution requirements of an MSI experiment depend on the type of analyte being examined. Mass accuracy and resolutive power rely heavily on instrumental capability, especially the kind of mass analyzer used. Whereas, spatial resolution is dependent on the quality of preparation of the sample and all the other steps prior to analysis in the mass spectrometer.

In regard to instrumentation, current mass analyzers can be categorized into three groups, with the Fourier transform ion cyclotron resonance (FTICR) possessing the highest resolutive capability of up to 3,000,000 FWHM (Smith, Kiss et al. 2013), followed by the orbitrap mass analyzers (resolution up to 240,000 FWHM, (Michalski, Damoc et al. 2012)) and the time-of-flight (TOF), ion trap, or any of the hybrids (usually containing two mass analyzers and a

collision cell such as triple-quadrupole, quadrupole-TOF, and TOF-TOF, with resolutive power typically between 1,000-25,000 FWHM).

TOF mass analyzers use a drift tube under vacuum, with the ion source held at one end and a microchannel plate detector placed on the other end. The m/z of the ions generated from the bombardment of the sample is obtained by measuring the time it takes for the ion to travel thru the field-free drift tube until they hit the detector (Cotter 1992). Initially all the ions that enter the drift tube have ideally the same kinetic energies, thus, the time it takes for each ion to hit the detector will be dependent on its mass with the lighter ions reaching first and the heavier ones going last. As such, TOF mass analyzers theoretically have no mass limit, although measurements only up to m/z 200 kDa have been reported for proteins. MSI is commonly performed using TOF mass analyzers because of the broad mass range afforded by these instruments, their rapid acquisition rate, and their compatibility with vacuum-based ion sources (Suckau, Resemann et al. 2003).

Quadrupole mass analyzers transmit ions by a filtering mechanism using oscillating electrical fields to selectively stabilize or destabilize ion trajectories as the ions pass through a radio frequency quadrupole field produced by 4 parallel rods (Miller and Denton 1986). Changing the potentials of the rods allows for the selective passing of a certain m/z range, allowing the ejection of qualified ions either continuously or in discrete successive pulses. Variations of this configuration such as in triple-quadrupoles allows for further discernment of ions (Yost and Boyd 1990). They are initially filtered in the first quadrupole, then enter a collision cell where additional fragmentation is performed, then pass through a second filter in the third quadrupole, which only selectively transmits fragments of the ion that has been subjected to fragmentation in the collision cell.

In contrast to quadrupoles which transmit ions by filtering them according to m/z, ion traps "trap" packets of ions in a three-dimensional space defined by a ring electrode between two endcap electrodes (Paul 1990). Varying the radio frequency potential (supplemental oscillatory excitation voltage) of the endcap electrodes allows for the destabilization of the orbits of the ions, so that those with low masses get ejected from the z-axis on to the detector, while the rest remain in orbit inside the trap. As the space of the trap is defined, only a limited range of m/z can be trapped at a time. Hence the working m/z range of ion traps is typically below 4000 Da, though interfacing it

with ion sources capable of producing multiple-charged ions makes it applicable to larger molecules such as proteins.

FTICR mass analyzers (Marshall, Hendrickson et al. 1998) measure mass using the image current the ion generates as it accelerates in a spiral trajectory held by a static magnetic field and pushed by a rapidly varying radio frequency electric field (Penning trap). As they do so, they form part of a circuit and detectors specifically placed at fixed positions in space record the electrical signal over time, generating the periodic signal. The frequency of the cycling of the ion is dependent on its m/z, and a Fourier tranformation deconvolution of the signal leads to this value. Repetitive measurement of the ion signal thus provides high sensitivity and resolutive power in this mass analyzer. Orbitrap mass analyzers, on the other hand, trap ions in an oscillating electric field produced by a central electrode working at constant voltage and outer electrodes at varying voltages (Zubarev and Makarov 2013). As in FTICR instruments, the image of the ion current generated as it oscillates back and forth on the central electrode is recorded and subjected to Fourier tranformation to obtain the m/z value.

As previously mentioned, spatial resolution is a microscopy factor dictated by the quality of sample preparation as well as the steps prior to introduction of the sample to the mass spectrometer. Degradation begins immediately upon procurement of biological samples as they often contain proteolytic enzymes that are activated once exposed to suitable conditions. As such, biopsy samples are frozen in liquid nitrogen and stored in -80°C to retard the proteolytic activity of these enzymes (Norris and Caprioli 2013). Rat brain samples prepared this way have been reported to be viable for one year, after which, degradation starts to become significant. Another method is to crosslink the proteins on the tissue using formaldehyde followed by embedding in parafilm (Maes, Broeckx et al. 2013) . This is the most common method of storing biopsy material for archiving, as the fixed sample can be stored at room temperature for years because the extensive crosslinking induced by formaldehyde hinders the enzymatic activity of the proteins on the tissue (Scicchitano, Dalmas et al. 2009). Once the samples have been procured, care in their handling during sectioning must also be done. For frozen samples, an abrupt increase in temperature during cryosectioning can lead to melting of the tissue and codensation of moisture on the surface, not only hastening enzymatic degradation but more importantly causing delocalization of the components. Also, the thickness of the sections could play a role in the sensitivity or the applicability of the ionization method that will be used. As a guide, the average

cell size between 10-20 µm is used depending on the tissue specimen, and single cell layers are generally prefered for imaging work (Ye, Greer et al. 2011).

Once the tissue sections have been prepared, the succeeding preparation steps are dependent on the type of ionization source that will be used. This can either be vacuum-dependent or ambient type.

Vacuum-Dependent Sources

Vacuum-dependent ion sources are those that necessitate the use of a vacuum system in order to function. The vacuum is essential because it is important that the ions produced do not collide with the background gases as they enter the detector, otherwise they lose their kinetic energies. The vacuum is also a means to prevent contamination by background gases that can be adsorbed during measurment.

Secondary Ion MS

In its simplest form, SIMS uses a primary beam of ions placed at a specific angle to the sample surface so that the collision of the ions with the sample produces a plume of secondary ions that are then collected and analyzed in the mass spectrometer (Pol, Strohalm et al. 2010). The ion beam can be generated using liquid metal ions, such as gold, Ga^+, Bi^+ or In^+ (Ye, Greer et al. 2011), or by electron ionization of noble gases such as $^{40}Ar^+$ or Xe^+, oxygen ($^{16}O^-$, $^{16}O_2{}^+$, $^{16}O_2{}^-$), ionized molecules ($SF_5{}^+$) and fullerene ($C_{60}{}^+$), or by surface ionization ($^{133}Cs^+$). As these primary ions possess high energies, their bombardment of the sample cause damage on the primary area of impact, thus making SIMS a hard ionization technique. The resultant high energy secondary ionization leads to the formation of very small fragments which are actually remnants of the analytes that were not on the direct path of the beam. SIMS works best at the low mass range, with a mass limit of m/z <4000. The size of the analyte that is ionized depends on the ion beam that is being used; large cluster ions such as $SF_5{}^+$ and $C_{60}{}^+$ are able to expel larger fragments from samples and thus these ion beams find particular biological applications notably for the analysis of lipids and small metabolites (Fletcher, Lockyer et al. 2007; Jones, Lockyer et al. 2007). Currently it is still difficult to analyze proteins using SIMS mainly because the ion beams mentioned are inefficient to expel such large molecular assemblies, though in some instances it might be possible to eject fragments of these proteins at low ion yields (**Figure 5** (Komatsu, Murayama et al. 2008)). Resampling of the same area on the sample permits the use of SIMS in depth profiling studies (Debois, Brunelle et al. 2007); however, the chemical damage induced by

the ion bombardment restricts depth profiling up to submicron levels, known as the “static limit” of the analysis.

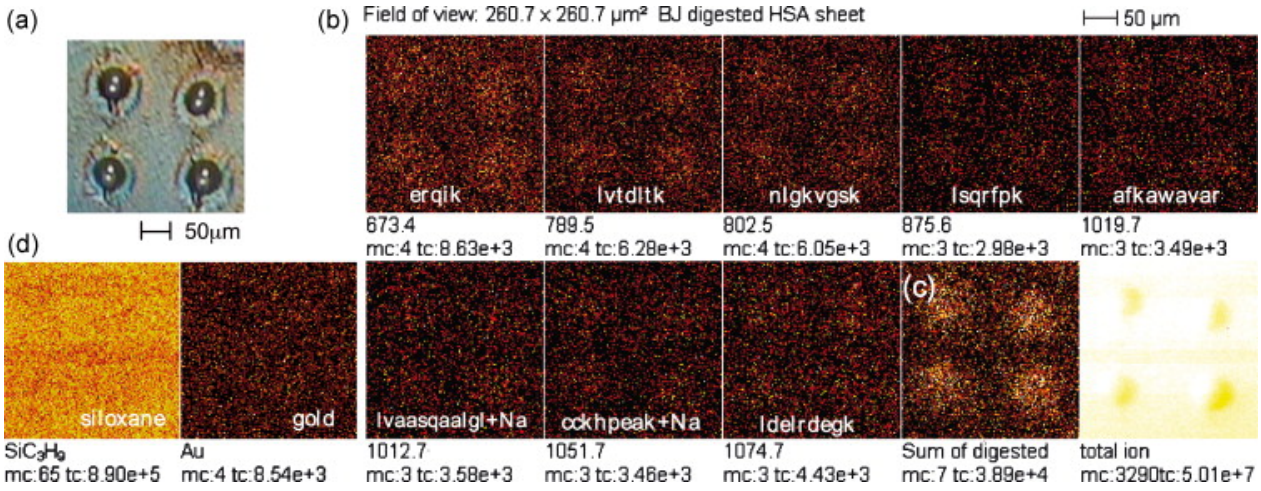


Figure 5. TOF-SIMS imaging of Human serum albumin (HSA) film sample.

(a) optical image of HSA film spotted with trypsin droplets using a modified BJ printer. (b) Ion images of digested fragment peptide ion peaks. (c) Gathered ion image of all digested fragment peptide ion images. (d) Ion images of non-digested material (siloxane molecule and gold). (Reprinted from (Komatsu, Murayama et al. 2008), with permission from Elsevier).

SIMS operating in static mode (Pol, Strohalm et al. 2010) currently offers nanometer resolution, the highest resolution possible for any MSI technique. This is made possible by pulsing the fluence of the primary ion beam, thus keeping it at a very low level thereby effectively reducing the area being sampled as well. The consequence of this though is the relatively low ion yield, leading to a much lower signal sensitivity. On the other hand, in dynamic SIMS, a higher energy, continuous primary ion beam is used, so that bombardment of the sample surface leads to its total destruction. As a consequence, more ions per pixel are produced and even smaller fragments are generated compared when using low energy ion beams.

In essence, no additional sample preparation steps are necessary for a SIMS analysis. Optional methods can improve the sensitivity of the MS acquisition depending on the application being pursued. For instance, a method called freeze fracturing can be done in order to lyse cells without disrupting the localization of their components by growing them on a conductive surface sandwiched with another one, freezing them in liquid nitrogen, and pulling apart the two pieces together (Roddy, Cannon et al. 2002). Freeze drying in liquid propane (Lanekoff, Phan et al. 2013) or crosslinking the cellular components with formaldehyde prior to analysis (Xia and Castner 2003) can also be done to preserve the localization of the components. For the analysis of

trypsinized proteins, it has been reported that the use of a liquid matrix containing trifluoroacetic acid enhances the sensitivity of the SIMS experiment (Nygren and Malmberg 2010).

Matrix-Assisted Laser Desorption/Ionization

MALDI is a soft-ionization technique that serves as the method of choice for detecting biomolecules directly from sample surfaces. In MALDI, the sample surface is coated with a matrix that serves as the primary energy absorbing compound (Karas, Bachmann et al. 1987). The surface is then irradiated with a laser beam passed the limit of the fluence threshold. Most of the energy from the incident laser beam is absorbed by the matrix, although some of it will eventually be transfered to the analytes on the sample surface, triggering their desorption and ionization. Eventually a plume of charged molecules, their alkali metal adducts, as well as neutral species is created on the irradiated surface, and the charged species are accelerated from the source towards the mass analyzer by application of an electric field. The resultant ionization generally produces singly charged species although sometimes doubly charged species can also be observed especially for larger molecules (Karas, Gluckmann et al. 2000; Karas and Kruger 2003). Conventional vacuum-based MALDI sources use UV lasers such as the N_2 or frequency-tripled neodymium-doped yttrium aluminium garnet (Nd:YAG) lasers operating at 337 nm and 355 nm respectively. The more recent development called atmospheric pressure MALDI (AP MALDI), a variation of this source that can operate without the need for a vacuum, runs on an IR laser operating at 2940 nm and takes advantage of the vibration of water molecules on biological samples in order to induce their desorption, but provides less sensitive detection (Li, Shrestha et al. 2007).

The matrix is of paramount importance in MALDI (Burton, Watson et al. 1997; Lietz, Gemperline et al. 2013). They are usually aromatic compounds that absorb energy at the aforementioned wavelengths preventing the premature degradation of the analytes prior to desorption and energy transfer steps (Li, Inutan et al. 2012). Also, during the process of deposition, matrix solvents extract the analytes from the surface and inhibit proteolytic activity of the enzymes present by denaturing them. The organic solvent is eventually evaporated by drying the slide under vacuum, leaving a homogeneous coating of matrix crystals on the surface doped with the analytes. When already in gas phase, the matrix acts in concert with the acid usually added to the solvent as a reservoir of protons that initiates and maintains proton, cation and electron tranfer processes eventually resulting in the formation of singly-charged species. Often

the choice of matrix determines the outcome of a MALDI analysis in terms of the type of analyte (**Figure 6**). Lipids are best detected using 2,5 dihydroxybenzoic acid (DHB), for example, while peptides and and proteins are best detected using α-cyano-4-hydroxy cinnamic acid (HCCA) and sinnapic acid (SA), respectively (Norris and Caprioli 2013). As these analytes usually occur at distinct mass ranges, typical MALDI analysis workflows usually also limit the mass range of analysis to focus on the particular type of analyte for enhanced sensitivity. Specialized matrices are also available for more specific kinds of analyses. For instance, the liquid ionic matrices derived from stoichiometric mixtures of aniline with the aforementioned matrices work best in the MS imaging of analytes on biological tissue samples (Meriaux, Franck et al. 2010). In-source decay fragmentation methodologies used in top-down proteomics MS specifically use 1,5-diaminonaphthalene to induce direct fragmentation of proteins (Sellami, Belgacem et al. 2012).

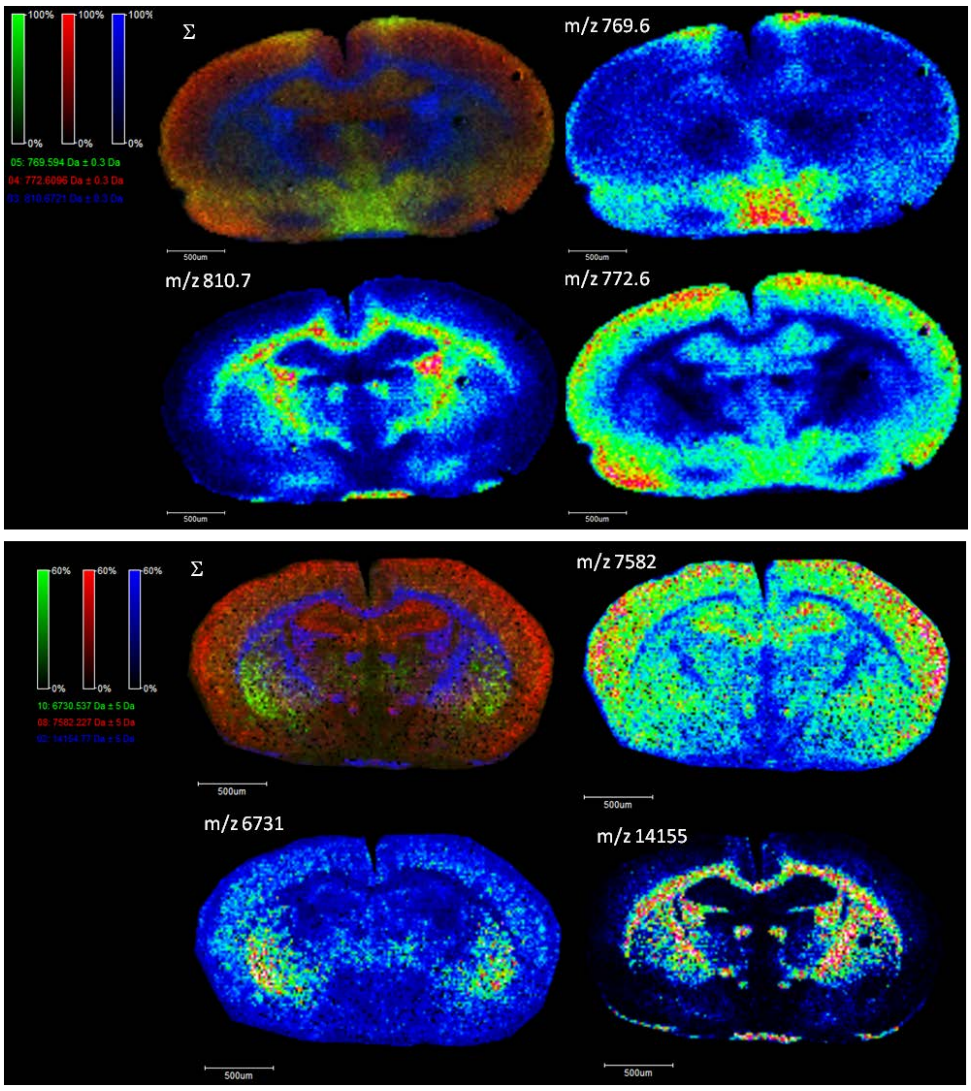


Figure 6. The choice of matrix depends on the analyte to be detected in MALDI MSI.

Images obtained from two consecutive rat brain sections. Top: imaging of lipids using DHB. Below: imaging of intact proteins using SA. In both panels, the upper left image next to the scale is the composite of the three m/z mapped.

For imaging, the matrix can be deposited to the sample surface by various means (Franck, Arafah et al. 2009; Rompp and Spengler 2013). The method used determines the size of the matrix crystals that are formed; for improved resolution it is desirable to have smaller, more uniform crystallization because the size of the crystal affects the size of the pixel, whereas for improved analyte extraction yield (and therefore, detection sensitivity) it is necessary to increase the exposure of the analyte surface to the matrix solution to facilitate the process (inadvertently creating larger matrix crystals). Matrix deposition can be effected by manual spotting using a micropipette. This can provide efficient co-crystallization of analyte with the matrix, although this is generally avoided in imaging because of the delocalization of the analyte within the entire pool of the matrix solvent in addition to poor reproducibility of the spotting. Manual spraying can also be done using a pneumatic sprayer to spray coat the entire section, but for consistent results, automated devices are generally preferred (Sugiura, Shimma et al. 2006). Among these, the ImagePrep device is very popular because it is capable of controlling the deposition chamber temperature, as well as tightly regulating the quality of spraying, amount of matrix deposited, and drying intervals between spraying. ImagePrep exposes the sample section to a matrix aerosol generated by vibrational vaporization. Other devices work based on other modes of sample deposition, such as electrospraying (Wei, Nolkrantz et al. 2004), sublimation (Hankin, Barkley et al. 2007) and microspotting (by acoustic deposition or ink jet spotting). Electrospraying and sublimation methods work best for analytes that do not need to be co-crystallized with the matrix (such as lipids) because there is no or very little amount of solvent used during the deposition. On the other hand microspotting generally leads to micrometer-sized spots (~150 µm diameter), with analyte diffusion determined by the size of the microspot. The loss of spatial resolution in microspotting is compensated for by the more efficient extraction and co-crystallization of the analyte and this method finds particular use in protein MS imaging.

Apart from matrix deposition, additional sample preparation steps are necessary for an effective MALDI-based imaging analysis. Depending on the analyte of interest, steps such as washing and interfering analyte depletion, *in situ* enzymatic digestion, and signal enhancement by chemical

derivatization, can be incorporated in the imaging workflow to improve the data generated. Washing or depletion (Grey, Chaurand et al. 2009) is done by incubating the tissue section in solvents that liberate components selectively. The selectivity of the wash depends on the chemical properties of the component that needs to be removed. For example, in the MS imaging of proteins, it is necessary to delipify tissue sections using organic solvents (Lemaire, Wisztorski et al. 2006). The amphoteric property of proteins renders them resistant against organic solvents and as such, they are retained on the tissue after the wash. Whereas doing the reverse, ie, wash with aqueous solvent to remove proteins, might not necessarily enhance lipid detection; they may serve to deplete the tissue of abundant proteins and salts that can otherwise mask the signal of the less abundant ones during the MS analysis. This has been observed in the analysis of integral membrane proteins in the eye lens, for example, where the depletion of crystallins by repeated washing with water exposes aquaporins (Grey, Chaurand et al. 2009). Enzymatic digestion is a means to generate peptides for bottom-up MS imaging strategies (Groseclose, Andersson et al. 2007); further details about protein identification using this strategy will be discussed later. Direct enzymatic digestion of a tissue section can be done by depositing the proteolytic enzyme using the same devices used for matrix deposition. As the enzyme will also be present on the tissue during MS analysis, care should be taken to minimize its signals by regulating the amount and concentration of the enzyme deposited, and by dissolving them in buffers that can be tolerated by the mass spectrometer. In cases where identification of the protein directly on tissue is necessary, chemical modifications that can orient the fragmentation of this protein to a specific ion series thereby enhancing the overall quality of the MS/MS spectrum generated can be done by treating the tissue with specific reagents prior to MS analysis (**Figure 7** (Franck, El Ayed et al. 2009)). These reagents can be deposited using the same devices previously described. In all cases, it must be noted that the overall spatial resolution of the image of the treated sample will be influenced by all the sample treatments used, and as such, the choice of the washes or the deposition of proteolytic enzymes or derivatization agents must be carefully regulated to avoid delocalization of the analytes.

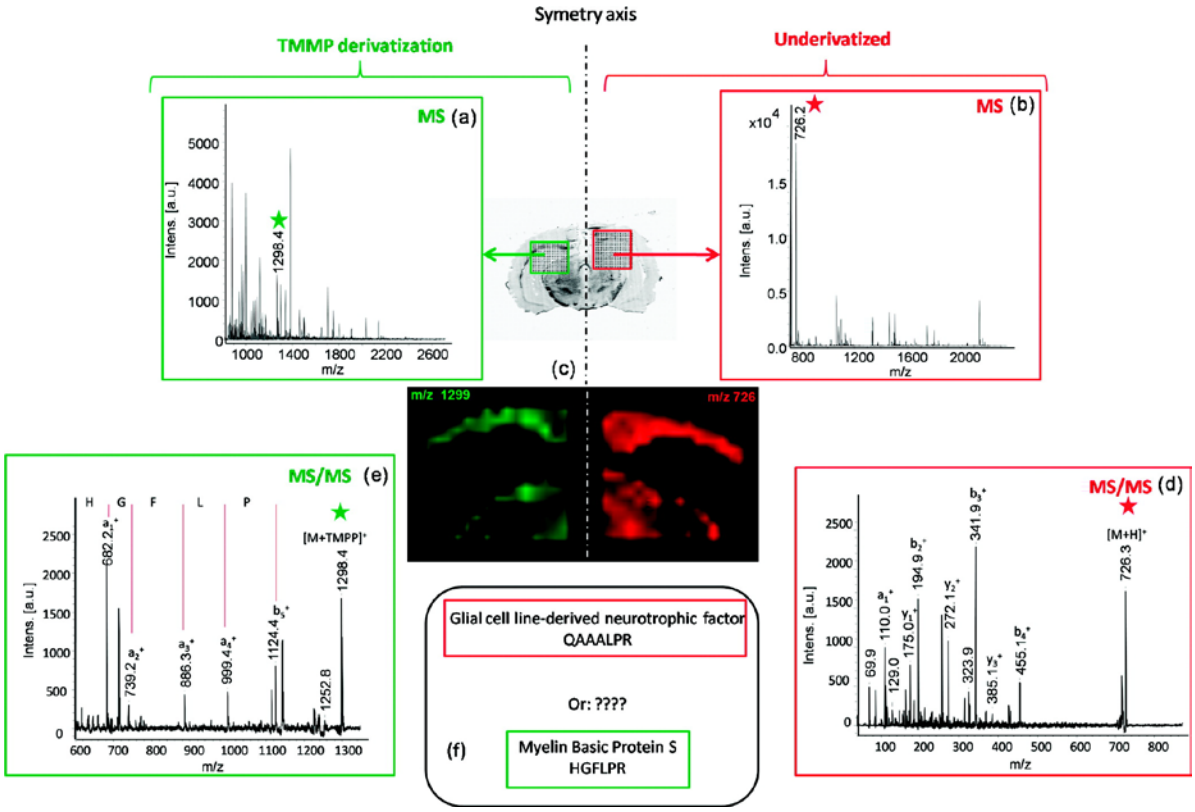


Figure 7. N-terminal derivatization improves on-tissue MS/MS identification of peptides.

MALDI MS spectra recorded after (a) enzymatic digestion followed by *N*-(succinimidyloxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP) derivatization and (b) enzymatic digestion. (c) Images of the digestion peptide at *m/z* 726 and its corresponding TMPP derivative at *m/z* 1299, showing an identical distribution for both ions. MALDI MS/MS spectra recorded from a rat brain tissue section in the same area after on-tissue trypsin digestion with and without TMPP derivatization (d and e, respectively). (f) Identification results from the databank using *R. norvegicus* taxonomy. Reproduced with permission from (Franck, El Ayed et al. 2009). Copyright (2009) American Chemical Society.

Ambient Sources

Several ambient ion sources have been used for MS imaging applications. Of these, the most extensively used are DESI (Eberlin, Ferreira et al. ; Wiseman, Ifa et al. 2006; Laskin, Heath et al. 2011) and LAESI (Nemes and Vertes 2007; Nemes and Vertes 2010). The major advantages of this type of imaging are that even lesser sample preparation steps are needed and ionization can be effected using ESI. As such, it is possible to generate multiply-charged ions which can be very useful if tandem MS analysis such single-reaction monitoring type imaging would be required. Their limitation though, is the low spatial resolution attainable, currently at about 180-200µm for DESI (Ye, Greer et al. 2011). (Spatial resolution as low as 40µm has been reported (Kertesz and

Van Berkel 2008), but its wider applications have yet to be demonstrated.) By eliminating the need for vacuum, ambient sources futher extend the applicability of MS imaging to real world situations. This however, presents its own unique challenges to MS imaging. For instance, it may not be fully possible to eliminate contamination due to the environment of the sample, paticularly if this environment cannot be controlled. Another challenge is the desorption, ionization and transmission of the analytes. Ambient sources are typically placed in front of the inlet of the mass spectrometer, with the desorbed analytes simply drawn into the inlet by virtue of gas flow into the capillary. Whereas in contrast, vacuum-based sources have been reported to have a maximum ion transmission of about 50%, in spite of more efficient and carefully controlled ion collection methods. It is highly doubtful that ambient sources are capable of matching this yield, although perhaps liquid extraction surface sampling type sources might be able to overcome this limitation by making multiple extractions on the same region prior to injection to the mass spectrometer. **Figures 8** and **9** show examples of ambient MS imaging of proteins.

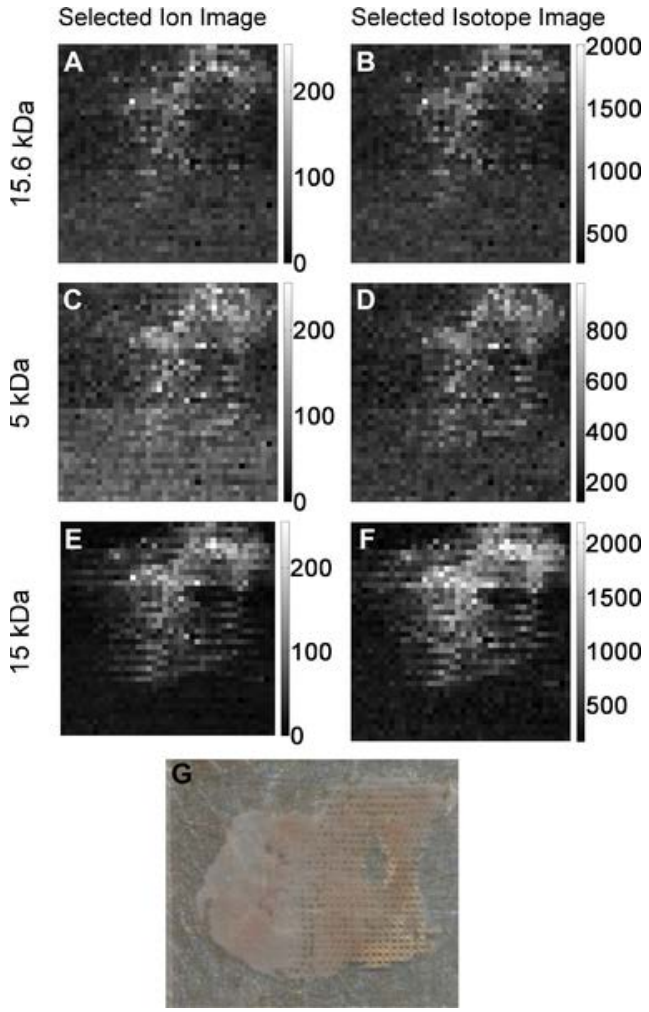


Figure 8. Top-Down LAESI-MS Imaging of intact proteins in a mouse lung section.

Optical (G), selected ion images (A, C, E) and selected isotope images (B, D, F) obtained using LAESI FT-ICR MS. The MS images show the distribution of the ions at m/z 822 (15.6 kDa: A, B), 827 (5 kDa: C, D) and 883 (15.6 kDa: E, F). (Reproduced with permission from (Kiss, Smith et al. 2014), Copyright (2014) John Wiley and Sons).

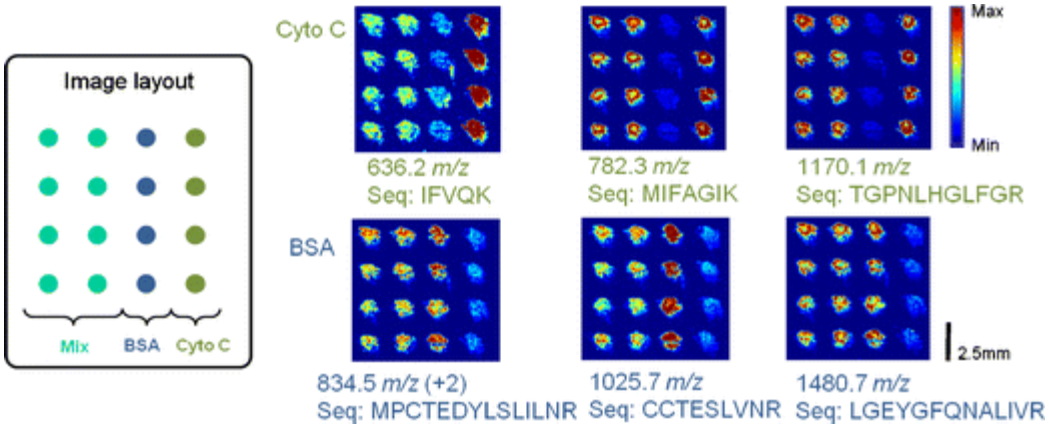


Figure 9. DESI-MS imaging of a binary combination of cytochrome C and BSA standards printed on Permanox surfaces.

Left shows the layout of the spotting. (Reprinted from (Rao, Celiz et al. 2013) with permission from Springer Science and Business Media).

Identification and Quantification of Proteins On Tissue by MS

Direct identification of proteins on tissue remains one of the key challenges in direct tissue methods of MS analysis. As such, it becomes the bottleneck of these methods aimed at the discovery of novel protein biomarkers for example, as understanding the biology of a disease and design of putative therapeutic drugs rely heavily on the identity and nature of these biomarkers. In contrast to small molecule metabolites, drugs, lipids and endogenous peptides, whose identification and quantification can be established using tandem MS and related techniques (SRM and so on), proteins present unique challenges owing to their significantly larger molecular masses. For instance, it is more difficult to fragment them using conventional fragmentation methods especially in the MALDI setting which yields singly-charged ions. An added complication is the existence of numerous protein isoforms and the presence of post-translational modifications. The former complicates an otherwise rather simple bottom-up analysis by demanding more identified peptides per protein ID in order to discern them better from their isoforms; in most instances particularly if the sample matrix is complex, it is already difficult to find one peptide for a low abundance protein. The latter on the other hand complicates the

MS/MS spectrum of a fragmented peptide by introducing masses not typically observed in that peptide's fragmentation pattern.

For a long time, MS developments have solely been improved for the qualitative assessment of specific proteins on samples. However, increasing applications demanding for the examination of protein expression levels necessitate the development of quantification strategies to monitor the changes of these levels especially in instances when the organism under study is being subjected to stress in disease states. These quantification strategies stress the importance of protein expression changes and relate these changes with each other so that an integrative view encompassing an entire protein network can be examined. As such, these strategies are therefore focused on the quantification of proteins *en masse*, in contrast to conventional methods aimed at detecting and quantifying only one or a few set of proteins.

Peptide Mass Fingerprinting

Peptide mass fingerprinting (PMF) is a method used to make protein ID assignments using mass spectra (Henzel, Billeci et al. 1993; James, Quadroni et al. 1993; Pappin, Hojrup et al. 1993; Yates, Speicher et al. 1993). An isolated protein is prepared either intact or digested, and subjected to ionization then mass analysis, so that the generated spectrum contains m/z peaks corresponding to the sequential fragmentation of the protein both along the N- and C-terminals. The mass peaks are tabulated and compared with pre-existing protein databases constructed from the theoretical fragmentation of a protein (termed *in silico* digestion) whose sequence has been determined from translation of the genome database. The specific sequence of the amino acids in a protein yields specific sequential fragmentation and this pattern is used to discern proteins from each other. After interrogation, a list of putative protein IDs is obtained and decisions on the protein ID assignment is based on the statistical confidence of the m/z alignment.

The strength of this approach is that it does not necessitate the use of the more rigorous *de novo* peptide sequencing method which is the traditionally used method for protein sequencing. However, it is important that the database used for spectral alignment must have a comprehensive protein list. Also, the method is very prone to error if the sample analyzed is impure. The presence of other proteins in the sample could interfere with the PMF generated by introducing artifact peaks that will decrease the confidence of the protein alignment. Also, not all proteins behave in exactly the same way and some do not yield very good PMFs. This can be due to poor

fragmentation or the presence of amino acids that tend to absorb incident ionizing energies more compared to others. To address this issue, it might be necessary to use chemical derivatization so that only a specific ion series containing intense and more or less complete fragmentation peaks will appear in the mass spectrum thus greatly improving unambiguous protein assignment.

A special means to generate protein PMFs is by analysis of the intact proteins directly using top-down approaches (Catherman, Skinner et al. 2014). Fragmentation of intact proteins is induced by subjecting them to other kinds of fragmentation mechanisms instead of the convential collision induced dissociation (Schey, Anderson et al. 2013; Kiss, Smith et al. 2014). In a MALDI-based platform, this can be achieved by spraying the tissue surface with 1,5-DAN and subjecting it to in-source decay (ISD, (Calligaris, Villard et al. 2011)). ISD is a phenomenon that occurs when the laser fluence is increased 15-20% above the analyte's threshold so that in the process proton transfer happens between the matrix molecules and the analyte's backbone C=O. This leads to fragmentation following either a radical pathway (forming c- and z- ion series) or a thermal pathway (forming x-, y- and a-, b- ion series). In electrsopray platforms, electron transfer dissociation (ETD, (Syka, Coon et al. 2004)) or electron capture dissociation (ECD, (Zubarev, Kelleher et al. 1998)) can be used to do top-down MS. Such methods uilize ion/electron or ion/ion interaction mechanisms to dissociate gas phase ions into their fragments (Hart-Smith 2014). The ions receive electrons to produce radical cations that initiate the fragmentation process. However, in contrast to CID, the fragmentation does not follow the typical energy redistribution that leads to the dissociation of the weakest bonds, but rather follows those similar to ISD. The success of these top-down approaches have been demonstrated in a variety of systems, the greatest advantage being that PTMs are retained and can be analyzed. However, they still remain somewhat restricted to single-protein or simple mixture analysis, although recent developments in separations will make these approaches a viable alternative to digestion-based ones.

Tandem MS

Tandem MS methods are those where a precursor ion from an initial full MS analysis is subjected to further fragmentation in order to create product ions in a subsequent MS/MS analysis (**Figure 10** (Louris, Wright et al. 1985)). The prototypical example of this is with the use of the triple quadrupole mass analyzer previously described, where the second quadrupole acts as a collision chamber where CID-type dissociation can occur (Yost and Enke 1977; De Hoffmann 1996). The

tandem MS method can also be realized using other mass analyzers. The most often used nowadays is the orbitrap hybrid composed of a linear ion trap or quadrupole mass filter that aligns precursor ions followed by a C-trap that bunches selected precursor ions similar to those in laser ion sources and sends them as packets to the orbitrap for high resolution analysis (Zubarev and Makarov 2013). With instruments such as the Q-exactive (a quadrupole mass filter-orbitrap hybrid), it is now possible to obtain up to 2,800 protein identifications from a HeLa cell lysate using a 30-min gradient (Ding, Jiang et al. 2013).

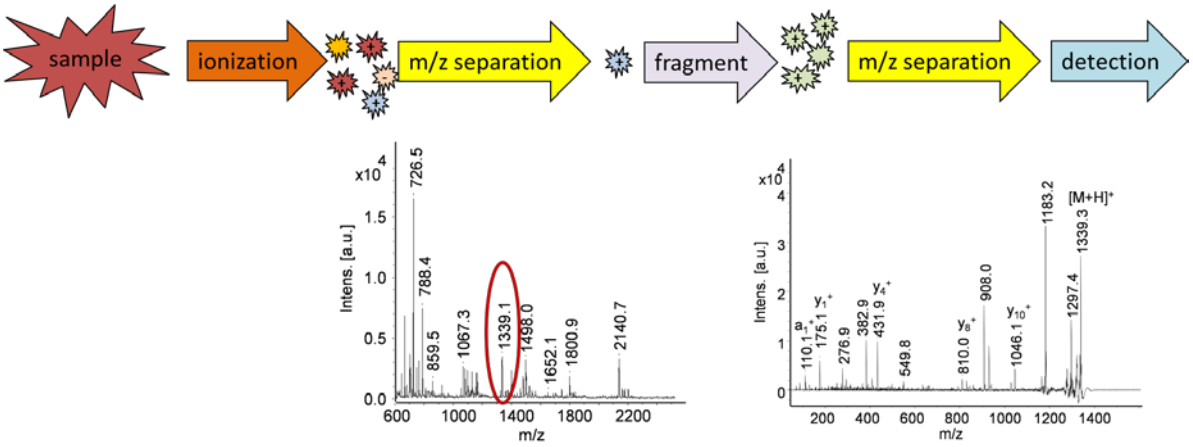


Figure 10. A tandem MS workflow.

After ionization, a precursor ion scan is obtained, where the ion of interest can be selected by filtering of the isolation window. Once the parent ion is isolated, it is subjected to collision-induced dissociation to yield daughter fragments that are then collected and resolved in a tandem mass analyzer. The product ion scan is then generated.

The strength of the tandem MS method of protein identification lies in the capability of hybrid orbitrap instruments to perform rapid data acquisition. A data-dependent (DDA) mode of acquisition (Mann, Hendrickson et al. 2001; Schmidt, Gehlenborg et al. 2008; Andrews, Dean et al. 2011; Johnson, Boyes et al. 2013)is typically implemented on nanoLC-interfaced orbitrap mass spectrometers. In this method, a survey scan at a user-defined mass range is first acquired usually at high resolution (50,000 at m/z 400 using the Q-exactive, (Michalski, Damoc et al. 2011)). Then a series of N MS/MS scans is performed on the top N most intense peaks obtained from the survey scan. The MS/MS scans are performed at a lower resolution (12,500 at m/z 400 for the Q-exactive), using either low CID fragmentation or the recently introduced high collision-induced dissociation (HCD). In Q-exactive instruments, the selection, accumulation and fragmentation of the ions in the quadrupole mass filter occurs simultaneously in parallel with the detection of transients in the orbitrap, so that the completion of a single top10 cycle only takes 1s.

As such, their use in proteomics allowed for a deeper probing of the proteome never before realized.

NanoLC-interfaced orbitrap mass spectrometers generally yield thousands of MS/MS spectra in a single chromatographic run. Individual inspection of these spectra is tedious, and bioinformatics methods to address this issue have been developed through the years. The tenet of protein identification in these methods can be explained as follows. Consider a peptide sequence shown in **Figure 11**, where it is shown that it can fragment in a variety of ways to yield different types of ion series during MS/MS. The most dominant of this is the b-ion series, where the ions are formed from amide bond cleavage on the C-terminal end of the peptide (Hunt, Yates et al. 1986). In an ideal situation, the sequence of the original peptide can be derived simply from calculating the mass differences of the peaks in a sequential manner from the high mass to the low mass end (or vice versa), and comparing these losses to the unique masses of the known amino acids, a method known as *de novo sequencing* (Medzihradszky and Chalkley 2013). However, in the real situation, many elements can compound a mass spectrum. For instance, the reverse sequence, the y-type ion series, which occurs if the charge is retained on the amine group during fragmentation, also appears in the spectrum and has an equal likelihood to appear (Biemann 1990). Sometimes the a-type ion series which are produced during the decarboxylation of the b-type ions can also be present. Further compounding the spectrum is the fact that the ion series may be incomplete because of variations in peak intensities due to the preference of the fragmentation to certain types of mechanisms.

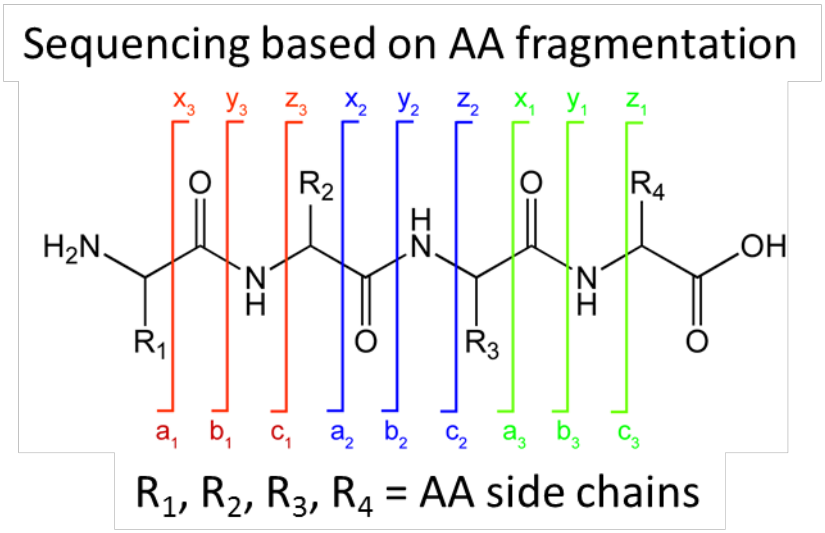


Figure 11. Peptide ion series nomenclature.

The most commonly observed peptide fragments in low energy collisions such as in ion trap, triple quadrupole and q-TOF mass spectrometers are the b, y and a ions. b-ions extend from the N-terminus while the y-ions extend from the C-terminus. a-ions are produced by decarboxylation of b-ions are diagnostic of their presence. c- and z-ions, on the other hand, are produced by peptide backbone fragmentation following a non-ergodic process when exposed to electrons of near thermal energies. Adapted from (Roepstorff and Fohlman 1984) and (Biemann 1988).

Nonetheless, since the probability of occurrence of the ion series is known, it is actually possible to produce a hypothetical spectrum considering all the factors mentioned. This system works if there is prior knowledge of the peptide sequence but fails, however, in discovery-driven approaches. The alternative approach in this case is to guess every amino acid sequence possible, build hypothetical spectra from these, and find the best match (Spengler 2004). Even if this is possible, it becomes impossibly difficult after 7-8 amino acids because of the increasing false positive rates. Even so, the idea to come up with hypothetical spectra for every possible sequence permutation became the basis for subsequent improvement. With the emergence of the whole genome sequencing era, it became apparent that the genome sequencing technique can also be applied to tandem MS data. Thus, instead of searching all possible peptide sequences, a significantly smaller genome database can instead be accessed. This translates to a 10,000 fold decrease in the amount of sequences that needs to be examined for a 20-mer set of non-redundant amino acid sequences currently available in the National Center for Biotechnology Information (NCBI) database.

With the possibility of doing hypothetical sequence matching came the construction of algorithms to systematically treat data (Johnson, Davis et al. 2005; Kapp and Schutz 2007; Cottrell 2011). SEQUEST, a search engine developed by Eng and Yates in 1994, made significant leaps in this field ((Eng, McCormack et al. 1994), **Figure 12**).

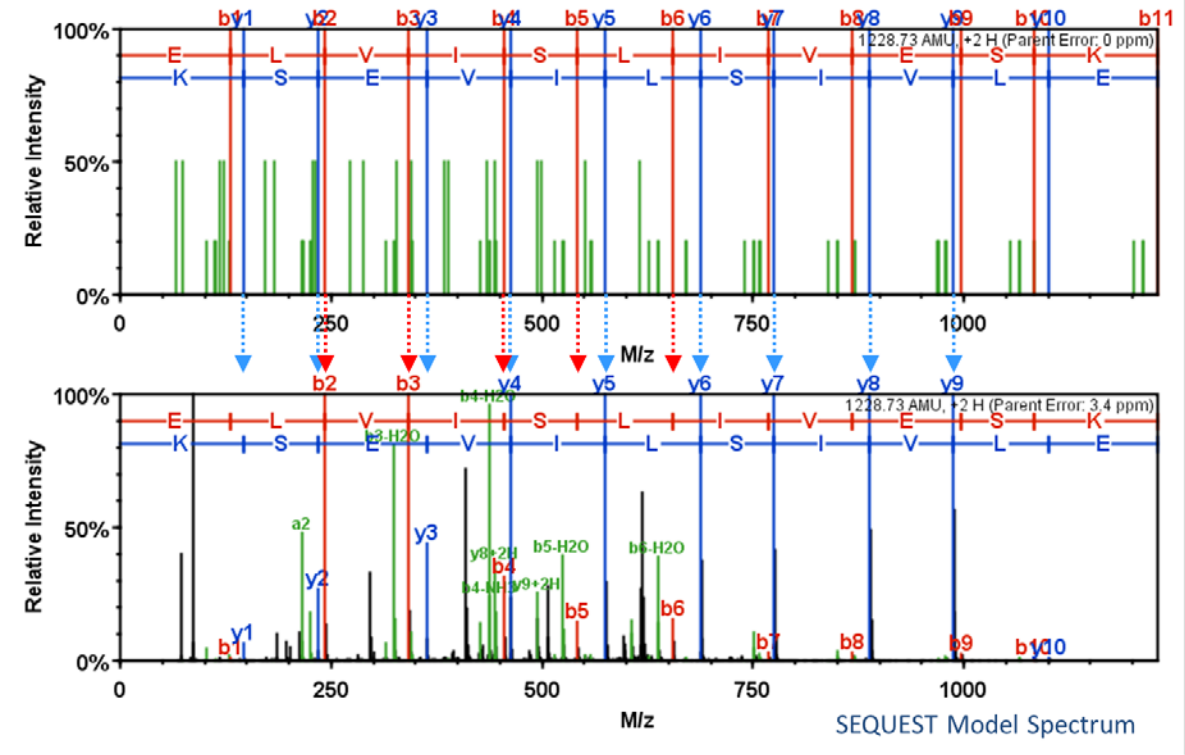


Figure 12. Matching spectrum with a hypothetical model.

The SEQUEST algorithm works by obtaining the peak overlaps between the normalized spectrum and the SEQUEST-generated model spectrum for a specific peptide sequence and taking its ratio against the peak overlaps generated when the spectrum is shifted 75 amu forward and back.

First they normalized spectra so that the inherent discontinuity in the intensity of peaks between actual and hypothetical spectra is reduced. Peak overlaps are then summed up (called cross-correlation). They next shifted the spectra 75 amu forward and back so that there will be no alignment, and used this term for background (called auto-correlation). Xcorr, the parameter used in SEQUEST to evaluate the accuracy of the peptide match, is then computed as the ratio of the cross-correlation of the average of the auto-correlation for the 75 amu range :

$$\mathrm{XCorr} = \frac{CrossCorr}{avg\left(AutoCorr_{\text{offset=-75 to 75}}\right)}$$

This implies that if the XCorr score is high, the direct comparison using overlapping peaks is significantly better than the background comparison using random overlaps. Another parameter called deltaCn measures how good an Xcorr of the best match is relative to that of the next best one :

$$\Delta C_n = \frac{XCorr_1 - XCorr_2}{XCorr_1}$$

Whereas Xcorr is a robust measure of the accuracy of the spectral match, deltaCn is a crude calculation of the goodness of this match, and as such other algorithms were developed. Mascot and X!Tandem are algorithms that focus more on the success of the relative score (Kapp, Schutz et al. 2005). Instead of just considering the best match relative to the second best one, X!Tandem considers the exponential decay distribution of lower scoring hits, assuming that they were all wrong. If the distribution represents the number of random matches for a given score, then linear fitting should be able to provide the expected number of random matches for a given score. As such, the probability that a given score (such as the one used for the best match) is random can be assessed (called the expected-value or E-value). Mascot attempts to arrive at the same result but using theoretical computations most possibly based on the number of identified peptides and the probability of occurrence of certain amino acids in the genomic database.

More recent developments in protein identification strategies benefit from the integration of several search algorithms to capitalize on the accuracy of the matching and the relativity of the scoring. This is done by converting the scores obtained for a protein identification from various search engine algorithms into probabilities. Note that it is meaningless to simply use a thresholding method when comparing the scores simply because they have been obtained using different parameters. For instance, SEQUEST considers relative peak intensities within a given spectrum, whereas X!Tandem considers only b/y-type ions instead of the entire model, and Mascot computes for theoretical p-values based on search engine space. One instance is the Protein Prophet algorithm (Nesvizhskii, Keller et al. 2003), which computes the true probability that a peptide identification is correct based on the comparison of the best match against all the other best matches in the entire sample. Doing so generates a bimodal distribution with the lower mode representing the most likely random matches (incorrect distribution) and the higher mode

representing the most likely probable ones (correct distribution). Using Bayesian statistics, the Protein Prophet computes for the probability of a score by obtaining the ratio of the number of matches of the correct distribution over the total distribution :

$$p(+\mid D) = \frac{p(D\mid +)p(+)}{p(D\mid +)p(+) \; + \; p(D\mid -)p(-)}$$

As the scores have been converted into probabilities, it now makes it easier to evaluate peptide matches from various search engine algorithms simply by setting thresholds. Commercial software such as Scaffold (Searle 2010) take advantage of these integrated algorithms to provide easy visualization and interpretation of large proteomic datasets, and their use is indespensable in current settings (**Figure 13**).

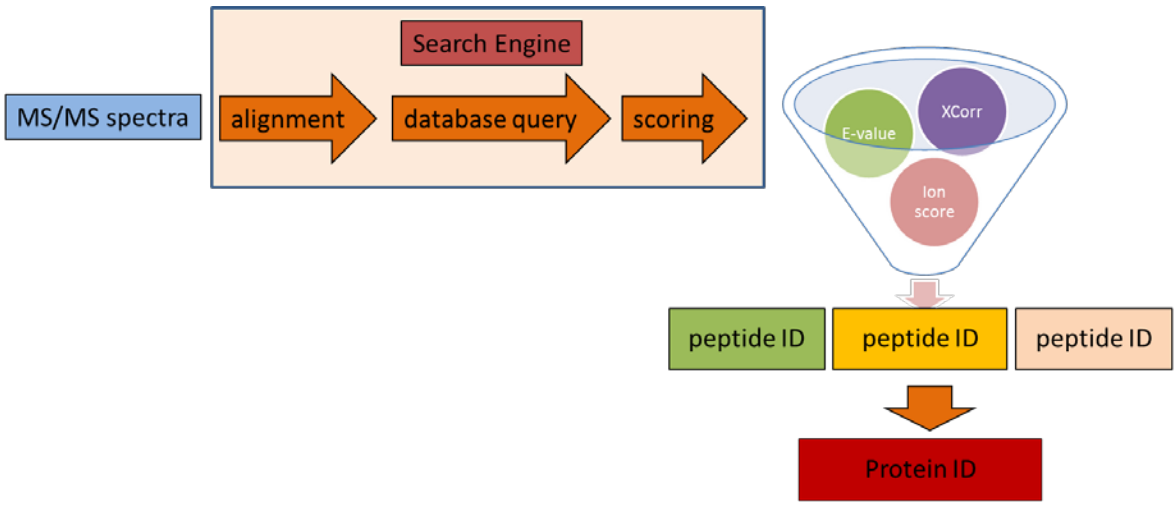


Figure 13. Example of a commercial pipeline for interpretation of large proteomic datasets.

After MS/MS spectra have been queried using search engines such as Mascot or Sequest, the protein identifications together with the scores provided by the search algorithms are exported. In Scaffold, the scores from various search algorithms are integrated using Peptide Prophet or other similar algorithms and are represented as peptide/protein probabilities. Additional features such as protein ID clustering, statistical analysis, or label-free quantification are also provided using a user-friendly interface.

ESI-MS/MS Databank Matching

One of the weaknesses of MSI in the analysis of proteins is the limited identification capabilities it can offer. Direct identification of intact proteins is currently limited at best to 30kDa (Franck, Longuespee et al. 2010), with the signals not even identified, but merely reported as m/z values. A simpler approach is to digest the proteins first and analyze the resulting peptides. In particular,

for MALDI MSI this is done by capitalizing on post source decay (PSD, (Spengler 1997)), a metastable type of fragmentation that occurs immediately after the ions exit the high field region of the source due to the excess internal energy that they possess by colliding with other ions or with gas. In MALDI-TOF/TOF instruments, a lift cell is employed in order to accelerate, collect and resolve these fragments and provides the MS/MS spectrum depending on the isolated m/z. Even so, PSD provides limited MS/MS spectra mainly because PSD fragmentation occurs after the analyte ion has left the source, leading to significant ion losses especially for larger m/z values (Durrant and Brown 2009). Increasing the laser energy may increase the amount of fragments that can be generated, but this is at the expense of losing spectral resolution. A further limitation is the innate complexity of the tissue matrix, which induces ion suppression. As such, direct MS/MS on tissue is scarcely performed, and often only leads to the identification of 10 or so abundant proteins or endogenous peptides. This can be further improved by introducing ion mobility to separate isobaric ions (Stauber, MacAleese et al. 2010), nonetheless, the single charge ion production of MALDI sources poses a serious drawback to begin with and limits the overall efficiency of the fragmentation in this method.

An alternative approach is done by subjecting an adjacent tissue section to protein extraction and nanoLC-MS analysis, generating a database of peptide and protein IDs that are then matched with the peptide masses obtained from the MSI of the other tissue section (Chaurand 2012; Mascini and Heeren 2012). For this approach to be effective, it is necessary to meet two criteria: 1) the accuracy of both MSI and LC-MS should be high, in order to avoid false identifications, and 2) the peptides in the tissue subjected to LC-MS analysis should be representative of those in the tissue subjected to MSI. The first criterion can be met by using high-accuracy instruments in both analyses, so that a maximum deviation of <10 ppm can be imposed during matching (Schober, Schramm et al. 2011). For MALDI-TOF instruments, such a criterion can be met by adding internal calibrants during the sample preparation (Gustafsson, Eddes et al. 2012). The second criterion, on the other hand, can be met by using the same sample block for protein extraction in both analyses. In addition to imposing a very low maximum deviation value, additional confirmatory analysis can also be performed to validate the protein identifications, such as by imuunohistochemistry, although such validation can be limited in terms of the number of proteins that can be validated due to expense.

It must be noted that although the generation of MS/MS databanks for matching with MSI data is an attractive means to provide unequivocal protein identifications, the approach does not really address the major limitation of MALDI-based approaches in that it is restricted to single charge ion production and thus limited protein identification utility (McDonnell, Corthals et al. 2010). Moreover, in the absence of reliable tandem MS spectra, correlation is highly dependent on the accuracy of the mass spectrometer and validation using complementary methods. Thus, there is still an on-going effort to merge the advantages of the two approaches into a single analysis.

Direct Quantification Methods

MS-based direct quantification methods pertain to those methods that allow the quantification of analytes directly on sample surfaces with minimal sample preparation and at the same time still keeping the analyte localization intact. This excludes stable isotope-based methods that entail extensive and rigorous sample preparative steps and where analyte localization is typically lost during sample processing. Quantification can be absolute or relative, and specific approaches have been developed depending on the type of direct MS analysis used.

Peak intensities in a mass spectrum do not necessarily reflect the amount of analyte in a sample, although variation in these intensities for the same analyte across multiple samples can be indicative of varying abundance (Nikolov, Schmidt et al. 2012). Although peak intensities cannot be harnessed directly, their chromatographic traces can be monitored by constructing the elution profile of the ion of interest post-acquistion. Each elution profile can be refered to as the feature of that ion, and is the sum of all peaks detected in all consecutive mass scans collected at the retention time dimension. As such a feature is the three-dimensional representation of the m/z, peak intensity, and retention time of an analyte ion. Features can be detected and measured using the extracted ion chromatogram (XIC, (Higgs, Knierman et al. 2005)). This method provides a simplest means to quantify analyte signal by simply plotting the total intensity of a selected ion for a specified m/z window at a given retention time. Here the retention time is the apex of the intensity of the feature (Sandin, Teleman et al. 2014). Detection and identification of features can be done simultaneously. In label-free methods (ie., quantitative methods that do not require stable isotope incorporation) utilizing tandem MS, the scanning MS mode provides the MS1 spectra where the XIC is extracted, while the data-dependent MS2 spectra provide the peptide identifications (Nahnsen, Bielow et al. 2013). Such feature-based detection methods are the basis of such quantification softwares as MaxQuant, Skyline and Spectrum Mill (see (Sandin, Teleman

et al. 2014) for the comprehensive list). These methods are extensively used because of the accuracy of the quantitative measurements they produce.

Another label-free method, called spectral counting, utilizes the linear relationship between the number of tandem MS/MS spectra used for protein identification and the abundance of that protein to derive relative quantitative information at the protein level (Neilson, Ali et al. 2011). This is possible because increased protein abundance leads to the generation of more tryptic peptides during the digestion step, and consequently results in increased protein coverage, number of unique peptide identifications, and eventually number of MS/MS spectra used for protein identification. Although less accurate than feature detection-based methods, spectral counting is more reproducible and provides a broader dynamic range (Zhu, Smith et al. 2009). Several improvements were done on this method, one of them corrects for the fact that larger proteins essentially generate more spectra compared to smaller ones. The normalized spectral abundance factor (NSAF, (Zybailov, Florens et al. 2007)) was defined to account for the variations in spectral count due to differences in potein length. Building on the spectral counting concept, absolute quantification strategies such as protein abundance index (PAI, (Rappsilber, Ryder et al. 2002)) and absolute protein expression (APEX, (Lu, Vogel et al. 2007)) have also been developed, among others; they mostly deal with the manner in which peptide spectra are being utilized to make quantitative inferences at the protein level. In PAI for example, a ratio of the identified peptides over the number of theoretical observable tryptic peptides is computed while in APEX, absolute protein concentration per cell is obtained.

In another approach, transitions that occur as one selected peak is subjected to tandem MS fragmentation can be followed, and these transitions can actually accurately determine analyte quantity. Selected reaction monitoring (SRM) is a targeted approach where precursor-product ion transitions are monitored over time to yield chronograms with the retention time and the intensity of the specific transition as coordinates (Lange, Picotti et al. 2008). Filtering at the two stages can effectively eliminate background ions, resulting in high selectivity. Unlike full scans, an SRM scan is conducted only at a narrow m/z range resulting in improved sensitivity and a linear response over a wide dynamic range. It is thus ideal for the quantification of less abundant components of the sample mixture.

In contrast to conventional shotgun proteomics approaches, SRM is a targeted method aimed at quantification of pre-selected peptides, the design of which depends on the peptide being examined. Prior knowledge of the transitions is necessary for an effective SRM measurement (Sandin, Teleman et al. 2014). Specifically, proteins of the desired protein set have to be selected and can be done, for example, by subjecting an aliquot of the sample to shotgun LC-MS. Next, peptides that provide good MS spectra and uniquely identify these proteins (called proteotypic peptides) have to be identified, usually by scanning through the shotgun data previously generated or by computational methods (Mallick, Schirle et al. 2007). Then, fragments that discriminate one peptide from another and provide optimal intensity during MS/MS have to be identified. Rigorous time and effort is needed to conduct a good SRM experiment, nonetheless, te meticulousness of this method pays off because once the method has been established, it can easily be implemented, used indefinitely, and multiplexed with other SRM methods.

In contrast to the methods previously discussed, MSI-based methods have been developed independently mainly because the heterogeneous nature of ionization has its own distinct challenges observable only in MSI (Gessel, Norris et al. 2014). A robust quantitative MSI method must take into consideration the fact that the analyte MS signal is dependent not only on the concentration but also on the efficiency by which the analyte has been ionized and also to its extractability from the sample. In matrix-based MSI, variations in matrix crystallization due to inhomogeneities present on the tissue further hamper the reproducibility of quantitative methods. Because of these limitations, current MSI-based quantification methods are mostly restricted to the analysis of drugs that have sufficient ionization yields and whose fragments can easily be monitored by SRM experiments if the need arises (Lietz, Gemperline et al. 2013). The most often used method nowadays relies on the addition of internal calibrants during the sample preparation steps, the MS responses of which are then used to make corrections during analyte concentration calculation (Gustafsson, Eddes et al. 2012). Internal calibrants used are usually stable isotope-labeled variants of the analyte. These are either deposited on the sample section together with the matrix, administered to the organism prior to sample procurement, or applied to the surface prior to tissue mounting. Varying concentrations of the internal standard on the tissue can be used to construct calibration curves that can then be used to deduce the absolute concentration of the analyte. Notable applications of this approach have been reported for MALDI MSI and ambient source imaging (Pirman, Reich et al. 2012; Vismeh, Waldon et al. 2012; Lanekoff, Thomas et al. 2014). Another approach, pioneered by Setou's group in 2010, uses a correction factor to account

for the signal intensity variations across different types of tissue (Hattori, Kajimura et al. 2010). Building upon this approach, a homologous correction factor called tissue extinction coefficient (TEC) can be determined for a specific analyte on various types of tissue and these factor is then taken into consideration during the calibration curve construction (Hamm, Bonnel et al. 2012). It is to be noted that, at the present, no imaging-based protein quantification strategies have yet been developed, owing to the aforementioned challenges.

PART 1 : IMPROVEMENT OF ON-TISSUE IDENTIFICATION STRATEGIES FOR MALDI IMAGING : TOWARDS SINGLE-PIXEL IDENTIFICATION

Chapter 1

Article 1

Development of Liquid Microjunction Extraction Strategy for Improving Protein Identification from Tissue Sections

Authors: Jusal Quanico, Julien Franck, Claire Dauly, Kerstin Strupat, Jocelyn Dupuy, Robert Day, Michel Salzet, Isabelle Fournier, Maxence Wisztorski

Article Status: published in Journal of Proteomics [Volume 79, pages 200-218, 2013]

Summary:

Experimental manipulations, including protocol optimization, sample preparation, and data gathering, were realized by Mr. Quanico. He also contributed to the design of experimental set-ups and decision making. In detail, my contributions include:

1. Implementation and optimization of the liquid extraction surface analysis strategy prior to its use in this proteomics analysis, first by coupling on-line with an ion trap MS instrument, then by off-line analysis;
2. Perform all MALDI-MS imaging experiments and MS imaging-related procedures such as sample preparation and enzymatic digestion, matrix application, and image acquisition and analysis; and
3. Perform the on-tissue microdigestion and microextraction using the microspotter and the developed liquid microjunction extraction protocol, followed by extract handling prior to nanoLC-MS analysis.
4. In terms of writing the manuscript, Mr. Quanico wrote the drafts which were then built upon in the subsequent writing of the final manuscript. He also edited the manuscript.

Resumé en français

La technique d'imagerie par spectrométrie de masse (MSI) MALDI a montré un grand potentiel pour la classification moléculaire et l'identification de marqueur de pathologies. Néanmoins, l'identification in situ des protéines reste très limitée due à la trop haute complexité des tissus. Par conséquent, seules les protéines les plus abondantes peuvent alors être identifiées.Les approches protéomiques conventionnelles quant à elles, permettent l'identification d'un grand nombre de protéines aussi bien majoritaires que minoritaires et de façon plus fiable mais la notion de localisation est alors perdue. Dans cet article est présenté une approche permettant l'identification de protéines d'abondance aussi bien élevée que faible à partir de coupes de tissu, tout en conservant la donnée de localisation. Cette techniquerepose sur une stratégie de micro-extraction par micro-jonction liquide sur des régions de coupes de tissus préalablement digérées à la trypsineCette stratégie s'est montrée très efficace pour extraire les peptides de digestion qui, après séparation par chromatographie liquide en phase inverse couplée à un instrument HR FTMS pour une analyse en mode MS/MS a permis l'identification moyenne de 1500 protéines à partir d'une région d'environ 650µm de diamètre correspondant à environ1900 cellules. Cette approche peut être facilement intégrée pour une approche de type MALDI-MSI donnant ainsi un fort potentiel pour des applications cliniques.

DEVELOPMENT OF LIQUID MICROJUNCTION EXTRACTION STRATEGY FOR IMPROVING PROTEIN IDENTIFICATION FROM TISSUE SECTIONS

Jusal Quanico[1,4]*, Julien Franck[1]*, Claire Dauly[2], Kerstin Strupat[3], Jocelyn Dupuy[2], Robert Day[4], Michel Salzet[1], Isabelle Fournier[1], Maxence Wisztorski[1]

[1]Université Lille Nord de France, Laboratoire de Spectrométrie de Masse Biologique Fondamentale et Appliquée - EA 4550, Bât SN3, 1er étage, Université de Lille 1, F-59655 Villeneuve d'Ascq, France

[2]Thermo Fisher Scientific France, 16 avenue du Québec, SILIC 765, 91963 Courtaboeuf Cedex, France

[3]Thermo Fisher Scientific, Hanna-Kunath-Strasse 11, 28199 Bremen, Germany

[4]Institut de pharmacologie de Sherbrooke, Université de Sherbrooke,Sherbrooke, Québec, J1H 5N4, Canada

Corresponding authors: Isabelle Fournier or Michel Salzet : Université Lille Nord de France, Laboratoire de Spectrométrie de Masse Biologique Fondamentale et Appliquée - EA 4550, Bât SN3, 1er étage, Université de Lille 1, F-59655 Villeneuve d'Ascq, France

Isabelle.fournier@univ-lille1.fr or michel.salzet@univ-lille1.fr

Tel: +33 3 2043 4194 , Fax: +33 3 2043 4054

*Coauthors

ABREVATIONS

MSI : Mass Spectrometry Imaging ;**RPLC :** Reverse Phase Liquid Chromatography; **LCM:** Laser Capture Microdissection; **ROI:** Region of Interest; **PCA:** Principal Component Analysis; **HC:** Hierarchical Clustering, **FFPE:** Formalin Fixed and Paraffin Embedded; **TOF:** Time-of-Flight; **ISD:** In Source Decay, **ETD:** Electron Transfer Dissociation; **DESI:** Desorption Electrospray Ionization; **LESA:** Liquid Extraction Surface Analysis; **DBS:** Dried Blot Spot; **LTQ:** Linear ion Trap; **MeOH:** Methanol; EtOH: Ethanol; **HCCA:**Δ-cyano-4-hydroxycinnamic acid; SA: Sinapinic Acid; **ANI:** Aniline; **HES:** Haematoxylin Eosin Safran; ITO: Indium Tin Oxyde; Nd:YAG: Neodymium-doped Yttrium Aluminium Garnet; **BSA:** Bovine Serum Albumin; CID: Collision induced Dissociation; **FWHM:** Full Width at Half Maximum; **amu.:** Atomic Mass unit; **FDR:** False Discovery Rate; **S/N:** Signal to Noise; **ID:** identification; **ppm:** Part Per Million; **LMJ-SSP:** Liquid Microjunction-Surface Sampling Probe; **SSSP:** Sealing Surface Sampling Probe; **CE:** Capillary Electrophoresis.

SUMMARY

MALDI Mass Spectrometry Imaging has shown important potential for molecular classification and pathology marker discovery. Direct structural analysis from tissue sections has shown limitations for protein identification because of the high degree of complexity of tissues. Only proteins of major abundance are identified this way. On the contrary, conventional proteomics approaches clearly allow for reliable identification of complex protein extracts but does not provide fine correlation with protein location in their original context. Here is presented an approach to obtain identification of proteins of various abundances while keeping their localization within the section. On-tissue trypsin digestion followed by micro-extraction using a liquid micro-junction interface is an efficient strategy to extract tryptic peptides and further identify the associated proteins off tissues. It was shown that conventional Reverse Phase Liquid Chromatography separation on the extracted material followed by MS/MS analysis on a HR FTMS instrument enabled the identification 1500 proteins on average with high confidence from an area of about 650 μm in diameter, which corresponds to an estimated number of 1900 cells in average. The approach can be easily integrated in the MALDI MSI workflow and should provide interesting insights for clinical applications.

INTRODUCTION

Identification of molecular components involved in cellular mechanisms and their fine regulation is a key point for understanding physiological processes occurring in living organisms. Cell dysfunction processes greatly impact signaling pathways by modifications of molecular players. In particular, proteins are largely affected by cellular changes. These changes have an impact on their regulation (activation/repression of protein synthesis), primary structure (protein cleavage by enzymes, post-translational modifications), spatial arrangement and ability to complex with other partners (directly related to changes in protein concentrations and primary structure or due to changes in the environment such as ionic strength or pH). Over the past 25 years, many efforts were given to gain protein structural information and quantification. To date, proteomics offers a wide range of high performance strategies, methodologies and instrumentation to reach this goal. However, if conventional proteomics offers access to a large number of identified proteins, there is increasing evidence suggesting that such strategies lack correlation of the identified proteins to their location in their original context (i.e. tissues). For tissues, laser capture microdissection (LCM) systems have been developed to collect cells of specific phenotypes based on morphological criteria recognition (Craven and Banks 2001; Craven and Banks 2003; Mustafa, Kros et al. 2008; Rodriguez, Espina et al. 2008). Currently, for cell collection from tissues, LCM is the better suited method. However, LCM remains technically somehow difficult to setup whatever the used technology. For proteomics, sample collection is time-consuming because of the large number of cells needed to be collected to permit detection given the instrument available. Indeed, few thousands up to several dozen thousands might have to be captured for proteomics analyses. Studies show a clear correlation of collected cell numbers with the number of subsequently identified proteins. Previous studies show identification of about 100-200 proteins for 500 cells, 500-700 proteins for 5000 cells and 1500-2500 proteins for 50 000-60 000 cells (Cha, Imielinski et al. 2010; Wang, Xu et al. 2010; Stingl, van Vilsteren et al. 2011; Thakur, Rejtar et al. 2011; Wisniewski, Ostasiewicz et al. 2011). However, automatic cell collection, through the use of morphological recognition dedicated tools, is not always very reliable and manual collection is very time consuming. Additionally, it must also be noted that cell collection for LCM is based on morphological criteria and thus does not guarantee that all cells are involved in the same physiological processes due to the cells' local microenvironment and cross-talk.

On the other hand, MALDI Mass Spectrometry Imaging (MSI) is a molecular imaging tool that can be used to study the spatial distribution of endogenous and exogenous compounds including, drugs (Khatib-Shahidi, Andersson et al. 2006; Bouslimani, Bec et al. 2010; Goodwin, Macintyre et al. 2010; Goodwin, Scullion et al. 2010; Solon, Schweitzer et al. 2010; Castellino, Groseclose et al. 2011; Goodwin, Mackay et al. 2011; Prideaux, Dartois et al. 2011), metabolites (Esquenazi, Coates et al. 2008; Miura, Fujimura et al. 2010; Van Dyck, Flammang et al. 2010; Porta, Grivet et al. 2011; Lee, Perdian et al. 2012; Sarsby, Towers et al. 2012; Yoshimura, Zaima et al. 2012), lipids (Chen, Hui et al. 2009; Djidja, Francese et al. 2009; Chen, Jiang et al. 2010; Rompp, Guenther et al. 2010; Colsch, Jackson et al. 2011; Delvolve, Colsch et al. 2011; Veloso, Astigarraga et al. 2011), peptides (Hanrieder, Ljungdahl et al. 2011; Meriaux, Arafah et al. 2011; Minerva, Boonen et al. 2011) and proteins (Seeley and Caprioli 2008; Grassl, Taylor et al. 2011) from both vegetal (Kaspar, Peukert et al. 2011) and animal models. MALDI MSI was found to be a powerful technology for many fields of research such as pharmaceuticals (Bonnel, Legouffe et al. 2011), biomarker discovery and tracking for various pathologies e.g. brain disorders (Stoeckli, Staab et al. 2002; Stauber, Lemaire et al. 2008; Hanrieder, Ljungdahl et al. 2011) or oncology (Lemaire, Menguellet et al. 2007; El Ayed, Bonnel et al. 2010; Rauser, Marquardt et al. 2010; Willems, van Remoortere et al. 2010; Balluff, Rauser et al. 2011; Le Faouder, Laouirem et al. 2011; Elsner, Rauser et al. 2012; Meding, Nitsche et al. 2012; Nipp, Elsner et al. 2012). To highlight the regulation of molecules with important functions related to a specific biological process by a simple MALDI MSI acquisition, these molecules have to be subjected to further characterization. For several years, many groups have been developing and improving on tissue protein identification strategies. In particular, on tissue Bottom-Up approaches through i*n situ* enzymatic digestion have been demonstrated to allow direct identification of proteins from both frozen (Groseclose, Andersson et al. 2007) and formalin fixed paraffin embedded (FFPE) (Lemaire, Desmons et al. 2007) tissues. In order to maintain localization of peptides throughout the digestion process, methodologies were developed to perform enzymatic digestion from discrete locations on the tissues using a micro-spotter (Groseclose, Andersson et al. 2007; Lemaire, Desmons et al. 2007; Franck, Arafah et al. 2009) or from whole tissue sections with micro-sprayer devices (Djidja, Claude et al. 2010; Gustafsson, Oehler et al. 2010). In this bottom-up approach, tryptic peptides are subsequently analyzed by MS^2 and identified upon databank interrogation. However, even if MS spectra on each pixel show hundreds of peptides, only few proteins are identified (Schober, Schramm et al. 2011) . One of the ways to increase the number of protein identifications is the improvement of tryptic peptide detection. This has been done by

incorporation of tissue treatments in the strategy, such as washing to minimize ion suppression effect (Franck, Arafah et al. 2009) . Moreover, peptides with a very close m/z that enter and get fragmented together at the collision cell preventing the identification were separated by the inclusion of a gas phase separation using ion mobility (McLean, Ridenour et al. 2007; Djidja, Claude et al. 2009; Stauber, MacAleese et al.). In addition, improvement of protein fragmentation to aid in identification was also examined by N-terminal derivatization. This orientates the fragmentation of tryptic peptides towards a specific ion series, increasing the protein identification score (Franck, El Ayed et al. 2009). All these efforts have led to better protein identification, however, the less abundant proteins which present significant biological activities, in particular pathologies, remain difficult to detect and identify. More recently, on-tissue top-down strategy (Debois, Bertrand et al. 2010; Zimmerman, Debois et al.) was introduced by means of In Source Decay (ISD) (Brown and Lennon 1995; Hardouin 2007) . This method of fragmentation is generally well-suited for purified proteins and leads to large sequence determination. However, *in situ* ISD has allowed the detection and identification only of the most abundant protein from the pixel analyzed (Debois, Bertrand et al. 2010) due to the high complexity of the sample.

Another strategy consists of combining MALDI-MSI and LC-MS/MS in a single workflow, allowing the improvement of protein identification. Basically, proteins are *in situ* digested, fractionated and then extracted prior to nanoLC separation, followed by MS/MS analysis for databank interrogation (Stauber, Lemaire et al. 2008). This approach was shown to improve protein identification, but since it was performed on a whole or half of a tissue section, information about protein localization is lost. This can be regained if back correlation to imaging data of tryptic peptides is done, as previously demonstrated for FFPE tissue samples (Lemaire, Desmons et al. 2007; Stauber, Lemaire et al. 2008). More recently, intact proteins were extracted from tissue prior to fractionation by ultracentrifugation, and, in combination with high resolution and accurate mass determination, has allowed the detection of about a hundred proteins (Schober, Schramm et al. 2011) . Although in this approach, the localization of proteins is lost, it can be regained by correlation with the tryptic peptide signals from the image of an adjacent section. The number of identified and localized proteins improved, but the approach continues to be limited by ion suppression effect and decreased sensitivity due to dominant signals from the most abundant proteins. Localized intact proteins provided by MALDI-MSI can also be identified after their extraction from a consecutive section prior to analysis by nanoLC-MS/MS using electron

transfer dissociation (ETD) as activation mode. This strategy was used for clinical applications in oncology and has lead to the identification of several markers (Rauser, Marquardt et al. 2010; Elsner, Rauser et al. 2012; Nipp, Elsner et al. 2012). However, only the highly soluble proteins were identified and localized in this manner, and the correlation was not accurate because the image was obtained using a MALDI TOF instrument. Moreover, proteins of interest could be diluted among the large amount of abundant proteins, especially if the region of interest is small compared to the size of the sample.

There is therefore a crucial need to develop a new strategy that allows protein identification on discrete regions of tissue sections, avoiding dilution of low abundant proteins. Recently, surface sampling using liquid microjunction was introduced demonstrating liquid-solid micro-extraction that can be coupled to any ionization sources (Van Berkel, Kertesz et al. 2008). The success of these methods as alternatives to typical ambient surface sampling/ionization sources is due to their greater extracting efficiency and larger surface sampling areas. For example, it has been shown that metabolites of a drug could be detected using liquid microjunction extraction, whereas they were unsuccessfully detected using DESI (Kertesz and Van Berkel 2010). In one of these methods called liquid extraction surface analysis (LESA) (Kertesz and Van Berkel 2010), the solvent can be made in contact with the sample surface by dispensing in a controlled fashion discrete droplets. Molecules extracted can be subjected to further pretreatments including desalting and separation steps prior to MS analysis.

In this paper, a new strategy is presented for protein identification from discrete locations on tissue sections determined by hierarchical clustering (HC) (Deininger, Ebert et al. 2008) that can be used to classify cell groups according to their molecular content profiles. Proteins are then extracted by liquid microjunction using LESA followed by nanoLC-MS/MS analysis. In particular, localized on-tissue enzymatic digestion followed by micro-extraction and subsequent analysis on an HR (high resolution) ESI MS instrument providing high accurate mass measurement was evaluated. This strategy can be integrated within the MALDI MSI workflow in order to perform supervised profiling on distinct cell groups having specific molecular profiles. This robust and fast new strategy shows to be a good alternative than LCM and leads to identification of a larger number of proteins with respect to the number of cells studied.

EXPERIMENTAL PROCEDURES

Chemicals

All chemicals were of the highest purity obtainable. Water, Formic Acid (FA), trifluoroacetic acid (TFA), Acetonitrile (ACN), methanol (MeOH) are purchased from Biosolve B.V. (Valkenswaard, the Netherlands). Ammonium bicarbonate (NH_4HCO_3), Ethanol (EtOH), α-Cyano-4-hydroxycinnamic acid (HCCA) (puriss), sinapinic acid (SA), aniline (ANI), and reagents for HES staining were purchased from Sigma-Aldrich (Saint-Quentin Fallavier, France). Trypsin (sequencing grade modified trypsin, porcine) was purchased from Promega (Charbonnieres, France). A standard digest of BSA was obtained from Bruker Daltonics (Bruker Care Online, Part-No. #217498) and used without any further purification.

Tissue sample preparation

Male Wistar rats at adult stage maintained under standard care were used for tissue collection. Animals were sacrificed by CO_2 gas and brains were dissected. Tissues were subsequently frozen into isopentane cooled at -50°C. Samples were then stored at -80°C until required. 10µm rat brain tissue sections were obtained using a cryostat CM1510S (Leica Microsystems, Nanterre, France) and thaw mounted onto indium-tin oxide (ITO)-coated conductive glass slides (Bruker Daltonics, Bremen, Germany). The tissue sections were vacuum-dried in a desiccator for10 min then soaked in 70% EtOH, 95% EtOH and chloroform for 30s each prior to vacuum dessication for 5min. The sections were then dried and stored at -20°C until use (Longuespee, Gagnon et al. 2013). For some sections, after MALDI analysis, the matrix was removed by washing the slides with 70% ethanol and the sections were HES-stained.

MALDI-TOF MSI Experiments

MALDI matrix was applied using the ImagePrep device (Bruker Daltonics, GmbH), by spraying SA/ANI matrix (1 equivalent of aniline (20µl for 5 ml) in 10 mg/ml in 70:30 ACN/H2O, 0.1%TFA of SA solution).Protein imaging was performed on an Ultraflex II MALDI-TOF/TOF instrument (Bruker Daltonics, GmbH) equipped with a smartbeam laser (Nd:YAG, 355nm). Protein mass spectra were acquired in linear positive ion mode within a mass range of m/z 3000-30000. Distance between raster points was set to 300µm and a total of 300 laser shots were accumulated at 200-Hz repetition rate for each pixel. Spectra were processed by baseline

correction and smoothing using FlexAnalysis 3.2 software (Bruker Daltonics, GmbH). Image analysis and data visualization were performed with the FlexImaging 2.1 software (Bruker Daltonics, GmbH).

For statistical analysis, the data set obtained from FlexImaging was loaded into the ClinProTools 2.2 software (Bruker Daltonics, GmbH) to conduct hierarchical clustering. Unsupervised clustering was selected with Euclidean as distance method and ward as linkage method. Results of hierarchical clustering were exported to FlexImaging 2.1 software to retrieve the areas with different profiles.

MALDI-Orbitrap MSI Experiments

Automatic trypsin digestion was performed using a high position accuracy chemical inkjet printer (CHIP-1000, Shimadzu Biotech, Kyoto, Japan) (Franck, Arafah et al. 2009). The printed array of the CHIP-1000 on the tissue section comprised of spots spaced by 250µm center-to-center. A total amount of 40 nL of solution containing 20µg/mL of trypsin in 50mM NH_4HCO_3 was applied on each spot. Five droplets of approximately 100pL were deposited on each spot per cycle. One hour was necessary to obtain the final volume and to perform sufficient digestion with no additional incubation time. After digestion, HCCA/ANI (1.5 equivalent of aniline (36µl for 5 ml) in 10mg/mL in 7:3 ACN/ water/ 0.1% TFA of HCCA solution.) was deposited on the same positions using the ChIP-1000, with the same parameters used during trypsin deposition.

MS imaging was conducted using a MALDI linear ion trap – orbitrap hybrid mass spectrometer (MALDI LTQ Orbitrap XL, Thermo Fisher Scientific). Spectra were acquired using Xcalibur 2.0.7 SP1 software (Thermo Fisher Scientific, Bremen, Germany). The LTQ Orbitrap instrument was operated in positive ion mode over a high mass range (m/z 800-4000). The instrument was equipped with an intermediate pressure (75 mTorr) MALDI source using a nitrogen laser (MNL 100; Lasertechnik Berlin), providing an output at 337 nm with a maximum energy of 80 µJ per pulse and a repetition rate of 60 Hz. The Orbitrap analyzer was used for MS data acquisition with a mass resolving power of 30,000 FWHM. Automatic gain control was turned off and other data acquisition conditions were as follows: laser power per pulse of 10 µJ, 25 laser pulses per spectra, and 2 scans averaged per pixel. Distance between raster points was set to 250µm. Image analysis and data visualization were performed with the Imagequest1.0.1 software (Thermo Fisher Scientific, Bremen, Germany).

On-tissue digestion and direct analysis

Digestion of a specific area on tissue section was performed using the CHIP-1000 previously described. The same parameters were used, except that this time the “fill plane” option was selected before trypsin deposition. After digestion, HCCA (10mg/mL in 7:3 ACN/ water/ 0.1% TFA) was deposited on the same positions using the CHIP-1000, with the same parameters as used during trypsin deposition. Direct tissue analysis was done on the MALDI Orbitrap XL used previously for MSI. Around 20 scans per digested point in FTMS mode in the range 600-4000 with a nominal -mass resolution of 100000 at m/z 400 were acquired.

Liquid microjunction extraction experiments

Micro-extraction experiments were performed using the TriVersa Nanomate platform (Advion BioSciences Inc., Ithaca, NY, USA), with the Liquid Extraction Surface Analysis (LESA) option (Kertesz and Van Berkel 2010). Extraction was performed by creating, on a specific location, a liquid microjunction where solid-liquid extraction can occur. The slide with the tissue section was mounted on a universal adapter plate and the precise X-Y position of the extraction site on the sample surface generated using LESA Points software. In most cases, 2µL of solvent was aspirated into the tip and used to extract peptides, with 0.6µl dispensed and made in contact with the tissue section for 30s. The solvent was then aspirated and dispensed onto a clean 0.2ml Eppendorf vial maintained at 4°C to avoid degradation. Here, unlike the conventional use of the microjunction extraction system where the extracted liquid is directly dispensed to the microfluidic nanoESI chip of the device; the extracted solution is collected in a tube. This allows performing several extraction cycles and pooling all collected solutions in the same vial to increase quantity of material collected. Iterations of extraction cycles were performed (3, 5 or 15) to determine the optimal number of extractions. Different extraction solvents were tested: ACN: 0.1% TFA aq (8:2, v/v), MeOH/0.1% TFA aq (7:3, v/v), 0.1% TFA aq, 1% TFA aq and EtOH/CH_3COOH/water (90:9:1, v/v/v). To determine which is the most efficient, 1µl of standard BSA tryptic digest solution was spotted on rat brain tissue sections and on this, micro-extraction was performed after the spots have completely dried. Collected extracts were subjected to MALDI-TOF analysis using the Ultraflex II instrument in reflector mode with HCCA as matrix, and the resulting PMFs were subjected to a MASCOT search algorithm (v2.3.01.241) against the SwissProt database (Swissprot_2011_08) using Biotools 3.1 software (Bruker Daltonics, Bremen, Germany). The mass tolerance was set to 0.2 Da. Three missed cleavages were allowed in order

to account for partial digestion. Cysteine carbamidomethylation was set as a fixed modification since the standard solution has undergone cysteine reduction/alkylation and that one tryptic peptide is expected to show carbamidomethylation (see product information). Methionine oxidation and tyrosine phosphorylation were set as variable modifications.

LC MS/MS

After micro-extraction, collected samples were dried under vacuum, reconstituted with 10µl of 0.1%TFA aq and subjected to solid phase extraction using a C18 Ziptip (Millipore,Saint-Quentin-en-Yvelines, France), eluted by 80% ACN, and dried under vacuum. Dried samples were suspended in water/5% acetonitrile/0.1% formic acid. Samples were separated by online reversed-phase chromatography using a Proxeon Easy-nLC system equipped with a Proxeon trap column (100 µm ID x 2 cm, Thermo Scientific) and C18 packed tip column (100 µm ID x 15 cm, Nikkyo Technos Co. Ltd). Peptides were separated using an increasing amount of acetonitrile (5%-40% over 110 minutes) and a flow rate of 300 nL/min. The LC eluent was electrosprayed directly from the analytical column and a voltage of 1.7 kV was applied via the liquid junction of the nanospray source. The chromatography system was coupled to a Thermo Scientific Orbitrap Elite mass spectrometer (ThermoFischer Scientific, France). The mass spectrometer was programmed to acquire in a data dependent mode. The survey scans were acquired in the Orbitrap mass analyzer operating at 120,000 (FWHM) resolving power. A mass range of 400 to 2000 m/z and a target of 1E6 ions were used for the survey scans. Precursors observed with an intensity over 500 counts were selected “on the fly” for ion trap collision-induced dissociation (CID) fragmentation with an isolation window of 2 amu and a normalized collision energy of 35%. A target of 5000 ions and a maximum injection time of 200 ms were used for CID MS^2 spectra. The method was set to analyze the top 20 most intense ions from the survey scan and a dynamic exclusion was enabled for 20 s.

Tandem mass spectra were processed with Thermo Scientific Proteome Discoverer software version 1.3. Spectra were searched against UniprotKB/Swiss-Prot (version January 2012) filtered with *Rattus norvegicus* taxonomy (41604 sequences) using the SEQUEST® algorithm. The search was performed choosing trypsin as the enzyme with two missed cleavages allowed. Precursor mass tolerance was 10 ppm and fragment mass tolerance was 0.5 Da. N-terminal acetylation, tyrosine phosphorylation, methionine oxidation and arginine deamidation were set as variable modifications. Peptide validation was performed with the Percolator algorithm. Peptides

were filtered based on a q-value below 0.01 which corresponds to a false discovery rate (FDR) of 1% which is referred in the following sections as “high confidence” identification.

RESULTS

On tissue direct identification of proteins by MALDI MSI

On-tissue identification of localized proteins was performed following the MALDI-MSI bottom-up approach. Trypsin was first deposited on an area of a control rat brain sagittal tissue section using a microspotter followed by depositing HCCA/ANI as matrix. **Figure 1** presents the results obtained through such a strategy combined to high spectral resolution analysis on a MALDI-LTQ Orbitrap instrument. **Figure 1a** shows the scanned image of the area analyzed on the tissue section showing the raster of micro spots. Images extracted from MALDI MSI data collected from this region are presented **Figure 1b** for three ions of distinct m/z at 1198.717 (green), 1612.950 (blue) and 2141.138 (red) respectively showing highly complementary distributions across the area. The tryptic digestion fingerprint MS spectrum extracted in one of the morphological region of the analyzed tissue (indicated by a dark circle) on **Figure 1b** is shown **Figure 1c**. The MS spectrum shows a complex digestion profile dominated by a few tens of major ions. The m/z list extracted from this spectrum after deisotoping and exclusion of trypsin autolysis ions (m/z 2211.1310; 2283.2016 and 1045.5776) shows the presence of around 350 monoisotopic ions in the m/z range [800-4000] with a S/N>5. Only 20 of these ions present a relative intensity >10% of the base peak, meaning that the 330 remaining ions are observed but are too low in intensity to generate MS^2 spectra that will lead to any identification of peptide. The experimental m/z of these 20 most intense ions were compared to the predicted m/z from *in silico* digestion of Myelin Basic Protein S (accession number P02688). **Table 1 in supplementary data 1** reports the results of the identification. Out of the top 20 most intense ions, 17 are attributed to Myelin Basic Protein S (P02688) among which 3 correspond to acetylated forms, 3 to isoforms 3 and 2 possible sodium and potassium adduct ions. The 3 other ions remained unidentified. In this approach, only few major proteins are identified although many more ions are observed. Similar results were recently published by Schober et al. (Schober, Schramm et al. 2011) . In their work they showed the possible identification of only 14 peptides belonging to 3 proteins, respectively Myelin Basic Protein, Tubulin Beta 5 Chain and Actin Cytoplasmic 1. Given the limited number of proteins identified directly at the tissue level, a new strategy was

developed based on micro-extraction allowing an efficient identification of proteins while keeping their spatial delocalization at the tissue level to the minimum.

LESA micro-extraction optimization

The device used for micro-extraction allows the deposition of solvents or reagents in a controlled manner by creating a liquid microjunction that permits extraction and subsequent uptake of the solvent containing the extracted material. The size of the contact area between the dispensed solvent and the tissue depends on the nature of the tissue surface, its preparation, and the physico-chemical properties of the solvent (polarity, hydrophobicity, surface tension). The bottom-up approach with on-tissue trypsin digestion followed by extraction of the digested peptides from the tissues using the liquid microjunction interface was first explored. Optimization of this approach, including the microjunction size, extraction time, number of extraction cycles and solvent mixture compositions, were performed using standard BSA spotted on rat brain sections and can be found in the **Supplementary Data 1.** Optimal extraction in terms of sequence coverage and peptide ion intensity was obtained using ACN: 0.1%TFAaq (8:2, v/v) as solvent and doing five extraction repetitions on the same spot with a contact time of 30 s. This protocol was used in all subsequent experiments.

In order to obtain a fine correlation between protein identification and tissue localization, reduction of the micro-extraction size is necessary. This parameter is tightly linked to system configuration which presently cannot be changed. The minimum volume of liquid that can be properly dispensed from the tip of the LESA is currently 0.6 µL, leading to formation of microjunction spots of 1.5-1.7 mm in diameter depending on the solvent and surface properties. By assuming that analyzed cells present an average size of 15 µm and are round-shaped, we can estimate that micro-extraction will be performed on approximately 10 000-13 000 cells. As it is difficult to further reduce the extraction size using the instrument, the size of the digested areas was reduced instead. This is possible using a micro-spotting piezoelectric device, which has been shown previously to deliver picoliter volumes of trypsin on small areas of tissue surfaces (Franck, Arafah et al. 2009).

Figure 2 presents the obtained results by comparing the full MS spectra recorded from two distinct locations on two consecutive tissue sections, after subjecting them to washing to remove most abundant lipids and peptides that can be co-extracted with the tryptic peptides. On the first section, only micro-extraction was done on the tissue (**Figure 2a**), whereas on the second section,

trypsin digestion was first done followed by micro-extraction (**Figure 2b**). Micro-extraction alone shows an area of the expected size (~1700 µm). On the tissue submitted to digestion by micro-spotting prior to micro-extraction, the 670 µm-sized spot corresponding to the enzymatic digestion can be observed on the tissue and are highlighted with gray circles (**Figure 2b**). As expected for the tissue first submitted to on-tissue digestion, complex profiles of ions are observed in the m/z 700-4000 range with observable differences according to the analyzed region of the tissue. On the other hand, on the tissue submitted to micro-extraction without any digestion, a low populated mass spectrum profile was observed (only one predominant ion in the spectrum, indicating that the peptides observed in **Figure 2b** were only coming from the digested portion of the spot. This is a particularly interesting feature since this allows the indirect reduction of the extraction area to a region of 670 µm. By applying the same rough calculation as previously (i.e. round shaped cells of about 15 µm diameter in average), the estimated number of studied cells can be reduced down to about 1900.

To further assess the method, a comparison of the MS spectra when the tryptic peptides were analyzed on tissue and when they were subjected to micro-extraction in the consecutive section at the same position (**Figure 3**). Straightforward MS/MS interrogation allowed the identification of dominant peaks in these spectra coming from peptide fragments of abundant proteins present in the rat brain, namely, tubulin, actin, myelin and hemoglobin. When the spectra are zoomed, due to the high spectral resolution, signals coming from the less abundant proteins become more apparent. However, it is difficult to subject these ions to MS/MS analysis due to low signal intensity and complexity (overlapping) that prevent the accurate selection of the precursor ion. Comparison of spectra obtained directly on tissues and after micro-extraction shows as expected a better S/N ratio for conventional MALDI analysis (an average of the S/N ratio for the 20 more intense peaks give 60 for direct tissue analysis and 700 for analysis after extraction). There is also a marked difference in the signal intensity and the number of detected peptides (118 peptides for on tissue MS and 675 for conventional MS after extraction, **Figure 3b**). However, it must be noted that signals showing the highest intensity are the same in both experiments. This is expected because of the additional desalting and concentrating step using a hydrophobic phase (C18 ziptip) in addition to the absence of interference coming from the tissue matrix that induces poor ionization yields. However, this remains unsatisfactory in terms of protein identification because the clear advantage of the developed strategy is to provide a solution that can be

subsequently processed by any possible proteomics strategies involving LC separation coupled to HR MS.

On tissue micro-digestion & micro-extraction followed by nanoLC-MS/MS analysis

Statistical analysis has proven to be a very useful approach to identify regions of different molecular content that can be specific to certain biological processes. This is exemplified in **Figure 4a** by the use of Hierarchical Clustering (HC) on the MALDI MSI data collected on a control rat brain sagittal tissue section prepared for protein imaging. Across the studied section, four distinct regions of specific molecular fingerprint are primarily differentiated. In a complete biological investigation, the succeeding step consists of the identification of the regulated proteins characterizing each of these regions. For this, on-tissue enzymatic digestion combined with micro-extraction on two discrete positions of two of these regions (**Figure 4b**) as determined previously by HC (**Figure 4a left**) and correlated to HES staining (**Figure 4a right**) was then performed in triplicate on three consecutive sections (S1, S2 and S3) thus, for a total of 6 samples (3 experiments for each of the position, S1P1, S2P1, S3P1 versus S1P2S2P2 and S3P2). Specifically, the two investigated positions are in the thalamic nuclei region (P1) and in the cerebellar nucleus (P2) according to the rat brain atlas (**Figure 4b right panels**).

The full MS spectra of micro-extracted samples from P1 and P2 are shown **Figure 4c**. Even though these spectra are dominated by the signals from abundant proteins, differences can be observed between the two positions. The tryptic peptide extracts were then collected and analyzed by online nanoLC coupled to an ESI-orbitrap Elite with high speed and high spectral resolution capabilities (see **Supplementary Data 1 Figure 2** for the base peak chromatograms). An average of 25 000 peptide ions was sequenced in MS/MS and, because of the high mass accuracy on the parent ion mass determination and the quality of MS/MS spectra in terms of type, number and intensity of fragment ions, many of these peptides led to protein identification. **Figure 5** presents examples of typical MS/MS spectra recorded during these experiments. Because of MS/MS spectra quality and parent ion MW determination accuracy, a large number of proteins were unambiguously identified with high confidence. To exemplify the complexity of the sample, in **Table 1** are listed all the ions with isobaric m/z 1383. Thanks to the efficiency of peptide separation, 8 different peptides are found to be corresponding to this nominal m/z with a maximum separation of 144 mDa in mass. Here, these peptides appear at distinct retention times. Even with high resolution capability of the instrument these peptides would have been very

difficult to identify from each other in the direct analysis without the separation dimension. Here, on the contrary we obtained high confidence identification for each of the 8 peptides with a good accuracy on the mass measurement.

Overall, the experiments led to the identification of a significant number of proteins for each sample and each position (>1000). The identified proteins can be found in the **Supplementary Data 2**, from which it is shown that 1517 proteins were identified for P1 from S1, 1554 proteins from S2 and 1582 from S3. **Figure 6** presents a Venn diagram showing the number of proteins identified for each position across the 3 replicates. 1405 proteins are common to the 3 sections; 11 are unique to S1P1, 17 to S2P1 and 40 to S3P1. By taking all the experiments for P1 we obtained a total of 1658 proteins identified (**Figure 6a**). For P2, 1289 proteins common to the 3 sections were identified, 28 being unique to S1P2, 4 to S2P2 and 74 to S3P2, amounting to 1761 protein. In most cases, less than 5% variation is observed between each position, demonstrating reproducibility. For each samples, identified proteins came from about 24500-25500 MS/MS recorded (depending on the experiment) and sent for databank interrogation. By merging data from the different samples (S1, S2, S3) a comparison of the results from the 2 analyzed positions can be made. Merging the 3 replicates and using a filter to select only high confidence identification, 1334 proteins from P1 and 1409 proteins from P2 were obtained (**Figure 6b**). Altogether, 1606 protein groups based on 5111 peptides were identified. This demonstrates the efficiency of the method, leading to identification of about 1300-1400 proteins with high confidence according to the position for an estimated equivalent number of 1900 cells with an average diameter of 15 µm.

Biological significance

Global analysis on the identified proteins shows that they present various cellular localization and molecular functions, and are implied in different biological processes. **Figure 7** presents a diagram of protein location in the cell for proteins identified from P2 (cerebellar nucleus). Most of the identified proteins are cytoplasmic (24.54%) and membrane proteins (18.54%), although to a less extent proteins from the nucleus, cytosol, cytoskeleton and mitochondria were also identified. Comparison of molecular functions and biological processes of proteins identified from P1 and P2 are given in **Table 2**. Looking at molecular functions specific to P1 or P2, the most important variations observed are between catalytic activity, metal ion and nucleotide binding. Major significant variations for the biological activities of the identified proteins

between the two positions are found for metabolic activity and cell organization/biogenesis and development (**Table 2**). Some tiny changes are also observed for transport, cell differentiation and cell communication. The number of proteins identified and involved in metabolomic activity is 46% higher in cerebral nucleus than in thalamus/hippocampus region one, whereas the ones implicated in cell organization is 9% higher in the latter.

Back correlation of proteins identified via micro-extraction nanoLC-ESI to MALDI MS Imaging data

The on-tissue micro-extraction method followed by nanoLC-ESI MS/MS analysis allowed for the identification of a large number of proteins. The generated data were then used to determine whether it would be possible to correlate this data to the MALDI MSI data since from the latter, only a few major proteins were identified. Because the mass measurements have been performed with high mass accuracy, tryptic peptide assignments observed in MALDI MSI can be directly compared with those determined under nanoLC-ESI MS/MS.

As an example, in the cortex region of the primary motor cortex on a consecutive section, on tissue microdigestion was performed with the same conditions as in the MALDI-MSI experiment and directly analyzed. Assignment of all tryptic peptide masses observed in the on tissue MALDI experiment is summarized in the table presented in **Supplementary Data 4**. The table gives the measured m/z list and the corresponding relative intensities of the observed tryptic peptides, which are classified by decreasing intensity. Also given are the m/z of tryptic peptides extracted from the nanoLC-ESI experiments, protein IDs (name and accession number), the mass difference (in ppm) between the two measurements, and the degree of confidence by which these peptides have been identified. From the on-tissue MALDI MSI spectrum extracted in the region of interest (**Supplementary data 1 Figure 3**), 176 tryptic peptides m/z can be extracted. On these 176 peptides, 120 can be assigned by correlation to nanoLC-ESI and 56 m/z remain unassigned. Of the 120 assigned peptides only 3 are found with low confidence identification. In terms of the mass difference in the measured peptides (**Table 3**), 41 peptides showed a mass difference <1 ppm, 34 within 1-2 ppm, 24 within 2-3 ppm, 12 within 3-4 ppm and only 9 showing a mass difference >5 ppm. By setting up to 3 ppm as a good acceptance for confident identification, 82.5% of the peptides were assigned. Data were compiled to group tryptic fragments of single proteins (unique accession entry) altogether, and are presented in **Supplementary Data 1 Table 3**. Altogether, 61 proteins with unique entry were identified, 20 of

which contain 2 or more peptide hits. If the proteins are grouped as shown **Table 4**, 45 distinct protein families can be reassigned to the MALDI MSI data. The most prevalent protein in the analyzed region from the tissue is Tubulin family which is identified with 38 fragments, followed by ATP Synthase, Hemoglobin, Myelin Basic Protein S, Sodium/Potassium-Transporting ATPase, Actin, Clathrin, Synapsin-1, Fructose-Biphosphate Adolase. However, 56 peptides remain unassigned and 18 of them with a relative intensity of more than 5%. This corresponds to 32.18% of the peptides with no assignment.

DISCUSSION

The obtained results demonstrate the power of liquid junction micro-extraction in identifying proteins from a section with respect to the location of extracted cells in the considered tissue. Previously the use of the liquid microjunction for the extraction and analysis of small molecules and lipids has been demonstrated (Eikel, Vavrek et al. 2011; Stegemann, Drozdov et al. 2011; Schadt, Kallbach et al. 2012), but the applicability of this technology was never shown for neither peptide nor protein analysis and in particular for the identification of proteins in Bottom-Up strategies combined to on tissue micro-extraction.

The combination of on tissue digestion using a microspotter followed by micro-extraction offers many advantages regarding the analysis of proteins. This strategy allows an efficient control of the digestion by accurate delivery of the trypsin-containing droplets, leading to the reduction of the digested region by half, compared when the digestion was done by LESA alone. In contrast to a bottom-up MSI strategy where the digestion is not as efficient due to the quick drying of the trypsin solution between each successive deposition, here, the digestion efficiency has greatly improved because the droplet is maintained during the entire incubation. Micro-extraction of tryptic peptides is more advantageous than intact proteins considering that the extraction of the latter required in most of the cases the use of lysis buffer containing detergents and chaotropes. These solutions cannot be handled by a microspotter because it requires their aspiration from the surface afterwards. Although it is possible to use LESA for this purpose, the high viscosity of the solutions prevents them from being accurately dispensed or aspirated in the current configuration.

Other ambient surface sampling techniques that are based on the solid-liquid extraction principle, such as the liquid microjunction surface sampling probe (LMJ-SSP) and the sealing surface sampling probe (SSSP) have been reported (Van Berkel, Kertesz et al. 2008). These methods operate in continuous mode. The sample surface is perpetually in contact with a probe that rasters

across the surface and as extraction occurs, the extracts are continuously delivered to the ESI source. Whereas this mode can be advantageous in surface scanning and ultimately, ESI imaging (Blatherwick, Van Berkel et al. 2011), the time of contact between the solvent and the surface is short and only a single pass on each position is being performed. As such, this method has low detection and at the moment is only restricted to small molecules. On the other hand, the liquid junction micro-extraction is a discontinuous extraction method. This is advantageous in increasing extraction yields by allowing the repetition of extractions on the same position. Moreover, this method permits the collection of extracts, enabling them to be further processed by means of conventional analytical strategies, e.g. various fractionation and/or separation methods (LC, CE, gel electrophoresis) and various MS instruments including non ambient MS sources like MALDI. This increases sample processing time but makes the complex mixture more amenable to MS analysis thereby increasing the number of molecules detected and identified.

The obtained results demonstrate that using this strategy, about 1500 proteins can be identified with high confidence and good reproducibility for a given position. Here we have shown that currently micro-extraction through liquid microjunction is achieved on a spot with a minimum size of about 1.5-1.7 mm in diameter using the conventional setup of the instrument. For an average cell diameter of 15 µm, this corresponds to a range of 10 000-12 000 cells. To improve correlation, the sampling area was decreased down to a spot of 670 µm in diameter corresponding to about 1900 cells. From this region, high confidence identification of an average of 1500 proteins was obtained. Currently, shotgun proteomics can lead to the identification of up to about 5000 unique proteins (Walther and Mann 2011) as extracted from organs or specific regions of organs. However, global extractions if efficient in the number of identified proteins do not allow correlation to specific type of cells in their specific tissue micro-environment. Over the past years, LCM combined to proteomics strategies has been demonstrated to present a high potential for identification of proteins from specific cell phenotypes (Besson, Pavageau et al. 2011; Gozal, Dammer et al. 2011). In 2010, Wang et al (Wang, Xu et al. 2010) examined the correlation between the number of protein identifications with the decreasing number of LCM-processed cells. In their strategy they were able to identify 167, 237, 491 and 619 proteins from 500, 1 000, 2 500 and 5 000 collected cells, respectively. Similar results were shown by Stingl et al in 2011 (Stingl, van Vilsteren et al. 2011). Protein identification performed out of 2000 epithelial or stromal micro-dissected cells led to an average of 366 identified protein-groups. Recently,

Wisniewski et al (Wisniewski, Ostasiewicz et al. 2011) demonstrated the possible identification of 1500 proteins from 1000 and 2500 from 10 000 HeLa cell cultures. From LCM micro-dissected mouse liver tissue sections, they obtained the identification of 3600-4400 proteins out of 20 000 laser capture cells. Thakur et al (Thakur, Rejtar et al. 2011), meanwhile, have identified about 1700 proteins from 10 000 initially collected cells of primary and metastatic breast cancer tissues by combination of SDS-PAGE separation and LC-MS/MS for an equivalent of 1 000 cells injected. In the present study we assume to have analyzed about 1900 cells (15 µm diameter) from 670 µm spots and identified an average of 1500 proteins. This demonstrates the interest of the present strategy which has a clear potential for protein identification with fine correlation to their micro-environment.

The biological significance of the data obtained points out the presence of specific protein signatures between the two brain localizations which are in line with their biological role in the brain. Cerebellar nucleus is known to modulate the outputs of other brain areas and to calibrate precision and timing of movements. The thalamic nucleus is involved in relaying information to and from the cerebral hemispheres, and is also involved in motivation, eating, drinking, defecation, and copulation. The high number of proteins identified in the cerebellar nucleus is in line with its high consumption of glycogen and its involvement in oxidative stress. On the other hand, the high expression level of proteins involved in cell organization can be explained by the great number of connections and cell projections occurring in thalamic nuclei. This is in line with the higher level of proteins implicated in cell communication, cell differentiation and regulation identified in this region compared to the cerebellar nucleus region. Altogether, the data confirm that specific protein signatures rely on the brain area and their canonical biological functions to the brain functioning.

Finally, the nanoLC-ESI data were correlated to those obtained from MALDI MSI. High spectral resolution and high mass accuracy measurements given by the FT instrument allow the assignment of the peptides observed on tissue from the tryptic digestion by simple m/z correlation. The good correlation of the data allows indirect back identification and subsequently protein distribution in tissues. A non-negligible number of peptides were not assigned from nanoLC-ESI data though. First, the MALDI MSI data were collected from a tissue section adjacent to the one used for micro-extraction experiments. Therefore, it is possible that not exactly the same type of cells were targeted, taking into account the heterogeneity of cell phenotypes in the brain. The observed discrepancy can also be due to the different ion production

source used (ESI vs. MALDI) in the 2 experiments. In fact, it is well-known that MALDI and ESI generally provide complementary data (Baeumlisberger, Rohmer et al. 2011) in terms of observed peptides because of the very different physico-chemical processes by which ions are formed in these two sources. The number of unassigned peptides could be decreased if the micro-extracted tryptic peptides are analyzed using the same MALDI-LTQ Orbitrap where the MALDI-MSI experiment was performed. However, in this configuration, low energy activation can only be obtained leading to a poor fragmentation yield coming from singly-charged peptides and, subsequently, poor MS/MS spectra. This implies that the correlation assignment can be ameliorated but this will be at the expense of protein identifications. An alternative would be the parallel analysis of the micro-extracted tryptic peptides in the MALDI-TOF/TOF instrument providing higher energy activation and therefore probably better protein identifications. However, the back correlation will suffer due to the lack of accurate mass measurement obtained from the MALDI-TOF.

Taken altogether, a workflow was designed allowing both imaging and classification of tissue regions leading to protein identifications and finally, back correlation to the MALDI MSI data (**Figure 8**).

CONCLUSION& PERSPECTIVES

Here, it was demonstrated that micro-extraction using the liquid microjunction is an efficient strategy for protein identification from tissue sections. An approach for micro-extraction on a well-controlled area (670 µm diameter), allowing fine correlation of identified proteins to their original location in the tissue, was established. This was achieved by performing on tissue digestion using micro-spotting followed by micro-extraction using the liquid microjunction interface. By applying such a strategy in combination with conventional shotgun analysis for simplification of the complex extract using RPLC separation with on-line coupling to high spectral resolution ESI FTMS, identification with high confidence of about 1500 proteins for an equivalent of about 1900 cells was obtained. This is largely in the range of what is currently achieved using LCM followed by proteomics analysis. This method demonstrates to be robust, fast and efficient and can be automated. This method was incorporated into the MALDI MSI workflow, by analyzing collected MALDI MSI data using statistical tools to determine cells that present a similar molecular content. This classification was then used to determine ROIs which were used on the tissue digestion and micro-extraction strategy to further identify the proteins present in these specific regions. For example in clinics, MALDI MSI will allow for classification of the cells according to their role in the physiopathological process compared to normal ones, and the proposed strategy presented here will lead to the identification of the proteins in these different regions. Even if the strategy has shown great potential, there are many further developments needed to be performed. First, the separation of micro-extracted tryptic digest can be improved, not only by the use of 1D RPLC separation but also by other methods, such as 2D LC or capillary electrophoresis (CE). In order to obtain complementary data, the extracts can be subjected to parallel MS analysis using ESI and MALDI sources. In order to improve the correlation of identification data to protein location, the reduction of both digested and extracted area will be investigated. This can be obtained through several means. Decreasing the volume of trypsin and extraction solution is one possibility. This can be achieved by mounting specific tips on the liquid microjunction extraction system and is to be tested on short term. Finally, it was observed that the number of proteins unique to a specific position is low (197 for P1 and 272 for P2) in comparison to all identified proteins. There are certainly numerous proteins that are present in both positions but have differential regulation that can be highlighted by implementing a quantification strategy such as stable isotope labeling (Brewis and Brennan 2010). Finally, the extraction of small, soluble intact proteins followed by their separation will

also be explored in order to obtain their identification using top-down strategies including ETD or ISD in the case of ESI or MALDI analysis respectively.

ACKNOWLEDGMENTS

This research was supported by University of Lille 1, French Region “Nord Pas de Calais” by ARCIR FEDER funds for FT MS instrument (I. Fournier), Ministère de l’Enseignement Supérieur et de la Recherche for funding via Institut Universitaire de France (I. Fournier).

TABLES

Table 1: Efficiency of Liquid Microjunction extraction combined to nanoLC MS and MS/MS analysis to identify isobaric ions. Summary of all ions showing nominal m/z 1383 with their corresponding retention time (RT), peptide sequence from identification by databank interrogation, mass difference (ΔM) between experimental measurement and theoretical mass and identification probability (PEP)

m/z	RT (min)	Sequence	ΔM [ppm]	PEP
1383.58135	60.43	GGAEQFMEETER	-2.35	3.2146E-03
D4AC23, RCG55994, isoform CRA_c, [D4AC23_RAT]				
1383.61296	52.03	GASQAGMTGYGMPR	-1.08	2.0589E-04
P37805, Transgelin-3, [TAGL3_RAT]				
1383.64470	70.25	EMLGLYTENEGK	-1.03	6.6262E-03
F1LRV4, Heat shock 70 kDa protein 4, [F1LRV4_RAT]				
1383.66802	83.72	ImNTFSVmPSPK	2.35	1.4294E-03
Q3KRE8, Tubulin beta-2B chain, [TBB2B_RAT]				
1383.67253	36.22	AALPAGEGESPEGAK	-1.89	1.6774E-03
F1LPP0, Amphiphysin_1, [F1LPP0_RAT]				
1383.69267	75.34	SVPEMDPAATIPR	-0.79	7.2579E-2
D4A6C9, Tom1l2, [D4A6C9_RAT]				
1383.70513	104.26	IFDAVGFTFPNR	-0.37	8.6078E-06
O35095, Neurochondrin, [NCDN_RAT]				
1383.72539	61.47	LYQSATQAVFQK	-0.97	1.4450E-03
Q08602, Geranylgeranyl transferase type-2 subunit alpha, [PGTA_RAT]				

Table 2: Molecular and biological processes of identified proteins using Liquid Microjunction extraction workflow for 2 different tissue regions. Comparison between position 1 (P1: thalamic nucleus) and position 2 (P2: cerebellar nucleus) of proteins identified with high confidence a) molecular functions and b) associated biological processes. In red are indicated molecular functions or biological processes showing the most pronounced difference comparing position 1 and 2.

a) Molecular Function	P1 Thalamic nuclei	P2 Cerebellar nucleus
Binding	**31%**	**24%**
Catalytic Activity	22%	24%
Metal Ion Binding	11%	13%
Nucleotide Binding	9%	12%
Enzyme Regulator Activity	5%	5%
Transporter Activity	5%	5%
DNA Binding	**4%**	**1%**
Signal Transducer Activity	3%	4%
Receptor Activity	3%	2%
RNA Binding	2%	3%
Structural Molecule Activity	2%	4%
Transcription Regulator Activity	2%	2%
Motor Activity	1%	1%

b) Biological Processes	P1 Thalamic nuclei	P2 Cerebellar nucleus
Regulation	19%	20%
Metabolic	**13%**	**28%**
Transport	12%	11%
Cell organization / Biogenesis	**11%**	**1%**
Response to Stimulus	10%	9%
Development	8%	6%
Cell Differentiation	5%	6%
Cell Communication	5%	6%
Cell Death	5%	4%
Cell Proliferation	3%	2%
Cellular Homeostasis	2%	1%
Cellular Component movement	2%	1%
Reproduction	2%	2%
Defense Response	2%	2%
Cell Division	1%	2%
Coagulation	1%	0%
Cell Growth	1%	1%

Table 3: Mass accuracyin m/z assignment for back correlation of Liquid Microjunction extraction data to MALDI MSI. Number of peptides with respect to the m/z accuracyin the assignment of peptides identified from the nanoLC-ESI-LTQ-orbitrap to the MALDI-LTQ-orbitrap MS Imaging data.

M/Z precision (ppm)	Number of peptides
<1ppm	41
1ppm<x<2ppm	34
2ppm<x<3ppm	24
3<x<4ppm	12
>5ppm	9

Table 4: List of protein families identified by back correlation of Liquid Microjunction extraction data to MALDI MSI. Number of peptides found and corresponding percentage with respect to the total number of observed peptides in the MALDI MS Imaging spectra sorted against each protein family identified by correlation of mass between nanoLC-ESI-LTQ-orbitrap and MALDI-LTQ-orbitrap on tissue data.

Protein family	Number of peptides fragments matching / protein family	% peptides / total number of peaks
Not assigned	**56**	**32.18**
Tubulin	38	21.84
ATP Synthase	10	5.75
Hemoglobin	6	3.45
Myelin Basic Protein S	5	2.87
Sodium/Potassium-transporting ATPase	5	2.87
Actin	4	2.30
Clathrin	4	2.30
Synapsin-1	4	2.30
Fructose-biphosphate aldolase	3	1.72
Cytochrome c oxydase	2	1.15
Dihydropyrimidinase-related protein	2	1.15
Heterogeneous nuclear ribonucleoproteins A2/B1	2	1.15
Histone	2	1.15
Phosphoglycerate kinase 1	2	1.15
Adenylate kinase	1	0.57
Anaphase Promoting Complex	1	0.57
ATP synthase-coupling factor	1	0.57
B-Cell receptor-associated protein	1	0.57
Brain Acid Soluble Protein	1	0.57
Breast Carcinoma-amplified sequence 1 homolog	1	0.57
Calcium/Calmodulin-dependent protein kinase	1	0.57
Complexin-1	1	0.57
Excision repair cross-complementing rodent repair deficiency	1	0.57
Glutamine synthase	1	0.57
GlutathioneS-transferase omega-1	1	0.57
Glyceraldehyde-3-phosphate dehydrogenase	1	0.57
Heat Shock Protein HSP 90-alpha	1	0.57
Hepatocyte Growth Factor-regulated tyrosine kinase substrate	1	0.57
Isoform 2 of MAGUK p55 subfamily	1	0.57
Leucine Rich repeat and IQ domain-containing protein 3	1	0.57
Malate dehydrogenase	1	0.57
Microtubule-associated protein	1	0.57
Mitogen-Activated protein kinase 8 interacting protein 3	1	0.57
Neurofilament heavy polypeptide	1	0.57

Protein bassoon	1	0.57
Purkinje Cell protein 4	1	0.57
Pyruvate kinase isozymes M1/M2	1	0.57
Rab GDP dissociation inhibitor alpha	1	0.57
Similar to Dystrophia myotonica WD repeat-containing protein	1	0.57
Sodium-dependent neutral amino acid transporter SLC6A17	1	0.57
Stathmin	1	0.57
UDP-glucose pyrophosphorylase 2	1	0.57
Vesicle associated membrane protein 2B	1	0.57
Vesicle -fusing ATPase	1	0.57
Hydroxysteroid dehydrogenase-like protein 2	1	0.57

FIGURES

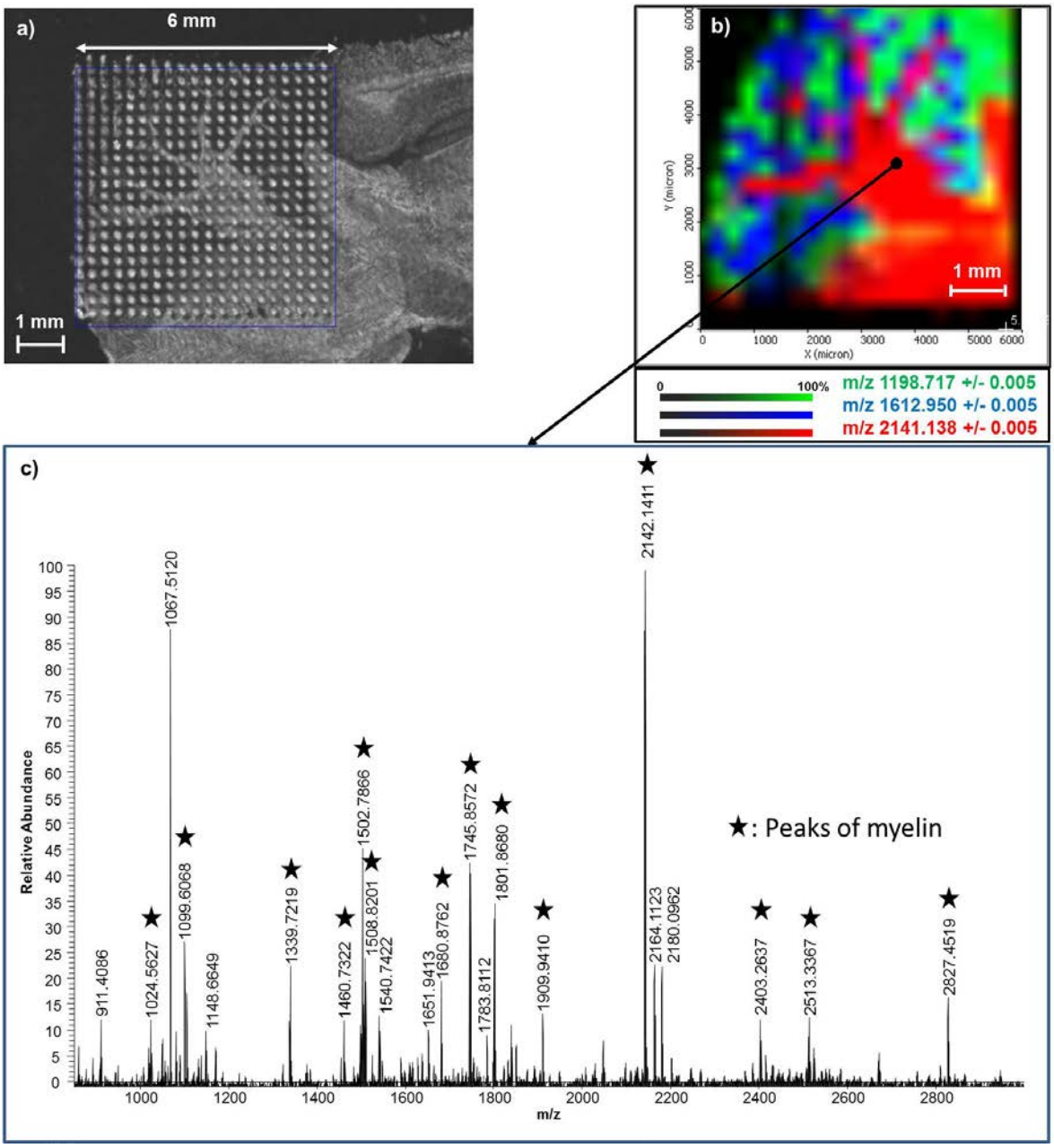


Figure 1: MALDI-LTQ-orbitrap MSI after on tissue trypsin digestion by micro-spotting. MALDI MSI sequence performed on a control rat brain sagittal tissue section in the cerebellum region on a MALDI-LTQ-orbitrap instrument after on tissue trypsin digestion and matrix deposition by micro-spotting. Images were acquired with a raster of 250 µm **a)** Scanned imageof a part of the tissue section showing the analyzed region corresponding to the pitch of micro-spots. **b)** Reconstructed composited molecular image based on the signals of ions at m/z 1198.717(green), 1612.950 (blue) and 2141.138 (red). **c)** MALDI MS spectrum recorded in the positive mode in the orbitrap analyzer extracted from a specific area of the analyzed tissue (indicated by a dark circle on Figure 2b).

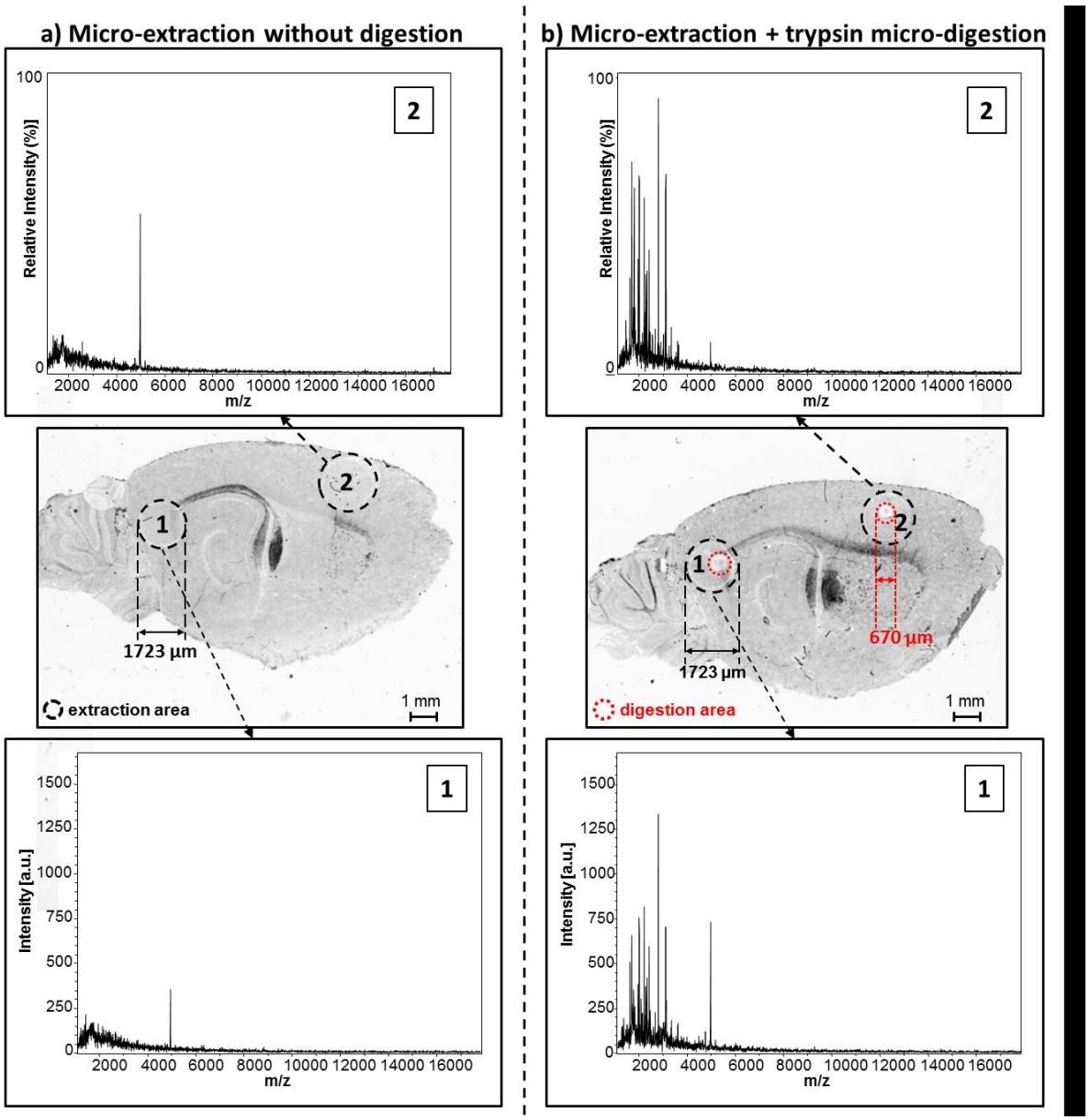


Figure 2: On tissue Liquid Microjunction extraction of trypsin digested or undigested regions. Comparison of MALDI-TOF MS spectra recorded in the linear positive mode from 2 consecutive rat brain sagittal tissue sections at two distinct locations in two regions (1, 2) of the tissue after micro-extraction in the optimized conditions for tryptic digestion peptides **a)** micro-extraction of non digested areas and **b)** micro-extraction after on tissue digestion on 300 µm spots. Black circles delimit the micro-extraction areas which are about 771 µm diameter and red circles define the position of the 300 µm trypsin digestion spots.

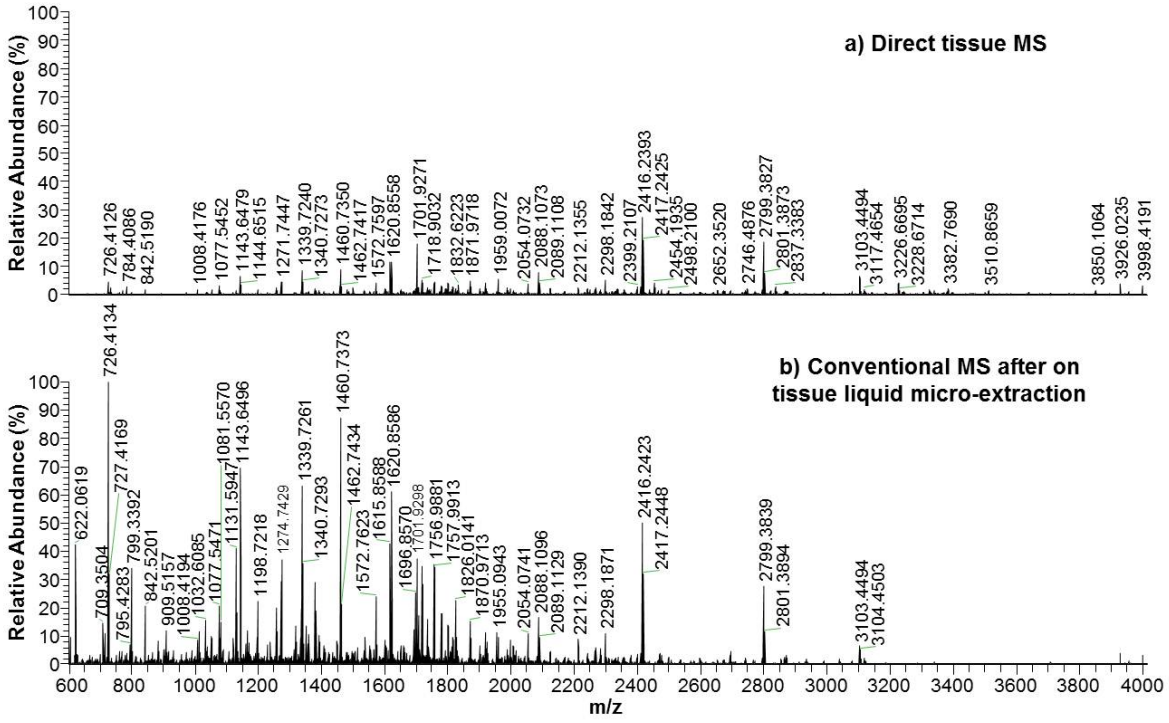


Figure 3: Direct on tissue versus conventional MALDI after Liquid Microjunction extraction analysis of trypsin digested regions. Comparison of the MS spectra recorded on an MALDI-LTQ-orbitrap instrument in the positive mode from 2 adjacent sagittal tissue sections of control rat brain at the same location (in the cerebellar nucleus region i.e. P2 position) after on trypsin digestion by 1 µL enzyme deposition using a micropipette and subsequent **a)** direct tissue analysis and **b)** after micro-extraction using liquid microjunction and conventional MALDI analysis. Mass Spectra are normalized against the same intensity scale.

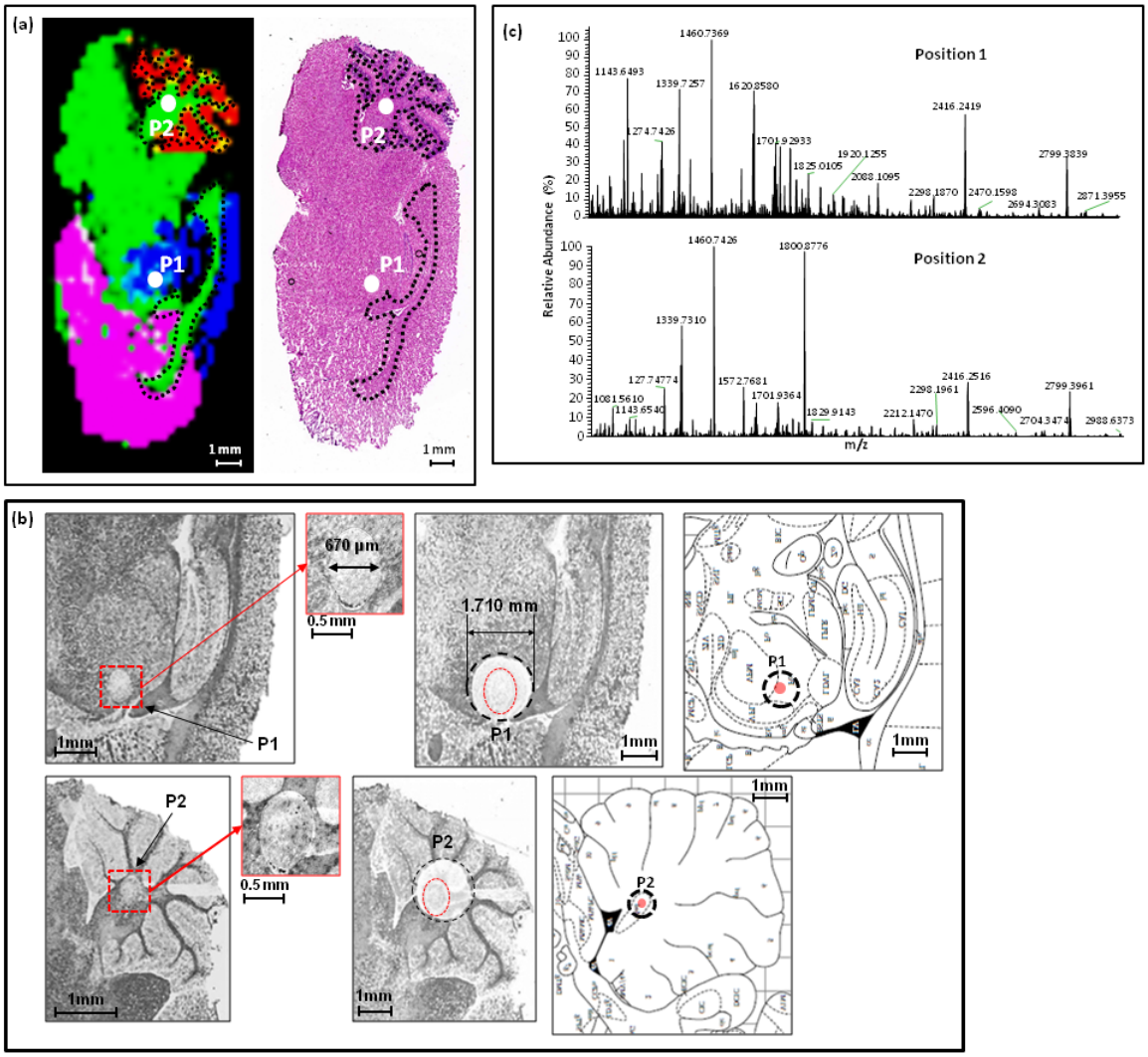


Figure 4: MALDI MSI molecular classification of tissue section. **a left**) Resultant image of the statistical analysis using Hierarchical Clustering (HC) showing areas of the section with different molecular signature. MALDI.**a right**) Haematoxylin Eosine (HE) staining of the same tissue section. The white circles represent the position of micro-extraction, position 1 is in the thalamic nucleus and position 2 is in cerebellar nucleus. **b**) Optical picture of the tissue section used for micro-extraction. Left: picture after trypsin digestion by micro-spotting in 2 distinct positions (P1, P2) and Right: picture after the micro-extraction step. **c**) Maps extracted from the rat brain atlas (sagittal view, lateral 0.9 mm panel) showing the regions used for extraction. Top, zoomed view of the panel on the cerebellum region showing that digestion/extraction was performed in the cerebellar nucleus and *Bottom*, zoomed view of the thalamic / hippocampus regions indicating that extraction was performed in the thalamic nuclei. **c)** MALDI-MS orbitrap spectra recorded in the positive full FT scan mode after micro-extractionfor position 1 (**d**) and for position 2 (e).

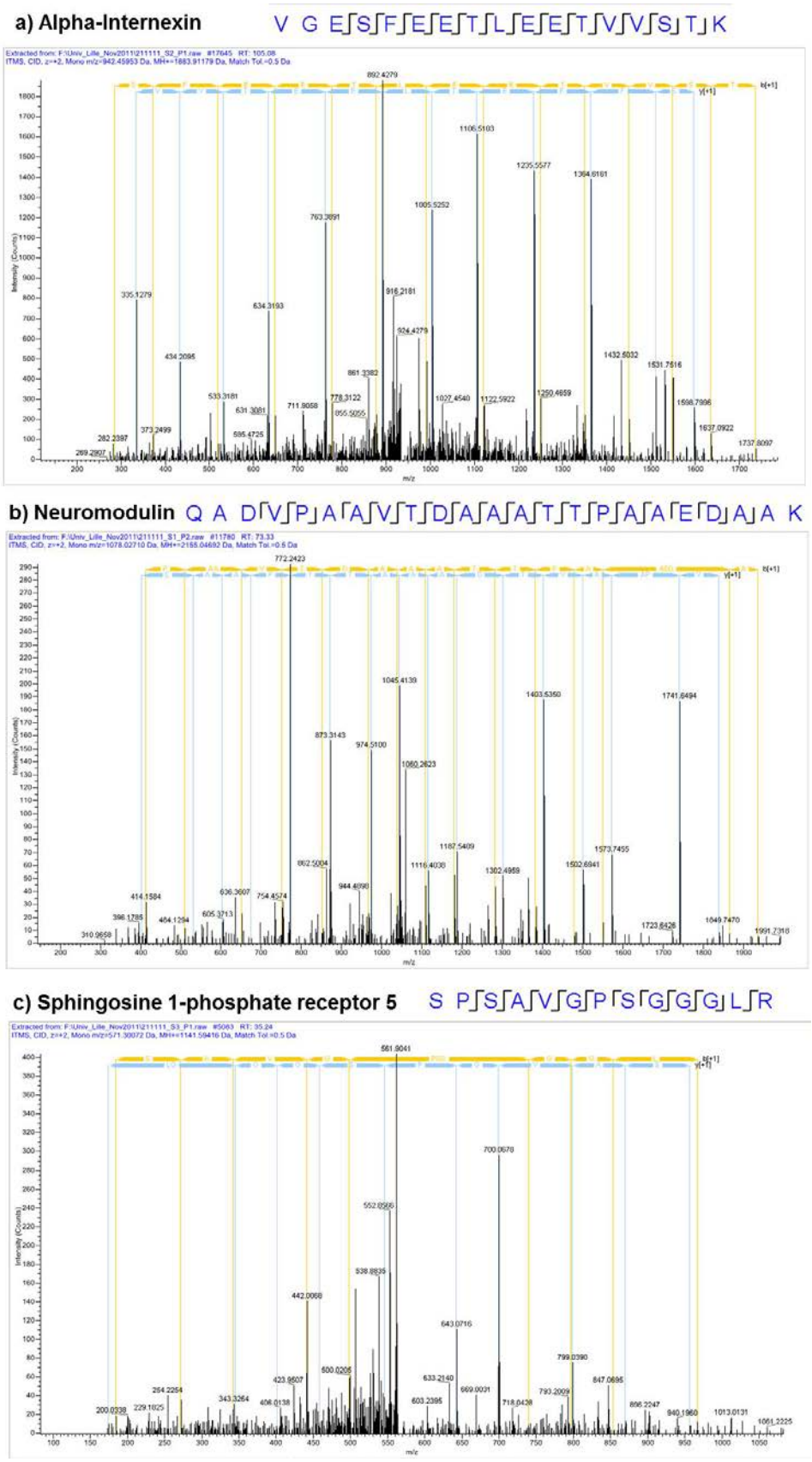


Figure 5: Extraction yields obtained through Liquid Microjunction extraction are sufficient to provide high quality MS/MS spectra. MS/MS spectra recorded in the positive mode during

the nanoLC-ESI-LTQ-orbitrap experiment for position 1 (thalamic nucleus) by selection of **a)** parent ion [M+2H]2+ at m/z 942.4595 (retention time 105.08 min) peptide attributed to Alpha Internexin, **b)** parent ion [M+2H]2+ at m/z 1078.0271 (retention time 73.33 min) peptide attributed to Neuromodulin and **c)** [M+2H]2+ at m/z 571.3007 (retention time 35.24 min) peptide attributed to Sphingosine 1-Phosphate Receptor 5. Peptides were activated by CID. Both fragmentation and detection was performed in the LTQ analyzer of the instrument.

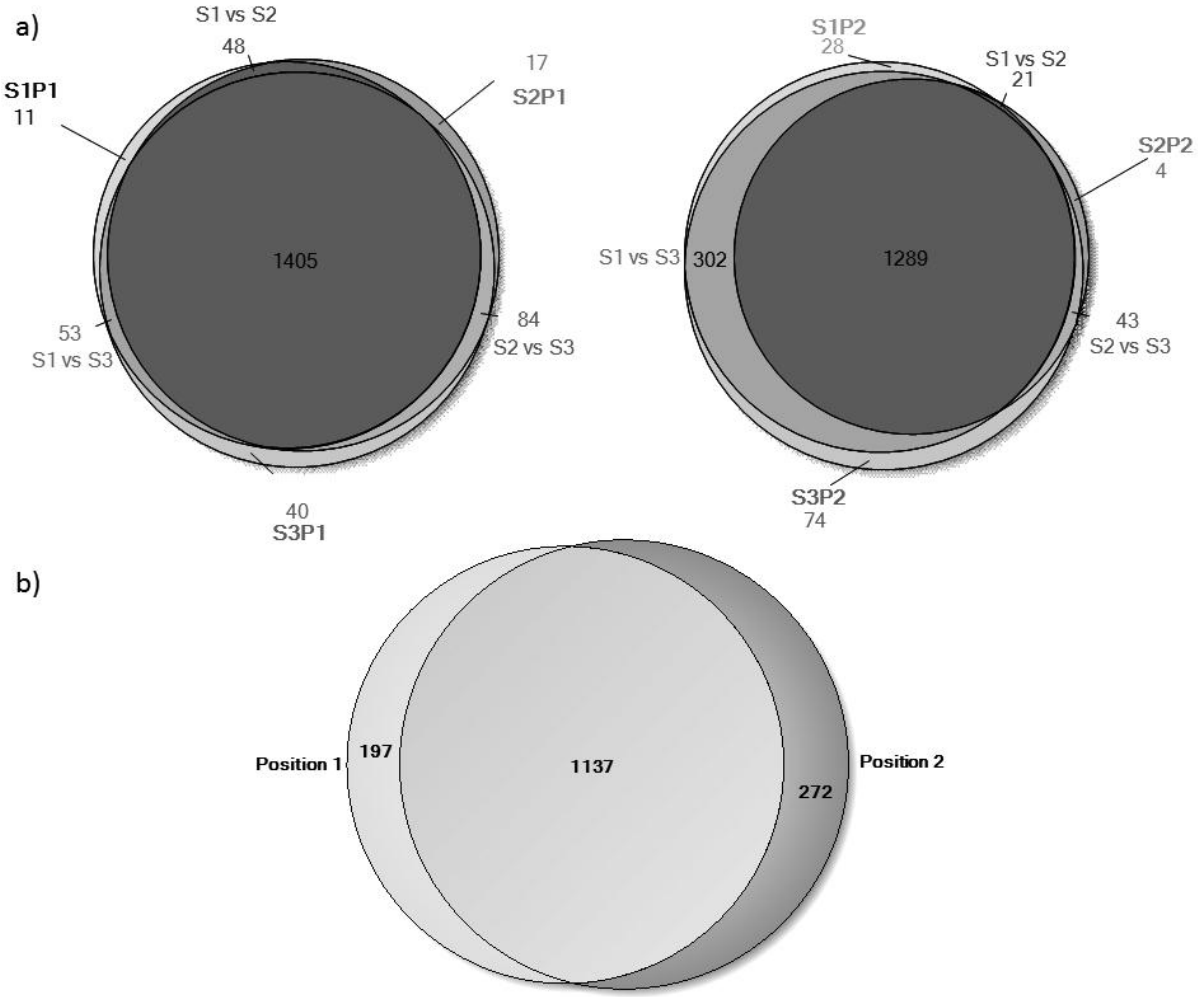


Figure 6: Number of identified proteins using Liquid Microjunction extraction workflow for 2 different tissue regions. Venn diagram representation of the number of proteins identified from position 1 (thalamic nucleus) and position 2 (cerebellar nucleus). Experiments were performed in triplicates from 3 consecutives tissue sections. On each replicate P1 and P2 were extracted and analyzed. **a)**Comparison for a specific position of the number of identified proteins from each sample(all confidence identification criterion). **b)** Global representation with merged data of the replicates for each position (high confidence identification criterion); 1137 proteins are identified from both position 1 and position 2, 197 proteins are unique to position 1 and 272 to position 2 for a total of 1606 proteins groups of proteins identified in total originating from 5111 high confident peptides.

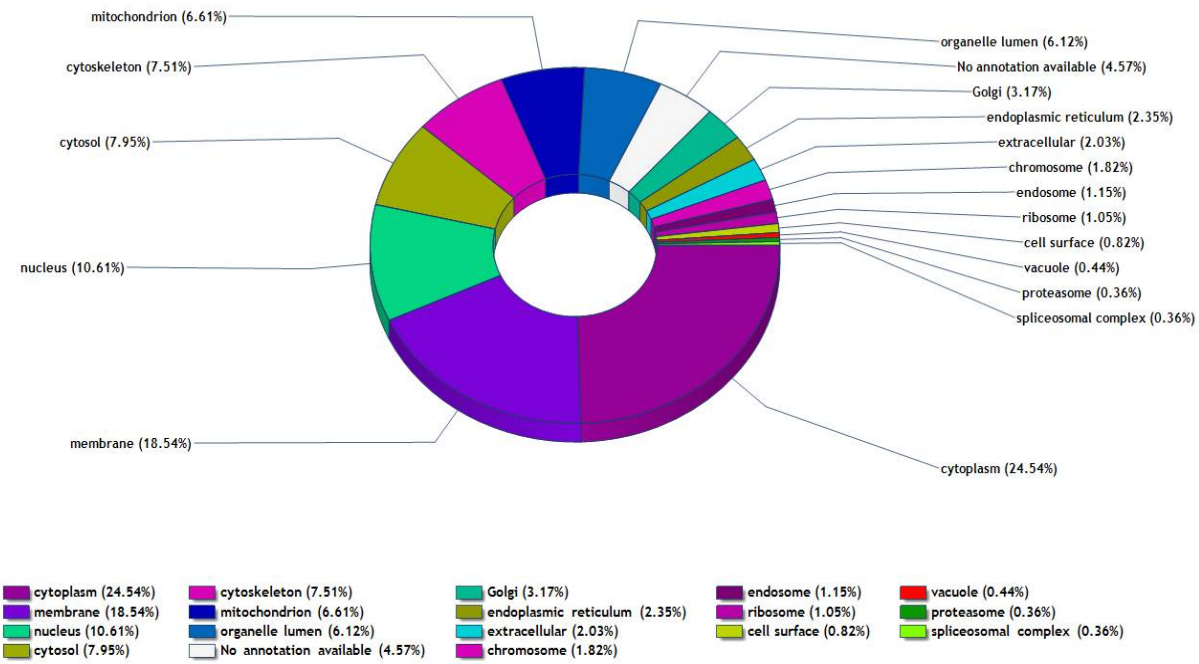


Figure 7: Cellular localization of identified proteins using Liquid Microjunction extraction workflow for 2 different tissue regions. Diagram plotting the relative distribution (%) of identified proteins according to their cellular location.

Figure 8: Markers discovery workflow. General workflow for markers discovery combining MALDI MSI and On tissue extraction-Off tissue analysis identification strategy.

SUPPLEMENTARY INFORMATION

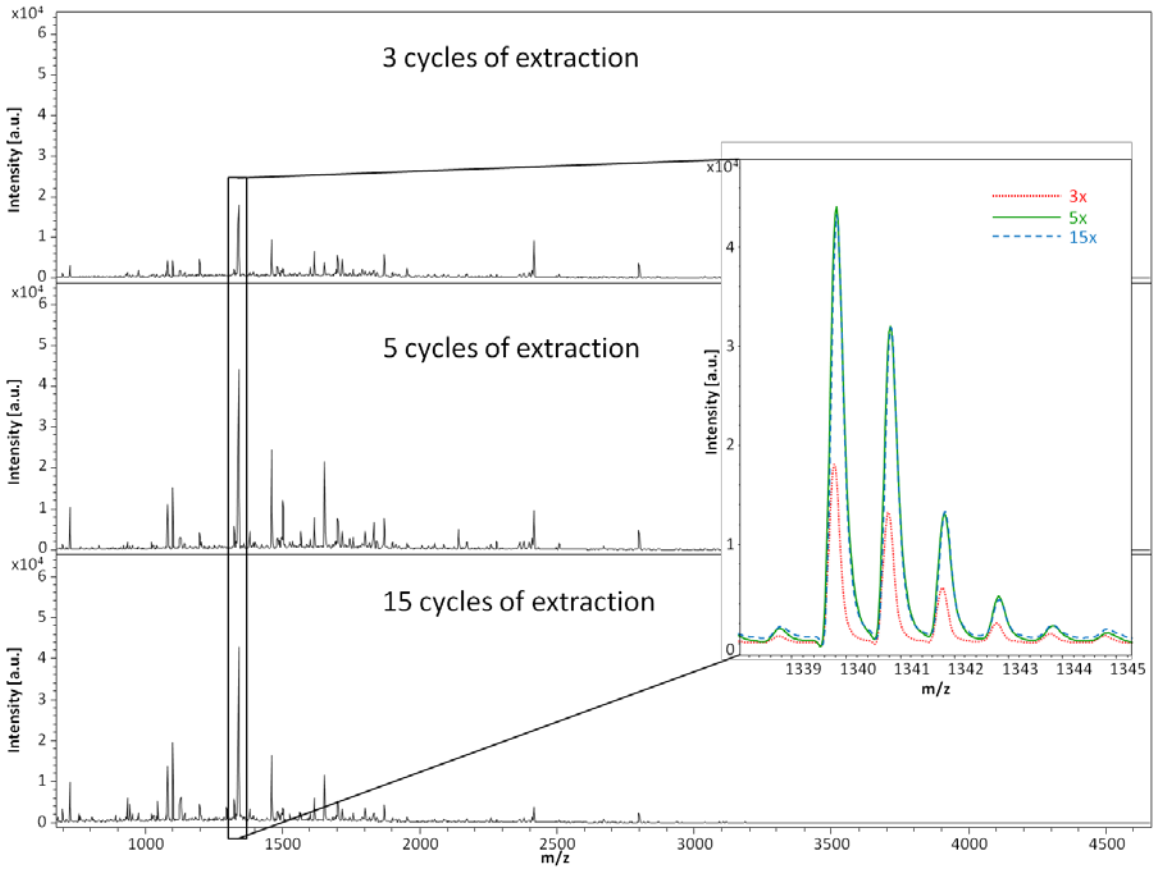
Optimization of micro-extraction of on tissue digested peptides

The micro-extraction parameters were first optimized in order to maximize the amount of peptides extracted. This includes the extraction time, the number of extraction cycles, and the type of solvent. Extraction time is an important parameter but cannot be varied on a large range. Long contact time (1 min or more) generally leads to evaporation of the liquid at the microjunction interface and too short times (15 s) do not provide efficient extraction. A good

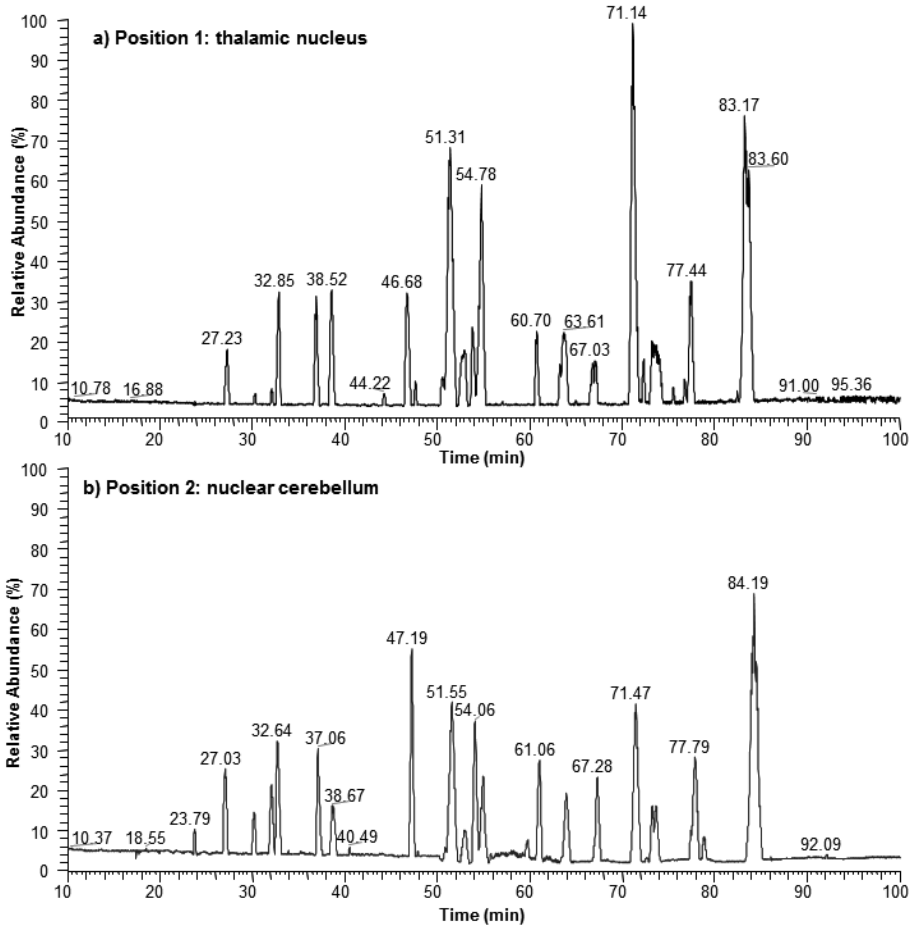
compromise for contact time is 30 s. Thus, it is more efficient to increase the number of extraction cycles than the extraction time. Different mixtures of solvents were tested to determine which provides the highest extraction yield. For these experiments we used a model protein (here Bovine Serum Albumin, BSA) digested in solution and deposited at a specific location on a rat brain section. 1 µL of trypsin-digested BSA solution was spotted on the tissue leading to an area of approximately 1.5 mm in diameter onto the tissue. After evaporation of the solution, micro-extraction was performed using solvents of varying compositions and repeated 5 times. Conventional solvents used for peptide extraction from tissue sections were initially tested, as mentioned in the **Experimental** section. After extraction the solutions were analyzed by conventional MALDI-TOF MS analysis and data were compared (**Supplementary Data 1 Table 2**). Globally the tested solutions were relatively efficient allowing for the observation of a number of BSA tryptic peptides. Only data for ACN: 0.1% TFAaq (8:2, v/v), MeOH/0.1% TFAaq (7:3, v/v) and 0.1% TFAaq are compiled in **Supplementary Data 1 Table 2** because those were the solvent mixtures providing the better extraction yields based on the number of tryptic peptides observed and sequence coverage. All the solvent mixtures successfully extracted the BSA tryptic peptides enabling its identification based on the PMF, but with different scores and variation in the sequence coverage and the number of fragments. The use of 0.1% TFAaq enabled the detection of 32 fragments corresponding to a sequence coverage of 52.4%. Using MeOH/0.1%TFAaq (7:3, v/v) improved the detection to 65.1% with 37 fragments. Finally ACN: 0.1%TFAaq (8:2, v/v) gave a better sequence coverage (73.6%) with 45 fragments. Similar results were obtained by extraction of a second model protein (Lysozyme C) digest spotted on tissue sections (data not shown). It must also be noted that some peptides are only specifically extracted by certain solvents. It can be therefore interesting to use different solvents successively during extraction to maximize the extraction efficiency.

The number of extraction cycles (n) was then examined. In this case, the experiments were performed on tryptic peptides from consecutive sagittal rat brain. On the same location for each section, 2 µL trypsin was spotted manually using a micropipette and incubated for 1h. After trypsin digestion, the sections were subjected to micro-extraction using n=3, 5 and 15 extraction cycles with ACN: 0.1%TFAaq (8:2, v/v), and the extracts subjected to MALDI-TOF MS analysis. The spectra quality was compared for n=3, 5 or 15 extraction cycles (**Supplementary Data 1 Figure 1**). Adequate spectra were obtained n=5 or 15 extractions, as determined by the increase of the S/N ratio of peaks. This effect is illustrated in the zoom on the ion at m/z 1340

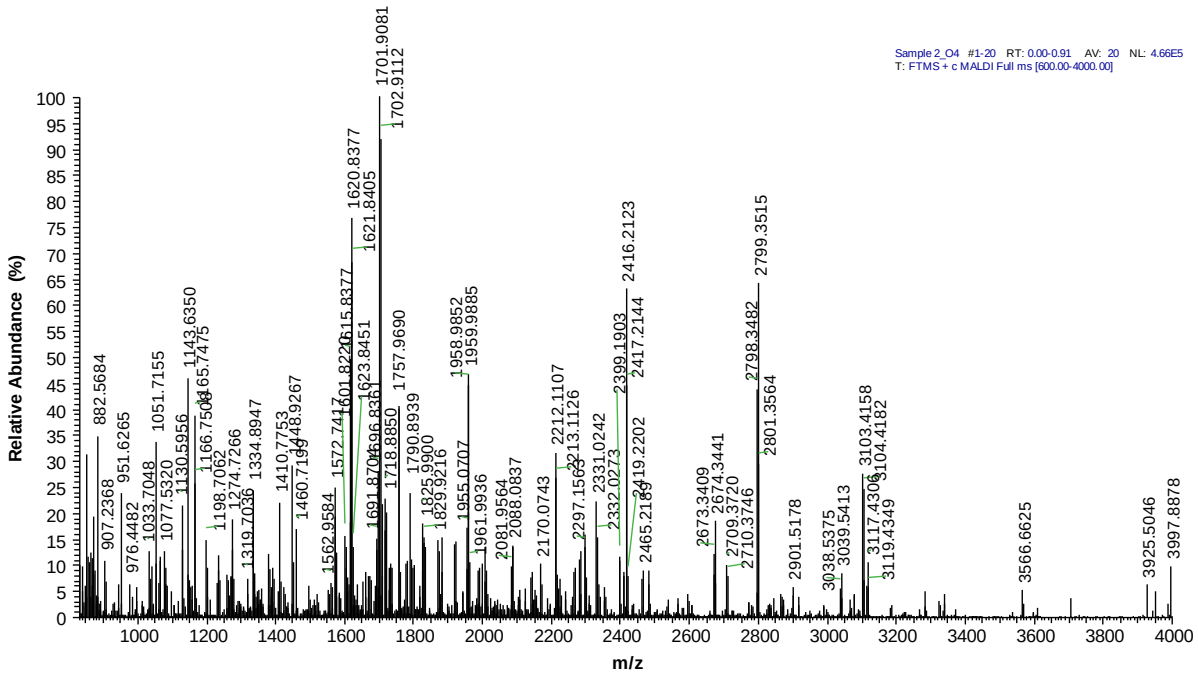
presented in the inset. Globally, the number of extracted peptides increased with the number of extraction cycles up to a maximum of 5 cycles and remained constant above this threshold. 64 peaks were detected in the MALDI TOF spectrum for the 3 extraction cycles, 84 for the 5 cycles and 80 for the 15 cycles, respectively. An increase in the intensity of ion signals from n=3 to 5 was also observed. For n=5 and 15, few significant differences were observed between the two spectra. As 15 extractions is time-consuming, the number of extraction cycles was fixed to 5. Taking into account all these parameters, a workflow for tryptic peptide extraction was designed using ACN: 0.1%TFA aq (8:2, v/v) as solvent and five extraction repetitions on the same spot with a contact time of 30 s.



Supplementary Data 1 Figure 1: Evolution of peptides extraction yields with the number of extraction cycles using Liquid Micro-junction extraction. MALDI-TOF MS mass spectra recorded in the positive reflectron mode in conventional MALDI analyses conditions (matrix: HCCA) of the solution micro-extracted from normal rat brain sagittal tissue sections after on tissue trypsin digestion. Each experiment is performed corresponding to a certain number of extraction is performed at the same location of consecutive tissue sections. Micro-extraction conditions are identical for all experiments (solvent: ACN/0.1% TFA in water (8:2, v/v), 2 µL solvent charged in the system tip and 0.6 µL pushed out to establish the liquid micro-junction, contact time 30s. **a)** 3, **b)** 5 and **c)** 15 cycles of extractions.



Supplementary Data 1 Figure 2 reports the base peak chromatograms for P1 compared to P2 extracted from the collected data. If a similar trend is observed between the two chromatograms, it should be noted that some differences could already be observed for the 2 positions.



Supplementary Data 1 Figure 3: On tissue MALDI MS spectrum used for back correlation of Liquid Micro-junction extraction data from direct tissue MALDI MS. MALDI-LTQ-orbitrap MS spectrum recorded in positive mode from a sagittal control rat brain tissue section after on tissue micro-digestion using micro-spotting in the cortex region of the primary motor cortex. Spectrum was recorded directly from the tissue section and shows the presence of 176 monoisotopic ions corresponding to tryptic peptides.

Supplementary Data 1 Table 1: Most of the top 20 tryptic peptides observed in MALDI MSI after on tissue digestion are attributed to Myelin Basic Protein S. List of the top 20 ions extracted with respect to their intensity from MS spectrum shown Figure 2c after deisotoping and exclusion of trypsin autolysis ions with a S/N>5 on the m/z range 800-4000. Experimental m/z are compared to the predicted m/z from *in silico* digestion of Myelin Basic Protein from databank. Out of 20, 15 of these ions are attributed to various forms of Myelin Basic Protein. Table give the corresponding peptide fragments, the mass precision and the modifications present on the protein sequence.

Measured m/z *On tissue measurement*	**Calculated m/z** *In silico digestion*	**Peptide fragment**	**DM** *(ppm)*	**Modifications**
2141.1377	2141.1138	104-121	11.16	
1067.51196	1067.5017	138-146	9.61	Isoform 3
1745.85718	1745.8387	11-26	10.59	
1502.78662	1502.771	91-103	10.39	
1800.86584	1800.8452	138-154	11.46	
2179.09424	2179.0696	104-121	11.30	$[M+K]^+$ ion
2163.10767	2163.0960	104-121	5.30	$[M+Na]^+$ ion
2826.44702	2826.4176	138-163	10.41	2x Acetyl
1508.82007	1508.804	2-14	10.65	2x Acetyl
1099.60681	1099.5967	2-10	9.19	2x Acetyl
1680.87622	1680.8564	132-146	11.79	Isoform 3
1339.72192	1339.7076	33-44	10.69	
2512.33472	2512.3055	101-121	11.63	
1909.94104	1909.9198	138-154	11.12	Isoform 3
2402.25928	2402.2323	33-54	11.23	
1838.81995	unassigned			
1540.74219	unassigned			
1105.46729	unassigned			
1460.73218	1460.7168	104-115	10.53	
1024.56270	1024.5534	122-131	9.08	

Supplementary Data 1 Table 2: Efficiency of Liquid Micro-junction extraction for tryptic peptides according the extraction solvent. Evaluation of digest peptides extracted using micro-junction micro-extraction for 3 different solvent compositions respectively ACN: 0.1% TFA in water (8:2, v/v), MeOH /0.1% TFA in water (7:3, v/v) and 0.1% TFA in water alone. A volume of 2 µL solvent was charged in the tip and 0.6 µL pushed out to create the micro-junction for each of the tested solvent mix. The number of extraction cycle was fixed to n=5. The up taken solution was then use for conventional MALDI-TOF analysis using HCCA as matrix. Table 2 summarize the m/z of the different digestion peptides observed, the position in the protein sequence of the digestion peptide for each condition and provide the databank interrogation (Mascot) score, the sequence coverage and the number of detected digestion fragment observed.

Calculated m/z *In Silico digestion*	**Peptide fragment**	**ACN/0.1% TFA (8:2 v/v)**	**MeOH/0.1% TFA (7:3 v/v)**	**0.1% TFA**
689.3352	236 - 241	x		
898.4982	483 - 489	x	x	
927.5078	161 - 167	x	x	x
1138.5353	499 - 507	x	x	x
1163.6822	66 - 75	x	x	x
1195.5683	223 - 232	x	x	x
1249.6661	35 - 44	x	x	x
1283.7548	361 - 371	x	x	x
1305.75	402 - 412	x	x	x
1349.5784	24-34			x
1399.7376	569 - 580	x	x	
1419.7466	89 - 100	x	x	x
1439.8673	360 - 371	x	x	x
1443.6998	286 - 297	x	x	
1479.851	421 - 433	x	x	x
1502.5195	375-386		x	
1511.7353	438 - 451	x	x	
1519.6849	360-371			x
1532.6815	298 - 309	x	x	x
1554.5584	387 - 399	x	x	x
1567.6411	347 - 359	x	x	
1576.6614	139 - 151	x	x	x
1624.8262	24 - 36	x		
1639.9965	437 - 451	x	x	x
1673.8249	118 - 130	x	x	x
1724.8834	469 - 482	x	x	
1733.9083	156 - 167	x	x	
1740.8667	469 - 482	x		
1747.7752	184 - 197	x	x	x
1749.738	267 - 280	x	x	x
1809.7696	156-168			x
1880.8521	508 - 523	x	x	x
1901.8012	123 - 138	x	x	x
1907.8541	529 - 544	x	x	x
1927.7362	581 - 597	x	x	x
1964.8838	452 - 468	x	x	x
2019.9224	139 - 155	x	x	x
2028.032	421 - 437	x	x	x
2044.9885	168 - 183	x	x	x

2205.0811	168 - 183	x		
2211.104	2-19	x	x	x
2247.9843	267 - 285	x	x	x
2458.2007	319 - 340	x	x	
2492.2818	45 - 65	x		
2541.1978	118 - 138	x	x	x
2612.1846	264 - 285	x		
2670.3188	2-23	x		x
2826.4231	2-24	x		
3511.7386	310 - 340	x		
Mascot Score		**191**	**156**	**106**
Sequence coverage (%)		**73.6**	**65.1**	**52.4**
Numbers of fragments		**45**	**37**	**32**

Supplementary Data 1 Table 3: List of individual proteins identified by back correlation of Liquid Micro-junction extraction data to MALDI MSI. Proteins assignment for the MS Imaging spectra in the ROI using mass correlation in between measured peptides after on tissue micro-extraction by nanoLC-ESI-LTQ-orbitrap and for MALDI direct tissue analysis. For each identified proteins that right column give the number of peptides assigned in the MALDI MS Imaging data.

Proteins Assession number (Uniprot)	Proteins name	Number of peptides fragments matching / protein
Not assigned		56
Q6P9T8	Tubulin beta -2C Chain	12
P68370	Tubulin Alpha-1A Chain	9
P85108	Tubulin Beta -2A Chain	8
P10719	ATP Synthase Subunit Beta, Mitochondrial	6
P02688	Myelin Basic Protein S	5
P11442	Clathrin heavy chain 1	4
P01946	Hemoglobin Subunit Alpha-1/2	4
P09951	Synapsin-1	4
Q4QQV0	Tubulin Beta-6	4
P60711	Actin Cytoplasmic 1	3
F1LP05	ATP Synthase Subunit Alpha	3
P47942	Dihydropyrimidinase-related protein 2	2
P05065	Fructose-biphosphate aldolase A	2
P02091	Hemoglobin Subunit Beta-1	2
A7VJC2	Heterogeneous nuclear ribonucleoproteins A2/B1	2
P16617	Phosphoglycerate kinase 1	2
P06685	Sodium/Potassium-transporting ATPase subunit aplha-1	2
P06686	Sodium/Potassium-transporting ATPase subunit aplha-2	2

P689897-2	Isoform 2 of Tubulin Beta-5 chain	2
P62738	Actin, aortic smooth muscle	1
Q6P2A5	Adenylate kinase	1
D4A205	Anaphase Promoting Complex Subunit 2	1
Q06647	ATP synthase subunit O, mitochondrial	1
P21571	ATP synthase-coupling factor 6, mitochondrial	1
Q6AY58	B-Cell receptor-associated protein 31	1
F1LNN9	Brain Acid Soluble Protein 1	1
Q3ZB98	Breast Carcinoma-amplified sequence 1 homolog (fragment)	1
P08413	Calcium/Calmodulin-dependent protein kinase type II subunit	1
P63041	Complexin-1	1
P00406	Cytochrome c oxydase subunit 2	1
P11951	Cytochrome c oxydase subunit 6C-2	1
D3ZZZ1	Excision repair cross-complementing rodent repair deficiency, complementation group 6 (Predicted)	1
P09117	Fructose-biphosphate aldolase C	1
P09606	Glutamine synthase	1
Q9Z339	GlutathioneS-transferase omega-1	1
P04797	Glyceraldehyde-3-phosphate dehydrogenase	1
P82995	Heat Shock Protein HSP 90-alpha	1
Q9JJ50	Hepatocyte Growth Factor-regulated tyrosine kinase substrate	1
P43278	Histone H1.0	1
P15865	histone H12	1
D3ZAA9	Isoform 2 of MAGUK p55 subfamily member 2	1
Q6AYL8	Leucine Rich repeat and IQ domain-containing protein 3	1
O88989	Malate dehydrogenase, cytoplasmic	1
F1MAQ5	Microtubule-associated protein	1
D4AD98	Mitogen-Activated protein kinase 8 interacting protein 3	1
P16884	Neurofilament heavy polypeptide	1
O88778	Protein bassoon	1
P63055	Purkinje Cell protein 4	1
P11980	Pyruvate kinase isozymes M1/M2	1
F1LNN9	Rab GDP dissociation inhibitor alpha	1
F1M3D2	Similar to Dystrophia myotonica WD repeat-containing protein	1
P06687	Sodium/Potassium-transporting ATPase subunit aplha-3	1
P31662	Sodium-dependent neutral amino acid transporter SLC6A17	1
P13668	Stathmin	1
Q5XIF6	Tubulin Aplha-4A Chain	1
Q4QRB4	Tubulin Beta-3 chain	1
P69897	Tubulin Beta-5 chain	1
F1LQ84	UDP-glucose pyrophosphorylase 2	1
F1LRP6	Vesicle associated membrane protein 2B	1

Q9QUL6	Vesicle -fusing ATPase	1
Q4V8F9	Hydroxysteroid dehydrogenase-like protein 2	1

Additional supplementary data (Supplementary Data 2, 3 and 4) can be found in the online version of the article at: http://dx.doi.org/10.1016/j.jprot.2012.11.025

Chapter 2

On-Line nanoLC-MS Coupling of Liquid Microjunction Microextraction

It was mentioned previously that the correlation of protein identification data to localization can be further improved by reducing the digested and extracted areas. To this end, a different approach is being pursued. It is to be noted that the liquid microjunction extraction can also be realized using other methods (Kertesz and Van Berkel 2010; Van Berkel and Kertesz 2013). In fact, any instrumentation capable of dispensing and aspirating microliter quantities of solvent is suitable for this purpose. In a past example, an HTC PAL autosampler (LEAP Technologies Inc., Carrboro, North Carolina) was interfaced with an LC-MS instrument for liquid handling during the extraction step. This way, propanolol and some of its metabolites were quantified on mouse whole body sections (Kertesz and Van Berkel 2010).

Materials and methods

Liquid Microjunction Instrumentation

Figure 14 shows another example, where a Spark 920 FAMOS autosampler (LC Packings Dionex, now Thermo Scientific, Bremmen, Germany) was interfaced with an EASY-nLC 1000 system (Thermo Scientific, Bremen, Germany) that was coupled to an LTQ orbitrap XL MS instrument. In this set-up, the syringe of the autosampler is connected to the valve S of the EASY-nLC device, and it replaces the tubing that connects the valve to the autosampler pump. A 100-μm capillary replaced the tubing in position 1 of the valve and this capillary is connected to the robotic arm that collects the sample from the autosampler vials. A cut falcon tube open at both ends covers the tissue section on one end and acts as a support on the other by preventing the piercing needle from completely descending towards the bottom of the sample plate.

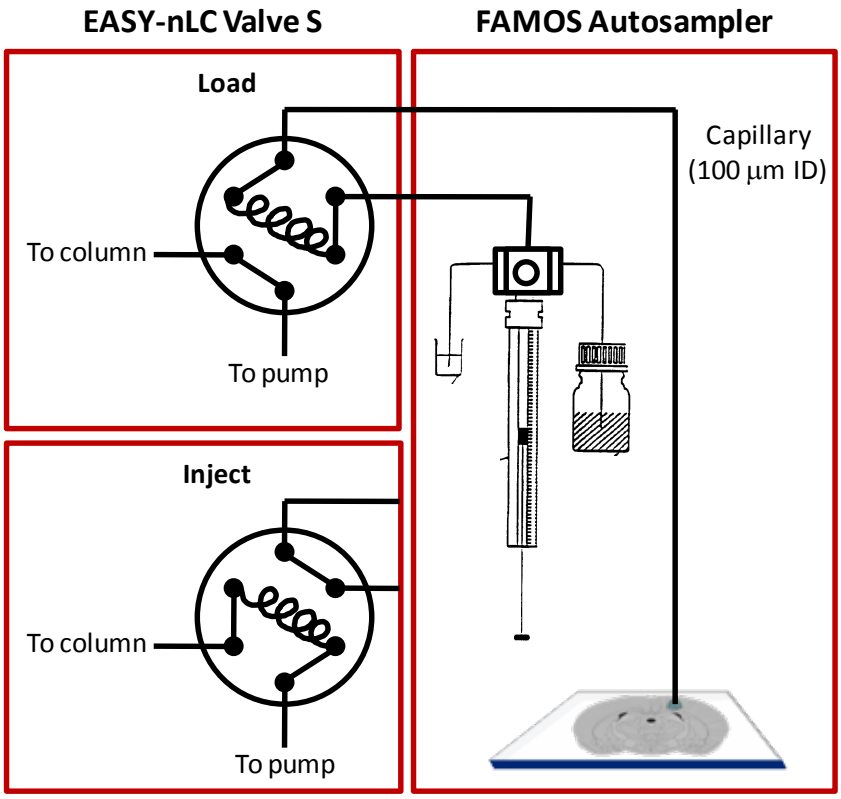


Figure 14. Coupling of the FAMOS autosampler with the valve S of Easy nLC.

Top-left and botom-left pictures show the valve position when in Load and Inject positions, respectively. The capillary that provides the liquid microjunction is connected to one end of a 20-µL sample loop, while the syringe pump of the FAMOS autosampler is connected to the other end. Solvent is aspirated from a reservoir (not shown) by the same capillary. Once extraction is finished, the extract is dispensed into an eppendorf tube placed in another reservoir. Upon attaining thedesired number of extractions, all the extracts are aspirated and loaded into the sample loop.

The FAMOS device was configured such that the “Template” and “Mix Method” options are active. Programming a series of steps in the autosampler permitting the automation of the entire liquid microjunction extraction sequence was performed using the “User Program” option. A sample sequence for this purpose is shown in **Table 2**.

Table 2. Programming sequence used to enable liquid microjunction extraction on the FAMOS autosampler.

Step	Command*	Volume	Location	Speed	Height
1	Valve Position	Load			
2	Aspirate	2 µL	Reagent A	1	10
3	Wait	20 s			
4	Aspirate	0 µL	Reagent A	1	10
5	Dispense	0.2 µL	Sample	1	16
6	Wait	5 s			
7	Dispense	0 µL	Sample	1	15
8	Wait	15 s			
9	Aspirate	0.2 µL	Sample	1	15
10	Wait	5 s			
11	Aspirate	0 µL	Sample	1	15
12	Dispense	0.2 µL	Reagent B	1	1
13	Wait	5 s			
14	Dispense	0 µL	Reagent B	1	1
15	Repeat	4X	Steps	2-14	
16	Aspirate	13 µL	Reagent B	3	1
17	Wait	10 min			
18	Aspirate	0 µL	Reagent B	1	1
19	Valve Position	Inject			
20	Needle Wash	50 µL			

*Commands are the functions that the autosampler executes. Location refers to the position of the robotic arm. Reagents A and B refer to positions of the solvent bottles located behind the sample plate, while Sample refers to the sample plate where the glass slide is mounted. The Sample location can be assigned depending on the type of sample plate used (96- or 384-well plates), and the position of extraction made precise by further adjusting manually. Speed and Height refer to the rate at which the syringe pump moves and the distance of the capillary from the bottom of the sample plate when it is being extended from the piercing needle.

nanoLC-MS

The EASY-nLC, on the other hand, was set-up as follows. The configuration was set to One-Way Contact Closure, and the data acquisition was started once the contact has been established. A method file was created and loaded in the EASY-nLC, independent of the chromatographic parameters that are being set in the Xcalibur method file. This method file instead dictates the pre-column and column equilibration volumes, flow rates and maximum pressure in the pumps in addition to the chromatographic parameters. Pre-column equilibration was set to 3 μL and the analytical column equilibration to 2 μL. The LTQ Orbitrap MS parameters, meanwhile, were still set using the Xcalibur method file. A spray voltage of 2 kV was used to generate the precursor ions, and scans were performed with a range of 300-1600 m/z. The LTQ orbitrap was operated in data acquisition mode performing MS/MS of the top 6 most intense ions from each survey scan. The resolving power of the Orbitrap mass analyzer was set to 60,000 FWHM at m/z 500 and a normalized collision energy of 35 % with an isolation window of 2 amu was used for the CID fragmentation of precursors whose intensities are over 500 counts. For the MS/MS, a target of 5000 ions and a maximum injection time of 200 ms were used. The dynamic exclusion was set to 30 s.

Data Analysis

Raw files obtained from nanoLC-MS were processed using Proteome Discoverer version 1.4.0.288 (Thermo Fisher Scientific, Bremmen, Germany). The protein identifications were obtained using the following parameters for the interrogation : Peptide Mass tolerance : ± 10 ppm, Fragment Mass tolerance : 0.6 Da, Max. Delta C_n = 0.05, Dynamic Modification : Oxidation/+ 15.995 Da (methionine), Enzyme : trypsin, 2 miscleavages, Protein FDR = 0.01. The searches were performed using the UniprotKB/Swiss-Prot database (version January 2012) filtered with Rattus norvegicus (41604 sequences) taxonomy using the SEQUEST HT algorithm. Raw files were also converted to *.mgf using MSConvert (Chambers, Maclean et al. 2012) and submitted to Mascot search as needed. For the Mascot searches, the following parameters were used : Taxonomy : other mammalia, Database : Uniprot, Enzyme : trypsin, 2 miscleavages, No fixed modifications, Variable modification : oxidation of methionine, Peptide tolerance : ± 10ppm, MS/MS tolerance : 0.6 Da, Peptide Charge : 2+, 3+ and 4+, Instrument : ESI-Trap and with the decoy search activated.

Results

Examination of BSA Standard

The performance of the configuration was evaluated using various samples. In the first instance, 0.5 µL of 10^{-4} M BSA in 50 mM bicarbonate buffer was spotted on glass slides and dried in the vacuum. The spots were extracted 5x with 0.1% FA dispensed at 0.2 µL each time and then collected and injected separately. Each extract was separated using a 30-min gradient, and the acquired *.raw files were converted to .mgf and submitted to Mascot interrogation. Results show successful identification of BSA, with the Mascot scores and Peptide matches decreasing as the number of extractions is increased **(Table 3)**. Examination of the extracted ion count of a selected peptide signal (m/z 722.32679) shows that after only three extractions, 99.7% of the peptide ion count has already been detected.

Table 3. Identification of BSA standard for each extraction using the FAMOS autosampler.

Extraction #	Mascot Score	Peptides Matched	% Sequence Coverage	Peak Area (m/z 722.32679)
1st	2729	141	62	98423539
2nd	849	57	50	11109431
3rd	787	41	47	2311063
4th	438	23	34	371891
5th	39	4	5	ND

Evaluation of Method on Rat Brain Section

In the examination of rat brain tissue sections, the size of the microjunction was first optimized to determine the minimum volume of solvent that can be dispensed which would provide the smallest sampling area possible. This was done as follows. First, the 10-µm tissue sections were soaked in 70% and 95% EtOH then in $CHCl_3$ for 30 s each and dried. Then 0.2, 0.3 or 0.4 µL of 0.1 % FA was dispensed at approximately the same position on the three consecutively cut sections. 0.2 µL is the smallest volume that can be dispensed by the autosampler syringe. Because of the small internal diameter of the capillary that dispenses the liquid on the tissue surface (100 µm), a waiting time of 5 s was set to allow the full release of the solvent before it touches the tissue surface. This in turn causes the droplet to become larger than the internal

diameter of the capillary. Unpublished accounts of other TriVersa Nanomate users suggested that the disposable tip can be made to touch the tissue surface so that very little liquid is expelled, thus reducing the sampling area equal in diameter to that of the pipette tip. In the case of the capillary though, this did not allow the extraction of surface components and was thus not pursued. Despite of this, the capillary was able to produce liquid microjunction diameters as small as 1200 μm from 0.2 μL of solvent dispensed **(Figure 15)**. This corresponds to a 30% decrease in the diameter of the surface extracted, when compared to the liquid microjunction produced by pipette tips in the conventional LESA set-up ((Quanico, Franck et al. 2013), see **Figures 2 & 4**).

Figure 15. Liquid microjunction sizes as a function of dispensed volume.

The capillary in the FAMOS autosampler is capable of dispensing 0.2 μL solvent, in contrast to the Triversa Nanomate where the tip can only dispense up to a minimum of 0.6 μL. Consequently, the size of the liquid microjunction formed is reduced.

Aside from reducing the size of the extracted area, the size of the microdigested area was also decreased. This was achieved by setting the CHIP 1000 microspotter to discrete mode, allowing the user to change the number of spots that will be microspotted per region specified. As such, the number of spots was set to 1 x 1, yielding only one spot on the marked region. The volume spotted was then set to 20 picoliters per spot. The number of droplets and waiting time were set depending on the preexisting conditions of the experiments and thus vary depending on the feasibility of forming the microdroplet prior to commencing the spotting. **Figure 16** shows an example of microdigested regions on the rat brain tissue section, showing that the diameter of microdigested spots can be decreased down to approximately 150 μm following this approach. This corresponds to about 78% reduction in microdigested region diameter. As in the LESA experiment, comparison of the extracts obtained from digested and undigested regions on the same position in a coronal rat brain section was performed, where it can be seen once again that

extraction only occurs in the digested region as shown by the very low number of tryptic peptide-based protein identifications in the undigested region (**Figure 17**). Examination of each extraction cycle by running them separately in nanoLC-MS showed that majority of the protein IDs (88%) were obtained from the first extraction alone. Successive extractions only contributed about 5 additional new identifications each time (**Figure 18**). In both experiments, only a 25-min gradient was employed for chromatographic separation, hence the low number of protein identifications even on the digested region. To determine if this indeed could be the case, the size of the microdigested area was increased in the succeeding experiment back to approximately 650 µm and the extracts ran in 1.5 and 4 h gradients using the same data acquisition parameters. Increasing the gradient separation, as expected, led to concomitant increases in the number of protein identifications (**Figure 19**).

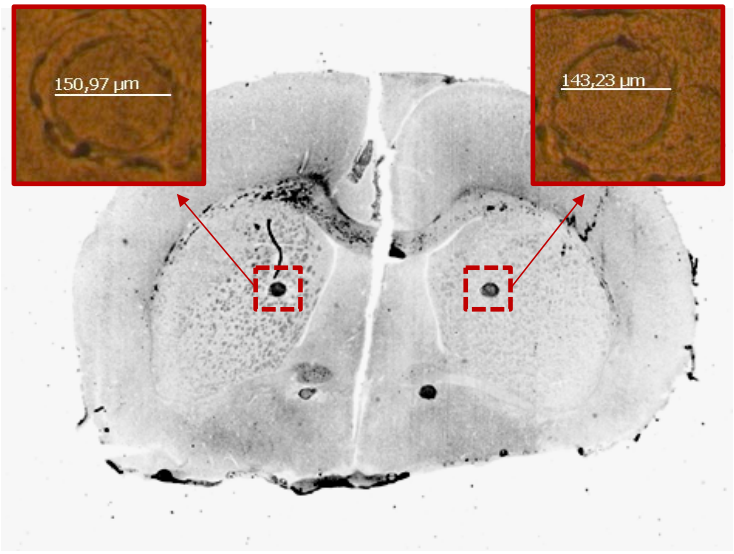


Figure 16. Sample microdigested regions reduced in diameter up to 150 µm.

Digestion was done by setting the number of spots in the digestion area (parameter required by the microspotter) to 1x1, corresponding to one droplet only.

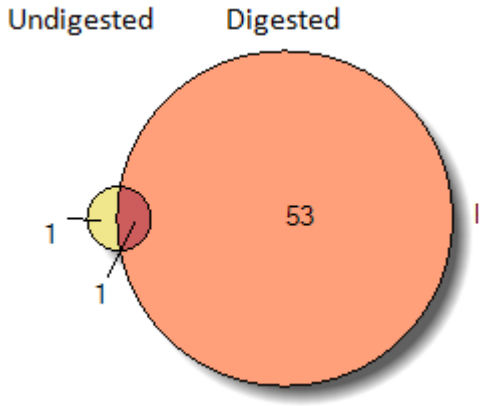


Figure 17. Comparison of protein IDs in the digested and undigested regions subjected to liquid junction microextraction using the FAMOS autosampler.

As expected, the undigested region yielded only a few protein identifications corresponding to endogenous neuropeptides. Whereas the digested region provided >50 protein IDs.

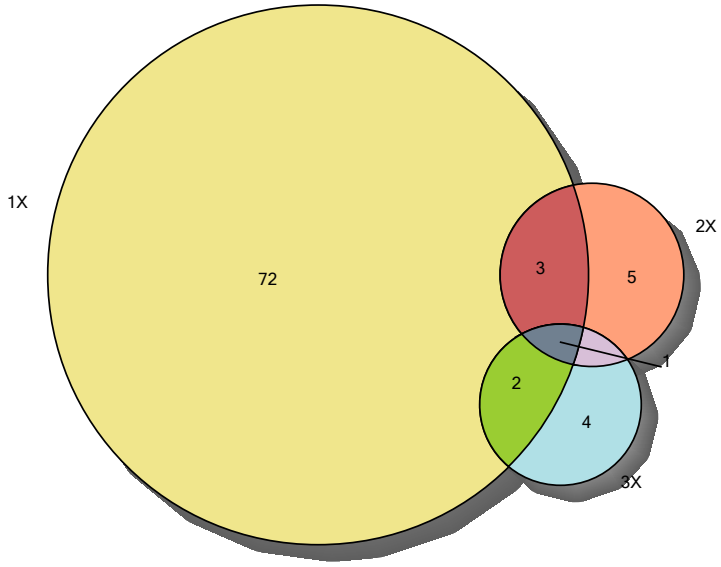


Figure 18. Protein IDs of successive extractions on the same spot using the FAMOS autosampler.

Most of the proteins extracted and identified correspond to those obtained during the first round of extraction. Nonetheless, to maximize the protein identification yield, five extractions were done on succeeding experiments.

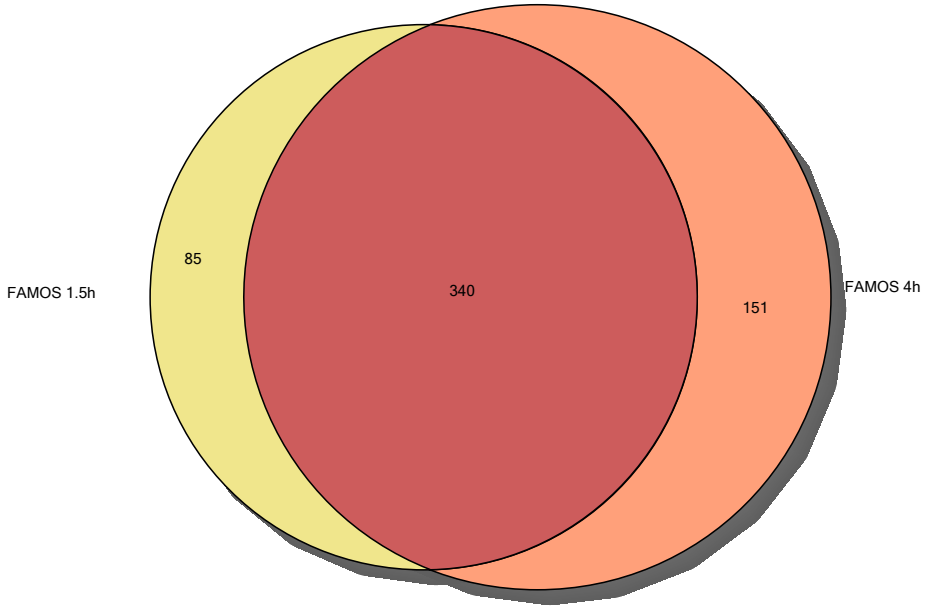


Figure 19. Increased chromatographic separation time increased the number of protein IDs for the same microextracted sample.

As expected, longer gradients (4h) yielded more protein IDs compared to shorter ones (1.5h).

Comparison of Performance Against LESA

Comparison of the performance of this FAMOS liquid microjunction extraction with that of LESA in terms of the number of protein IDs from the same digested position on the rat brain section was also evaluated using 1.5 h and 4 h gradients. In both conditions, the top 6 method was used for data-dependent MS/MS acquisition, and the protein identifications were constrained with a protein false discovery rate of less than 1%. The extracts that were subjected to protein identification were obtained using the sample preparation procedures that were optimized for each approach. This means that for the LESA approach, which was optimized for off-line analysis, extraction was done using 80% ACN/0.1% FA and desalting with Ziptips was performed. The FAMOS approach, on the other hand, was optimized for on-line coupling with a tandem nanoLC-MS instrument. As such, for the FAMOS approach, only 0.1% FA was used for extraction and no clean-up was done as it is being injected directly to the nanoLC. **Table 4** summarizes the extraction parameters used for the two approaches.

Table 4. Parameters used for comparing microextraction using the TriVersa NanoMate (LESA) and the FAMOS autosampler.

Parameter	LESA	FAMOS
Extraction solvent	80% ACN/0.1% TFA	0.1% FA
Dispensed solvent volume per cycle	0.6µL	0.2µL
Number of Extraction Cycles	5	5
Microjunction contact time	15s	15s
Ziptip?	Yes	No
Trypsin Concentration	20ug/mL	20ug/mL
CHIP Digestion size*	674µm	648µm

*Refers to diameter of digested spot

Results show that, in both gradients, more proteins were identified with the online approach than with the conventional LESA instrument (**Figure 20**). The number of identifications was significantly lower compared when using an Orbitrap Elite instrument, as expected. Although 0.1% FA does not extract as much components compared when an organic modifier is present in the solvent, the former is still sufficient to provide about 500 protein identifications using the FAMOS, given the acquisition limitations of the LTQ Orbitrap instrument. It is to be noted too that the volume of extracting solvent is almost 70% smaller when using the FAMOS; this not only reduces the size of the extraction area but also improves the extraction efficiency. Finally, elimination of sample pretreatments prior to injection to the nanoLC could also contribute to the higher protein identification yield. Although it is beneficial to desalt samples as it improves column lifetimes, minimizes adduct formation and leads to high protein recovery (Salplachta, Rehulka et al. 2004), desalting can also be a setback as small hydrophilic peptides can elute with 0.1% TFA leading to protein losses (Tannu, Wu et al. 2004). This also greatly shortens the sample preparation time as the drying steps using the speedvac were eliminated.

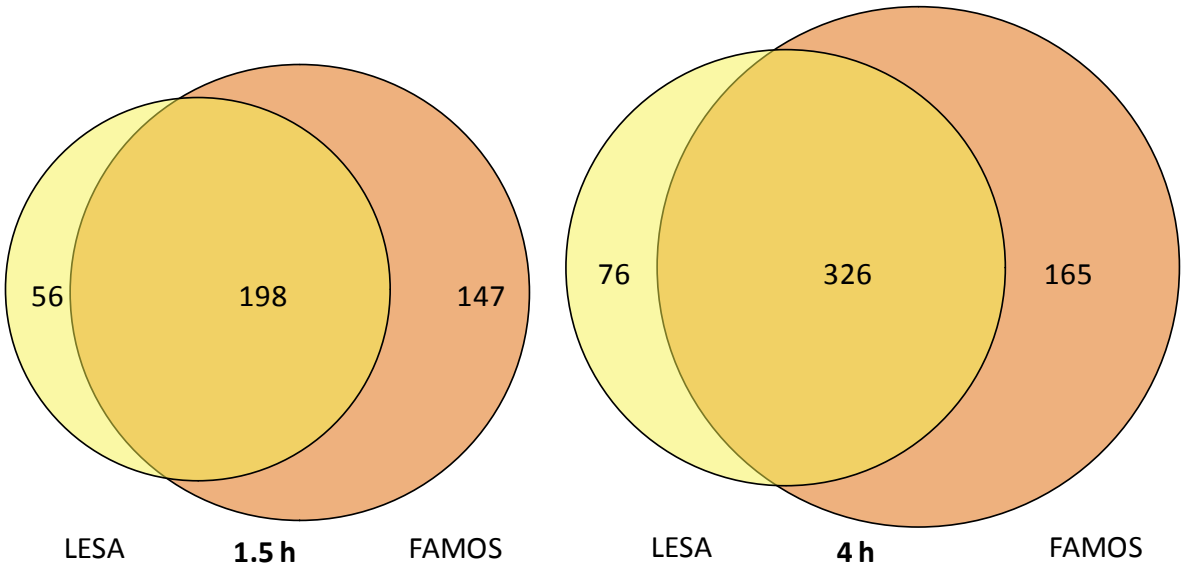


Figure 20. Comparison of protein IDs obtained using TriVersa NanoMate and the FAMOS autosampler.

Protein IDs were obtained after 1.5h and 4h chromatographic gradients and using the optimized parameters for the two methods.

The use of an autosampling device to do liquid microjunction extraction, as previously mentioned, is not novel. However, it is deemed that the capability of doing on-line direct tissue sampling has not been fully explored. Notable examples have been reported for small molecules such as drugs and their metabolites and protein standards (Schadt, Kallbach et al. 2012; Van Berkel and Kertesz 2013), as was also previously mentioned. It is now being presented here that this approach can in fact be applied for direct on-tissue microproteomics aiming at full automation from sample preparation until data acquisition.

Related devices capable of continuous flow extraction, such as the liquid microjunction surface sampling probe (Van Berkel and Kertesz 2013) and the sealing surface sampling probe (Luftmann 2004) have also been reported. Recently (Elnaggar, Barbier et al. 2011), it has been demonstrated that extraction using the LMJ SSP device is more efficient compared to discrete droplet-type microextraction because the liquid junction can be maintained for extended periods of time and fresh solvent continually replenishes the probe so that continuous extraction of the surface can occur. Its main disadvantage though is that the continuous flow implies dilution of the sample, and as such, signal profiles tend to be maximal during the initial stages of the extraction and diminish as a function of extraction time (Van Berkel and Kertesz 2013). This was overcome by collecting the extracts in a sample loop prior to injection to the column. Using a 20-

μm loop, the authors report extraction times up to 8 minutes, given a flow rate of 25 μL/min. This implies however that the size of the sample loop will limit the volume of solvent containing the extracted material that can be collected. Meanwhile, to ameliorate the extraction efficiency, multiple cycles of extraction are performed on the discrete droplet-type microextraction. Several reports (Eikel, Vavrek et al. 2011; Walworth, ElNaggar et al. 2011) indicate that repeated cycles could lead to increase in the size of the microextracted area, and that the small size of the droplet limits the dissolution of compounds to the liquid interface. Results of this bottom-up proteomics approach indicate that the latter is not the case, since close to a thousand proteins were in fact identified from just five cycles of extraction on a sample surface as small as 650 μm. Also with this approach, increases in the size of the microextracted area are not critical because the microdigested region ultimately determines the size of the extractable area. Perhaps the advantage that can be gleaned from the use of smaller microjunction areas then is that the volume of solvent used will also be small, thus, more cycles of extraction can be performed, if really desired. In the experiments, up to 5 cycles of extraction were done, using 0.2 μL of solvent each time. They were dispensed in a sample reservoir containing 13 μL of 0.1% FA, and after all cycles of extraction have been completed, 13 μL was injected into a 25 μL-sample loop. Note that the loop was not filled to capacity. During sample loading, the missing liquid was compensated for by solvent A (0.1% FA) and then injected to the column. If the need arises, additional extraction cycles can be done further since there is still an excess of the loop volume that needs to be filled. Even though the volume is very small and can thus limit the solubility of peptides, this was overcome because the repeated cycles of extraction leads to wetting of the sampling surface and thus improves the extraction efficiency with each consecutive cycle.

Perhaps the biggest limitation of the current bottom-up, liquid junction microextraction approach is the efficiency of the digestion process. Compared to conventional sample digestion techniques, *in situ* digestion has its distinct challenges. For instance, proteins *in situ* present structural conformations that render them less susceptible to proteolysis compared when they are in solution. Also, the amount of proteins on micrometer-thick tissue sections is significantly lower compared to whole organs (McDonnell, Corthals et al. 2010); their amounts can therefore be easily overwhelmed by successive addition of proteolytic enzymes leading to autolysis. Also, successive deposition of the proteolytic enzyme can easily lead to loss of spatial resolution because of the diffusion of the solvent dependent on the properties of the surface. **Supplementary Figure 1** (see Annex) shows a full MS survey scan from t= 56.28 min of the

1.5h gradient run using the FAMOS autosampler previously shown in **Figure 20**. The MS scan shows isotopic patterns of intact proteins, suspiciously corresponding to ubiquitin, an abundant protein. The presence of undigested proteins is indicative of incomplete digestion, which is expected given the short digestion time (1h), unregulated temperature, and small volume of trypsin in contact with the surface. Inspection of the sequence coverage of the proteins identified also indicates, in general, low coverage and number of peptides identified per protein, with an average of 47% having sequence coverage of less than 5% and having only 1 peptide. In fact, if the number of peptides is set to 2, the number of protein identifications is greatly reduced to 253 for the sample ran in a 4h gradient using the FAMOS autosampler. This is also unsurprising, since no antigen retrieval, chaotrope, or reducing agent was used to improve the digestion. The extent of digestion depends on a variety of factors. The presence of disulfide bonds between two adjacent cysteine residues plays an important role in protein stability, as it renders them less susceptible to denaturation and subsequently degradation (Betz 1993). As such, reduction of these bonds by reducing agents such as DTT, beta-mercaptoethanol, or TCEP is often incorporated in conventional protein digestion protocols, coupled with cysteine capping using iodoacetamide to prevent the re-oxidation of these residues. Introduction of chaotropes by heating to denaturing temperatures, addition of high concentrations of urea or guanidinium hydrochloride, addition of organic solvents, etc. is also imperative in the development of protein digestion protocols as the denaturation of proteins allows for the exposure of otherwise buried lysine and arginine residues thereby facilitating their cleavage.

Performing efficient digestion at the least possible time still remains a challenge not just for this approach but for comprehensive proteomics in general, and it has been examined extensively (Hustoft, Reubsaet et al. 2011). The authors finally concluded that none of the methods accelerating digestion (such as infrared energy, microwave energy, aqueous/organic systems or filter-aided sample preparation (FASP)) are as efficient as performing in-solution digestion for 12-16h (overnight). For *in situ* digestion, recent developments have been reported, namely by Harris *et al* and Jiao *et al* (Harris, Nicklay et al. 2013; Jiao, Miao et al. 2013). Harris *et al* used preformed, sub-millimeter-sized hydrogels with trapped trypsin to facilitate surface digestion once it is deposited on the tissue section. Although the approach provides a means to allow the use of multiple successive experimental workflows on the same tissue section, only 212 proteins were reportedly identified given that the hydrogel was incubated for extended periods of time (6-12h) at 50°C. Meanwhile Jiao *et al* immobilized trypsin on graphene oxide and deposited the

particles *in situ* prior to matrix deposition and MALDI MSI. Compared to other supports, graphene oxide presents notable structure and chemical features that improves protein digestion efficiency and makes it amenable to MALDI MSI.

Perhaps the most promising approach is by performing digestion under pressure such as in on-line digestion systems like the FOLDS set-up (Lopez-Ferrer, Petritis et al. 2010). High pressure causes proteins to get denatured, making them more susceptible to proteolysis. Coupling of on-tissue extraction with high-pressure digestion systems can be done, for instance, by performing intact protein extraction and then injecting them into an immobilized pepsin column interfaced with a nanoLC-MS instrument. This brings about the question of whether it is possible to perform intact protein extraction using the liquid junction extraction approach. Initial attempts were done by loading 1% SDS on the tips for LESA extraction, however, it was difficult to accurately aspirate or dispense the solution because it is highly viscous. Once the solution is dispensed, it rapidly spreads on the tissue thus creating a larger spot extracted. Furthermore, removal of the detergent is impossible without having to subject the extract to SDS-PAGE separation, FASP, or tube-gel clean-up and digestion (Lu and Zhu 2005; Wisniewski, Zougman et al. 2009).

Protein extraction in the sub-millimeter scale has its advantages and disadvantages. On the one hand it is possible to obtain a localized distribution of proteins up to a level of spatial resolution of about several cells. This is expected to provide a much better estimation of morphological changes in the tissue since a more homogeneous population of cells is sampled. The ultimate goal of on-tissue microextraction is to be able to sample single cells *in situ*, since it has been established that individual variation is apparent even for a population of what was otherwise perceived as homogeneous cells. However, this has its own unique consequences and limitations. For one, the number of DNA copies of a single protein is limited per cell. This means that the amounts of protein expressed per unit cell is significantly lower compared to those from a cell population. This imposes a much lower limit of detection to mass spectrometers. In the presence of more abundant proteins such as albumin, these lowly-expressed proteins become undetectable because of predominant ion suppression effect. Extensive work by Sweedler's group and others (Nemes, Rubakhin et al. 2013) on the metabolomic profiling of a single cell have shown that even with sophisticated instrumentation, one is able to monitor only a small portion of the 100-1000 small molecules that are expected to be produced in single cells, mainly because these molecules are detected in their natural abundances (ie., without the use of amplification procedures such as those employed in DNA quantification). In this regard, it might be ideal to

restrict spatial resolution up to a certain limit so that there still remains a certain degree of protein enrichment that will be sufficient for detection. This limitation will be imposed by the nature of the tissue that is being studied, and will vary depending on its extent of homogeneity. Note that this should be similar in principle to laser capture microdissection, where enrichment is done by collecting thousands of cells of identical phenotype at the expense of losing localization because of the need to collect them in several places on the tissue. Also, it might be ideal to continue with this approach because, in contrast to single cell analysis, the cell microenvironment is being considered when following this approach.

The other complication is the complexity of proteins owing to their bigger sizes relative to other types of biological molecules. Because current mass spectrometers are designed to detect masses within the low mass range, there is a bias against high-molecular weight proteins, although this is currently being overcome with the design of new generation mass spectrometers. Although it is desirable to do protein identification on the basis of their in source decay fragmentation alone as in bottom-up proteomics approaches, current in-source decay instruments lack precursor ion detection which renders them useless for comprehensive proteomics analysis.

Conclusions – Part 1

Develop alternative methods to ameliorate MSI protein identification strategies.

This can be achieved by developing a localized protein extraction method based on the liquid microjunction approach. The approach is similar in principle to the generation of protein ID databases that are then used for matching with peaks in MSI spectra. In contrast though, the protein database will be obtained from localized regions on the tissue section. This will be done by performing *in situ* tryptic digestion on those regions, followed by extracting the peptides using the liquid microjunction approach, then subjecting the extracts to shotgun nanoLC MS.

This section demonstrated the applicability of the liquid microjunction for localized microproteomics analysis directly on tissue, highlighting its advantages and disadvantages in this context. Coupled with on-tissue microdigestion, the liquid microjunction strategy provides a robust means to examine proteins of limited abundance by extracting and identifying them from individual regions of the tissue. In return, the identified low abundance proteins were used to discern the distinct regions from each other, showing also the applicability of this strategy for tissue profiling. In spite of the small region (and consequently low number of cells) sampled, the liquid microjunction strategy can provide a large number of protein identifications. As such, it can serve as a suitable alternative to whole tissue/organ protein extraction for database generation in MSI back-correlation methods.

PART 2 : TOWARD QUANTIFICATION-BASED MS IMAGING

Chapter 3

Article 2

Quantification-Based Mass Spectrometry Imaging of Proteins by Parafilm Assisted Microdissection

Authors: Julien Franck, Jusal Quanico, Maxence Wisztorski, Robert Day, Michel Salzet and Isabelle Fournier

Article Status: published in Analytical Chemistry [Volume 85(17) pages 8127-34, 2013]

Summary: Contributed in the design of experiments and their implementation, interpretation of data, and writing of the manuscript. In detail, my contributions include:

1. Preparation of tissue sections and microdissection of regions of interest, followed by protein extraction, digestion and subsequent nanoLC-MS.
2. Performed all mass spectrometry imaging experiments and related techniques such as hierarchical clustering and image processing.
3. Analysis of chromatographic and protein identification data, quantification by spectral counting, generation of PAM images based on spectral counts.

In terms of writing the manuscript, Mr. Quanico wrote and revised the drafts together with Mr. Franck.

Résumé en français

La technique d'imagerie par spectrométrie de masse (MSI) MALDI fut présentée comme une bonne stratégie pour identifier les régions présentant des phénotypes spécifiques basés sur les différences de profils moléculaires. Les protéines présentes dans différentes régions peuvent être identifiées *in situ* mais le nombre de protéines identifiables reste très faible comparé aux approches protéomiques basées sur la MS. Pour outrepasser ce problème, une nouvelle stratégie impliquant la microdissection manuelle de régions de sections de tissus montés sur des lames histologiques recouvertes de parafilm M, fut étudiée dans cet article. L'extraction et le fractionnement de protéines à partir de régions d'intérêts a permis d'identifier plus de 1000 protéines venant de chacune des microdissections. L'intérêt de cette approche réside dans son faible coût et la facilité d'exciser des zones de l'ordre du millimètre. Ainsi il possible d'identifier des protéines venant de cellules de phénotype spécifique mis en évidence par la classification moléculaire à partir d'expérience de MALDI-MSI. Cette approche peut être appliquée à l'ensemble d'une coupe de tissu afin de générer des images moléculaire de protéines basée sur la quantification sans marquage. Cette approche a été appliquée sur une coupe de tissu monté sur une lame recouverte de parafilm M, qui a ensuite été découpée en morceaux réguliers. L'identification et la quantification relative des peptides selon la stratégie du comptage de spectres obtenus par nanoLC MS & MS/MSa permis de reconstruites des images de protéines. Cette stratégie de MSI basée sur la quantification offre la possibilité de pouvoir cartographier un grand nombre de protéines, qu'elles soient en grande ou en faible abondance.

QUANTIFICATION BASED MASS SPECTROMETRY IMAGING OF PROTEINS BY PARAFILM ASSISTED MICRODISSECTION

Julien Franck[1], Jusal Quanico[1,2], Maxence Wisztorski[1], Robert Day[2], Michel Salzet[1] and Isabelle Fournier[1*]

[1]Laboratoire de Spectrométrie de Masse Biologique Fondamentale et Appliquée -EA 4550, Université de Lille 1, Bât SN3, 1er étage, F-59655 Villeneuve D'Ascq, France

[2]Institut de pharmacologie de Sherbrooke, Département de chirurgie/service d'urologie, Faculté de Médicine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, Québec, J1H 5N4, Canada

ABSTRACT

MALDI-MSI was presented as a good strategy to highlight regions presenting specific phenotypes based on molecular content. The proteins present in the different areas can be identified by MALDI MSI, however, the number of protein identifications remains low in comparison with classical MS-based proteomics approaches. To overcome this, a new strategy, involving the microdissection of tissue sections mounted on parafilm M-covered glass slides is presented. Extraction and fractionation of proteins from a specific region of interest was investigated, leading to the identification of more than 1000 proteins from each microdissected piece. The strength of this cheap technique lies in the facile excision of millimeter-sized portions from the tissue allowing for the identification of proteins from cells of a specific phenotype obtained from the MALDI MS imaging based molecular classification using hierarchical clustering. This approach can be extended to whole tissue sections in order to generate images of the section based on label-free quantification obtained from identification data. As a proof of concept we have studied a tissue mounted on a parafilm M-covered glass slide, cut it into regular pieces and submitted each piece to identification and quantification according to the developed PAM method. Images were then reconstructed by relative quantification of identified proteins based on spectral counting of the peptides analyzed by nanoLC-MS & MS/MS. This strategy of quantification-based MSI offers new possibilities for mapping a large number of high and low abundance proteins.

*: Corresponding author: Prof. Isabelle Fournier, Laboratoire de Spectrométrie de Masse Biologique Fondamentale et Appliquée (FABMS) - EA 4550, Bât SN3, 1er étage, Université Lille 1, F-59655, Villeneuve D'Ascq, France. Tel: +33 (0)3 2043 4194, Fax: +33 (0)3 2043 4054, email: isabelle.fournier@univ-lille1.fr

INTRODUCTION

To date, MALDI Mass Spectrometry Imaging (MALDI-MSI) is widely used to map the distribution of compounds within tissue sections. Different classes of endogenous or exogenous molecules are detected namely drugs (Drexler, Tannehill-Gregg et al. 2011; Kallback, Shariatgorji et al. 2012; Shanta, Kim et al. 2012), metabolites (Trim, Francese et al. 2009; Sugiura and Setou 2010; Shanta, Kim et al. 2012), lipids (Hankin, Farias et al. 2011; Landgraf, Garrett et al. 2011; Murphy, Hankin et al. 2011), peptides (Goodwin, Pennington et al. 2008; Sole-Domenech, Johansson et al. 2010; Bruand, Sistla et al. 2011) and proteins (Grey, Chaurand et al. 2009; Franck, Longuespee et al. 2010; Clemis, Smith et al. 2012) from organs or even whole body animals (WBA). The strength of MALDI MSI lies in the detection and localization of hundreds of compounds without any labeling in one acquisition step. The technique is now widely used in many fields of research for various applications such as DMPK or ADME studies in pharmacology (Prideaux, Staab et al. 2010; Goodwin, Iverson et al. 2012; Takai, Tanaka et al. 2012), and biomarker investigation in oncology (Lemaire, Menguellet et al. 2007; Elsner, Rauser et al. 2012; Longuespee, Gagnon et al. 2013) or brain disorders such as Parkinson and Alzheimer diseases (Stoeckli, Staab et al. 2002; Stauber, Lemaire et al. 2008; Strobel, Haerle et al. 2009). It has been shown to be a well-suited tool to highlight changes in biomolecule regulation as the organisms undergo alterations in their physiological state. If corresponding molecules are identified, valuable information can be obtained on the signaling pathways involved which can then be used to deduce the underlying biological mechanisms. Identification of imaged molecules is, thereof, essential for the understanding of biological phenomena.

Since several years many groups have developed strategies allowing the direct identification from tissue sections of the different classes of molecules. This has been successfully demonstrated for drugs, lipids and peptides (Cerruti, Benabdellah et al. 2012; Gagnon, Franck et al. 2012). For proteins, however, conventional ion activation methods failed to produce fragmentation in MALDI from singly-charged state ions. By performing In-Source Decay (ISD), a single-step Top-Down strategy has been developed for protein identification using 1,5-Diaminonaphthalene (1,5-DAN) as matrix (Bonnel, Longuespee et al. 2011; Debois, Smargiasso et al. 2013). Likewise, a Bottom-Up strategy employing *in situ* enzymatic digestion based on the deposition of a proteolytic enzyme on the surface of the whole section or on discrete regions has been developed (Franck, El Ayed et al. 2009; Gustafsson, Oehler et al. 2011). However, in both cases,

only the major proteins are identified because the signal from less abundant proteins is hampered by the complexity of the sample. To overcome this, proteins can be *in situ*-digested and the resulting peptides microextracted prior to their analysis by nanoLC-MS & MS/MS minimizing the ion suppression effects by addition of a separation step and concentrating the peptides. This strategy allows the identification of many proteins including the less abundant ones, but the localization is lost. This can be regained by doing a back correlation of the m/z of the off-tissue identified peptides to those of the tryptic peptides detected by MALDI-MSI. This strategy was described in the case of oncology (El Ayed, Bonnel et al. 2010) and neurodegenerative diseases(Stauber, Lemaire et al. 2008). However, the inaccurate mass measurement from the two different instruments used could be detrimental. Recently, Schober et al. have updated this strategy by using a high accuracy mass spectrometer (Schober, Schramm et al. 2011). Nevertheless, even if the procedure is now very attractive due to the accurate mass measurement available, the analysis is still limited mainly by the extent of the digestion efficiency on the tissue section in imaging mode and by the dominating presence of abundant proteins which tend to suppress the signal of the low abundant ones.

An alternative strategy would consist of extracting all proteins on a restricted region of interest (ROI of around several hundred μm^2) allowing for their identification in a complete proteomics approach involving enzymatic digestion and nanoLC-MS/MS. In this strategy, many proteins could be highlighted and their localization maintained. To date, Laser Capture Microdissection (LCM) is largely used to study proteins from cells of a defined phenotype which are automatically or manually selected according to morphology criterion. However, LCM requires a relatively large number of cells in order to reach the sufficient amount of proteins needed for a conventional proteomics investigation (1500-2500 proteins identified from 50 000-60 000 cells collected) (Wisniewski, Dus et al. 2012). This makes the technique meticulous and time-consuming especially if the cell collection has to be performed manually. On gross tissues, the morphological changes are also observable and often, a single histological analysis to discern between normal and diseased tissues is sufficient especially when dealing with multiple samples. However, gross tissue analysis can be unspecific especially if only restricted areas are affected by the disease. Thus, a compromise between the highly specific LCM and gross tissue analysis can be performed. One possible approach would be the use of liquid microjunction-microextraction, which was already presented as an efficient strategy to identify proteins from discrete locations on a tissue section (Quanico, Franck et al. 2013; Wisztorski, Fatou et al. 2013). In this approach,

proteins were first digested *in situ* using a microspotter and the resulting peptides were extracted by liquid microjunction prior to their analysis using nanoLC-HRMS & MS/MS. From a region of approximately 600-700 µm in diameter, up to 1500 proteins were identified showing the great interest in this new strategy for biomarker hunting and understanding of physiopathological processes. This strategy was applied to Formalin Fixed and Paraffin Embedded (FFPE)(Wisztorski, Fatou et al. 2013) tissues for studying the fimbria region of tubo-ovarian cancer samples allowing the identification of 983 and 792 non-redundant proteins in the benign and cancer regions, respectively.

Recently, Kristiansen and coworkers (Kristiansen 2010) have presented an approach where samples are manually microdissected under microscope using a needle and the dissected tissue was then soaked in a cell lysis buffer. This strategy has presented good results in a very short time while maintaining the localization within the tissue. Unlike LCM, this procedure allows the analysis of cells and also their micro-environment in a high throughput mode and different class of biomolecules could be analyzed.

Here is presented another strategy where tissue sections are mounted on a parafilm M-covered glass slide allowing for their manual microdissection under microscope using a scalpel. In parallel, a consecutive section is subjected to MALDI-MSI for hierarchical clustering (HC), providing information on the molecular content of groups of cells. It is then possible to distinguish ROIs that can be manually microdissected and the small pieces of parafilm M that contain the ROIs are then soaked in buffer followed by digestion of the proteins. The resulting peptides are then analyzed using nanoLC-HR MS & MS/MS providing very accurate mass measurement allowing for high confidence identification of thousands of proteins. The complete sequence of this new procedure called Parafilm-Assisted Microdissection (PAM) is presented in Figure 1**.** Because simple identification does not provide any information on protein abundance variations, we have examined protein quantification based on a label-free strategy using spectral counting. By combining all these in an integrated approach, we present a new concept of molecular imaging by MS where images of each identified proteins can be reconstructed based on label-free quantification data.

EXPERIMENTAL

Reagents

Reagents used for the experiments are listed and accessible in Supplementary information file.

Tissue preparation & MS Imaging

Detailed protocols for tissue dissection, sectioning and preparation for MALDI MS Imaging as well as statistical analysis (Hierarchical clustering) of MS imaging can be found in Supplementary Information.

Microextraction & digestion

The consecutive 10-µm sections of the rat brain adjacent to the one subjected to MALDI MS imaging were mounted on a polylysine glass slide covered with a layer of parafilm M (Pechiney Platic Packaging, Chicago, Illinois). For this, the parafilm M was manually stretched and placed on the surface of the polylysine glass slide. According to the hierarchical clustering, regions of interest were then selected and cut using a scalpel. The pieces of parafilm M containing the tissue were then placed in a tube containing 20 µL of a solution of DTT (50 mM) in NH_4HCO_3 buffer (50 mM) (pH=8) and heated for 15 min at 55°C (incubation temperature was limited to 55 °C only to prevent the parafilm M from melting). After cooling, 20 µL of a solution of IAA (100 mM) in NH_4HCO_3 buffer (50 mM) were added and the mixture was incubated for 15 min at room temperature in the dark. 25 µL of a solution of trypsin (20 µg/mL) in NH_4HCO_3 (50 mM) were then added and the sample was incubated overnight at 37 °C.

Intact Protein Extraction and SDS-PAGE

Intact proteins from the PAM pieces were extracted by incubating the pieces in 20 µL of 50 mM bicarbonate buffer containing 50 mM DTT and 1% SDS at 55 °C for 15 min. The extracts were then loaded on an 11.5% polyacrylamide gel and separated at 70 V for 15 min and then 150 V until the dye front reaches the other end of the gel. After migration, the gel was incubated in the gel fixative solution for 20 min and stained with colloidal Coomassie brilliant blue overnight. The stain was then removed by washing the gel three times with distilled deionized water.

NanoLC-MS/MS

Samples were desalted on a C-18 Ziptip (Millipore,Saint-Quentin-en-Yvelines, France), dried under vacuum and then resuspended in AcN/0.1% FA, 5:95, v/v). The samples were separated by online reversed-phase chromatography using a Thermo Scientific Proxeon Easy-nLC system equipped with a Proxeon trap column (100 µm ID x 2 cm, Thermo Scientific) and C18 packed tip column (100 µm ID x 15 cm, NikkyoTechnos Co. Ltd). Elution was carried out using an increasing gradient of AcN (5% to 40% over 110 minutes) and a flow rate of 300 nL/min. A voltage of 1.7 kV was applied via the liquid junction of the nanospray source. The chromatography system was coupled to a Thermo Scientific Orbitrap Elite mass spectrometer programmed to acquire in a data-dependent mode. The survey scans were acquired in the Orbitrap mass analyzer operated at 120,000 (FWHM) resolving power. A mass range of 400 to 2000 m/z and a target of 1E6 ions were used for the survey scans. Precursor ions observed with an intensity over 500 counts were selected "on the fly" for ion trap collision-induced dissociation (CID) fragmentation with an isolation window of 2 u. and a normalized collision energy of 35%. A target of 5000 ions and a maximum injection time of 200 ms were used for MS^2 spectra. The method was set to analyze the top 20 most intense ions from the survey scan and a dynamic exclusion was enabled for 20 s.

Tandem mass spectra were processed with Thermo Scientific Proteome Discoverer software version 1.3. Spectra were searched against UniprotKB/Swiss-Prot (version January 2012) filtered with *Rattus norvegicus* (41604 sequences) taxonomy using the SEQUEST® algorithm (version 1.3.0.339). The search was performed choosing trypsin as the enzyme with one missed cleavage allowed. Precursor mass tolerance was 10 ppm and fragment mass tolerance was 0.5 Da. N-terminal acetylation, phosphorylation of tyrosine, serine and cysteine, methionine oxydation and arginine deamidation were set as variable modifications. Peptide validation was performed with the Percolator algorithm by filtering based on a q-value below 0.01, which corresponds to a false discovery rate (FDR) of 1%. The proteins identified using Sequest were then further examined using Scaffold version 3.6.2 (Proteome Software Inc., Portland, Oregon). Scaffold uses the Peptide Prophet algorithm (Keller, Nesvizhskii et al. 2002) to validate the peptide identifications and the Protein Prophet algorithm (Nesvizhskii, Keller et al. 2003) to validate protein identifications. Peptide and protein identifications were accepted if they can be established at a confidence level greater than 95.0% and 99.0%, respectively, as specified by these algorithms. Proteins sharing the same peptides and for which the peptide IDs cannot be discerned were

grouped together to satisfy the principles of parsimony. Only protein IDs containing 2 or more peptides were considered.

Quantification-based MS Imaging approach

A 10-µm rat brain tissue section was mounted on a parafilm M covered glass slide and half of the tissue was microdissected manually using a scapel under microscope to produce 60 equal portions of approximately 1.5 mm^2 in size. The section was then scanned and each piece numbered on the scanned photo so that they can be used later for mapping the proteins on the tissue. Each tissue piece on the scanned picture was also measured for use in data normalization. Tissue pieces were then excised and labeled with their corresponding number on the scanned photo. From each piece, intact proteins were digested as described before, and the obtained tryptic peptides were desalted using C18 Ziptips. Samples were separated using a Thermo Scientific Proxeon Easy-nLC system equipped with a Proxeon trap column (100 µm ID x 2 cm, Thermo Scientific) and C18 packed tip column (75 µm ID x 15 cm, Acclaim Pepmap 300 Thermo Scientific). Additional information regarding the separation and identification of the proteins can be found in the Supplementary Information. After loading the MSF files for all the parafilm M-mounted samples in Scaffold, normalized total spectrum counts based on the total number of spectra associated with the protein inclusive of the spectra from the group members (as generated by clustering), were obtained for each piece. In Microsoft Excel software version 2007, the resulting values were then tabulated to the numbers that corresponds to the positions of the PAM pieces on the tissue section. Additionally, the Excel cells were selected in order to reconstruct the shape of tissue section. To generate maps for each identified protein, the Excel cells were subjected to the Conditional Formatting tool using the Three-Color Gradient Scale function following a coloring scheme with gray corresponding to the lowest value (0 spectral counts), yellow to the 50% value, and orange to the maximum value. This scaling method was done individually for each protein being mapped, so that the yellow and orange colors always correspond to the 50% and maximum values of the spectrum counts of that protein.

RESULTS & DISCUSSION

MALDI-MSI & molecular histology of rat brain

Classification of molecular profiles by molecular histology provided by a MALDI MSI experiment after hierarchical clustering (HC) will highlight cells of different phenotypes. Since

we are interested in studying proteins we should logically perform HC from MALDI MSI protein data. However, MALDI MSI of proteins requires specific preparation procedures that can be time consuming such as tissue washing (Lemaire, Wisztorski et al. 2006) and matrix deposition. In fact, due to their high molecular weight, protein signals can often be masked by abundant lipid signals. More important than experimental time, since proteins require true incorporation into matrix crystals during the crystallization process for MALDI, contrary to small molecules such as lipids, they also require a matrix deposition method where a minimum volume of solvent is spotted or sprayed. This makes impossible the use of systems allowing very fine and homogenous crystal layer deposition where spray is very dry. This also as consequence impairs the spatial resolution of experiments. On the other hand, for lipids, the matrix can be deposited in various ways (spray coating (Gustafsson, Oehler et al.)) or even solvent-free methods (Murphy, Hankin et al. 2011)) generating a gain in spatial resolution and do not require tissue treatment prior to matrix deposition allowing for a gain of time. Because MALDI MSI of lipids can be performed under simpler preparation methods and under higher spatial resolution conditions, lipid expression variations was examined to determine whether it correlates to protein expression variations. For this, two consecutive sections were subjected to MALDI-MSI experiments for the detection of lipids on one hand and proteins on the other. Based on the detection of both classes of biomolecules, hierarchical clustering (HC) was established. HC results for lipids and proteins are presented in Figure 2. Using the same algorithm for both lipids and proteins allowed the classification of the hippocampus, corpus callosum and thalamus regions. However, the ventromedial hypothalamic nuclei can only be discerned in the cluster for the lipids; in the case of protein clustering, these regions are indistinguishable from the thalamus. This could be attributed to the slight delocalization of proteins on the section after washing or even matrix deposition. Indeed, as presented in Figure S-1**,** the m/z 810.7 present a distribution in the corpus calosum, thalamus, ventromedial hypothalamic nuclei and mediodorsal nuclei while m/z 772.6 is localized within the hypothalamus, brain stem, hippocampus and amygdala, and m/z 769.6 is localized in the primary and secondary motor cortex and hypothalamus. This clearly shows that according cell phenotypes there are clear changes in the molecular content that affect drastically various biomolecule classes (including most abundant ones) with a correlation between them. This gives similar classification by HC whatever the classification is based on (lipids, peptides or proteins). We can even observe that classification is more precise based on lipid data due to the preparation of sample. Thus, ROIs can be determined based on lipids and used further for

proteins. Therefore, based on the lipid clusters, 3 ROIs were selected for protein analysis using the PAM strategy.

Microextractions from rat brain

In order to establish the efficiency of the new microextraction strategy based on the PAM approach, the three different ROIs selected from HC data were excised under the microscope from an adjacent rat brain section (Figure 3A). ROI1 corresponds to the CA3 and dentate gyrus of the hippocampus, the ROI2 is more focused to the external capsule of the corpus callosum while the ROI3 is dedicated to the lateral globuspallidus and the caudate putamen (striatum) (Figure 3B). The dimensions of the ROIs are shown in Figure 3A allowing calculation of the corresponding areas, respectively (ROI1 = 3.4 mm^2, ROI2 = 2.86 mm^2, ROI3 = 1.77 mm^2). Proteins from these ROIs were then reduced, alkylated and digested overnight by trypsin and the generated peptides were separated using a nanoLC device coupled to a HR mass spectrometer. MS/MS allows the identification of approximately 1140 proteins when the IDs from all the regions are summed up (Figure 3C). Moreover, 36 proteins were only detected in ROI1, 31 in ROI2 while 40 proteins where detected in ROI3 as shown in the Venn diagram in Figure 3C. The corresponding base peak chromatograms and the detailed list of protein IDs can be found on the Figure S-2 and Supplementary Information-2 respectively. It must be noted that the ROIs don't have the same size; this was intentional to demonstrate the versatility in the size and shape of the regions that can be sampled. For comparative studies it is recommended to analyze ROIs of the same sizes so that the variation in protein content will be reflective of the type of cells present and consequently of the variation in protein expression. This implies that it is not necessary to obtain regions of uniform cell content. In tissues where there is innate heterogeneity of cells or induction of such when the tissue is in a disease state, it is important not only to show the variation in protein expression but also to show that there is a change in cell phenotype. This is evident for example in the case of prostate adenocarcinomas where the progression from the early to late stages of the disease is concomitant with the increase of cancer cells and their invasion of nearby glands leading to loss of cell differentiation (Tomlins, Rubin et al. 2006).

With minimal amount of starting materials, more than 1000 protein IDs are still obtainable and very specific proteins can be highlighted on localized regions. For instance, examination of the protein IDs from ROI1 shows proteins specifically expressed in this region, such as gamma-synuclein (found in the cerebellar cortex, thalamus, hypothalamus, and the CA1, CA2, CA3 and

CA4 regions of the hippocampus) (Buchman, Hunter et al. 1998), and protein unc-13 homolog B (found predominantly in the cerebral cortex, hippocampus and striatum) (Augustin, Betz et al. 1999) for example. From ROI3, diacylglycerol kinase beta (confined to the caudate putamen, the accumbens nucleus and the olfactory tubercle) (Hozumi, Fukaya et al. 2008) and cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A (detected in striatum and testis at the protein level) (Fujishige, Kotera et al. 1999). However, in a biological investigation, experiments would have to be repeated at least 3 times in order to improve the accuracy on the quantified proteins by label-free quantification.

Moreover, the technique is very easy to implement. Indeed, in comparison with MSI of proteins that requires specific instrumentation for sample preparation, user's skills and is generally time consuming, no sample pretreatments such as lipid and peptide removal are necessary. This strategy only requires the transfer of a thin layer tissue section on a parafilm M-covered glass slide, materials which are readily available in the laboratory. In traditional proteomics approaches, entire organs are homogenized and interesting proteins that are generally expressed in a limited number of cells are diluted, limiting their detection, and even if they were detected, their localization within the tissue is lost. Here the strategy offers many advantages including reduced sample preparation time, extraction from restricted regions that minimize dilution of less abundant proteins, and high throughput capabilities. Moreover, digestion of proteins is well controlled and additional steps including reduction and alkylation of cysteine residues to improve protein detection can be done, both of which remain difficult to implement directly on a tissue section in the case of a MALDI-MSI experiment.

This procedure allows a high throughput analysis of samples but with regards to the hydrophobicity of the solution of trypsin, digestion of hydrophobic proteins such as membrane proteins is limited. Extraction of proteins with detergents such as Triton X, CTAB, SDS, BAC, CHAPS and OGP was thus performed, using microdissected portions coming from the symmetric regions of three adjacent tissue sections. These detergents have specific features including differences in charge (zwitterionic, neutral, negative or positive). The samples were incubated with each detergent solution and the resulting extracts were submitted to SDS-PAGE separation and stained with Coomassie blue, as shown in Figure-S3. According to the limit of detection of this staining procedure (50-100ng of proteins) (Miller, Crawford et al. 2006), it can be seen that whatever the detergent used, many proteins are detected up to 250 kDa with a high sensitivity.

Extraction of intact proteins from the micro-dissected samples can be advantageous to increase the number of identified proteins since this approach will provide access to more hydrophobic proteins and also enables to add a dimension of fractionation prior to the shot-gun analysis. Indeed proteins are firstl separated by 1D SDS PAGE, then the gels are divided into several bands and subjected to in-gel digestion followed by LC MS & MS/MS analysis.

PAM from entire tissue sections

Since proteins can be identified from specific regions of interest using the PAM approach, this strategy can be extended to an entire section. Then, the extracted proteins from each piece of tissue can be digested and analyzed by nanoLC-MS & MS/MS. After identification, proteins can be then quantified by label free quantification such as spectral counting. It is then possible to assimilate each PAM piece as an image pixel and use these quantification data for image reconstruction in order to provide protein imaging based on quantification data.. This approach was thus applied to the individual analysis of 60 PAM pieces (Figure 4) of approximately identical size (~1.5 mm^2) from an entire rat brain tissue section cut in half along the axis of symmetry. For further use in image reconstruction, the positions where the PAM pieces have been obtained were noted. The tryptic peptides were then separated on a 2-h gradient chromatographic separation and analyzed by MS & MS/MS on an LTQ Orbitrap XL instrument. The generated spectra were then subjected to Sequest querying followed by verification with Scaffold. Then, label-free quantification by spectral counting, available on Scaffold, was used to normalize the spectra and provide quantitative information per protein per PAM piece. The images were then constructed by plotting in Excel the normalized spectral count of each protein as observed for each PAM sample with the corresponding position on the tissue. Thus, each grid represents one PAM piece, and its color depends on the value of the spectral count. Figure 4 shows representative images for some of the proteins identified using this approach. Despite the poor spatial resolution of these images compared to conventional MALDI MSI, it can still be gleaned that there is preferential distribution of the proteins across the analyzed sections with certain proteins showing highly localized distribution. In the strategy we were able to identify 1066 unique proteins with a 1% FDR, 2 peptides minimum, all of these being available for quantification-based imaging. It appears that the PAM strategy applied to entire tissue section allows the detection and localization of high abundant proteins as well as less abundant ones. Validation of the spatial mapping of protein abundances performed in this study was confirmed by the presence of oxytocin and vasopressin peptides in media preoptic area, in the

paraventricular nucleus of the hypothalamus and the anterior hypothalamus which are known to contain these peptides (Theodosis 2002). In this context, we focus our attention on neuropeptide precursors and detect in thalamus or hypothalamus area chromogranins (chromogranin B, chromogranin C), granulins, proenkephalins as well as a specific enzyme involved in their catabolism, the angiotensin-converting enzyme in the thalamus area. For molecules involved in immune response (IL16, macrophage migration factor) or neurotrophic molecules (glia maturation factor beta, pleiotrophin, basigin), their localization is specific to the protein concerned. IL16 is localized in the supraoptic nucleus. Pleiotrophin is found in the optical area and also in corpus callosum. The tumor suppressor factors CD82 or leucine-rich glioma inactivated factor (LGI1) also have specific localization which are in line with previous data (Wu, Orozco et al. 2009). LGI1 is known to be produced in the prefrontal cortex and amygdala. All these data reflect that the PAM procedure for protein quantification is in line with immunocytochemical data but bring now quantification of these molecules into the tissue. More protein distributions are found in Figure-S4. All proteins identified are listed in a table presented in Supplementary Information-2.

Spatial mapping of protein abundances by voxelation was previously demonstrated by Petyuk and coworkers (Liu and Smith 2003; Petyuk, Qian et al. 2007). In their strategy, they have adapted a bladder-based voxelation device for gene expression to study protein expression and compared gene and protein expression. Using this mechanical voxelation device they were able to produce 1 mm^3 (1 mm x 1mm x 1mm) voxels. For quantification of the identified proteins they used ^{18}O isotope labeling of tryptic digestion products. The ^{18}O labeling was performed from the consecutive tissue slice to the one used for voxelation, and spiked in each voxelation sample for use as reference for quantification. By PAM of tissue sections we have developed a strategy of higher spatial resolution using a much easier and straightforward approach. Indeed, we are able to produce 0.0225 mm^3 (1.5 mm x 1.5 mm x 10 µm) voxels which is 44.4% fold less. Moreover, using micro-dissection of thin tissue sections we can possibly even manually reach 1mm x 1mm x 10 µm voxels (i.e. 0.01 mm^3) which is 100 fold less in volume and 100 fold higher in resolution. The present approach also offers the advantage of being compatible with all types of samples including FFPE samples and thereof easily usable on hospital library samples.

The PAM strategy presented here provides some new interesting features for MSI of proteins as it offers the possibility to identify and localize directly proteins on tissue without the need of *in*

situ MS/MS which has revealed certain limitations in the number and dynamic range of identified proteins. Neither does it require back correlation which is based on the construction of in-house tryptic peptide databanks (Lemaire, Desmons et al. 2007; Schober, Schramm et al. 2011), nor an adjacent section to be *in situ* digested for comparison with image data, which is, for the moment, difficult to control with automatic devices which might lead to irreproducible digestion and cannot guarantee a similar cell content. The back correlation strategy can also suffer from ion suppression effect indirectly because the localization of the proteins is based on tryptic peptide detection by MALDI-MSI. Given the complex nature of the tissue, the ionization efficiency can be greatly decreased, thus leading to less number of peptide signals that can be correlated to protein identification generated by LC-MS & MS/MS. Indeed, it was previously shown that around 300-400 tryptic peptides could be detected using a MALDI-LTQ-orbitrap instrument on one pixel of a tissue section subjected to an *in situ* enzymatic digestion (Quanico, Franck et al. 2013). However, after micro extraction of these tryptic peptides by liquid micro-junction, 25,000 MS/MS were performed when the extract was analyzed by nanoLC-HR MS & MS/MS leading to the identification of approximately 3800 unique tryptic peptides. Here the proteins are identified off tissue, minimizing the ion suppression effect. Their distribution is based on spectral counting and as such, several identified peptides were used to map them out. This is in stark contrast to MS imaging, where the protein mapping is usually based on single peptides (Seeley and Caprioli 2008). Additionally, here, the proteins are not only identified based on more than one single peptide but are also localized using the same criteria. This means that if a protein is not detected on a PAM region, it is possible to assume that it is not due to the ion suppression effect coming from the tissue but rather because the protein is not present or with intensity beyond the limit of detection of the instrument.

CONCLUSION

The technique presented here provides a cheap, convenient and rapid microdissection method that can be used to probe protein expression changes in discrete, localized regions of the tissue defined by molecular histology. Although first developed on a rat brain tissue section only, the strategy is expected to find broader applications, especially when dealing with multiple samples as in the field of biomarker hunting. It was also introduced here that the PAM procedure can be extended to a whole section in order to do indirect molecular imaging based on the identification and subsequent quantification by label-free method of proteins out of the tissue. This new

strategy offers many advantages including the identification and localization of more than a thousand proteins in contrast to MS-based imaging strategies where ion suppression effect greatly limits the back correlation. The strength of the approach is its capability to identify and localize proteins simultaneously based on the identification of more than 2 unique peptides. For the moment, the distribution of the proteins was done using spectral counting as label free quantification. In order to get a more precise distribution of the intensity of a protein, an alternative quantitative approach based on isotopic tags (Gygi, Rist et al. 1999) can be implemented prior to nanoLC-MS & MS/MS. This will allow a more accurate relative quantification of the proteins region by region and thus of the entire section.

The new concept of protein imaging constructed based on protein quantification is expected to give a better description of protein distribution and abundance because more proteins can be identified with high confidence and imaged from a single tissue section. Such a different way of performing MSI can bring important improvement for biological and clinical applications in the understanding of regulation pathways in physiological and physiopathological contexts.

ACKNOWLEDGEMENTS

Supported by grants from University Lille 1 (BQR and PhD grant for J. Quanico), University of Sherbrooke (PhD grant for J. Quanico), French Region “Nord Pas de Calais” by ARCIR FEDER funding on the FT MS instrument (Prof. I. Fournier), Ministère de l’Enseignement Supérieur et de la Recherche *via* Institut Universitaire de France (Prof. I. Fournier) and a funding from "Comité du Nord de la Ligue Nationale contre le Cancer" (M. Wisztorski). This work was under collaboration with Thermo Fisher Scientific.

FIGURE CAPTIONS

Figure 1: Microdissection strategy used for identification of proteins from discrete location of a tissue section

Figure 2: (A) Cluster image obtained the MALDI MSI of proteins on a rat brain section at 300µm and (B) the corresponding dendrogram. (C) Cluster image obtained the MALDI MSI of lipids on the consecutive section at 300µm and (D) the corresponding dendrogram.

Figure 3: (A) Optical image of a rat brain section after PAM, showing the 3 ROIs. (B) The positions based the rat brain atlas. (C) Venn diagram of the protein IDs from the 3 ROIs after Sequest interrogation Scaffold validation.

Figure 4: Optical image of the rat brain tissue section (right panel), half of which was subjected to PAM (left panel) and reconstructed distribution of representative proteins based on the spectral count provided by nanoLC-MS/MS analysis.

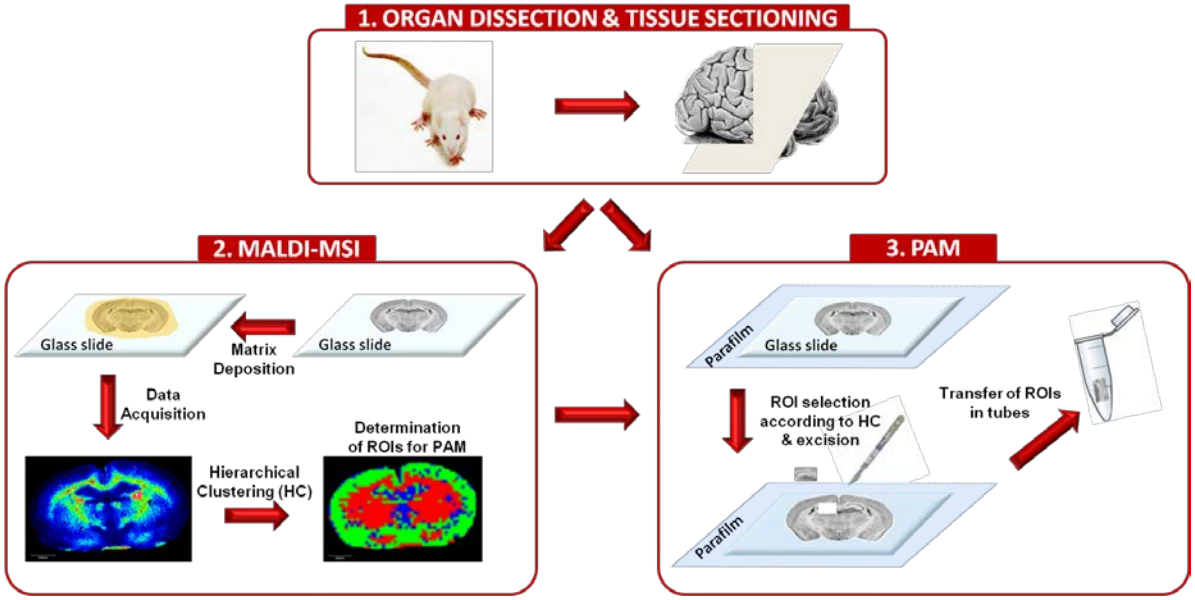


FIGURE 1

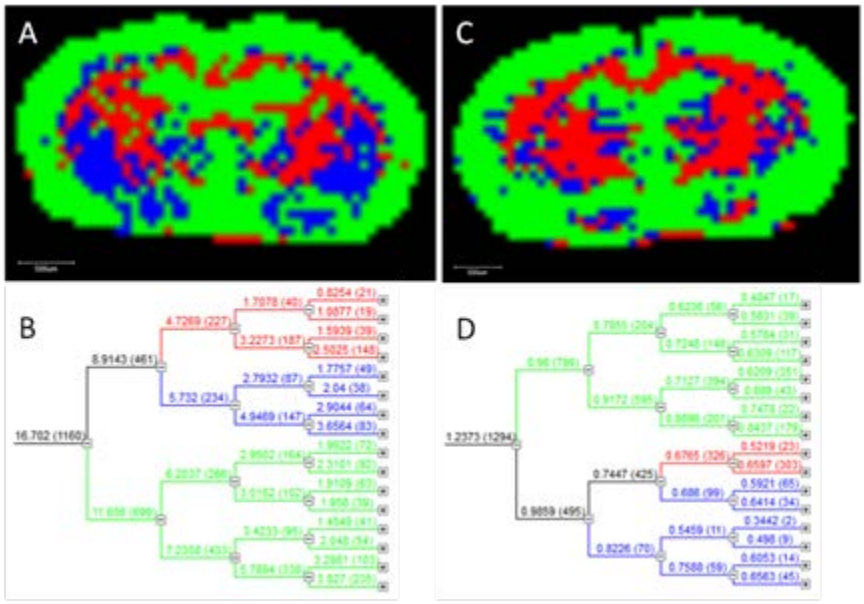


FIGURE 2

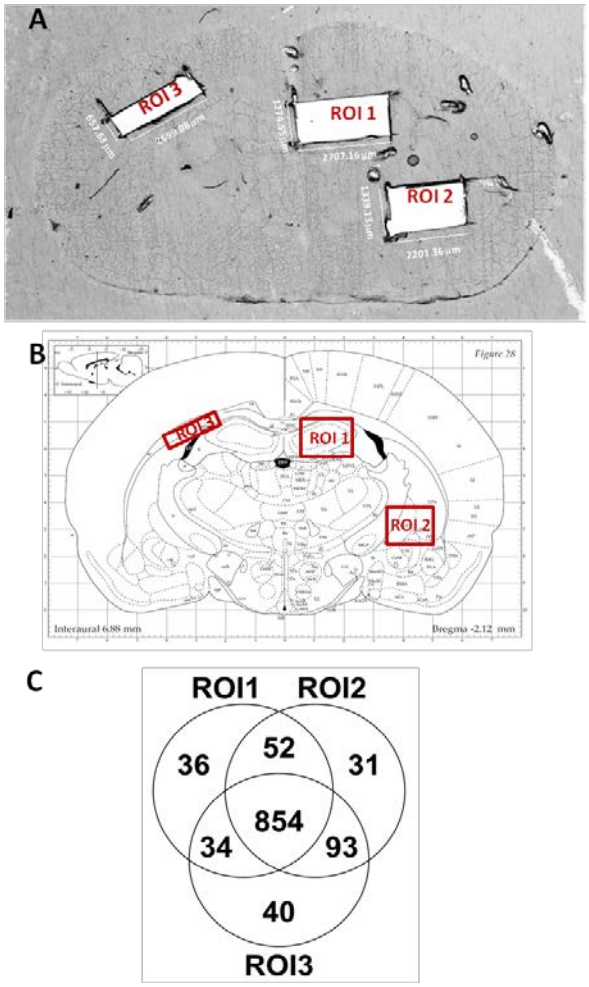


FIGURE 3

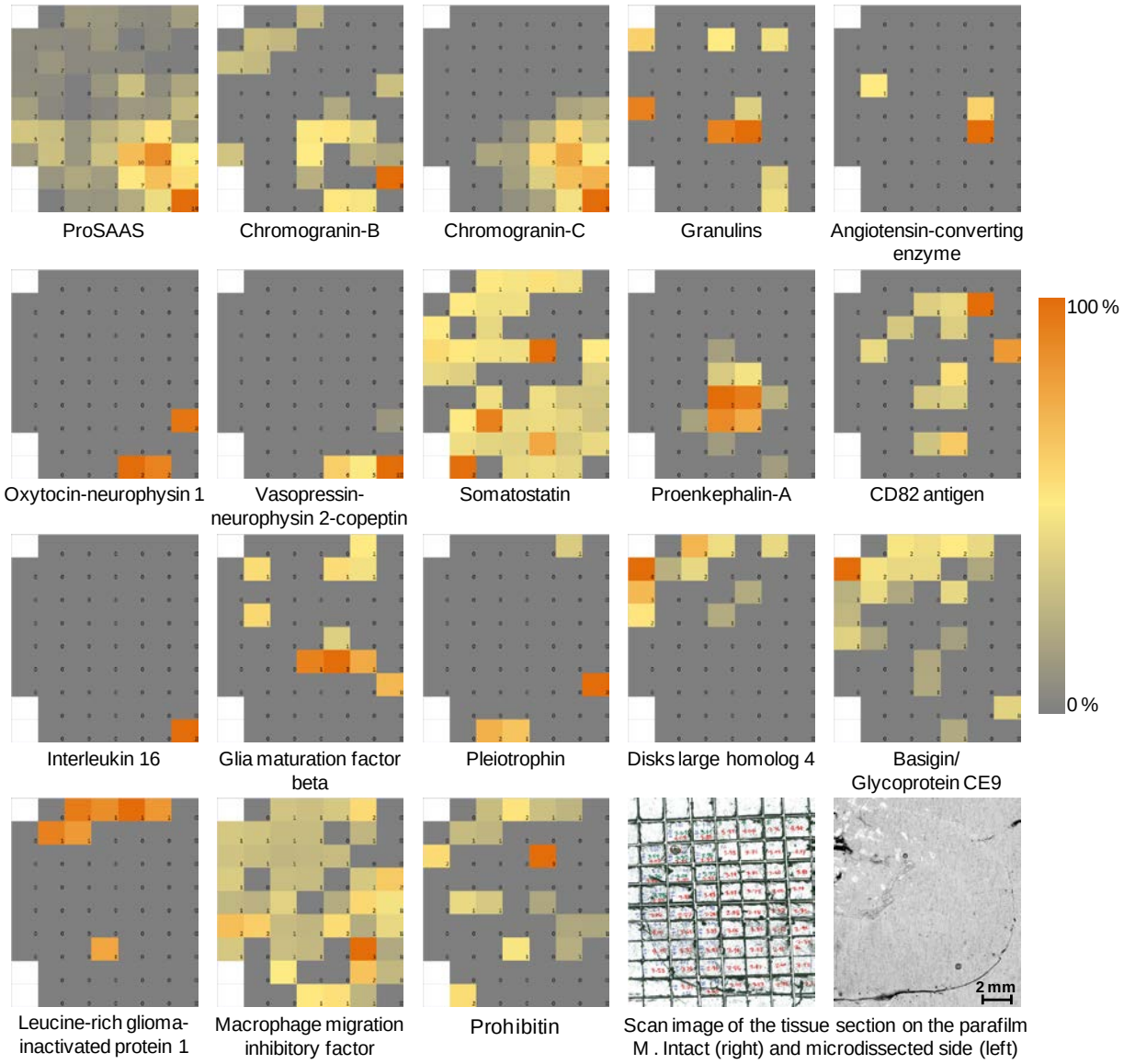
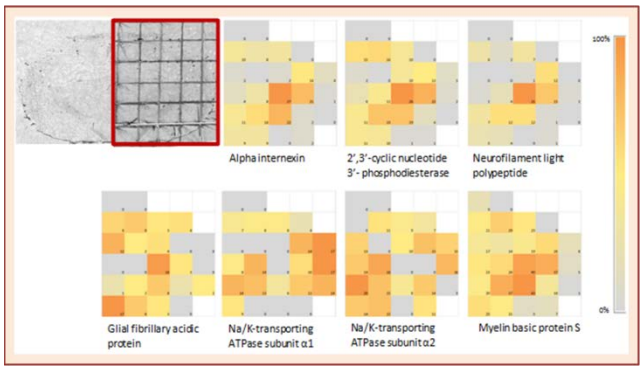


FIGURE 4

FOR TOC ONLY



SUPPLEMENTARY INFORMATION

TABLE OF CONTENTS

MATERIAL & METHODS SUPPLEMENTS

Chemicals

HPLC grade methanol (MeOH), acetonitrile (ACN), water and AR grade trifluoroacetic acid (TFA) were obtained from Biosolve B. V. (Valkensvaard, Netherlands). Ammonium bicarbonate (NH_4HCO_3), ethanol (EtOH), α-cyano-4-hydroxycinnamic acid (HCCA), 2,5-dihydrobenzoic

acid (2,5-DHB), DL-dithiothreitol (DTT), iodoacetamide (IAA), cetrimonium bromide (CTAB), benzalkonium chloride (BAC), octyl β-D-glucopyranoside (OGP) and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) were obtained from Sigma (Saint-Quentin Fallavier, France). Tetramethylethylenediamine (TEMED), ammonium persulfate (APS), sodium dodecyl sulfate (SDS) and Triton-X 100 were purchased from Bio-Rad (Marnes La Cocquette, France). Sequencing grade, modified porcine trypsin was obtained from Promega (Charbonnieres, France).

Rat Brain Dissection

Male Wistar rats were used and were treated in accordance to the guidelines of the Institutional Animal Care and Use Committee (IACUC) of University of Lille 1. Extraction of the brain was performed as described previously (Fournier, Day et al. 2003). Briefly, the rats were anesthetized with carbon dioxide and the collected brains snap-frozen in isopentane (–50°C). The frozen tissues were then stored at –80°C until use.

Mass Spectrometry Imaging (MSI)

10-µm sections were cut using a cryostat (Leica Microsystems, Nanterre, France) and subjected to lipid MSI as described previously (Meriaux, Franck et al. 2010). These were finger-thawed on ITO-coated slides (BrukerDaltonics, Bremem, Germany) and dried for 15 min in a dessicator. 2,5-DHB (20 mg/mL, dissolved in 7:3 EtOH/0.1% TFA, v/v) was then sprayed on the tissue surface as micro droplets using an electrospray nebulizer attached to a 500-µL syringe and the flow rate set to 2.5 µL/min. The nebulizer was moved uniformly throughout the tissue surface for 10 min to ensure uniform matrix deposition. MS spectral images were acquired using an Ultraflex II mass spectrometer equipped with a Smart beam (Nd:YAG, 355nm) laser having a repetition rate up to 200 Hz (BrukerDaltonics, Bremen, Germany). The images were acquired in positive reflector mode at 100 or 300-µm resolution at a mass range of m/z 200-2000, and the obtained spectra were averaged from 500 laser shots per pixel. Images obtained were analyzed using Fleximaging 2.1 (BrukerDaltonics, Bremen, Germany). For the imaging of intact rat brain proteins, consecutive 10-µm rat brain sections were subjected to a series of washing steps using 70% and 95% EtOH/H_2O then with $CHCl_3$ to deplete abundant proteins, endogenous peptides and lipids. Matrix deposition was performed using ImagePrep (Bruker Daltonics, Bremen, Germany), as described previously using sinapinic acid/aniline liquid ionic matrix (Franck,

Arafah et al. 2009). Uniform deposition of the matrix on the tissue was ensured by monitoring the tension of each deposition step online with a computer attached to the ImagePrep instrument. Using UltraFlex II, images were acquired in positive mode at 100 or 300-μm resolution at a mass range of m/z 300-20000, with each spectra being an average of 500 shots per pixel.

Hierarchical clustering

For statistical analysis, the data set obtained from the rat brain samples by FlexImaging was loaded in ClinProTools 2.2 software (BrukerDaltonics, Bremen, Germany) to conduct hierarchical clustering. The spectra were initially prepared by setting the resolution to 800 and the m/z range from 200 to 900 (for lipids) and 3000 to 17000 (for proteins). Savitsky-Golay smoothing was employed using 2.0 m/z spectral width and 5 cycles. The signal-to-noise (S/N) ratio of the total average spectrum was set 3.0. Unsupervised clustering was selected with Correlation as distance method and Average as linkage method. Results of the hierarchical clustering were exported to FlexImaging 2.1 software to reconstruct the areas with different profiles.

NanoLC MS/MS Parameters for Quantification-based MS Imaging Approach

The samples were separated by online reversed-phase chromatography using a Thermo Scientific Proxeon Easy-nLC system equipped with a Proxeon trap column (100 μm ID x 2 cm, Thermo Scientific) and C18 packed tip column (100 μm ID x 10 cm, NikkyoTechnos Co. Ltd). Elution was carried out using an increasing gradient of AcN (5% to 30% over 230 minutes) and a flow rate of 300 nL/min. A voltage of 1.6 kV was applied to the needle of the nanospray source. The chromatography system was coupled to a Thermo Scientific LTQ Orbitrap XL mass spectrometer programmed to acquire in data dependent mode. The survey scans were acquired in the Orbitrap mass analyzer operated at 60 000 (FWHM) resolving power. A mass range of 300 to 1600 m/z and a target of 1E6 ions were used for the survey scans. Precursor ions observed with an intensity over 500 counts were selected “on the fly” for ion trap collision-induced dissociation (CID) fragmentation with an isolation window of 2 a.m.u. and a normalized collision energy of 35%. A target of 5000 ions and a maximum injection time of 200 ms were used for MS^2 spectra. The method was set to analyze the top 20 most intense ions from the survey scan and a dynamic exclusion was enabled for 30s. SEQUEST was used for database querying as was done previously. For peptide and protein validation Scaffold version 4.0.5 was used, taking advantage

of the new local FDR-based scoring algorithm and the clustering feature, the latter allowing the grouping of proteins sharing similar peptide evidence (Searle 2010). As previously, default parameters were used.

FIGURE SUPPLEMENTS

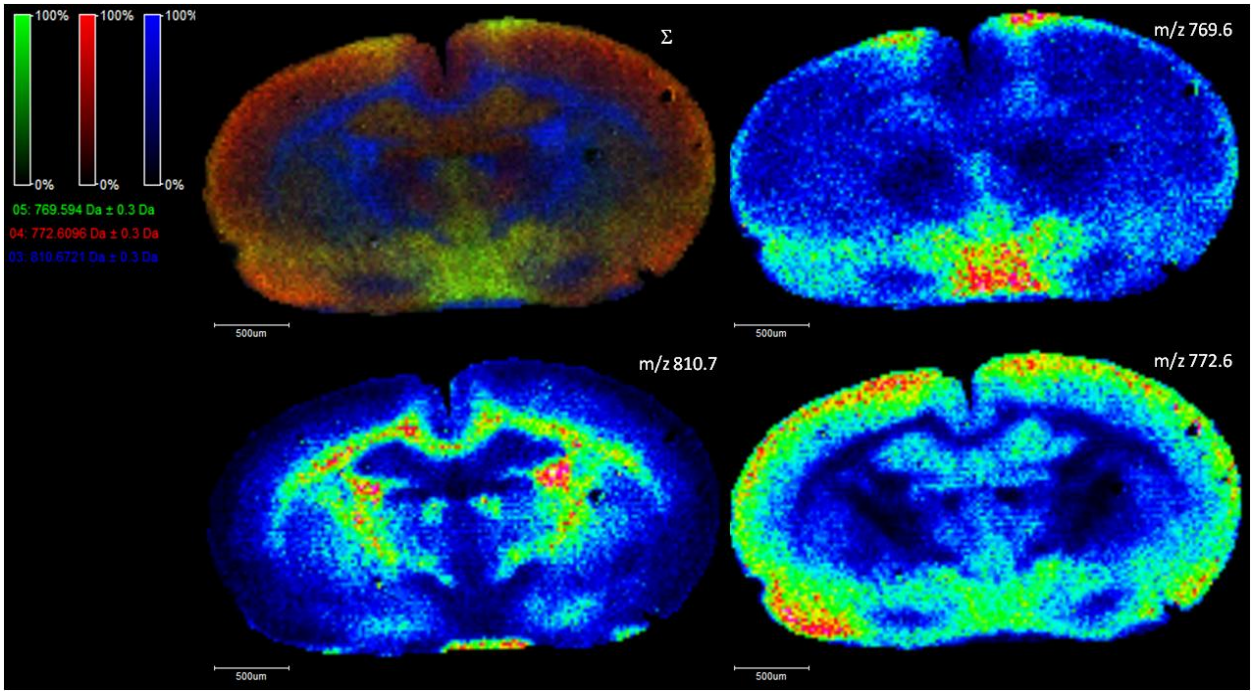


Figure S-1: Lipid molecular images of a rat brain tissue section, ∑ being the composite image of m/z 769.6, 772.6 and 810.7.

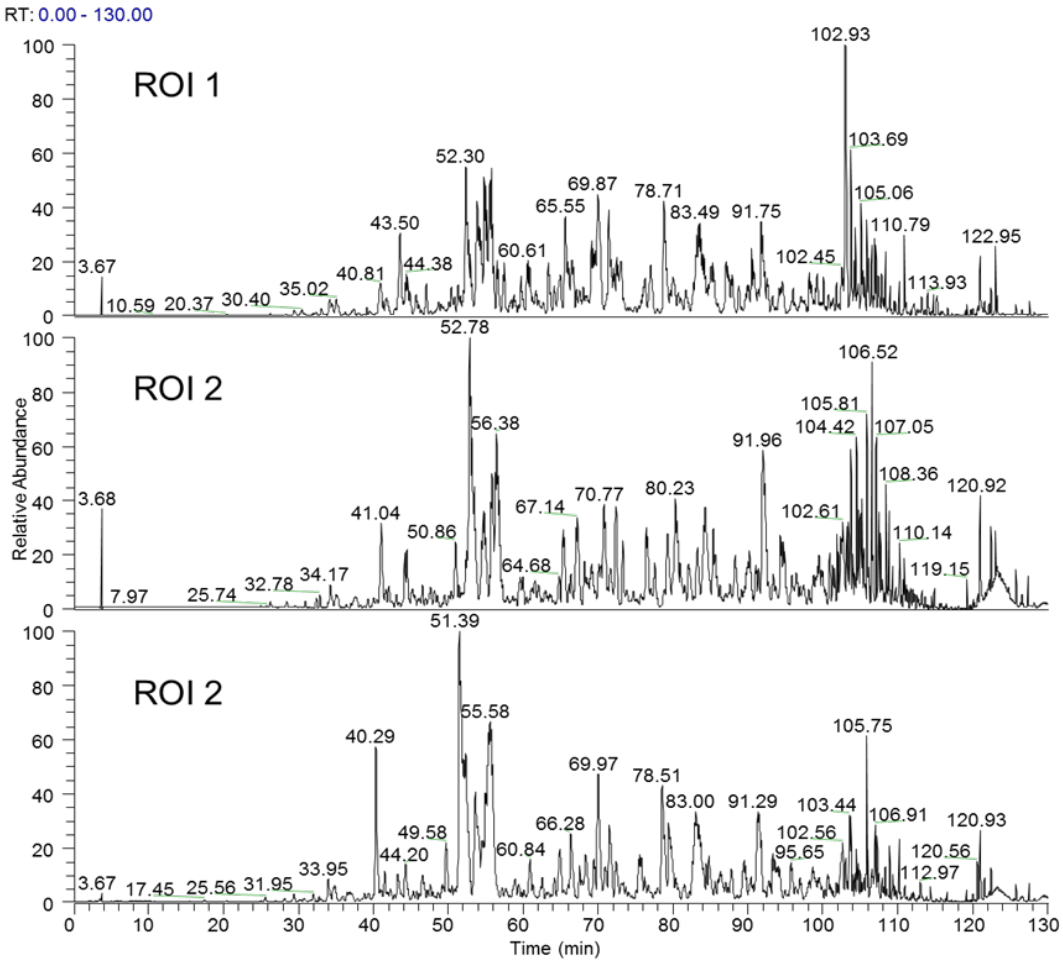


Figure S-2: Base peak chromatogram for each of the 3 different ROI's microdissected using the PAMM method and analyzed by nanoLC-MS & MS/MS

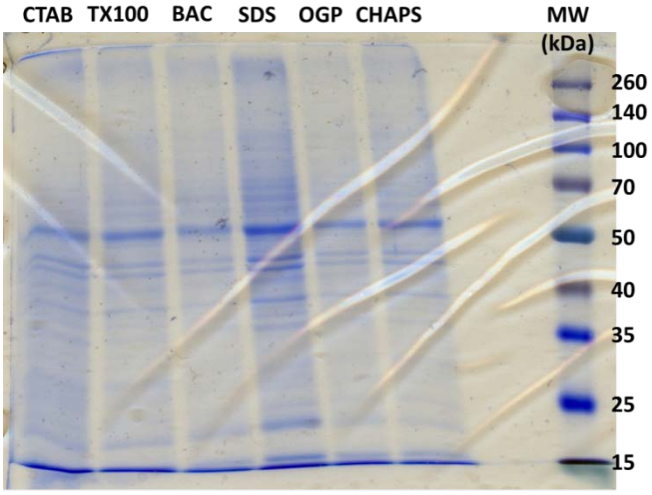


Figure S-3: Scan of the SDS-PAGE Gel of the intact protein extracts using PAMM approach and employing various detergents.

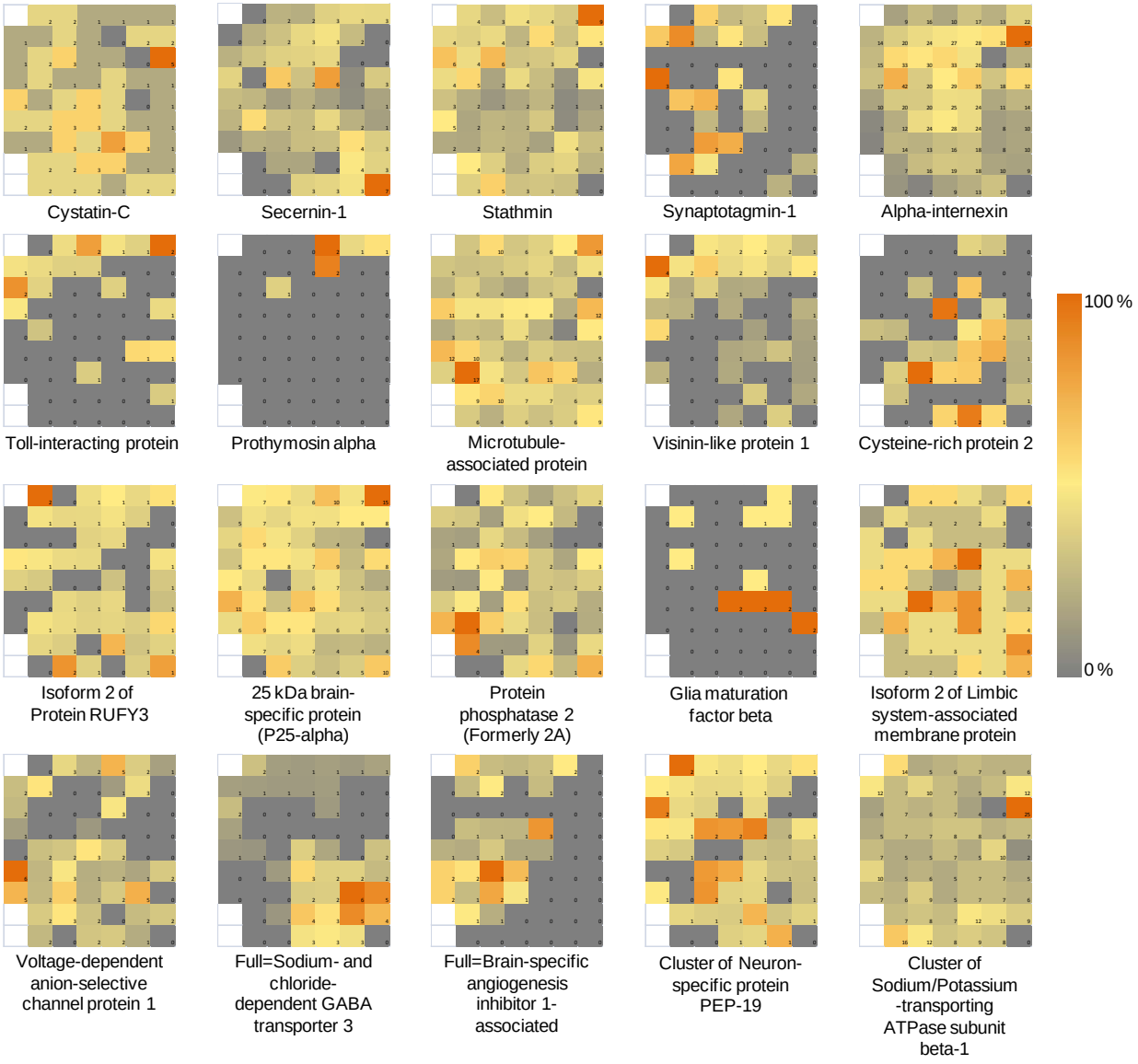


Figure S-4: Example of identified proteins distributions reconstructed based on spectral counting label free quantification method.

Additional supplementary data (Supplementary Information-2) can be found in the online version of the article at: http://pubs.acs.org/doi/suppl/10.1021/ac4009397

Chapter 4

Parafilm-Assisted Microdissection: Application to Prostate Cancer

The early diagnosis of prostate cancer remains a challenge because of the lack of reliable biomarkers that can effectively describe this disease independent of sample heterogeneity due to genetic variation and environmental exposure. Current methods for prostate cancer detection comprised of digital rectal exam (DRE), prostate-specific antigen (PSA) measurement, and investigation by transrectal ultrasound (TRUS, (Stavridis, Saidi et al. 2010)) are being used for routine screening prior to undergoing more invasive follow-up procedures such as a biopsy. However, the use of these screening methods has been controversial because overdiagnosis of insignificant tumors led to unnecessary treatment and known harms associated with the diagnostic tests (Dijkstra, Mulders et al. 2014; Saman, Lemieux et al. 2014). Also, DRE screening often detects prostate cancer only at advanced stages. Although PSA is already a US FDA-approved protein biomarker for prostate cancer, nonetheless, increasing evidence suggests that it is not cancer-specific as elevated PSA levels are also observed during prostatitis, prostatic intraepithelial neoplasia (PIN), benign prostatic hypertrophy, or even recent ejaculation (Brawer 2005; McVary 2007; Xia, Cui et al. 2012; Rajaei, Momeni et al. 2013; Stephan, Ralla et al. 2014). Furthermore, although elevated PSA levels correlate with prostate cancer incidence, PSA alone or in tandem with other screening methods cannot be used to provide reliable prognostic indicators on which decisions to undergo treatment despite of its deleterious effects heavily rely on (Wolf, Wender et al. 2010). Clearly, there is a need for the establishment of biomarkers that can discern prostate cancer in its aggressive stages, in order to discriminate patients requiring immediate therapy from those whose likelihood of benefit from treatment is low.

To this end, mass spectrometry (MS) is increasingly becoming a method of choice for protein biomarker discovery. Specifically, MS-based shotgun proteomics has been demonstrated to allow the nearly complete characterization of large proteomes by taking advantage of robust developments in rapid data acquisition, high resolution MS, increased sensitivity and rapid duty cycle. To date, for example, it has been demonstrated on the yeast proteome (Mann, Kulak et al. 2013), and it is expected that full characterization of the human proteome will soon follow (Wilhelm, Schlegl et al. 2014). MS-based shotgun proteomics is a powerful tool to detect differences in protein expression levels at low-attomole detection limits particularly when using Q Exactive systems. Its major limitation for screening purposes though is that it is low-

throughput, because of extensive purification procedures during sample clean-up, as well as long chromatographic gradients required for efficient separation, and rigorous interpretation of large datasets that are generated. Direct MS-based methods such as liquid extraction surface analysis (LESA, (Quanico, Franck et al. 2013)) and desorption electrospray ionization (DESI, (Liu and Ouyang)) to name a few, as well as matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS, (Anderson and Kodukula 2014)) in both profiling and imaging modes, have been developed to minimize sample preparation and allow mapping of protein distributions but these methods are restricted at best to tissue profiling studies rather than discovery of candidate biomarkers because of their limited sensitivity and/or protein identification capabilities considering that they were not aimed at complete proteome characterization.

To be able to analyze more samples at a much faster rate using the MS-based shotgun proteomics approach, we have recently developed the parafilm-assisted microdissection technique (PAM) in order to sample regions of interest (ROIs) directly from tissue sections, the ROIs having been identified from the MALDI MSI analysis of an adjacent section (Franck, Quanico et al. 2013). With this method, we were able to analyze multiple samples in a sequential manner, allowing us to generate heat maps of protein distributions across an entire tissue section by assigning each PAM piece as a voxel of that image. The PAM approach provided a means to obtain millimeter-sized portions that can be easily subjected to protein extraction and digestion procedures without the use of additional rigorous sample purification steps. This time we exploit the advantage of using the PAM approach in profiling mode to analyze prostate tumor samples, with the objective of finding candidate protein biomarkers that can be further validated on a larger cohort of samples.

Materials and methods

Tissue Specimens

Fresh frozen tumors embedded on OCT polymer were retrieved from archived specimens of a previous study (D'Anjou, Routhier et al. 2011). These specimens have been obtained with informed consent from patients in the Centre Hospitalier of University of Sherbooke and the tumor and benign regions have been identified by an experienced pathologist. These were confirmed, upon receipt of the samples, by dissecting 10-µm sections using a cryomicrotome (Leica Microsystems, Nantere, France), subjecting them to HES staining, and verifying the actual positions of the regions.

MALDI-MSI Analysis

The consecutive 10-µm section adjacent to the one subjected to HES staining was used for MALDI MSI analysis of lipids. The section was mounted on an ITO-coated conductive slide and dried under vacuum for 5 min. 2,5-Dihydroxybenzoic acid (DHB) was used as matrix and prepared at a concentration of 20 µg/mL in 70 :30 ethanol/0.1 % TFA. The matrix solution was sprayed uniformly throughout the entire tissue section using a discarded electrospray nebulizer. The nebulizer is attached to a 500-µL syringe that is driven by a pump at a flow rate of 2.5 µL/min. The fine mist generated by the spray head ensures the deposition of micrometer-sized matrix droplets onto the tissue surface. Spraying was performed for 10 min and the formation of matrix crystals was confirmed by examination under a light microscope.

The prepared sections were examined using an UltraFlex II instrument equipped with a Smart beam (Nd:YAG, 355nm) laser having a repetition rate up to 200 Hz (BrukerDaltonics, Bremen, Germany). The images were acquired in positive reflector mode at 100-µm resolution at a mass range of m/z 200-1500, and the obtained spectra were averaged from 500 laser shots per pixel. Images obtained were analyzed using Fleximaging 2.1 (BrukerDaltonics, Bremen, Germany) to map the entre section. Regions of interest (ROIs) were defined on the benign and tumor regions basing from the HES optical images, and the spectra from these regions were exported to ClinPro Tools version 3.0 (Bruker Daltonics, Bremen, Germany). A limit of less than 1000 spectra was exported per region. These representative spectra were used for the hierarchical clustering analysis using the following parameters. The spectra were initially prepared by setting the resolution to 800 and the m/z range from 200 to 900. Savitsky-Golay smoothing was employed using 2.0 m/z spectral width and 5 cycles, and the signal-to-noise (S/N) ratio of the total average spectrum was set 3.0. Unsupervised clustering was selected with Correlation and Average as distance and linkage methods, respectively. The results were then exported back to FlexImaging and the spectra of the branches of the dendrogram overlapped with the optical image of the tissue section. This method verifies the visually-defined benign and tumor regions as well as made them more precise.

Protein Extraction and nanoLC-MS

The consecutive section adjacent to the one subjected to MALDI MSI was used for the proteomics experiments. These sections were mounted on parafilm-covered glass slides, dried under vacuum, and subjected to the PAM procedure previously described (Franck, Quanico et al.

2013). The approximately 1.5 mm^2 pieces were incubated in 25 µL of 50 mM dithiothreitol (DTT) for 15 min at 55°C and then 25 µL of 150 mM iodoacetamide (IAA) for 15 min in the dark. The entire mixture was digested overnight using 20 µL of 20 µg/mL sequencing grade modified trypsin (Promega, Lyon, France) at 37°C. The extracts were then dried *in vacuo* and reconstituted with 10 µL 0.1% TFA for desalting using C18 Ziptips (Millepore, Saint-Quentin-en-Yvelines, France). These were again dried and resuspened in 5% ACN/0.1% FA prior to injecting to the nanoLC-MS instrument.

Separation of the sample components was done using an online reversed-phase chromatographic system (Thermo Scientific Proxeon Easy-nLC II) equipped with a Proxeon trap column (100 µm ID x 2 cm, Thermo Scientific) and C18 packed tip column (100 µm ID x 15 cm, NikkyoTechnos Co. Ltd). Elution was carried out using an increasing gradient of AcN (5% to 40% over 110 minutes) and a flow rate of 300 nL/min. A voltage of 1.7 kV was applied via the liquid junction of the nanospray source. This was interfaced to a Thermo Scientific Orbitrap Elite mass spectrometer programmed to acquire in data-dependent mode. The survey scans were acquired in the Orbitrap mass analyzer operated at 120,000 (FWHM) resolving power. A mass range of 400 to 2000 m/z and a target of 1E6 ions were used for the survey scans. Precursor ions observed with an intensity over 500 counts were selected “on the fly” for ion trap collision-induced dissociation (CID) fragmentation with an isolation window of 2 u and a normalized collision energy of 35%. A target of 5000 ions and a maximum injection time of 200 ms were used for MS2 spectra. The method was set to analyze the top 20 most intense ions from the survey scan and a dynamic exclusion was enabled for 20 s.

Protein Identification

The MS/MS spectra were analyzed with Sequest (Thermo Fisher Scientific, San Jose, CA, USA) version 1.4.1.14. Spectra were searched against the Uniprot database (version November 2013) filtered with *Homo sapiens* (122786 sequences) taxonomy using trypsin as digestion enzyme (one missed cleavage). Sequest was searched with a fragment ion tolerance of 0.100 Da and a parent ion tolerance of 10.0 ppm. Carbamidomethylation of cysteine and oxidation of methionine were set as fixed and variable modifications, respectively.

Scaffold (version 4.0.5, Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they

could be established at a given probability to achieve an FDR less than 5.0 % by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at a given probability to achieve an FDR less than 2.0 % and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Keller et al. 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony .

The protein IDs for each sample were loaded into Perspectives version 1.0.3 (Proteome Software Inc., Portland, Oregon) using the same FDR thresholds used in Scaffold and the decoys were removed from the final list. Clustering was performed to group together proteins with any shared evidence and for which the peptide IDs cannot be discerned. Doing so produced 1221 and 1227 protein groups for data generated by the Orbitrap Elite and Q-Exactive instruments, respectively. The total weighted spectral counts of the protein groups were then obtained and subjected to Fisher's Exact Test. Only proteins with $p<0.05$ were considered to have significant differential expression between the benign and tumor group datasets. The protein list was then exported in Excel and the $\log_2$ fold change was computed from the weighted spectral counts (tumor/benign). The protein list was further filtered by removing IDs with $\log_2$ fold change = 0 and selecting IDs which were detected in at least 60% of the benign or tumor samples (for samples with replicates, the spectral count was averaged). For the dataset obtained using the Orbitrap Elite instrument, this corresponds to detection in at least 3 of 4 benign or tumor samples, while for the dataset obtained using the Q-Exactive, this corresponds to at least 6 out of 9 benign or tumor samples.

Protein-Protein Interaction (PPI) Network

The differentially expressed protein list identified from the two datasets were combined and used as seed terms for the construction of the PPI network using STRING (Search Tool for the Retrieval of Interacting Genes) version 9.1. String is a database that generates and scores both physical and functional PPIs from various sources based on their neighborhood, gene fusions, co-occurrence, co-expression, experiments and literature mining. The extended network was constructed from the list using a confidence score = 0.7, corresponding to high level of confidence.

Network Construction and Analysis

The PPI network was visualized using Cytoscape network visualization software version 3.1.1 (Shannon, Markiel et al. 2003) and analyzed using the Network Analyzer plug-in (Assenov,

Ramirez et al. 2008). The topology of the network was visualized by mapping the Degree and Betweenness centrality parameters, corresponding to node size and node color, respectively. Degree pertains to the number of edges that links a given node, so that nodes possessing a high Degree value may represent hub genes. Betweenness centrality, on the other hand, reflects the importance of a node based on the number of shortest path lengths that passes this node. Meanwhile, the thickness of the edges was depicted using the edge betweenness values.

The constructed network was then subjected to Markov Clustering (MCL) algorithm using the ClusterMaker plug-in to determine the major clusters. The granularity parameter was set to default value (2.5) and the EdgeBetweenness parameter was used as the source for array data. With this parameter, the edge cut-off was set to 73.042. Major clusters identified in this manner were visualized using Betweenness centrality for node size and expression value for node color. For the expression values, a binary scale was used to represent the $\log_2$ fold change values: 0 for $\log_2$ fold change less than 0 (underexpressed), and 1 for those greater than 0 (overexpressed).

Gene Ontology Analysis

The differentially expressed proteins identified from the two datasets were uploaded to the ClueGO application (version 2.1.2, (Bindea, Mlecnik et al. 2009)) to check for enriched pathways in the benign and tumor regions. Functionally-grouped annotated networks were generated using the following settings. The organism was set to *Homo sapiens.* The gene ontology (GO, (Consortium 2010)) terms were accessed from the following ontologies/pathways : GO_Biological Process, GO_Molecular Function, GO_Cellular Component and GO_Immune System Process (ontology updated 5/23/2014), Kyoto Encyclopedia of Genes and Genomes (KEGG, updated 5/24/2014), Reactome Pathway database (updated 5/24/2014), and WikiPathways (updated 5/24/2014). All types of evidence were used (see http://www.geneontology.org/GO/evidence.shtml). The Use GO Term Fusion option was selected, and only pathways with $pV \leq 0.05$ were accepted. The significance of each term or group was calculated using the right-sided hypergeometric test corrected using Bonferroni step-down correction. The kappa score was set to 0.5 and for network specificity, the GO tree levels were restricted at 6-13 (medium-detailed specificity) and for each cluster a minimum of 5 genes and 7% of the gene population was set. GO terms were grouped with an initial group size of 3 and 50% for group merge. The remaining parameters were set to default.

The gene identifiers of the major clusters identified by MCL clustering during network analysis were also examined using the Cancer Gene Index feature available in the Reactome FI plug-in (Milacic, Haw et al. 2012). The tree of National Cancer Institute (NCI) disease terms was loaded and the term “neoplasm” was selected to highlight gene identifiers that have been annotated for this term and its sub-terms. To map selected over-represented pathways to Reactome events, The Pathway Analysis tool was used.

Results

Identification of ROIs by MALDI MSI

In this study, MALDI MSI was used to define regions of interest (ROI) on both the benign and tumor regions of the prostate tissue sections that can then be subjected to the PAM technique. Although the spatial resoluton used was only 100 µm, it was sufficient to provide images with discernible localized tumors that coincide very well with those identified by HES staining (**Figure 21a** and **b**). But more than just providing optical images, the data generated are actually molecular images. By subjecting the mass spectral information of these molecules to HC analysis and creating images from the resulting clusters (**Figure 21c** and **d**), it is possible to produce more simplified images where ROIs are defined based on statistical inference gained from the similarity of the cellular contents of neighboring cells. This greatly simplifies the selection criteria used to identify ROIs, in contrast to looking at cell morphologies which can at times be subjective or misleading and depend on the experience of the experimenter. In the case of prostate tissue, the latter task easily becomes complicated because of the innate heterogeneity of the tissue.

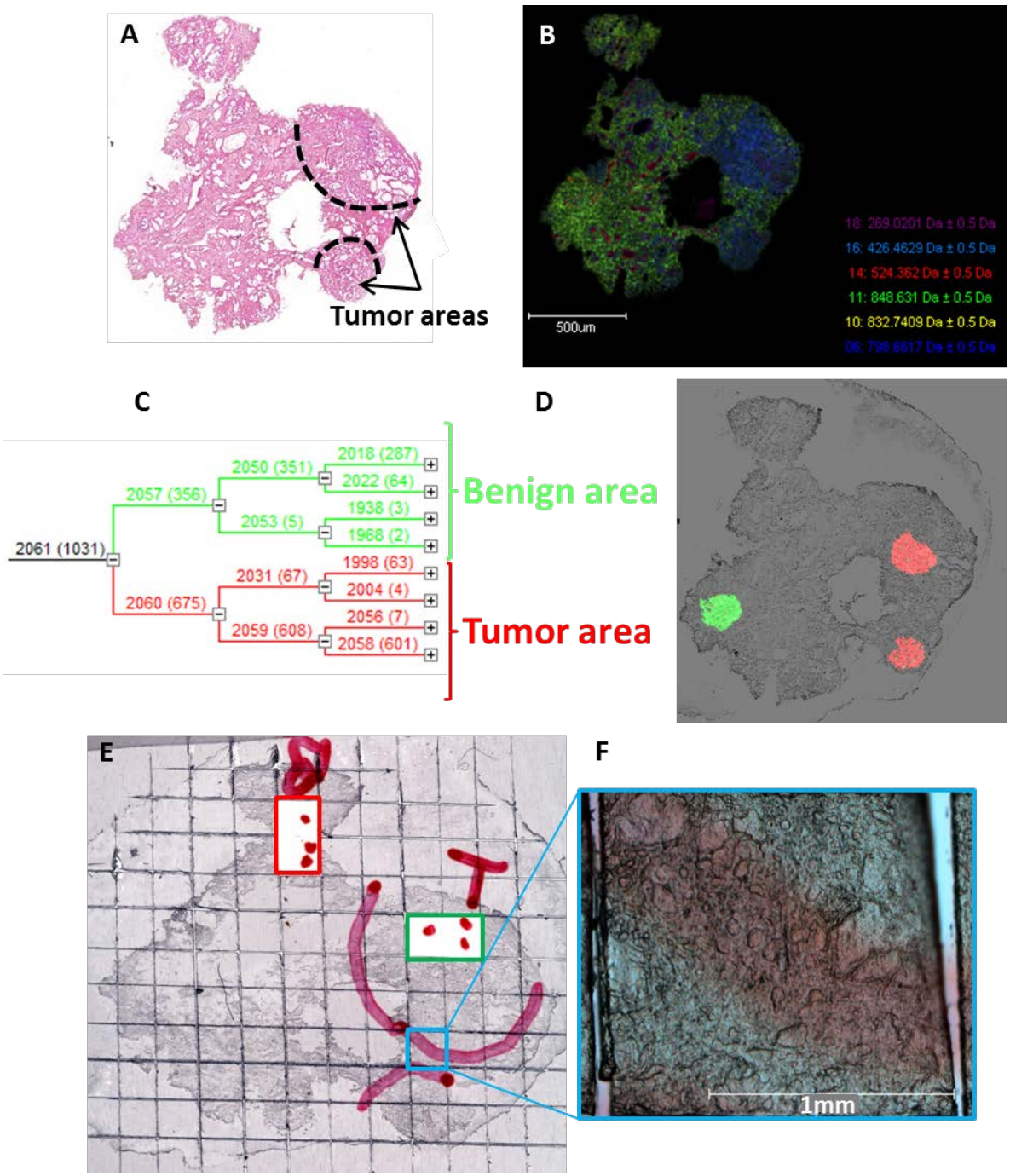


Figure 21. Hierarchical clustering of prostate tumor images obtained by MALDI MSI to define regions of interest for parafilm-assisted microdissection (PAM).

A) Tumor regions of observed in the HES stained tissue section. B) Composite image of selected peaks from the averaged mass spectrum taken from the adjacent section, showing masses specifically defining the tumor regions. C) Dendrogram of the clustering analysis splitting into two branches corresponding to the spectra of the benign and tumor regions. D) Clustered spectra plotted back to the image. E) Sampled regions on the adjacent section, and F) a detail of a single PAM piece.

It has previously been shown that lipid MS images are sufficient alternatives to protein MS images as both show similar component localization but with lipid imaging being done much faster and at better spatial resolution (Franck, Quanico et al. 2013). Lipids, in contrast to proteins, do not need to be incorporated into the matrix crystals for them to be desorbed in MALDI. Thus, it is possible to generate smaller matrix crystals without compromising the efficiency of the ionization/desorption process. Hence, lipid imaging was employed for ROI identification. Sample MS composite and peak intensity images for four prostate sections are shown in **Figure 22**, where it can be seen that for each sample, different peaks were used to localize the tumor regions. There is almost no consensus in the peaks observed in both regions in all samples, and yet HC analysis (**Figure 22**, bottom images) clearly shows that the regions are distinct from each other once the entire spectra of the molecular contents were taken into account. This suggests that despite of the heterogeneous nature of individual tissues (whose molecular expression profiles are dependent not only on the individual genetic factors but also on environmental exposure), their overall molecular make-up nonetheless distinguishes healthy tissues from those under disease states.

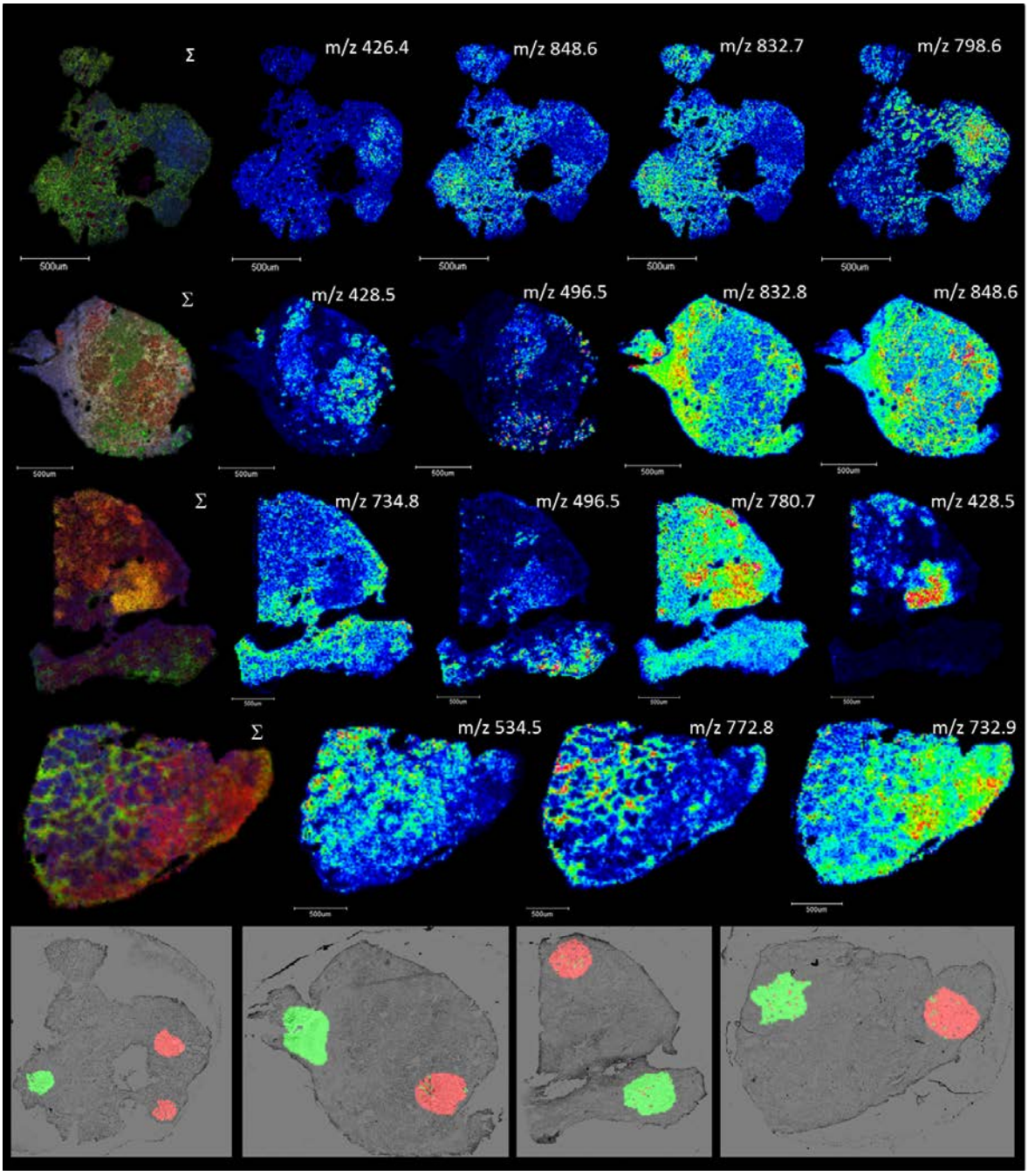


Figure 22. Sample MS images of lipids and resultant clustering images of prostate tissues.

The image on the left of each panel is the composite image of the selected peaks. The bottom panel shows the resultant clustering images, with the green and red clusters corresponding to benign and tumor areas, respectively.

Protein Expression Levels

The detection of differential protein expression is one of the fundamental challenges in shotgun proteomics experiments because subsequent steps such as the identification of underlying

mechanisms and the selection of druggable targets depend on the reliability of this stage. Specifically, the major problem is in the assessment of individual variability within a sample set. To this end, Scaffold and Perspective are complementary software that can provide a means to evaluate protein expression levels in terms of spectral counts, and introduce statistics into the dataset. Scaffold is used to validate protein IDs obtained from initial Sequest interrogations. Its LFDR algorithm allows one to maximize the number of protein IDs at an optimal false discovery rate. Perspective is then used to deal with large protein datasets and to allow the user to categorize the samples further depending on pre-existing information. Emphasis is given on this added categorization feature, as it provides control over the grouping of the datasets needed when implementing a statistical method. For datasets of small number of replicates, the Fisher's exact test provides a means to compute exactly the significance of a deviation from the null hypothesis. It can be used to compute the p value of each protein across sample categories. To further screen the differentially expressed protein candidates, only those with non-zero spectral counts in at least 60% cases in at least one sample group were selected, and only those with fold change ≥ 2 or ≤ 0.5 were used for network analysis.

Initial nanoLC-MS analysis of four prostate tissue samples subjected to the PAM technique and using an Orbitrap Elite instrument yielded 1221 protein identifications, 374 of which have Fisher's Test p-values ≤ 0.05. A total of 122 proteins were identified to have differential expression after filtering using the fold change and detectability across samples criteria, 79 of which were classifed as over-expressed and 43 as under-expressed. In this experiment, single PAM pieces were taken from the benign and tumor regions. In the succeeding experiment using the Q-exactive, two PAM pieces from each region were taken from 8 prostate tissue samples and one from a sample of limited size. This yielded 1251 protein IDs, 485 of which fit the Fisher's test criterion. Applying the filtering parameters yielded 208 differentially expressed proteins of which, 135 were overexpressed and 73 underexpressed. Thus, a total of 273 differentially expressed proteins are reported in this work (**Table 5**).

Table 5. Over- and under-expressed proteins and their fold change values.

Protein Name	Log2 Fold	Protein Name	Log2 Fold	Protein Name	Log2 Fold	Protein Name	Log2 Fold
GOLM1	4	RPSA	1	ECH1	1	COL6A2	-1
PRKDC	4	FASN	1	PTBP1	1	C3	-1
POSTN	4	PGLS	1	ANXA3	1	COL18A1	-1
RPL23	4	LAMP1	1	RPS19	1	GNAI2	-1
GDF15	4	FSTL1	1	RPS27A	1	CP	-1
HYOU1	4	SUB1	1	MDH2	1	ACPP	-1
CTNNA1	4	SRSF3	1	RPS4X	1	OGN	-1
PSME2	3	HSPD1	1	CKAP4	1	CRYZ	-1
PARP1	3	DBNL	1	RPLP2	1	VTN	-1
HADHA	3	TTN	1	RPL15	1	FGG	-1
DAK	3	NPM1	1	PSAP	1	LMOD1	-1
TRIM28	3	COPG	1	MT1G	1	PALLD	-1
ST13	2	UBA1	1	RPL18	1	HMGB1	-1
RPL31	2	CCT2	1	P4HB	1	DPYSL2	-1
RPN1	2	S100A9	1	LTF	1	FGA	-1
LAMP2	2	PPP2CA	1	IQGAP2	1	APOH	-1
DECR1	2	RPL18A	1	DHRS7	1	CORO1A	-1
LAP3	2	SET	1	PRDX3	1	LAMB2	-1
DYNC1H1	2	VCP	1	PHB	1	EMILIN1	-1
ANXA6	2	COPA	1	DDAH2	1	SERPINA1	-1
GAA	2	HSPE1	1	MAOB	1	PTRF	-1
CIRBP	2	FAM129A	1	LMNA	1	GDI1	-1
ATIC	2	TXN	1	H3F3B	1	LAMA4	-1
GRHPR	2	AHNAK	1	LCP1	1	HSPB1	-1
C1QBP	2	TPI1	1	CDC42	1	CFH	-1
FBP1	2	FUCA1	1	RPS27A	1	NID1	-1
ETFB	2	HSD17B4	1	ALDH6A1	1	ELN	-1
puf	2	CANX	1	HNRNPM	1	HSPB6	-1
PPP2R4	2	GPI	1	COPB2	1	EHD2	-1
RPL13	2	RPL7A	1	RPS13	1	LASP1	-1
UQCRC1	2	NCL	1	ECHS1	1	SERPINC1	-1
MCCC2	2	RBMX	1	CALR	1	C4B	-1
ASAH1	2	RPL14	1	TLN1	1	PTGR1	-1
ANP32A	2	RNPEP	1	RAN	1	CLU	-1
PRDX5	2	PRKCSH	1	PDIA6	1	AMBP	-1
RPL8	2	XRCC5	1	ATP1A1	1	HRG	-1
C4A	2	ACAT1	1	YWHAZ	1	CAV1	-1
NUMA1	2	PARK7	1	PCBP1	1	A1BG	-1
TKT	2	LPP	1	EIF4A1	1	CCT8	-1
APEX1	2	CPE	1	HSPA8	1	EPHX2	-1
NACA	2	COL12A1	1	RDH11	1	FHL1	-1
SEPT9	2	TAGLN	1	GLUD1	1	FBLN1	-1
DHX9	2	GOT2	1	ANXA1	-1	CFB	-1
TGFBI	2	RPS12	1	SYNPO2	-1	ARMC5	-1
MARCKS	2	DDX39B	1	SORBS1	-1	COL14A1	-1
NANS	1	RPS8	1	CSRP1	-1	SYNE1	-1
TPP1	1	HMGN2	1	MYO1C	-1	VWA1	-1
PRDX6	1	HNRNPC	1	APOA1	-1	SERPINA3	-1
IQGAP1	1	VCL	1	DPYSL3	-1	LAMA5	-1
EEF2	1	ERP29	1	CALD1	-1	DSP	-1
HEXB	1	HNRNPU	1	TNS1	-1	PLG	-2
CTSD	1	LTA4H	1	IGL@	-1	KNG1	-2
PDIA4	1	H2AFY	1	PHGDH	-1	FN1	-2
PTMS	1	PABPC1	1	GC	-1	DCN	-2
SCARB2	1	TNC	1	LGALS3BP	-1	EFEMP1	-2
S100A8	1	RPL6	1	APOA2	-1	CAPG	-2
NDUFS1	1	EEF1G	1	SYNM	-1	FBLN5	-2
ACLY	1	UGDH	1	RSU1	-1	KANK2	-2
PDLIM5	1	CMPK1	1	TPSB2	-1	COL3A1	-2
COPB1	1	CORO1B	1	FERMT2	-1	TINAGL1	-2
RPL23A	1	TXNDC17	1	HSPG2	-1	F2	-2
XRCC6	1	GNB2L1	1	PDLIM7	-1	ADIRF	-2
MSN	1	HSP90AA1	1	PPP2R1A	-1	GSTM3	-3
RPS18	1	APRT	1	FBN1	-1	OLFML1	-3
DDX17	1	PHB2	1	GSTP1	-1	A2M	-3
HNRNPK	1	ALDH7A1	1	ALDH4A1	-1	NID2	-3
UQCRC2	1	RPL12	1	FGB	-1	MAP1B	-5
PTGES3	1	AKR1A1	1	LAMC1	-1	MSMB	-5

Negative values denote under-expression. Cells are color-coded so that the most intense colors represent 16 to 32-fold change, medium shades represent 4 to 8-fold change, while the lightest shades represent 2-fold change.

Examination of the overexpressed proteins with 16-32 fold change identified previously reported proteins implicated in prostate cancer processes. This includes growth differentiation factor 15 (GDF15), hypoxia upregulated protein 1 (HYOU1), periostin (POSTN) and poly(ADP-ribose) polymerase 1 (PARP1). GDF15 or prostate-derived factor (PDF) is a divergent member of the transforming growth factor-β (TGFβ) superfamily that promotes androgen receptor (AR)-positive tumor progression by stimulating cell proliferation through the ERK1/2 signalling pathway (Chen, Karan et al. 2007). HYOU1, also known as oxygen regulated protein 150 (Orp150) is a heat shock protein involved in humoral response, whose secretion promotes cytoprotection by inducing macrophage-stimulated elaboration of endothelial cells and angiogenesis factors, as well as suppression of programmed cell death due to oxygen depravation, and facilitation of post-translational transport and processing of the vascular endothelial growth factor (VEGF) through the endoplasmic reticulum (Miyagi, Hori et al. 2002; Stojadinovic, Hooke et al. 2007). POSTN is an extracellular matrix protein that can activate the Akt/PKB (protein kinase B) pathway to facilitate tumorigenesis (Gillan, Matei et al. 2002). PARP1, a nuclear enzyme, has a well-documented role in DNA damage repair by catalyzing the poly(ADP-ribose) ylation (PARylation) of target substrates using NAD^+ as a cofactor (Griffin, Curtin et al. 1995). By conferring resistance to genotoxic effects to cancer cells, PARP1 could thus decrease their sensitive towards DNA-damaging antitumor agents. In addition, PARP1 could also be recruited to sites of AR, restoring AR function in castration-resistant prostate cancer (Schiewer, Goodwin et al. 2012).

Proteins implicated in other cancers were also identified, such as hydroxyacyl-coenzyme A (CoA) dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit (HADHA) and tripartite motif containing 28 (TRIM28). HADHA is a component of the mitochondrial trifunctional protein complex which is involved in the beta-oxidation cycle of long-chain acyl-CoAs. Its enzymatic activity is required in the beta-oxidation of polyunsaturated fatty acids, whose inhibitory effects on the growth of tumor cells are well-documented (Dai, Shen et al. 2013). Thus, increased expression of HADHA has been implicated in cisplatin resistance via loss of polyunsaturated fatty acids in lung cancer cell lines (Kageyama, Nagashio et al. 2011). TRIM28 is a member of the tripartite motif family, an evolutionarily conserved family of transcription co-factors that have been implicated in various pathologies and cancers because of its pleotropic effects on both normal and tumor cells. It has been demonstrated to have dual roles in non-small cell lung cancer cells, playing a tumor-suppressor role in the early

stage of adenocarcinoma and contributing to the epithelial-to-mesenchymal transition (EMT) during cancer metastasis (Chen, Munoz-Antonia et al. 2014). Promotion of EMT by TRIM28 has also been reported for colorectal cancer (Fitzgerald, Sheehan et al. 2013).

In addition, overexpression of DNA-dependent protein kinase catalytic subunit (DNA-PK or PRKDC) was also observed, in agreement with findings on the immunohistochemical staining of malignant epithelial cells of prostate carcinoma *in situ* and invasive prostate carcinoma (Moll, Lau et al. 1999). The overexpression, as well as underexpression, of PRKDC has also been reported in various other cancers (Hsu, Zhang et al. 2012), although the precise explanation for this variation and overexpression contradictory to its major function remains elusive. PRKDC is a member of the phophatidyl inositol 3-kinase related kinases (PIKK) superfamily, which is involved in the DNA double strand break response via non-homologous end joining during the G0 and G1 cell cycle stage (Shaheen, Znojek et al. 2011). As such, it is a critical regulator of DNA repair due to double strand breaks induced by ionizing radiation. Increased PRKDC levels in tumor cells could serve a protective role by rendering them less susceptible to radiation-based therapy.

Proteosome activator complex 28 subunit beta (PA28β or PSME2) was also shown to exhibit 16-fold overexpression in the tumor region. The PA28 activator complex is a member of the major histocompatibility complex (MHC) which enhances the generation of class 1 binding peptides by activating the catalytic complex containing γ-interferon (IFN-γ). In this sense, PA28 is important in the antigen presentation via MHC class 1 molecules and enhances antitumor activity and resorption; its impairement is thus beneficial for cancer cells, as has been demonstrated in the loss of human leukocyte antigen class 1 expression in human colon cancer cells (Miyagi, Tatsumi et al. 2003). Surprisingly though, proteasome activity is often higher in malignant tumors, suggesting that the absence of antitumor immune reaction might be due to the diminished function of the MHC class 1 itself (Kondakova, Spirina et al. 2014). It has been proposed that the high proteasome activity is either because of enhanced activity of cellular processes, many participants of which are proteasome subunits, or due to increased violations induced by cancer processes on the feedback between proteasome activity and transcription of their genes (Kondakova, Spirina et al. 2014).

Highly underexpressed proteins are also noted in the list. This includes the prostate cancer risk-modulating factor glutathione S-transferase mu 3 (GSTM3), as well as β-microseminoprotein (MSMB), one of the highly secreted proteins of the prostate whose function has not yet been elucidated (Xu, Valtonen-Andre et al. 2010; Hoogland, Dahlman et al. 2011; Haiman, Stram et al. 2012). GSTM3 is a member of the glutathione S-transferases, a multigenic family of enzymes that catalyzes the conjugation of reactive oxygen species (ROs) during oxidative damage in prostate carcinogenesis (Loktionov, Watson et al. 2001). Evidence shows that polymorphism of several genes encoding for GSTM3 leads to the production of inactive allelic variants, conferring variable efficiencies to carcinogen metabolism (Medeiros, Vasconcelos et al. 2004; Kesarwani, Singh et al. 2009). Also included are the serum protease inhibitor α-2-macroglobulin (A2M), whose decreased levels could indicate prostate cancer metastasis to the bone (Kanoh, Ohtani et al. 2001), nidogen 2 (NID2), whose hypermethylation has been implicated to favor invasion and tumor metastasis in oral cavity squamous cell carcinoma (Guerrero-Preston, Soudry et al. 2011) and has also been reported for lung, colon, and gastric cancers (Ulazzi, Sabbioni et al. 2007; Geng, Sun et al. 2012), and α-catenin (CTNNA1), whose gene deletion supports its role as a recessive cancer gene (Bignell, Greenman et al. 2010; Majewski, Kluijt et al. 2013).

Pathway and GO Analysis

The differentially expressed proteins were examined further by establishing the predicted protein-protein interactions (PPI) using STRING, plotting the PPIs using Cytoscape, analyzing the network parameters using NetworkAnalyzer, and performing MCL clustering and GO analysis using clusterMaker and ClueGO, respectively. This analysis was performed in order to determine which pathways were significantly perturbed in the tumor regions. This approach is in contrast with the differential expression analysis used previously, where individual protein markers were independently selected. The advantage is that the results are less prone to individual genetic heterogeneity that greatly decreases the discriminative power of individual markers (Dittrich, Klau et al. 2008; Lee, Chuang et al. 2008; Frohlich 2011). Also, proteins and their genes rarely act alone, thus, for their cellular functions to be understood, they have to be examined in the context of the pathways where their roles interplay with those of other proteins (Dittrich, Klau et al. 2008; Barrenas, Chavali et al. 2012; Yang, Daigle et al. 2012). The other advantage is that, with network analysis, we can perform MCL clustering to identify core modules and locate hubs (Nair, Ghatge et al. 2014). Hubs are large nodes in the network that have many connections with the smaller nodes; their centrality denotes that they have more functional relevance compared to

the other proteins/genes (Jeong, Mason et al. 2001; Ummanni, Mundt et al. 2011). It has been demonstrated, for example, that deletion of gene hubs could lead to increased frequency of occurrence of disease phenotypes compared to the deletion of less-connected ones (Barabasi and Oltvai 2004). Hubs are essential for keeping the integrity of networks because they are involved in the relay of information, thus, proteins occupying these positions become more promising biomarker candidates for more precise disease classification (Penrod and Moore 2014).

To determine which pathways were enriched in the benign and tumor regions of the prostate samples, the differentially expressed protein list was analyzed using ClueGO. ClueGO can integrate GO terms, KEGG/BioCarta pathways, Reactome and WikiPathways terms and can create a functionally grouped GO/pathway network using kappa statistics for term linkage (Bindea, Mlecnik et al. 2009), in contrast to other GO analysis softwares, such as BiNGO or PIPE, which generate hierarchical trees based on assessment of overrepresented GO terms (Maere, Heymans et al. 2005; Ramos, Shannon et al. 2008).

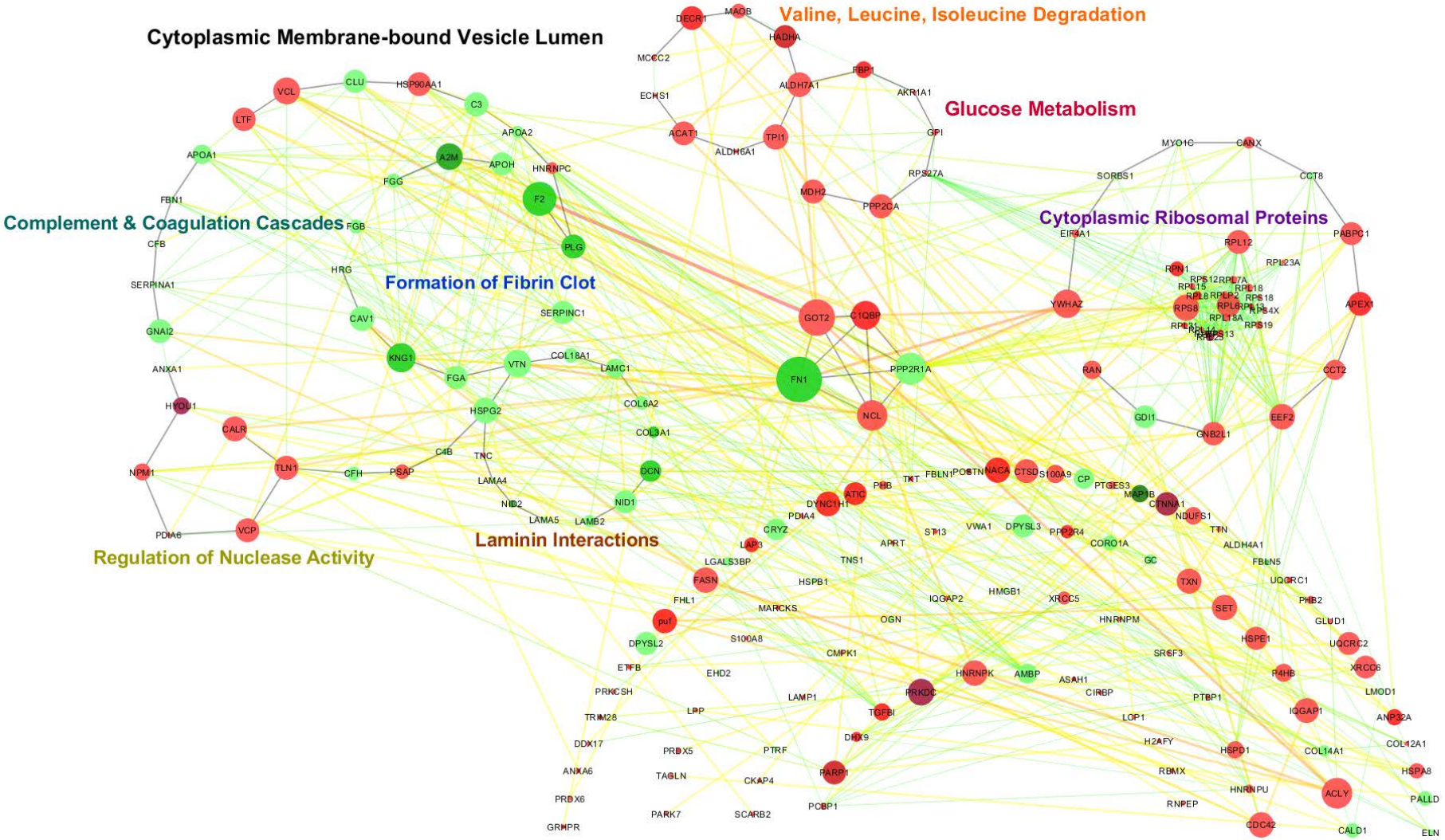

Cytoplasmic Membrane-bound Vesicle Lumen
Valine, Leucine, Isoleucine Degradation
Glucose Metabolism
Complement & Coagulation Cascades
Formation of Fibrin Clot
Cytoplasmic Ribosomal Proteins
Laminin Interactions
Regulation of Nuclease Activity
DECR1
MACB
HADHA
MCCC2
ECHS1
RBP1
ALDH7A1
AKR1A1
ACAT1
TPI1
ALDH6A1
GPI
MDH2
PPP2CA
RPS27A
MYO1C
CANX
SORBS1
CCT8
EIF4A1
RPL12
PABPC1
RPL23A
RPN1
YWHAZ
APEX1
RAN
CCT2
GDI1
EEF2
GNB2L1
VCL
CLU
HSP90AA1
C3
LTF
APOA2
APOA1
A2M
APOH
HNRNPC
FGG
FBN1
F2
FGB
CFB
PLG
HRG
SERPINA1
SERPINC1
CAV1
GNAI2
KNG1
ANXA1
VTN
COL18A1
LAMC1
FGA
HYOU1
HSPG2
COL6A2
CALR
COL3A1
C4B
TNC
NPM1
TLN1
CFH
PSAP
LAMA4
DCN
NID2
NID1
LAMA5
LAMB2
PDIA6
VCP
GOT2
C1QBP
FN1
PPP2R1A
NCL
NACA
CTSD
S100A9
CP
PTGES3
MAPRE
CTNNA1
NDUFS1
TTN
ALDH4A1
FBLN5
UQCRC1
PHB2
GLUD1
UQCRC2
XRCC6
P4HB
LMOD1
IQGAP1
ANP32A
COL12A1
HSPA8
PALLD
ELN
CALD1
ACLY
HNRNPU
RNPEP
CDC42
RBMX
H2AFY
HSPD1
COL14A1
PTBP1
LOP1
CIRBP
ASAH1
AMBP
HNRNPK
PRKDC
LAMP1
TGFBI
LPP
TRIM28
PRKCSH
ETFB
EHD2
DPYSL2
S100A8
CMPK1
SR6F3
HSPE1
OGN
HNRNPM
puf
FHL1
MARCKS
IQGAP2
HMGB1
XRCC5
SET
TXN
FASN
HSPB1
LGALS3BP
TNS1
APRT
CORO1A
GC
LAP3
CRYZ
PDIA4
DYNC1H1
ATIC
PHB
TKT
FBLN1
POSTN
VWA1
DPYSL3
PPP2R4
ST13
GRHPR
PARK7
SCARB2
PCBP1
PARP1
CKAP4
TAGLN
ANXA6
PRDX6
PRDX5
PTRF
DDX17
DHX9

Figure 23. Regulatory network plotted using Betweenness Centrality (node size) and log_2Fold Change (color) using the differentially expressed protein dataset.

Color scale: green = downregulation, red = upregulation. Color intensity denotes extent of down- or upregulation in the tumor region.

Results of the network and GO analysis are summarized in **Figure 23** and **Supplementary Table 1** (see **Annex**). After constructing the regulatory network using the PPIs and mapping using degree and betweenness centrality, the proteins were grouped according to the GO terms that they were used to identify. The node colors were then changed to log_2Fold Change to show the distinctly perturbed pathways based not only on protein member overrepresentation, but also expression. As expected, distinct pathways were perturbed. Majority of the upregulated proteins in the tumor region are involved in cytoplasmic ribosomal protein synthesis and function, including 22 ribosomal proteins. The overexpression of cytoplasmic ribosomal proteins is concomitant with the increased metabolic activity necessary to drive the proliferation of neoplastic cells (Teng, Thomas et al. 2013). The amount of cell components must necessarily be increased in order to support the survival and proper functioning of daughter cells. In tumor cells, there is a close relationship between proliferation and ribosomal protein synthesis, because tumor suppressors and proto-oncogenes that stimulate proliferation also up-regulate ribosome synthesis in accordance with the altered cell cycle progression rate (Montanaro, Trere et al. 2008; Hein, Hannan et al. 2013). Ribosomal proteins have also been proposed to actively mediate certain aspects of tumorigenesis due to the extra-ribosomal cellular functions that they possess independent of protein biosynthesis (Ziemiecki, Muller et al. 1990; Naora 1999; Shenoy, Kessel et al. 2012; Teng, Thomas et al. 2013). Likewise, the enrichment of metabolism-related terms in the tumor region, such as valine, leucine and isoleucine degradation (KEGG:00280) and glucose metabolism (REACTOME:723), is correlated with increased cell proliferation. Such processes can be utilized by the tumor cells to drive energy production and to provide metabolites needed for cellular processes and protection against oxidative stress (Feron 2009; Vander Heiden, Cantley et al. 2009). Meanwhile, the ClueGO analysis of the underregulated terms showed enrichment of proteins involved in complement and coagulation pathways (WP:558, REACTOME:2051), essential components of the immune system used in the eradication of pathogens and clot formation; such diminished immune system activity is essential for continued survival of tumor cells. Laminin interactions (REACTOME:169262) and cytoplasmic membrane-

bounded vesicle lumen (GO:0060205) were also underregulated, in accord with the alterations in the extracellular matrix to compromise the basement membrane that acts as a barrier for cancer cell invasion during EMT (Lu, Weaver et al. 2012).

To identify the hubs in the network, MCL clustering was performed (**Figure 24**). Five modules were identified from the clusters and the hubs were pulled together and placed at the center of the regulatory network to highlight them (**Figure 23** and **Supplementary Figure 2**, see **Annex**). Exploring further, the cancer gene indices (CGI) of the modules were loaded using Reactome. CGI contains data on 6,955 human genes, nearly 12,000 NCI Thesaurus cancer disease terms, and 2,180 unique pharmacological compounds from the NCI Thesaurus. The gene-disease and gene-compound associations were extracted from over 92 million analyzed sentences of nearly 20 million abstracts. The resource was last updated in June, 2009 (Milacic, Haw et al. 2012). As shown in **Figure 25**, the highlighted nodes (yellow) indicate proteins that have been annotated with the term "neoplasm". Neoplasms are abnormal growths of tissue that may or may not necessarily produce a mass (Fragoso, de Coronado et al. 2004). Expectedly, majority of these hubs have well established roles in the regulation of cancer processes, as supported by their CGI annotations. For example, fibronectin 1 (FN1) is known for its role in suppressing prostate cancer cell migration via fibronectin matrix-mediated cohesion (Jia, Entersz et al. 2012). Nucleolin (NCL) is a multifunctional ubiquitously expressed protein that acts as a shuttle between the cell surface and the cytoplasm. Cell surface expressed NCL is overexpressed in various cancers, and its ligands, such as midkine and plieotrophin among others, have been implicated in tumorigenesis and angiogenesis (Hovanessian, Soundaramourty et al. 2010; Birmpas, Briand et al. 2012). PPP2R1A codes for the 65kDa subunit Aα isoform of serine/threonine-protein phosphatase 2A, a major serine/threonine phosphatase involved in various cellular processes such as metabolism, apoptosis, DNA replication, transcription and translation (Chen, Mai et al. 2013), and it has been reported that it is a target of missense mutations in high grade serious endometrial tumors as well as ovarian tumors but at lower frequencies (McConechy, Anglesio et al. 2011). Complement 1q-binding protein (C1QBP) can affect respiration in the mitochondria where it is predominantly expressed by shifting from oxidative phosphorylation to glycolysis (Amamoto, Yagi et al. 2011). This phenomenon, called the Warburg effect, is typical in carcinogenesis. In addition, inhibition of C1QBP leads to the shutdown of the mitochondrial membrane potential leading to the production of reactive oxygen species (ROs) that in turn can arrest the G1/S phase

of the cell cycle and lead to a succession of events leading to cell death, hence the supposition that C1QBP possesses anti-apoptotic activity (McGee, Douglas et al. 2011).

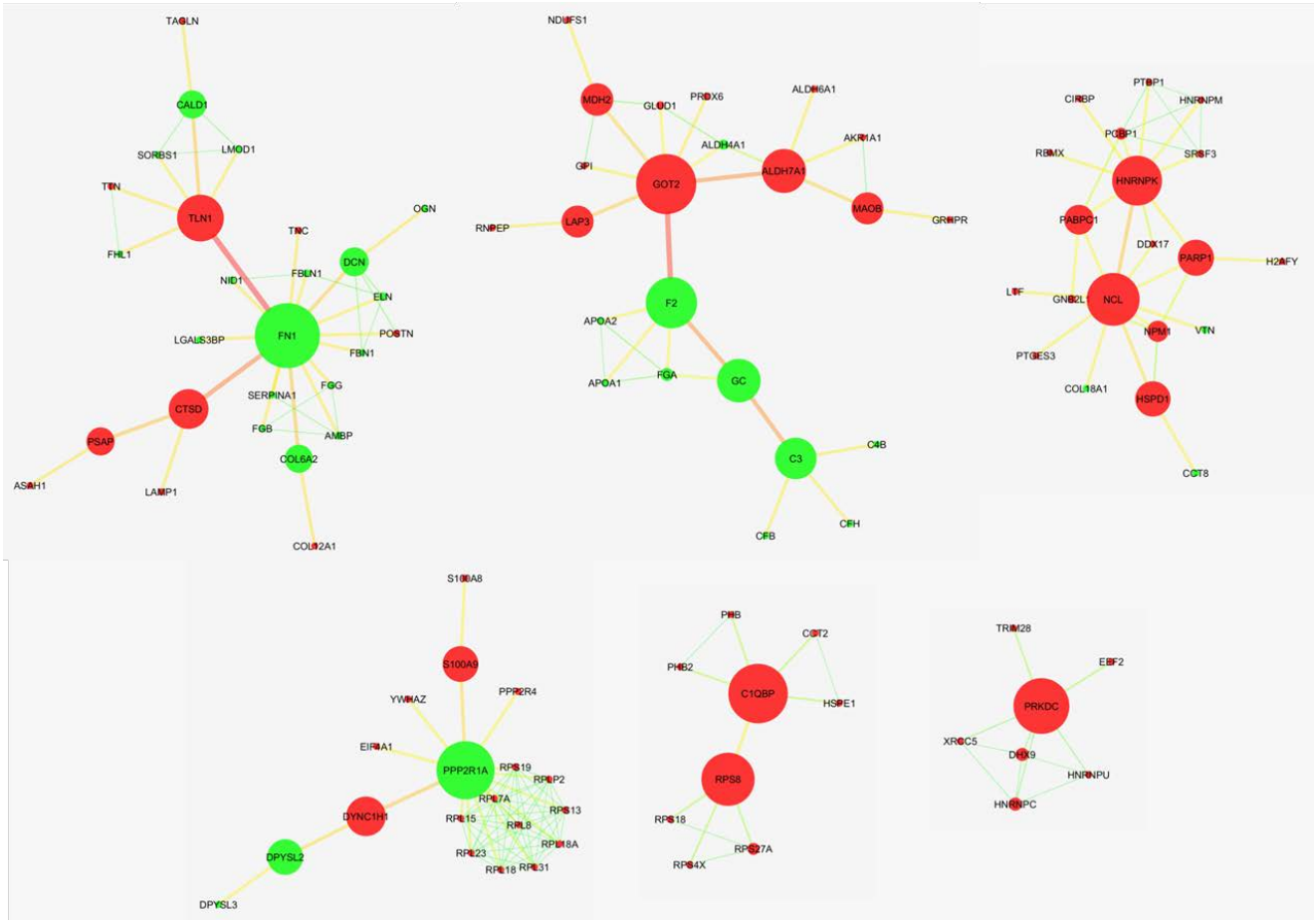


Figure 24. Top MCL clusters obtained from the PPI network, showing the hubs where most nodes intersect.

Betweenness centrality and normalized $\log_2$Fold change were used to plot the node size and color, respectively. Green and red nodes indicate under- and overregulated proteins, respectively.

Of the five hubs identified, only glutamate oxaloacetate transferase 2 (GOT2) did not possess CGI annotations, suggesting that this protein is probably involved in non-neoplastic processes. GOT2 encodes for the mitochondrial isozyme of aspartate aminotransferase (AAT), and catalyzes the conversion of oxaloacetate to aspartate. GOT2, together with mitochondrial malate dehydrogenase 2 (MDH2), comprises the malate-aspartate shuttle of the glycolysis pathway. The malate-aspartate shuttle is known to be active in neoplastic cells of several tumor types and believed to account for about 20% of the total respiratory rate {Thornburg, 2008 #2760}.

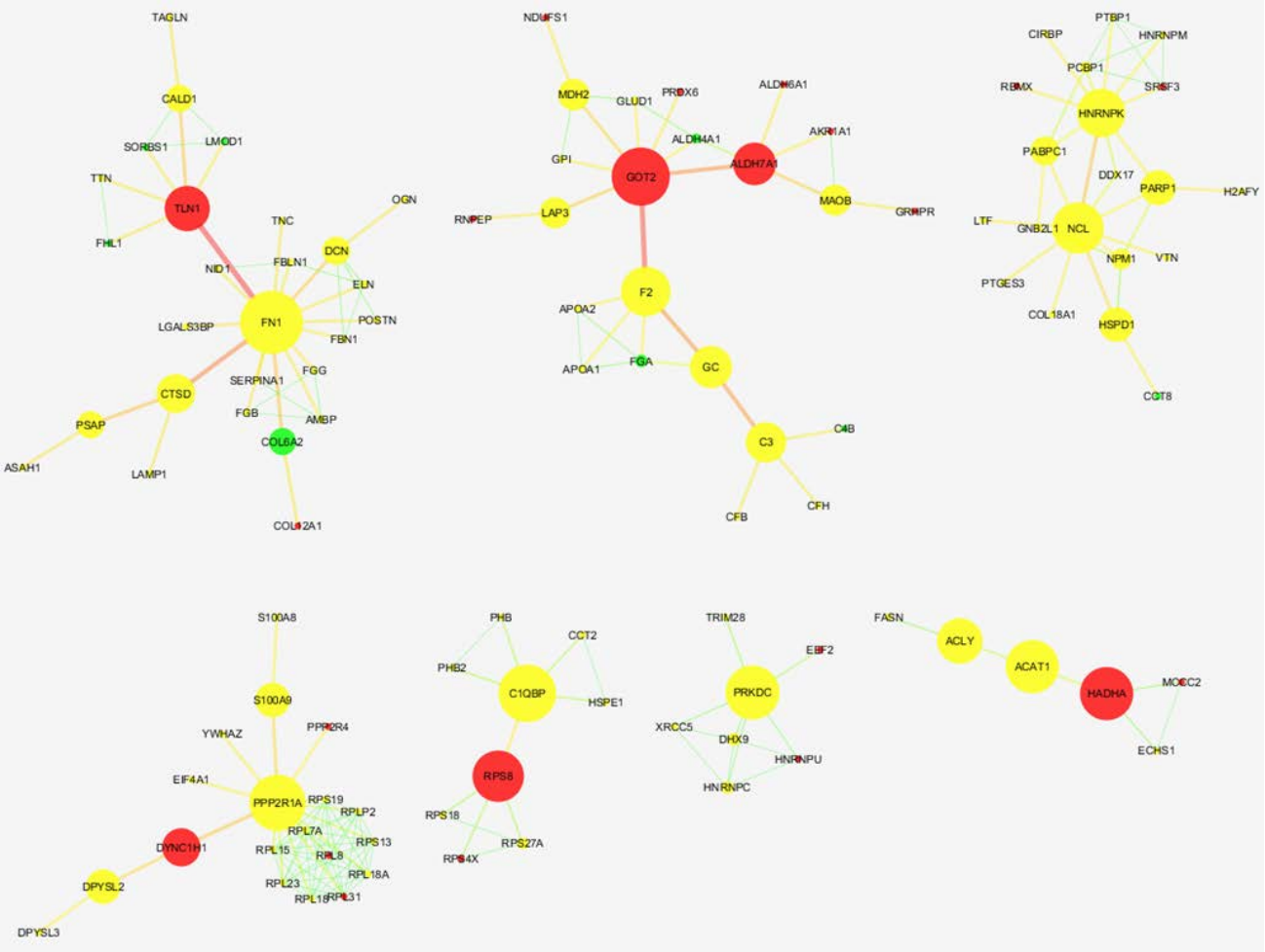


Figure 25. Mapping of Cancer Gene Index-annotated proteins using the term "neoplasm" (nodes in yellow) in the modules.

GOT2 is the only hub that does not have CGI annotations.

In summary, the current work provided a simple means to evaluate protein expression data by making use of a combination of MSI, PAM and nanoLC-MS methods to sample and analyze regions directly from tissue specimens without extensive sample preparation or purification. A combination of rigid protein identification and spectral counting parameters followed by robust statistical methods were then applied to obtain high quality differential protein expression data. Examination of the differentially expressed proteins showed established biomarkers for prostate cancer diagnosis, as well as potential candidates for further verification. Apart from individual protein expression analysis, we also used network and GO analyses to identify FN1, GOT2, NCL, PPP2R1A and C1QBP as key regulators of prostate cancer among the hundreds of proteins

that we identified to have differential expression in the tumor region. This network-based approach supports existing evidence on the active roles of these proteins in carcinogensis, and allowed us to visualize the relative importance of the protein interactions based on well-defined topological parameters that may aid in the better understanding of the underlying mechanisms of the disease and provide avenues for specific protein therapeutic targets. Results of this work need to be further validated to realize their potential in the clinical setting.

Conclusions – Part 2

This part described the development of the PAM approach for the examination of proteins directly on tissue sections. Simple mounting of tissue sections on parafilm made them amenable to direct manipulation, so that specific ROIs can easily be obtained and subjected to conventional proteomic analysis. The approach can be used in profiling and imaging modes. Profiling of regions in a rat brain section allowed for the identification of proteins of limited abundance in these regions. By extending the PAM approach to entire tissue sections, images can be generated based on spectral count data, so that the proteins mapped in this case were not imaged from only one or several peptide peaks as in conventional MSI, but from the same set of spectra that were used to identify the protein. The implication for imaging is two-fold. Now, we can map protein distribution with more confidence compared to back correlation-based MSI strategies because protein identification is done together with protein localization. Second, we can now obtain better inferences from imaging data, particularly for proteins, because we now look at localized differences based on protein expression changes themselves (which are quantifiable, by label-free methods and others), not just peak intensities (which only provide qualitative information). As the strategy is still in development, it remains to be demonstrated how the individual PAM pieces have to be decreased in size without compromising the protein identification yield.

Profiling of prostate tumors using the PAM strategy made possible the detection of protein expression changes across the benign and cancerous regions of the samples. Results led to the identification of established diagnostic biomarkers for prostate cancer. Moreover, the analysis yielded potential biomarkers which warrant further examination. These experiments highlight the advantage of localized microproteomic analysis offered by the PAM strategy; we thus expect to find more applications of this strategy in our future proteomics studies.

DISCUSSION

In the previous sections, it has been demonstrated that direct MS methods can be used in obtaining qualitative and quantitative information of proteins in tissue sections. This information ranges from protein identification, quantification, and localization. The idea of protein localization on tissue is a tenet of MS imaging. However, because of its inherent limitations, MS imaging can provide identification and quantification of only a limited set of compound species (small molecules, drugs, metabolites and lipids). Even if it has been applied to proteins also, still the information generated is restricted only to the most abundant proteins, which greatly limits its applicability in proteome-wide investigations. For instance its protein biomarker success fails in comparison to conventional proteomics approaches, where attomole protein expression changes can be routinely examined nowadays. Even then, MSI can provide localization information, which is often lost in conventional MS-based proteomics due to extensive sample purification.

In order to address this limited protein identification by MSI, we constructed protein ID databases from adjacent sections and used this to increase protein identifcations on MSI data. This approach of using high-resolution data from LC-MS for correlation with high-resolution images in MSI is similar to what other groups have previously done. However, in contrast, our approach restricts database generation only to selected regions on the tissue sample. These localized protein datasets can potentially further refine correlation because it can detect proteins of low abundance in that region alone. But more importantly in these experiments, we have demonstrated that it is possible to obtain a huge set of protein identifications on localized regions at the sub-millimeter level. To date, this amount can only be matched using LCM-based proteomics methods, but with significantly more cells sampled. This approach deviates from our initial objective of improving protein identification in MSI-based platforms, but the promising result warrants further examination. Notably, the use of the liquid microjunction extraction approach in this study to extract proteins on localized regions in profiling mode can be extended to imaging mode, thus presenting the idea of being able to simultaneously identify and image a large set of proteins in a single analysis.

Simultaneously performing imaging and protein identification is attractive because it eliminates the need of generating a database of identifications from adjacent sections, provides deeper

proteome coverage, and allows the generation of quantitative and distribution data in a single analysis. However, the major drawback of these approaches is the limited spatial information that can be obtained, which is reliant on the extent of decreasing the digestion spot size. Hence we resorted to microspotting of the proteolytic enzyme instead of performing digestion using the liquid microjunction, which is also possible, but provides lesser resolution and poor digestion efficiency (data not shown). To this end, alternatives that can produce smaller liquid microjunction spots were examined. Here, we used a capillary tubing with 100µm internal diameter to generate the liquid microjunction of smaller droplet sizes not attainable when using the current configuration of the TriVersa Nanomate. As the instrument can be directly interfaced with the nanoLC, we also examined the possibility of directly injecting the sample to the column even without sample clean-up. Results showed that this approach leads to more protein identifications, probably because a number of peptides have been lost in the stationary phase of ziptips. In this approach, the final spatial resolution was still determined by the size of the digested spot because the microjunction size attained was still large; we therefore are in the process of searching for alternatives to further decrease the size of the microjunction without compromising sensitivity.

In parallel with the liquid microjunction approach, we also developed the PAM strategy to examine tissue sections directly. This is simply a facile method of sampling ROIs directly on tissue sections for in-solution processing with a nanoLC-MS instrument. Thus we can perform profiling experiments using this approach by determining ROIs through an MSI experiment on an adjacent section, and selectively sample through the section by PAM using the determined ROI. Perhaps the major advantage of this approach can be seen when it is performed in imaging mode, which is realized by processing all the individual PAM pieces. Here we demonstrate that simultaneous identification and quantification of proteins can be done by taking advantage of the label-free quantification method spectral counting. Using spectral data in place of peak intensities is advantageous because it provides a direct correlation of spectral information that has been used for protein identification with protein localization. This is in stark contrast with MSI approaches, where proteins are identified and localized based on individual peak assignments alone, whose tenability is largely dependent in the resolutive power of the mass spectrometer used. Multiple spectra not only provide reliable protein identification but in this case also confident localization. At the present, we only focused on the development of this concept, but in the future we expect to address the limited spatial resolution and reproducibility of this approach. In addition, we applied

the PAM technique on the profiling of prostate tumors. The high rate of protein identifications obtained echoes the results we got from the liquid microjunction extraction of rat brain tissues. Label-free quantification allowed the detection of differential expression levels of proteins that have already been implicated in protate cancer, as well as those that might serve as potential candidates for further studies.

At this point I also would like to mention that we are currently developing a strategy to merge the benefits that we have gleaned using both the liquid microjunction and PAM approaches. **Figure 31** shows the schematic of this approach. This project is particularly focused on systematic automatization, multiple region sampling, and short chromatographic analyses. We aim to extend the application of this workflow in Bottom-Up proteomics approaches, on diverse types of biological samples including archived specimens.

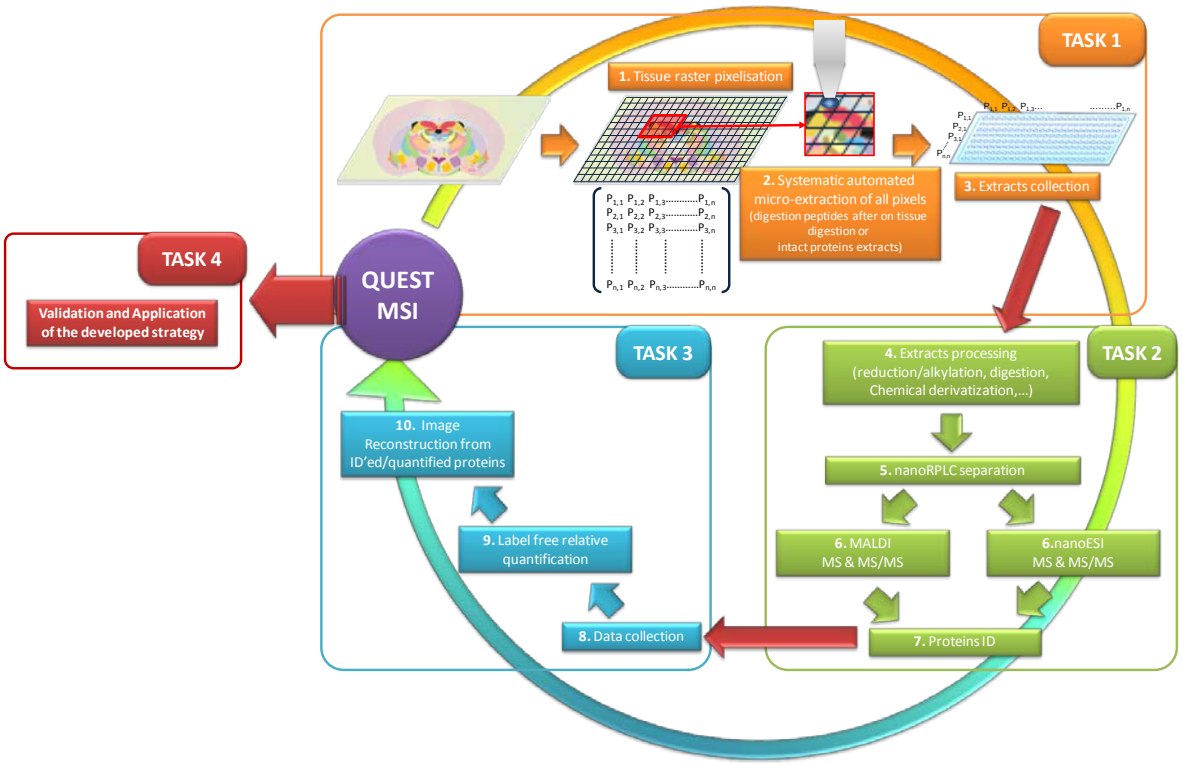


Figure 26. Workflow of the project called QUEST-MSI (Quantitative Highly-Sensitive Mass Spectrometry Imaging of Protein) currently being proposed based on the results of this work.

Taken together, these developed on-tissue direct MS proteomics strategies presented data not previously achieved with related methods, sparking the interest in their further development. Their utitlity in real-life samples is expected to increase in the near future.

LIST OF REFERENCES

Alpmann, A. and G. Morlock (2006). "Improved online coupling of planar chromatography with electrospray mass spectrometry: extraction of zones from glass plates." Anal Bioanal Chem **386**(5): 1543-51.

Amamoto, R., M. Yagi, et al. (2011). "Mitochondrial p32/C1QBP is highly expressed in prostate cancer and is associated with shorter prostate-specific antigen relapse time after radical prostatectomy." Cancer Sci **102**(3): 639-47.

Anderson, D. C. and K. Kodukula (2014). "Biomarkers in pharmacology and drug discovery." Biochem Pharmacol **87**(1): 172-88.

Andrews, G. L., R. A. Dean, et al. (2011). "Improving proteome coverage on a LTQ-Orbitrap using design of experiments." J Am Soc Mass Spectrom **22**(4): 773-83.

Aranda, M. and G. Morlock (2007). "New method for caffeine quantification by planar chromatography coupled with electropray ionization mass spectrometry using stable isotope dilution analysis." Rapid Commun Mass Spectrom **21**(7): 1297-303.

Assenov, Y., F. Ramirez, et al. (2008). "Computing topological parameters of biological networks." Bioinformatics **24**(2): 282-4.

Augustin, I., A. Betz, et al. (1999). "Differential expression of two novel Munc13 proteins in rat brain." Biochem J **337 (Pt 3)**: 363-71.

Badu-Tawiah, A. K., L. S. Eberlin, et al. (2013). "Chemical aspects of the extractive methods of ambient ionization mass spectrometry." Annu Rev Phys Chem **64**: 481-505.

Baeumlisberger, D., M. Rohmer, et al. (2011). "Simple dual-spotting procedure enhances nLC-MALDI MS/MS analysis of digests with less specific enzymes." J Proteome Res **10**(6): 2889-94.

Balluff, B., S. Rauser, et al. (2011). "MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer." Am J Pathol **179**(6): 2720-9.

Barabasi, A. L. and Z. N. Oltvai (2004). "Network biology: understanding the cell's functional organization." Nat Rev Genet **5**(2): 101-13.

Barrenas, F., S. Chavali, et al. (2012). "Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms." Genome Biol **13**(6): R46.

Bereman, M. S. and D. C. Muddiman (2007). "Detection of attomole amounts of analyte by desorption electrospray ionization mass spectrometry (DESI-MS) determined using fluorescence spectroscopy." J Am Soc Mass Spectrom **18**(6): 1093-6.

Besson, D., A. H. Pavageau, et al. (2011). "A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker." Mol Cell Proteomics **10**(12): M111 009712.

Betz, S. F. (1993). "Disulfide bonds and the stability of globular proteins." Protein Sci **2**(10): 1551-8.

Biemann, K. (1988). "Contributions of mass spectrometry to peptide and protein structure." Biomed Environ Mass Spectrom **16**(1-12): 99-111.

Biemann, K. (1990). "Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation." Methods Enzymol **193**: 455-79.

Bignell, G. R., C. D. Greenman, et al. (2010). "Signatures of mutation and selection in the cancer genome." Nature **463**(7283): 893-8.

Bindea, G., B. Mlecnik, et al. (2009). "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks." Bioinformatics **25**(8): 1091-3.

Birmpas, C., J. P. Briand, et al. (2012). "Nucleolin mediates the antiangiogenesis effect of the pseudopeptide N6L." BMC Cell Biol **13**: 32.

Blatherwick, E. Q., G. J. Van Berkel, et al. (2011). "Utility of spatially-resolved atmospheric pressure surface sampling and ionization techniques as alternatives to mass spectrometric imaging (MSI) in drug metabolism." Xenobiotica **41**(8): 720-34.

Bonnel, D., R. Legouffe, et al. (2011). "MALDI imaging techniques dedicated to drug-distribution studies." Bioanalysis **3**(12): 1399-406.

Bonnel, D., R. Longuespee, et al. (2011). "Multivariate analyses for biomarkers hunting and validation through on-tissue bottom-up or in-source decay in MALDI-MSI: application to prostate cancer." Anal Bioanal Chem **401**(1): 149-65.

Bouslimani, A., N. Bec, et al. (2010). "Matrix-assisted laser desorption/ionization imaging mass spectrometry of oxaliplatin derivatives in heated intraoperative chemotherapy (HIPEC)-like treated rat kidney." Rapid Commun Mass Spectrom **24**(4): 415-21.

Brawer, M. K. (2005). "Prostatic intraepithelial neoplasia: an overview." Rev Urol **7 Suppl 3**: S11-8.

Brewis, I. A. and P. Brennan (2010). "Proteomics technologies for the global identification and quantification of proteins." Adv Protein Chem Struct Biol **80**: 1-44.

Brown, R. S. and J. J. Lennon (1995). "Sequence-specific fragmentation of matrix-assisted laser-desorbed protein/peptide ions." Anal Chem **67**(21): 3990-9.

Bruand, J., S. Sistla, et al. (2011). "Automated querying and identification of novel peptides using MALDI mass spectrometric imaging." J Proteome Res **10**(4): 1915-28.

Buchman, V. L., H. J. Hunter, et al. (1998). "Persyn, a member of the synuclein family, has a distinct pattern of expression in the developing nervous system." J Neurosci **18**(22): 9335-41.

Burton, R. D., C. H. Watson, et al. (1997). "Proton affinities of eight matrices used for matrix-assited laser desorption/ionization." Rapid Commun Mass Spectrom **11**: 443-446.

Calligaris, D., C. Villard, et al. (2011). "Advances in top-down proteomics for disease biomarker discovery." J Proteomics **74**(7): 920-34.

Castellino, S., M. R. Groseclose, et al. (2011). "MALDI imaging mass spectrometry: bridging biology and chemistry in drug development." Bioanalysis **3**(21): 2427-41.

Catherman, A. D., O. S. Skinner, et al. (2014). "Top Down proteomics: facts and perspectives." Biochem Biophys Res Commun **445**(4): 683-93.

Cerruti, C. D., F. Benabdellah, et al. (2012). "MALDI Imaging and Structural Analysis of Rat Brain Lipid Negative Ions with 9-Aminoacridine Matrix." Anal Chem **84**(5): 2164-71.

Cha, S., M. B. Imielinski, et al. (2010). "In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology." Mol Cell Proteomics **9**(11): 2529-44.

Chambers, M. C., B. Maclean, et al. (2012). "A cross-platform toolkit for mass spectrometry and proteomics." Nat Biotechnol **30**(10): 918-20.

Chaurand, P. (2012). "Imaging mass spectrometry of thin tissue sections: a decade of collective efforts." J Proteomics **75**(16): 4883-92.

Chen, H. and R. Zenobi (2008). "Neutral desorption sampling of biological surfaces for rapid chemical characterization by extractive electrospray ionization mass spectrometry." Nat Protoc **3**(9): 1467-75.

Chen, H. F., J. R. Mai, et al. (2013). "Role of a novel functional variant in the PPP2R1A promoter on the regulation of PP2A-Aalpha and the risk of hepatocellular carcinoma." PLoS One **8**(3): e59574.

Chen, L., T. Munoz-Antonia, et al. (2014). "Trim28 Contributes to EMT via Regulation of E-Cadherin and N-Cadherin in Lung Cancer Cell Lines." PLoS One **9**(7): e101040.

Chen, R., L. Hui, et al. (2009). "Three dimensional mapping of neuropeptides and lipids in crustacean brain by mass spectral imaging." J Am Soc Mass Spectrom **20**(6): 1068-77.

Chen, R., X. Jiang, et al. (2010). "Mass spectral analysis of neuropeptide expression and distribution in the nervous system of the lobster Homarus americanus." J Proteome Res **9**(2): 818-32.

Chen, S. J., D. Karan, et al. (2007). "Prostate-derived factor as a paracrine and autocrine factor for the proliferation of androgen receptor-positive human prostate cancer cells." Prostate **67**(5): 557-71.

Clemis, E. J., D. S. Smith, et al. (2012). "Quantitation of spatially-localized proteins in tissue samples using MALDI-MRM imaging." Anal Chem **84**(8): 3514-22.

Colsch, B., S. N. Jackson, et al. (2011). "Molecular Microscopy of Brain Gangliosides: Illustrating their Distribution in Hippocampal Cell Layers." ACS Chem Neurosci **2**(4): 213-222.

Consortium (2010). "The Gene Ontology in 2010: extensions and refinements." Nucleic Acids Res **38**(Database issue): D331-5.

Coon, J. J., H. A. Steele, et al. (2002). "Laser desorption-atmospheric pressure chemical ionization: a novel ion source for the direct coupling of polyacrylamide gel electrophoresis to mass spectrometry." J Mass Spectrom **37**(11): 1163-7.

Cotter, R. J. (1992). "Time-of-flight mass spectrometry for the structural analysis of biological molecules." Anal Chem **64**(21): 1027A-1039A.

Cottrell, J. S. (2011). "Protein identification using MS/MS data." J Proteomics **74**(10): 1842-51.

Craven, R. A. and R. E. Banks (2001). "Laser capture microdissection and proteomics: possibilities and limitation." Proteomics **1**(10): 1200-4.

Craven, R. A. and R. E. Banks (2003). "Laser capture microdissection for proteome analysis." Curr Protoc Protein Sci **Chapter 22**: Unit 22 3.

D'Anjou, F., S. Routhier, et al. (2011). "Molecular Validation of PACE4 as a Target in Prostate Cancer." Transl Oncol **4**(3): 157-72.

Dai, J., J. Shen, et al. (2013). "Effects of polyunsaturated fatty acids on the growth of gastric cancer cells in vitro." Lipids Health Dis **12**(1): 71.

De Hoffmann, E. (1996). "Tandem mass spectrometry: a Primer." J Mass Spectrom **31**(2): 129-37.

Debois, D., V. Bertrand, et al. (2010). "MALDI-in source decay applied to mass spectrometry imaging: a new tool for protein identification." Anal Chem **82**(10): 4036-45.

Debois, D., A. Brunelle, et al. (2007). "Attempts for molecular depth profiling directly on a rat brain tissue section using fullerene and bismuth cluster ion beams." Int J Mass Spectrom **260**: 115-20.

Debois, D., N. Smargiasso, et al. (2013). "MALDI in-source decay, from sequencing to imaging." Top Curr Chem **331**: 117-41.

Deininger, S. O., M. P. Ebert, et al. (2008). "MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers." J Proteome Res **7**(12): 5230-6.

Delvolve, A. M., B. Colsch, et al. (2011). "Highlighting anatomical sub-structures in rat brain tissue using lipid imaging." Anal Methods **3**(8): 1729-1736.

Dijkstra, S., P. F. A. Mulders, et al. (2014). "Clinical use of novel urine and blood based prostate cancer biomarkers: A review." Clin Biochem **47**: 889-896.

Ding, C., J. Jiang, et al. (2013). "A fast workflow for identification and quantification of proteomes." Mol Cell Proteomics **12**(8): 2370-80.

Dittrich, M. T., G. W. Klau, et al. (2008). "Identifying functional modules in protein-protein interaction networks: an integrated exact approach." Bioinformatics **24**(13): i223-31.

Djidja, M. C., E. Claude, et al. (2010). "Novel molecular tumour classification using MALDI-mass spectrometry imaging of tissue micro-array." Anal Bioanal Chem **397**(2): 587-601.

Djidja, M. C., E. Claude, et al. (2009). "MALDI-ion mobility separation-mass spectrometry imaging of glucose-regulated protein 78 kDa (Grp78) in human formalin-fixed, paraffin-embedded pancreatic adenocarcinoma tissue sections." J Proteome Res **8**(10): 4876-84.

Djidja, M. C., S. Francese, et al. (2009). "Detergent addition to tryptic digests and ion mobility separation prior to MS/MS improves peptide yield and protein identification for in situ proteomic investigation of frozen and formalin-fixed paraffin-embedded adenocarcinoma tissue sections." Proteomics **9**(10): 2750-63.

Drexler, D. M., S. H. Tannehill-Gregg, et al. (2011). "Utility of quantitative whole-body autoradiography (QWBA) and imaging mass spectrometry (IMS) by matrix-assisted laser desorption/ionization (MALDI) in the assessment of ocular distribution of drugs." J Pharmacol Toxicol Methods **63**(2): 205-8.

Durrant, E. E. and R. S. Brown (2009). "Wavelength Dependence On The Level Of Post-Source Metastable Ion Decay Observed In Infrared Matrix-Assisted Laser Desorption Ionization." Int J Mass Spectrom **287**(1-3): 119-127.

Eberlin, L. S., C. R. Ferreira, et al. "Desorption electrospray ionization mass spectrometry for lipid characterization and biological tissue imaging." Biochim Biophys Acta **1811**(11): 946-60.

Edwards, R. L., A. J. Creese, et al. (2011). "Hemoglobin variant analysis via direct surface sampling of dried blood spots coupled with high-resolution mass spectrometry." Anal Chem **83**(6): 2265-70.

Eikel, D., M. Vavrek, et al. (2011). "Liquid extraction surface analysis mass spectrometry (LESA-MS) as a novel profiling tool for drug distribution and metabolism analysis: the terfenadine example." Rapid Commun Mass Spectrom **25**(23): 3587-96.

El Ayed, M., D. Bonnel, et al. (2010). "MALDI imaging mass spectrometry in ovarian cancer for tracking, identifying, and validating biomarkers." Med Sci Monit **16**(8): BR233-45.

Ellis, S. R., S. H. Brown, et al. (2013). "Surface analysis of lipids by mass spectrometry: more than just imaging." Prog Lipid Res **52**(4): 329-53.

Elnaggar, M. S., C. Barbier, et al. (2011). "Liquid microjunction surface sampling probe fluid dynamics: computational and experimental analysis of coaxial intercapillary positioning effects on sample manipulation." J Am Soc Mass Spectrom **22**(7): 1157-66.

Elsner, M., S. Rauser, et al. (2012). "MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma." J Proteomics.

Emmert-Buck, M. R., R. F. Bonner, et al. (1996). "Laser capture microdissection." Science **274**(5289): 998-1001.

Eng, J. K., A. L. McCormack, et al. (1994). "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database." J Am Soc Mass Spectrom **5**(11): 976-89.

Esquenazi, E., C. Coates, et al. (2008). "Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging." Mol Biosyst **4**(6): 562-70.

Fenn, J. B., M. Mann, et al. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.

Fenyo, D. and R. C. Beavis (2008). "Informatics development: challenges and solutions for MALDI mass spectrometry." Mass Spectrom Rev **27**(1): 1-19.

Feron, O. (2009). "Pyruvate into lactate and back: from the Warburg effect to symbiotic energy fuel exchange in cancer cells." Radiother Oncol **92**(3): 329-33.

Fitzgerald, S., K. M. Sheehan, et al. (2013). "Relationship between epithelial and stromal TRIM28 expression predicts survival in colorectal cancer patients." J Gastroenterol Hepatol **28**: 967-974.

Fletcher, J. S., N. P. Lockyer, et al. (2007). "TOF-SIMS 3D biomolecular imaging of Xenopus laevis oocytes using buckminsterfullerene (C60) primary ions." Anal Chem **79**(6): 2199-206.

Ford, M. J., M. A. Deibel, et al. (2005). "Quantitative thin-layer chromatography/mass spectrometry analysis of caffeine using a surface sampling probe electrospray ionization tandem mass spectrometry system." Anal Chem **77**(14): 4385-9.

Fournier, I., R. Day, et al. (2003). "Direct analysis of neuropeptides by in situ MALDI-TOF mass spectrometry in the rat brain." Neuro Endocrinol Lett **24**(1-2): 9-14.

Fournier, I., M. Wisztorski, et al. (2008). "Tissue imaging using MALDI-MS: a new frontier of histopathology proteomics." Expert Rev Proteomics **5**(3): 413-24.

Fragoso, G., S. de Coronado, et al. (2004). "Overview and utilization of the NCI thesaurus." Comp Funct Genomics **5**(8): 648-54.

Franck, J., K. Arafah, et al. (2009). "Improving tissue preparation for matrix-assisted laser desorption ionization mass spectrometry imaging. Part 1: using microspotting." Anal Chem **81**(19): 8193-202.

Franck, J., K. Arafah, et al. (2009). "MALDI imaging mass spectrometry: state of the art technology in clinical proteomics." Mol Cell Proteomics **8**(9): 2023-33.

Franck, J., M. El Ayed, et al. (2009). "On-tissue N-terminal peptide derivatizations for enhancing protein identification in MALDI mass spectrometric imaging strategies." Anal Chem **81**(20): 8305-17.

Franck, J., R. Longuespee, et al. (2010). "MALDI mass spectrometry imaging of proteins exceeding 30,000 daltons." Med Sci Monit **16**(9): BR293-9.

Franck, J., J. Quanico, et al. (2013). "Quantification-based mass spectrometry imaging of proteins by parafilm assisted microdissection." Anal Chem **85**(17): 8127-34.

Frohlich, H. (2011). "Network based consensus gene signatures for biomarker discovery in breast cancer." PLoS One **6**(10): e25364.

Fujishige, K., J. Kotera, et al. (1999). "Striatum- and testis-specific phosphodiesterase PDE10A isolation and characterization of a rat PDE10A." Eur J Biochem **266**(3): 1118-27.

Gagnon, H., J. Franck, et al. (2012). "TARGETED MASS spectrometry Imaging: Specific Targeting Mass Spectrometry imaging technologies from history to perspective." Prog Histochem Cytochem **47**(3): 133-74.

Ge, Y., I. N. Rybakova, et al. (2009). "Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state." Proc Natl Acad Sci U S A **106**(31): 12658-63.

Geiger, T., A. Velic, et al. (2013). "Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse." Mol Cell Proteomics **12**(6): 1709-22.

Geng, J., J. Sun, et al. (2012). "Methylation status of NEUROG2 and NID2 improves the diagnosis of stage I NSCLC." Oncol Lett **3**(4): 901-906.

Gessel, M. M., J. L. Norris, et al. (2014). "MALDI imaging mass spectrometry: Spatial molecular analysis to enable a new age of discovery." J Proteomics.

Giavalisco, P., J. Hummel, et al. (2008). "High-resolution direct infusion-based mass spectrometry in combination with whole 13C metabolome isotope labeling allows unambiguous assignment of chemical sum formulas." Anal Chem **80**(24): 9417-25.

Gillan, L., D. Matei, et al. (2002). "Periostin secreted by epithelial ovarian carcinoma is a ligand for alpha(V)beta(3) and alpha(V)beta(5) integrins and promotes cell motility." Cancer Res **62**(18): 5358-64.

Goodwin, R. J., S. L. Iverson, et al. (2012). "The significance of ambient-temperature on pharmaceutical and endogenous compound abundance and distribution in tissues sections when analyzed by matrix-assisted laser desorption/ionization mass spectrometry imaging." Rapid Commun Mass Spectrom **26**(5): 494-8.

Goodwin, R. J., L. Macintyre, et al. (2010). "A solvent-free matrix application method for matrix-assisted laser desorption/ionization imaging of small molecules." Rapid Commun Mass Spectrom **24**(11): 1682-6.

Goodwin, R. J., C. L. Mackay, et al. (2011). "Qualitative and quantitative MALDI imaging of the positron emission tomography ligands raclopride (a D2 dopamine antagonist) and SCH 23390 (a D1 dopamine antagonist) in rat brain tissue sections using a solvent-free dry matrix application method." Anal Chem **83**(24): 9694-701.

Goodwin, R. J., S. R. Pennington, et al. (2008). "Protein and peptides in pictures: imaging with MALDI mass spectrometry." Proteomics **8**(18): 3785-800.

Goodwin, R. J., P. Scullion, et al. (2010). "Use of a solvent-free dry matrix coating for quantitative matrix-assisted laser desorption ionization imaging of 4-bromophenyl-1,4-diazabicyclo(3.2.2)nonane-4-carboxylate in rat brain and quantitative analysis of the drug from laser microdissected tissue regions." Anal Chem **82**(9): 3868-73.

Gozal, Y. M., E. B. Dammer, et al. (2011). "Proteomic analysis of hippocampal dentate granule cells in frontotemporal lobar degeneration: application of laser capture technology." Front Neurol **2**: 24.

Grassl, J., N. L. Taylor, et al. (2011). "Matrix-assisted laser desorption/ionisation mass spectrometry imaging and its development for plant protein imaging." Plant Methods **7**(1): 21.

Grey, A. C., P. Chaurand, et al. (2009). "MALDI imaging mass spectrometry of integral membrane proteins from ocular lens and retinal tissue." J Proteome Res **8**(7): 3278-83.

Griffin, R. J., N. J. Curtin, et al. (1995). "The role of inhibitors of poly(ADP-ribose) polymerase as resistance-modifying agents in cancer therapy." Biochimie **77**(6): 408-22.

Griffiths, W. J. and Y. Wang (2009). "Mass spectrometry: from proteomics to metabolomics and lipidomics." Chem Soc Rev **38**(7): 1882-96.

Groseclose, M. R., M. Andersson, et al. (2007). "Identification of proteins directly from tissue: in situ tryptic digestions coupled with imaging mass spectrometry." J Mass Spectrom **42**(2): 254-62.

Gross, J. H. (2014). "Direct analysis in real time--a critical review on DART-MS." Anal Bioanal Chem **406**(1): 63-80.

Guerrero-Preston, R., E. Soudry, et al. (2011). "NID2 and HOXA9 promoter hypermethylation as biomarkers for prevention and early detection in oral cavity squamous cell carcinoma tissues and saliva." Cancer Prev Res (Phila) **4**(7): 1061-72.

Gustafsson, J. O., J. S. Eddes, et al. (2012). "Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS." J Proteomics **75**(16): 5093-105.

Gustafsson, J. O., M. K. Oehler, et al. (2010). "Citric acid antigen retrieval (CAAR) for tryptic peptide imaging directly on archived formalin-fixed paraffin-embedded tissue." J Proteome Res **9**(9): 4315-28.

Gustafsson, J. O., M. K. Oehler, et al. (2011). "MALDI Imaging Mass Spectrometry (MALDI-IMS)-Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis." Int J Mol Sci **12**(1): 773-94.

Gygi, S. P., B. Rist, et al. (1999). "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags." Nat Biotechnol **17**(10): 994-9.
Haapala, M., J. Pol, et al. (2007). "Desorption atmospheric pressure photoionization." Anal Chem **79**(20): 7867-72.
Haddad, R., R. Sparrapan, et al. (2006). "Desorption sonic spray ionization for (high) voltage-free ambient mass spectrometry." Rapid Commun Mass Spectrom **20**(19): 2901-5.
Haiman, C. A., D. O. Stram, et al. (2012). "Levels of beta-microseminoprotein in blood and risk of prostate cancer in multiple populations." J Natl Cancer Inst **105**(3): 237-43.
Hamm, G., D. Bonnel, et al. (2012). "Quantitative mass spectrometry imaging of propranolol and olanzapine using tissue extinction calculation as normalization factor." J Proteomics **75**(16): 4952-61.
Hankin, J. A., R. M. Barkley, et al. (2007). "Sublimation as a method of matrix application for mass spectrometric imaging." J Am Soc Mass Spectrom **18**(9): 1646-52.
Hankin, J. A., S. E. Farias, et al. (2011). "MALDI mass spectrometric imaging of lipids in rat brain injury models." J Am Soc Mass Spectrom **22**(6): 1014-21.
Hanrieder, J., A. Ljungdahl, et al. (2011). "L-DOPA-induced dyskinesia is associated with regional increase of striatal dynorphin peptides as elucidated by imaging mass spectrometry." Mol Cell Proteomics **10**(10): M111 009308.
Hardouin, J. (2007). "Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry." Mass Spectrom Rev **26**(5): 672-82.
Harris, G. A., J. J. Nicklay, et al. (2013). "Localized in situ hydrogel-mediated protein digestion and extraction technique for on-tissue analysis." Anal Chem **85**(5): 2717-23.
Hart-Smith, G. (2014). "A review of electron-capture and electron-transfer dissociation tandem mass spectrometry in polymer chemistry." Anal Chim Acta **808**: 44-55.
Hattori, K., M. Kajimura, et al. (2010). "Paradoxical ATP elevation in ischemic penumbra revealed by quantitative imaging mass spectrometry." Antioxid Redox Signal **13**(8): 1157-67.
Hein, N., K. M. Hannan, et al. (2013). "The nucleolus: an emerging target for cancer therapy." Trends Mol Med **19**(11): 643-54.
Henzel, W. J., T. M. Billeci, et al. (1993). "Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases." Proc Natl Acad Sci U S A **90**(11): 5011-5.
Higgs, R. E., M. D. Knierman, et al. (2005). "Comprehensive label-free method for the relative quantification of proteins from biological samples." J Proteome Res **4**(4): 1442-50.
Hoogland, A. M., A. Dahlman, et al. (2011). "Cysteine-rich secretory protein 3 and beta-microseminoprotein on prostate cancer needle biopsies do not have predictive value for subsequent prostatectomy outcome." BJU Int **108**(8): 1356-62.
Hovanessian, A. G., C. Soundaramourty, et al. (2010). "Surface expressed nucleolin is constantly induced in tumor cells to mediate calcium-dependent ligand internalization." PLoS One **5**(12): e15787.
Hozumi, Y., M. Fukaya, et al. (2008). "Diacylglycerol kinase beta accumulates on the perisynaptic site of medium spiny neurons in the striatum." Eur J Neurosci **28**(12): 2409-22.
Hsu, F. M., S. Zhang, et al. (2012). "Role of DNA-dependent protein kinase catalytic subunit in cancer development and treatment." Transl Cancer Res **1**(1): 22-34.
Huang, M. Z., S. C. Cheng, et al. (2011). "Ambient ionization mass spectrometry: a tutorial." Anal Chim Acta **702**(1): 1-15.
Hunt, D. F., J. R. Yates, 3rd, et al. (1986). "Protein sequencing by tandem mass spectrometry." Proc Natl Acad Sci U S A **83**(17): 6233-7.

Hustoft, H. K., L. Reubsaet, et al. (2011). "Critical assessment of accelerating trypsination methods." J Pharm Biomed Anal **56**(5): 1069-78.

Ishihama, Y. (2005). "Proteomic LC-MS systems using nanoscale liquid chromatography with tandem mass spectrometry." J Chromatogr A **1067**(1-2): 73-83.

James, P., M. Quadroni, et al. (1993). "Protein identification by mass profile fingerprinting." Biochem Biophys Res Commun **195**(1): 58-64.

Jeong, H., S. P. Mason, et al. (2001). "Lethality and centrality in protein networks." Nature **411**(6833): 41-2.

Jia, D., I. Entersz, et al. (2012). "Fibronectin matrix-mediated cohesion suppresses invasion of prostate cancer cells." BMC Cancer **12**: 94.

Jiao, J., A. Miao, et al. (2013). "Realization of on-tissue protein identification by highly efficient in situ digestion with graphene-immobilized trypsin for MALDI imaging analysis." Analyst **138**(6): 1645-8.

Johnson, D., B. Boyes, et al. (2013). "Optimization of data-dependent acquisition parameters for coupling high-speed separations with LC-MS/MS for protein identifications." J Biomol Tech **24**(2): 62-72.

Johnson, R. S., M. T. Davis, et al. (2005). "Informatics for protein identification by mass spectrometry." Methods **35**(3): 223-36.

Jones, E. A., N. P. Lockyer, et al. (2007). "Mass spectral analysis and imaging of tissue by ToF-SIMS--The role of buckminsterfullerene, C_{60}^+, primary ions." Int J Mass Spectrom **260**: 146-57.

Jungmann, J. H. and R. M. Heeren (2012). "Emerging technologies in mass spectrometry imaging." J Proteomics **75**(16): 5077-92.

Kageyama, T., R. Nagashio, et al. (2011). "HADHA is a potential predictor of response to platinum-based chemotherapy for lung cancer." Asian Pac J Cancer Prev **12**(12): 3457-63.

Kallback, P., M. Shariatgorji, et al. (2012). "Novel mass spectrometry imaging software assisting labeled normalization and quantitation of drugs and neuropeptides directly in tissue sections." J Proteomics **75**(16): 4941-51.

Kanoh, Y., N. Ohtani, et al. (2001). "Levels of alpha 2 macroglobulin can predict bone metastases in prostate cancer." Anticancer Res **21**(1B): 551-6.

Kapp, E. and F. Schutz (2007). "Overview of tandem mass spectrometry (MS/MS) database search algorithms." Curr Protoc Protein Sci **Chapter 25**: Unit25 2.

Kapp, E. A., F. Schutz, et al. (2005). "An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis." Proteomics **5**(13): 3475-90.

Karas, M., D. Bachmann, et al. (1987). "Matrix-assisted ultraviolet laser desorption of non-volatile compounds." Int J Mass Spectrom Ion Processes **78**: 53-68.

Karas, M., M. Gluckmann, et al. (2000). "Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors." J Mass Spectrom **35**(1): 1-12.

Karas, M. and F. Hillenkamp (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." Anal Chem **60**(20): 2299-301.

Karas, M. and R. Kruger (2003). "Ion formation in MALDI: the cluster ionization mechanism." Chem Rev **103**(2): 427-40.

Kaspar, S., M. Peukert, et al. (2011). "MALDI-imaging mass spectrometry - An emerging technique in plant biology." Proteomics **11**(9): 1840-50.

Keller, A., A. I. Nesvizhskii, et al. (2002). "Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search." Anal Chem **74**(20): 5383-92.

Kertesz, V., M. J. Ford, et al. (2005). "Automation of a surface sampling probe/electrospray mass spectrometry system." Anal Chem **77**(22): 7183-9.

Kertesz, V. and G. J. Van Berkel (2008). "Improved imaging resolution in desorption electrospray ionization mass spectrometry." Rapid Commun Mass Spectrom **22**(17): 2639-44.

Kertesz, V. and G. J. Van Berkel (2010). "Fully automated liquid extraction-based surface sampling and ionization using a chip-based robotic nanoelectrospray platform." J Mass Spectrom **45**(3): 252-60.

Kertesz, V. and G. J. Van Berkel (2010). "Liquid microjunction surface sampling coupled with high-pressure liquid chromatography-electrospray ionization-mass spectrometry for analysis of drugs and metabolites in whole-body thin tissue sections." Anal Chem **82**(14): 5917-21.

Kertesz, V., G. J. Van Berkel, et al. (2008). "Comparison of drug distribution images from whole-body thin tissue sections obtained using desorption electrospray ionization tandem mass spectrometry and autoradiography." Anal Chem **80**(13): 5168-77.

Kesarwani, P., R. Singh, et al. (2009). "Association of GSTM3 intron 6 variant with cigarette smoking, tobacco chewing and alcohol as modifier factors for prostate cancer risk." Arch Toxicol **83**(4): 351-6.

Khatib-Shahidi, S., M. Andersson, et al. (2006). "Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry." Anal Chem **78**(18): 6448-56.

Kiss, A., D. F. Smith, et al. (2014). "Top-down mass spectrometry imaging of intact proteins by laser ablation ESI FT-ICR MS." Proteomics **14**(10): 1283-9.

Komatsu, M., Y. Murayama, et al. (2008). "Protein fragment imaging using ink jet printing digestion technique." Appl Surf Sci **255**: 1162-64.

Kondakova, I. V., L. V. Spirina, et al. (2014). "Chymotrypsin-like activity and subunit composition of proteasomes in human cancers." Mol. Biol. (Moscow, Russ. Fed., Engl. Ed.) **48**(3): 384-389.

Kristiansen, G. (2010). "Manual microdissection." Methods Mol Biol **576**: 31-8.

Landgraf, R. R., T. J. Garrett, et al. (2011). "Considerations for quantification of lipids in nerve tissue using matrix-assisted laser desorption/ionization mass spectrometric imaging." Rapid Commun Mass Spectrom **25**(20): 3178-84.

Lanekoff, I., N. T. Phan, et al. (2013). "Mass spectrometry imaging of freeze-dried membrane phospholipids of dividing Tetrahymena pyriformis." Surf Interface Anal **45**(1): 211-214.

Lanekoff, I., M. Thomas, et al. (2014). "Shotgun approach for quantitative imaging of phospholipids using nanospray desorption electrospray ionization mass spectrometry." Anal Chem **86**(3): 1872-80.

Lange, V., P. Picotti, et al. (2008). "Selected reaction monitoring for quantitative proteomics: a tutorial." Mol Syst Biol **4**: 222.

Laskin, J., B. S. Heath, et al. (2011). "Tissue imaging using nanospray desorption electrospray ionization mass spectrometry." Anal Chem **84**(1): 141-8.

Le Faouder, J., S. Laouirem, et al. (2011). "Imaging mass spectrometry provides fingerprints for distinguishing hepatocellular carcinoma from cirrhosis." J Proteome Res **10**(8): 3755-65.

Lee, E., H. Y. Chuang, et al. (2008). "Inferring pathway activity toward precise disease classification." PLoS Comput Biol **4**(11): e1000217.

Lee, Y. J., D. C. Perdian, et al. (2012). "Use of mass spectrometry for imaging metabolites in plants." Plant J **70**(1): 81-95.

Lemaire, R., A. Desmons, et al. (2007). "Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections." J Proteome Res **6**(4): 1295-305.

Lemaire, R., S. A. Menguellet, et al. (2007). "Specific MALDI imaging and profiling for biomarker hunting and validation: fragment of the 11S proteasome activator complex, Reg alpha fragment, is a new potential ovary cancer biomarker." J Proteome Res **6**(11): 4127-34.

Lemaire, R., M. Wisztorski, et al. (2006). "MALDI-MS direct tissue analysis of proteins: Improving signal sensitivity using organic treatments." Analytical Chemistry **78**(20): 7145-7153.

Li, J., E. D. Inutan, et al. (2012). "Matrix assisted ionization: new aromatic and nonaromatic matrix compounds producing multiply charged lipid, peptide, and protein ions in the positive and negative mode observed directly from surfaces." J Am Soc Mass Spectrom **23**(10): 1625-43.

Li, Y., B. Shrestha, et al. (2007). "Atmospheric pressure molecular imaging by infrared MALDI mass spectrometry." Anal Chem **79**(2): 523-32.

Lietz, C. B., E. Gemperline, et al. (2013). "Qualitative and quantitative mass spectrometry imaging of drugs and metabolites." Adv Drug Deliv Rev **65**(8): 1074-85.

Liu, D. and D. J. Smith (2003). "Voxelation and gene expression tomography for the acquisition of 3-D gene expression maps in the brain." Methods **31**(4): 317-25.

Liu, J. and Z. Ouyang "Mass spectrometry imaging for biomedical applications." Anal Bioanal Chem **405**(17): 5645-53.

Liu, Y., R. Huttenhain, et al. (2013). "Mass spectrometric protein maps for biomarker discovery and clinical research." Expert Rev Mol Diagn **13**(8): 811-25.

Lohne, J. J., W. C. Andersen, et al. (2012). "Laser diode thermal desorption mass spectrometry for the analysis of quinolone antibiotic residues in aquacultured seafood." Rapid Commun Mass Spectrom **26**(24): 2854-64.

Loktionov, A., M. A. Watson, et al. (2001). "Glutathione-S-transferase gene polymorphisms in colorectal cancer patients: interaction between GSTM1 and GSTM3 allele variants as a risk-modulating factor." Carcinogenesis **22**(7): 1053-60.

Longuespee, R., H. Gagnon, et al. (2013). "Proteomic analyses of serous and endometrioid epithelial ovarian cancers - cases studies - molecular insights of a possible histological etiology of serous ovarian cancer." Proteomics Clin Appl **7**(5-6): 337-54.

Lopez-Ferrer, D., K. Petritis, et al. (2010). "Pressurized pepsin digestion in proteomics: an automatable alternative to trypsin for integrated top-down bottom-up proteomics." Mol Cell Proteomics **10**(2): M110 001479.

Louris, J. N., L. G. Wright, et al. (1985). "New scan modes accessed with a hybrid mass spectrometer." Anal Chem **57**: 2918-24.

Lu, P., C. Vogel, et al. (2007). "Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation." Nat Biotechnol **25**(1): 117-24.

Lu, P., V. M. Weaver, et al. (2012). "The extracellular matrix: a dynamic niche in cancer progression." J Cell Biol **196**(4): 395-406.

Lu, X. and H. Zhu (2005). "Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins." Mol Cell Proteomics **4**(12): 1948-58.

Luftmann, H. (2004). "A simple device for the extraction of TLC spots: direct coupling with an electrospray mass spectrometer." Anal Bioanal Chem **378**(4): 964-8.

Luftmann, H., M. Aranda, et al. (2007). "Automated interface for hyphenation of planar chromatography with mass spectrometry." Rapid Commun Mass Spectrom **21**(23): 3772-6.

Luosujarvi, L., V. Arvola, et al. (2008). "Desorption and ionization mechanisms in desorption atmospheric pressure photoionization." Anal Chem **80**(19): 7460-6.

Maere, S., K. Heymans, et al. (2005). "BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks." Bioinformatics **21**(16): 3448-9.
Maes, E., V. Broeckx, et al. (2013). "Analysis of the formalin-fixed paraffin-embedded tissue proteome: pitfalls, challenges, and future prospectives." Amino Acids **45**(2): 205-18.
Majewski, I. J., I. Kluijt, et al. (2013). "An alpha-E-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer." J Pathol **229**(4): 621-9.
Mallick, P., M. Schirle, et al. (2007). "Computational prediction of proteotypic peptides for quantitative proteomics." Nat Biotechnol **25**(1): 125-31.
Mann, M., R. C. Hendrickson, et al. (2001). "Analysis of proteins and proteomes by mass spectrometry." Annu Rev Biochem **70**: 437-73.
Mann, M., N. A. Kulak, et al. (2013). "The coming age of complete, accurate, and ubiquitous proteomes." Mol Cell **49**(4): 583-90.
Marshall, A. G. and C. L. Hendrickson (2008). "High-resolution mass spectrometers." Annu Rev Anal Chem (Palo Alto Calif) **1**: 579-99.
Marshall, A. G., C. L. Hendrickson, et al. (1998). "Fourier transform ion cyclotron resonance mass spectrometry: a primer." Mass Spectrom Rev **17**(1): 1-35.
Marshall, P., V. Toteu-Djomte, et al. (2010). "Correlation of skin blanching and percutaneous absorption for glucocorticoid receptor agonists by matrix-assisted laser desorption ionization mass spectrometry imaging and liquid extraction surface analysis with nanoelectrospray ionization mass spectrometry." Anal Chem **82**(18): 7787-94.
Mascini, N. E. and R. M. A. Heeren (2012). "Protein identification in mass-spectrometry imaging." Trends Anal Chem **40**: 28-37.
McConechy, M. K., M. S. Anglesio, et al. (2011). "Subtype-specific mutation of PPP2R1A in endometrial and ovarian carcinomas." J Pathol **223**(5): 567-73.
McDonald, W. H. and J. R. Yates, 3rd (2002). "Shotgun proteomics and biomarker discovery." Dis Markers **18**(2): 99-105.
McDonnell, L. A., G. L. Corthals, et al. (2010). "Peptide and protein imaging mass spectrometry in cancer research." J Proteomics **73**(10): 1921-44.
McGee, A. M., D. L. Douglas, et al. (2011). "The mitochondrial protein C1qbp promotes cell proliferation, migration and resistance to cell death." Cell Cycle **10**(23): 4119-27.
McLean, J. A., W. B. Ridenour, et al. (2007). "Profiling and imaging of tissues by imaging ion mobility-mass spectrometry." J Mass Spectrom **42**(8): 1099-105.
McVary, K. T. (2007). "A review of combination therapy in patients with benign prostatic hyperplasia." Clin Ther **29**(3): 387-98.
Medeiros, R., A. Vasconcelos, et al. (2004). "Metabolic susceptibility genes and prostate cancer risk in a southern European population: the role of glutathione S-transferases GSTM1, GSTM3, and GSTT1 genetic polymorphisms." Prostate **58**(4): 414-20.
Meding, S., U. Nitsche, et al. (2012). "Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging." J Proteome Res **11**(3): 1996-2003.
Medzihradszky, K. F. and R. J. Chalkley (2013). "Lessons in *de novo* peptide sequencing by tandem mass spectrometry." Mass Spectrom Rev.
Meissner, F. and M. Mann (2014). "Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology." Nat Immunol **15**(2): 112-7.
Meriaux, C., K. Arafah, et al. (2011). "Multiple changes in peptide and lipid expression associated with regeneration in the nervous system of the medicinal leech." PLoS One **6**(4): e18359.
Meriaux, C., J. Franck, et al. (2010). "Liquid ionic matrixes for MALDI mass spectrometry imaging of lipids." J Proteomics **73**(6): 1204-18.

Miao, Z. and H. Chen (2009). "Direct analysis of liquid samples by desorption electrospray ionization-mass spectrometry (DESI-MS)." J Am Soc Mass Spectrom **20**(1): 10-9.
Michalski, A., E. Damoc, et al. (2011). "Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer." Mol Cell Proteomics **10**(9): M111 011015.
Michalski, A., E. Damoc, et al. (2012). "Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes." Mol Cell Proteomics **11**(3): O111 013698.
Milacic, M., R. Haw, et al. (2012). "Annotating cancer variants and anti-cancer therapeutics in reactome." Cancers (Basel) **4**(4): 1180-211.
Miller, I., J. Crawford, et al. (2006). "Protein stains for proteomic applications: which, when, why?" Proteomics **6**(20): 5385-408.
Miller, P. E. and M. B. Denton (1986). "The Quadrupole Mass Filter: Basic Operating Concepts." J Chem Educ **63**(7): 617-22.
Minerva, L., K. Boonen, et al. (2011). "Linking mass spectrometric imaging and traditional peptidomics: a validation in the obese mouse model." Anal Chem **83**(20): 7682-91.
Miura, D., Y. Fujimura, et al. (2010). "Ultrahighly sensitive in situ metabolomic imaging for visualizing spatiotemporal metabolic behaviors." Anal Chem **82**(23): 9789-96.
Miyagi, T., O. Hori, et al. (2002). "Antitumor effect of reduction of 150-kDa oxygen-regulated protein expression on human prostate cancer cells." Int J Urol **9**(10): 577-85.
Miyagi, T., T. Tatsumi, et al. (2003). "Impaired expression of proteasome subunits and human leukocyte antigens class I in human colon cancer cells." J Gastroenterol Hepatol **18**(1): 32-40.
Modestov, A. D., S. Srebnik, et al. (2001). "Scanning capillary microscopy/mass spectrometry for mapping spatial electrochemical activity of electrodes." Anal Chem **73**(17): 4229-40.
Moll, U., R. Lau, et al. (1999). "DNA-PK, the DNA-activated protein kinase, is differentially expressed in normal and malignant human tissues." Oncogene **18**(20): 3114-26.
Montanaro, L., D. Trere, et al. (2008). "Nucleolus, ribosomes, and cancer." Am J Pathol **173**(2): 301-10.
Montowska, M., W. Rao, et al. (2014). "Tryptic Digestion Coupled with Ambient Desorption Electrospray Ionization and Liquid Extraction Surface Analysis Mass Spectrometry Enabling Identification of Skeletal Muscle Proteins in Mixtures and Distinguishing between Beef, Pork, Horse, Chicken, and Turkey Meat." Anal Chem **86**(9): 4479-87.
Morlock, G. E. and U. Jautz (2008). "Comparison of two different plunger geometries for HPTLC-MS coupling via an extractor-based interface." J Planar Chromatogr - Mod TLC **21**(5): 367-71.
Murphy, R. C., J. A. Hankin, et al. (2011). "MALDI imaging of lipids after matrix sublimation/deposition." Biochim Biophys Acta **1811**(11): 970-975.
Mustafa, D., J. M. Kros, et al. (2008). "Combining laser capture microdissection and proteomics techniques." Methods Mol Biol **428**: 159-78.
Nahnsen, S., C. Bielow, et al. (2013). "Tools for label-free peptide quantification." Mol Cell Proteomics **12**(3): 549-56.
Nair, J., M. Ghatge, et al. (2014). "Network analysis of inflammatory genes and their transcriptional regulators in coronary artery disease." PLoS One **9**(4): e94328.
Naora, H. (1999). "Involvement of ribosomal proteins in regulating cell growth and apoptosis: translational modulation or recruitment for extraribosomal activity?" Immunol Cell Biol **77**(3): 197-205.
Neilson, K. A., N. A. Ali, et al. (2011). "Less label, more free: approaches in label-free quantitative mass spectrometry." Proteomics **11**(4): 535-53.

Nemes, P., S. S. Rubakhin, et al. (2013). "Qualitative and quantitative metabolomic investigation of single neurons by capillary electrophoresis electrospray ionization mass spectrometry." Nat Protoc **8**(4): 783-99.

Nemes, P. and A. Vertes (2007). "Laser ablation electrospray ionization for atmospheric pressure, in vivo, and imaging mass spectrometry." Anal Chem **79**(21): 8098-106.

Nemes, P. and A. Vertes (2010). "Atmospheric-pressure molecular imaging of biological tissues and biofilms by LAESI mass spectrometry." J Vis Exp(43).

Nesvizhskii, A. I., A. Keller, et al. (2003). "A statistical model for identifying proteins by tandem mass spectrometry." Anal Chem **75**(17): 4646-58.

Nikolov, M., C. Schmidt, et al. (2012). "Quantitative mass spectrometry-based proteomics: an overview." Methods Mol Biol **893**: 85-100.

Nipp, M., M. Elsner, et al. (2012). "S100-A10, thioredoxin, and S100-A6 as biomarkers of papillary thyroid carcinoma with lymph node metastasis identified by MALDI imaging." J Mol Med (Berl) **90**(2): 163-74.

Norris, J. L. and R. M. Caprioli (2013). "Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research." Chem Rev **113**(4): 2309-42.

Nygren, H. and P. Malmberg (2010). "High-resolution imaging and proteomics of peptide fragments by TOF-SIMS." Proteomics **10**(8): 1694-8.

Pappin, D. J., P. Hojrup, et al. (1993). "Rapid identification of proteins by peptide-mass fingerprinting." Curr Biol **3**(6): 327-32.

Pasilis, S. P., V. Kertesz, et al. (2008). "HPTLC/DESI-MS imaging of tryptic protein digests separated in two dimensions." J Mass Spectrom **43**(12): 1627-35.

Paul, W. (1990). "Electromagnetic traps for charged and neutral particles." Rev Mod Phys **62**(3): 531-40.

Penrod, N. M. and J. H. Moore (2014). "Influence networks based on coexpression improve drug target discovery for the development of novel cancer therapeutics." BMC Syst Biol **8**: 12.

Petyuk, V. A., W. J. Qian, et al. (2007). "Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry." Genome Res **17**(3): 328-36.

Picotti, P., M. Clement-Ziza, et al. (2013). "A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis." Nature **494**(7436): 266-70.

Pirman, D. A., R. F. Reich, et al. (2012). "Quantitative MALDI tandem mass spectrometric imaging of cocaine from brain tissue with a deuterated internal standard." Anal Chem **85**(2): 1081-9.

Pol, J., M. Strohalm, et al. (2010). "Molecular mass spectrometry imaging in biomedical and life science research." Histochem Cell Biol **134**(5): 423-43.

Porta, T., C. Grivet, et al. (2011). "Single hair cocaine consumption monitoring by mass spectrometric imaging." Anal Chem **83**(11): 4266-72.

Prideaux, B., V. Dartois, et al. (2011). "High-sensitivity MALDI-MRM-MS imaging of moxifloxacin distribution in tuberculosis-infected rabbit lungs and granulomatous lesions." Anal Chem **83**(6): 2112-8.

Prideaux, B., D. Staab, et al. (2010). "Applications of MALDI-MSI to pharmaceutical research." Methods Mol Biol **656**: 405-13.

Quanico, J., J. Franck, et al. (2013). "Development of liquid microjunction extraction strategy for improving protein identification from tissue sections
" J Proteomics **79**: 200-218.

Rajaei, M., A. Momeni, et al. (2013). "Effect of ejaculation on serum prostate specific antigen level in screening and non-screening population." J Res Med Sci **18**(5): 387-90.

Ramos, H., P. Shannon, et al. (2008). "The protein information and property explorer: an easy-to-use, rich-client web application for the management and functional analysis of proteomic data." Bioinformatics **24**(18): 2110-1.

Rao, W., A. D. Celiz, et al. (2013). "Ambient DESI and LESA-MS analysis of proteins adsorbed to a biomaterial surface using in-situ surface tryptic digestion." J Am Soc Mass Spectrom **24**(12): 1927-36.

Rappsilber, J., U. Ryder, et al. (2002). "Large-scale proteomic analysis of the human spliceosome." Genome Res **12**(8): 1231-45.

Ratcliffe, L. V., F. J. Rutten, et al. (2007). "Surface analysis under ambient conditions using plasma-assisted desorption/ionization mass spectrometry." Anal Chem **79**(16): 6094-101.

Rauser, S., C. Marquardt, et al. (2010). "Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry." J Proteome Res **9**(4): 1854-63.

Roddy, T. P., D. M. Cannon, Jr., et al. (2002). "Identification of cellular sections with imaging mass spectrometry following freeze fracture." Anal Chem **74**(16): 4020-6.

Rodriguez, A. S., B. H. Espina, et al. (2008). "Automated laser capture microdissection for tissue proteomics." Methods Mol Biol **441**: 71-90.

Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.

Rompp, A., S. Guenther, et al. (2010). "Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging." Angew Chem Int Ed Engl **49**(22): 3834-8.

Rompp, A. and B. Spengler (2013). "Mass spectrometry imaging with high resolution in mass and space." Histochem Cell Biol **139**(6): 759-83.

Rubakhin, S. S., J. C. Jurchen, et al. (2005). "Imaging mass spectrometry: fundamentals and applications to drug discovery." Drug Discov Today **10**(12): 823-37.

Salplachta, J., P. Rehulka, et al. (2004). "Identification of proteins by combination of size-exclusion chromatography with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and comparison of some desalting procedures for both intact proteins and their tryptic digests." J Mass Spectrom **39**(12): 1395-401.

Saman, D. M., A. M. Lemieux, et al. (2014). "A review of the current epidemiology and treatment options for prostate cancer." Dis Mon **60**(4): 150-4.

Sandin, M., J. Teleman, et al. (2014). "Data processing methods and quality control strategies for label-free LC-MS protein quantification." Biochim Biophys Acta **1844**(1 Pt A): 29-41.

Sarsby, J., M. W. Towers, et al. (2012). "Mass spectrometry imaging of glucosinolates in Arabidopsis flowers and siliques." Phytochemistry **77**: 110-8.

Schadt, S., S. Kallbach, et al. (2012). "Investigation of figopitant and its metabolites in rat tissue by combining whole-body autoradiography with liquid extraction surface analysis mass spectrometry." Drug Metab Dispos **40**(3): 419-25.

Schey, K. L., D. M. Anderson, et al. (2013). "Spatially-directed protein identification from tissue sections by top-down LC-MS/MS with electron transfer dissociation." Anal Chem **85**(14): 6767-74.

Schiewer, M. J., J. F. Goodwin, et al. (2012). "Dual roles of PARP-1 promote cancer growth and progression." Cancer Discov **2**(12): 1134-49.

Schmidt, A., N. Gehlenborg, et al. (2008). "An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures." Mol Cell Proteomics **7**(11): 2138-50.

Schober, Y., T. Schramm, et al. (2011). "Protein identification by accurate mass matrix-assisted laser desorption/ionization imaging of tryptic peptides." Rapid Commun Mass Spectrom **25**(17): 2475-83.

Schultz, G. A., T. N. Corso, et al. (2000). "A fully integrated monolithic microchip electrospray device for mass spectrometry." Anal Chem **72**(17): 4058-63.

Scicchitano, M. S., D. A. Dalmas, et al. (2009). "Protein extraction of formalin-fixed, paraffin-embedded tissue enables robust proteomic profiles by mass spectrometry." J Histochem Cytochem **57**(9): 849-60.

Scigelova, M., M. Hornshaw, et al. (2011). "Fourier transform mass spectrometry." Mol Cell Proteomics **10**(7): M111 009431.

Searle, B. C. (2010). "Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies." Proteomics **10**(6): 1265-9.

Seeley, E. H. and R. M. Caprioli (2008). "Molecular imaging of proteins in tissues by mass spectrometry." Proc Natl Acad Sci U S A **105**(47): 18126-31.

Sellami, L., O. Belgacem, et al. (2012). "In-source decay and pseudo tandem mass spectrometry fragmentation processes of entire high mass proteins on a hybrid vacuum matrix-assisted laser desorption ionization-quadrupole ion-trap time-of-flight mass spectrometer." Anal Chem **84**(12): 5180-5.

Shaheen, F. S., P. Znojek, et al. (2011). "Targeting the DNA double strand break repair machinery in prostate cancer." PLoS One **6**(5): e20311.

Shannon, P., A. Markiel, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res **13**(11): 2498-504.

Shanta, S. R., T. Y. Kim, et al. (2012). "A new combination MALDI matrix for small molecule analysis: application to imaging mass spectrometry for drugs and metabolites." Analyst **137**(24): 5757-62.

Shenoy, N., R. Kessel, et al. (2012). "Alterations in the ribosomal machinery in cancer and hematologic disorders." J Hematol Oncol **5**: 32.

Shiea, J., M. Z. Huang, et al. (2005). "Electrospray-assisted laser desorption/ionization mass spectrometry for direct ambient analysis of solids." Rapid Commun Mass Spectrom **19**(24): 3701-4.

Shrestha, B., P. Nemes, et al. (2010). "Direct analysis of lipids and small metabolites in mouse brain tissue by AP IR-MALDI and reactive LAESI mass spectrometry." Analyst **135**(4): 751-8.

Smith, D. F., D. P. Kilgour, et al. (2013). "Absorption mode FTICR mass spectrometry imaging." Anal Chem **85**(23): 11180-4.

Smith, D. F., A. Kiss, et al. (2013). "High mass accuracy and high mass resolving power FT-ICR secondary ion mass spectrometry for biological tissue imaging." Anal Bioanal Chem **405**(18): 6069-76.

Sole-Domenech, S., B. Johansson, et al. (2010). "Analysis of opioid and amyloid peptides using time-of-flight secondary ion mass spectrometry." Anal Chem **82**(5): 1964-74.

Solon, E. G., A. Schweitzer, et al. (2010). "Autoradiography, MALDI-MS, and SIMS-MS imaging in pharmaceutical discovery and development." AAPS J **12**(1): 11-26.

Spengler, B. (1997). "Post-source decay analysis in Matrix-Assisted Laser desorption/Ionization Mass Spectrometry of biomolecules." J Mass Spectrom **32**: 1019-36.

Spengler, B. (2004). "De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass determination by fourier transform ion cyclotron resonance mass spectrometry." J Am Soc Mass Spectrom **15**(5): 703-14.

Stahlman, M., C. S. Ejsing, et al. (2009). "High-throughput shotgun lipidomics by quadrupole time-of-flight mass spectrometry." J Chromatogr B Analyt Technol Biomed Life Sci **877**(26): 2664-72.

Stauber, J., R. Lemaire, et al. (2008). "MALDI imaging of formalin-fixed paraffin-embedded tissues: application to model animals of Parkinson disease for biomarker hunting." J Proteome Res **7**(3): 969-78.

Stauber, J., L. MacAleese, et al. (2010). "On-tissue protein identification and imaging by MALDI-ion mobility mass spectrometry." J Am Soc Mass Spectrom **21**(3): 338-47.

Stavridis, S., S. Saidi, et al. (2010). "Screening for prostate cancer: a controversy or fact." Hippokratia **14**(3): 170-5.

Stegemann, C., I. Drozdov, et al. (2011). "Comparative lipidomics profiling of human atherosclerotic plaques." Circ Cardiovasc Genet **4**(3): 232-42.

Stephan, C., B. Ralla, et al. (2014). "Prostate-specific antigen and other serum and urine markers in prostate cancer." Biochim Biophys Acta **1846**(1): 99-112.

Stingl, C., F. G. van Vilsteren, et al. (2011). "Reproducibility of protein identification of selected cell types in Barrett's esophagus analyzed by combining laser-capture microdissection and mass spectrometry." J Proteome Res **10**(1): 288-98.

Stoeckli, M., D. Staab, et al. (2002). "Molecular imaging of amyloid beta peptides in mouse brain sections using mass spectrometry." Anal Biochem **311**(1): 33-9.

Stojadinovic, A., J. A. Hooke, et al. (2007). "HYOU1/Orp150 expression in breast cancer." Med Sci Monit **13**(11): BR231-239.

Strobel, K., S. K. Haerle, et al. (2009). "Head and neck squamous cell carcinoma (HNSCC)--detection of synchronous primaries with (18)F-FDG-PET/CT." Eur J Nucl Med Mol Imaging **36**(6): 919-27.

Stumpo, K. A. and V. N. Reinhold (2010). "The N-glycome of human plasma." J Proteome Res **9**(9): 4823-30.

Suckau, D., A. Resemann, et al. (2003). "A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics." Anal Bioanal Chem **376**(7): 952-65.

Sugiura, Y. and M. Setou (2010). "Imaging mass spectrometry for visualization of drug and endogenous metabolite distribution: toward in situ pharmacometabolomes." J Neuroimmune Pharmacol **5**(1): 31-43.

Sugiura, Y., S. Shimma, et al. (2006). "Two-step matrix application technique to improve ionization efficiency for matrix-assisted laser desorption/ionization in imaging mass spectrometry." Anal Chem **78**(24): 8227-35.

Swan, A. L., A. Mobasheri, et al. "Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology." OMICS **17**(12): 595-610.

Syka, J. E., J. J. Coon, et al. (2004). "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry." Proc Natl Acad Sci U S A **101**(26): 9528-33.

Takai, N., Y. Tanaka, et al. (2012). "Quantitative analysis of pharmaceutical drug distribution in multiple organs by imaging mass spectrometry." Rapid Commun Mass Spectrom **26**(13): 1549-56.

Takats, Z., J. M. Wiseman, et al. (2005). "Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology." J Mass Spectrom **40**(10): 1261-75.

Takats, Z., J. M. Wiseman, et al. (2004). "Mass spectrometry sampling under ambient conditions with desorption electrospray ionization." Science **306**(5695): 471-3.

Tannu, N. S., J. Wu, et al. (2004). "Paraffin-wax-coated plates as matrix-assisted laser desorption/ionization sample support for high-throughput identification of proteins by peptide mass fingerprinting." Anal Biochem **327**(2): 222-32.

Teng, T., G. Thomas, et al. (2013). "Growth control and ribosomopathies." Curr Opin Genet Dev **23**(1): 63-71.

Thakur, D., T. Rejtar, et al. (2011). "Microproteomic analysis of 10,000 laser captured microdissected breast tumor cells using short-range sodium dodecyl sulfate-polyacrylamide gel electrophoresis and porous layer open tubular liquid chromatography tandem mass spectrometry." J Chromatogr A **1218**(45): 8168-74.

Theodosis, D. T. (2002). "Oxytocin-secreting neurons: A physiological model of morphological neuronal and glial plasticity in the adult hypothalamus." Front Neuroendocrinol **23**(1): 101-35.

Tomlins, S. A., M. A. Rubin, et al. (2006). "Integrative biology of prostate cancer progression." Annu Rev Pathol **1**: 243-71.

Trim, P. J., S. Francese, et al. (2009). "Imaging mass spectrometry for the assessment of drugs and metabolites in tissue." Bioanalysis **1**(2): 309-19.

Ulazzi, L., S. Sabbioni, et al. (2007). "Nidogen 1 and 2 gene promoters are aberrantly methylated in human gastrointestinal cancer." Mol Cancer **6**: 17.

Ummanni, R., F. Mundt, et al. (2011). "Identification of clinically relevant protein targets in prostate cancer with 2D-DIGE coupled mass spectrometry and systems biology network platform." PLoS One **6**(2): e16833.

Van Berkel, G. J., M. J. Ford, et al. (2006). "Evaluation of a surface-sampling probe electrospray mass spectrometry system for the analysis of surface-deposited and affinity-captured proteins." Rapid Commun Mass Spectrom **20**(7): 1144-52.

Van Berkel, G. J. and V. Kertesz (2009). "Application of a liquid extraction based sealing surface sampling probe for mass spectrometric analysis of dried blood spots and mouse whole-body thin tissue sections." Anal Chem **81**(21): 9146-52.

Van Berkel, G. J. and V. Kertesz (2013). "Continuous-flow liquid microjunction surface sampling probe connected on-line with high performance liquid chromatography/mass spectrometry for spatially resolved analysis of small molecules and proteins." Rapid Commun Mass Spectrom **27**: 1329-34.

Van Berkel, G. J., V. Kertesz, et al. (2008). "Liquid microjunction surface sampling probe electrospray mass spectrometry for detection of drugs and metabolites in thin tissue sections." J Mass Spectrom **43**(4): 500-8.

Van Berkel, G. J., S. P. Pasilis, et al. (2008). "Established and emerging atmospheric pressure surface sampling/ionization techniques for mass spectrometry." J Mass Spectrom **43**(9): 1161-80.

Van Berkel, G. J., A. D. Sanchez, et al. (2002). "Thin-layer chromatography and electrospray mass spectrometry coupled using a surface sampling probe." Anal Chem **74**(24): 6216-23.

Van Dyck, S., P. Flammang, et al. (2010). "Localization of secondary metabolites in marine invertebrates: contribution of MALDI MSI for the study of saponins in Cuvierian tubules of H. forskali." PLoS One **5**(11): e13923.

Vander Heiden, M. G., L. C. Cantley, et al. (2009). "Understanding the Warburg effect: the metabolic requirements of cell proliferation." Science **324**(5930): 1029-33.

Veloso, A., E. Astigarraga, et al. (2011). "Anatomical distribution of lipids in human brain cortex by imaging mass spectrometry." J Am Soc Mass Spectrom **22**(2): 329-38.

Venter, A., M. Nefliu, et al. (2008). "Ambient desorption ionization mass spectrometry." Trac-Trend Anal Chem **27**(4): 284-290.

Vismeh, R., D. J. Waldon, et al. (2012). "Localization and quantification of drugs in animal tissues by use of desorption electrospray ionization mass spectrometry imaging." Anal Chem **84**(12): 5439-45.

Wachs, T. and J. Henion (2001). "Electrospray device for coupling microscale separations and other miniaturized devices with electrospray mass spectrometry." Anal Chem **73**(3): 632-8.

Wachs, T. and J. Henion (2003). "A device for automated direct sampling and quantitation from solid-phase sorbent extraction cards by electrospray tandem mass spectrometry." Anal Chem **75**(7): 1769-75.

Walther, D. M. and M. Mann (2011). "Accurate quantification of more than 4000 mouse tissue proteins reveals minimal proteome changes during aging." Mol Cell Proteomics **10**(2): M110 004523.

Walworth, M. J., M. S. ElNaggar, et al. (2011). "Direct sampling and analysis from solid-phase extraction cards using an automated liquid extraction surface analysis nanoelectrospray mass spectrometry system." Rapid Commun Mass Spectrom **25**(17): 2389-96.

Walworth, M. J., J. J. Stankovich, et al. (2011). "Hydrophobic treatment enabling analysis of wettable surfaces using a liquid microjunction surface sampling probe/electrospray ionization-mass spectrometry system." Anal Chem **83**(2): 591-7.

Wang, N., M. Xu, et al. (2010). "Development of mass spectrometry-based shotgun method for proteome analysis of 500 to 5000 cancer cells." Anal Chem **82**(6): 2262-71.

Weckwerth, W. (2008). "Integration of metabolomics and proteomics in molecular plant physiology--coping with the complexity by data-dimensionality reduction." Physiol Plant **132**(2): 176-89.

Wei, H., K. Nolkrantz, et al. (2004). "Electrospray sample deposition for matrix-assisted laser desorption/ionization (MALDI) and atmospheric pressure MALDI mass spectrometry with attomole detection limits." Rapid Commun Mass Spectrom **18**(11): 1193-200.

Wilhelm, M., J. Schlegl, et al. (2014). "Mass-spectrometry-based draft of the human proteome." Nature **509**(7502): 582-7.

Wilkins, M. R., C. Pasquali, et al. (1996). "From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis." Biotechnology (N Y) **14**(1): 61-5.

Willems, S. M., A. van Remoortere, et al. (2010). "Imaging mass spectrometry of myxoid sarcomas identifies proteins and lipids specific to tumour type and grade, and reveals biochemical intratumour heterogeneity." J Pathol **222**(4): 400-9.

Wiseman, J. M., D. R. Ifa, et al. (2006). "Tissue imaging at atmospheric pressure using Desorption Electrospray Ionization (DESI) Mass Spectrometry." Angew Chem Int Ed Engl **45**: 7188-92.

Wisniewski, J. R., K. Dus, et al. (2012). "Proteomic workflow for analysis of archival formalin fixed and paraffin embedded clinical samples to a depth of 10,000 proteins." Proteomics Clin Appl.

Wisniewski, J. R., K. Dus, et al. (2013). "Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins." Proteomics Clin Appl **7**(3-4): 225-33.

Wisniewski, J. R., P. Ostasiewicz, et al. (2011). "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**(7): 3040-9.

Wisniewski, J. R., A. Zougman, et al. (2009). "Universal sample preparation method for proteome analysis." Nat Methods **6**(5): 359-62.

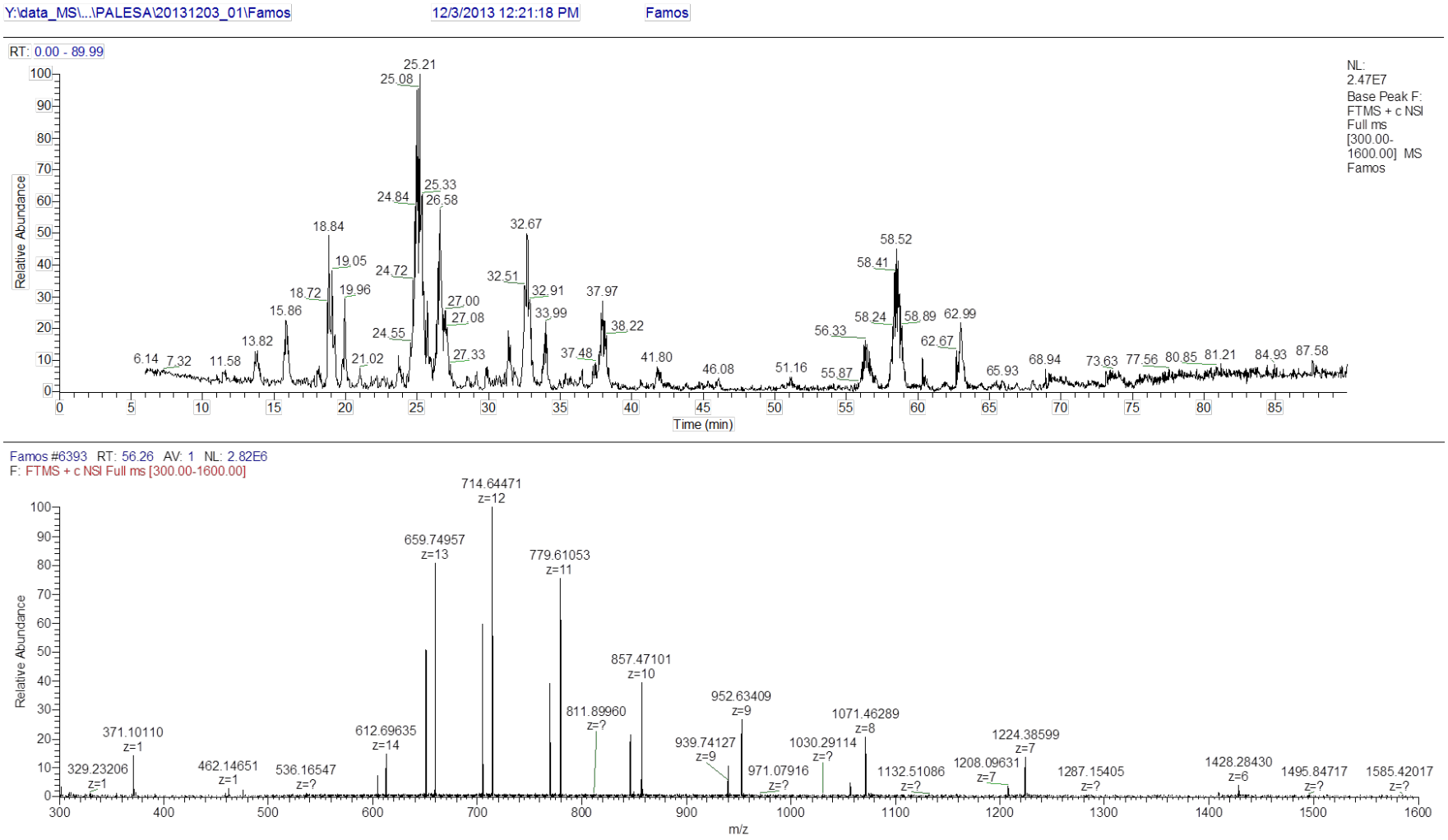
Wisztorski, M., B. Fatou, et al. (2013). "Microproteomics by liquid extraction surface analysis: application to FFPE tissue to study the fimbria region of tubo-ovarian cancer." Proteomics Clin Appl **7**(3-4): 234-40.

Wolf, A. M., R. C. Wender, et al. (2010). "American Cancer Society guideline for the early detection of prostate cancer: update 2010." CA Cancer J Clin **60**(2): 70-98.

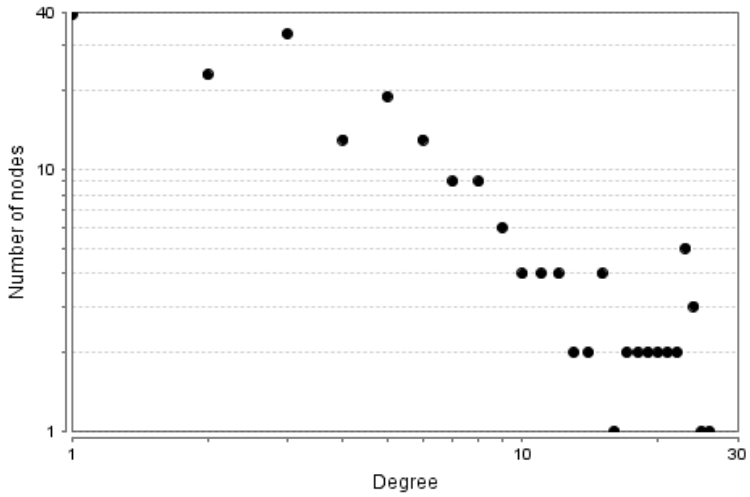
Wu, C., C. Orozco, et al. (2009). "BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources." Genome Biol **10**(11): R130.

Xia, N. and D. G. Castner (2003). "Preserving the structure of adsorbed protein films for time-of-flight secondary ion mass spectrometry analysis." J Biomed Mater Res A **67**(1): 179-90.
Xia, S. J., D. Cui, et al. (2012). "An overview of prostate diseases and their characteristics specific to Asian men." Asian J Androl **14**(3): 458-64.
Xu, X., C. Valtonen-Andre, et al. (2010). "Polymorphisms at the Microseminoprotein-beta locus associated with physiologic variation in beta-microseminoprotein and prostate-specific antigen levels." Cancer Epidemiol Biomarkers Prev **19**(8): 2035-42.
Yamashita, M. and J. B. Fenn (1984). "Electrospray ion source. Another variation on the free-jet theme." J Phys Chem **88**(20): 4451-59.
Yang, R., B. J. Daigle, Jr., et al. (2012). "Core module biomarker identification with network exploration for breast cancer metastasis." BMC Bioinformatics **13**: 12.
Yates, J. R., 3rd, S. Speicher, et al. (1993). "Peptide mass maps: a highly informative approach to protein identification." Anal Biochem **214**(2): 397-408.
Yates, J. R., C. I. Ruse, et al. (2009). "Proteomics by mass spectrometry: approaches, advances, and applications." Annu Rev Biomed Eng **11**: 49-79.
Ye, H., T. Greer, et al. (2011). "From pixel to voxel: a deeper view of biological tissue by 3D mass spectral imaging." Bioanalysis **3**(3): 313-32.
Yoshimura, Y., N. Zaima, et al. (2012). "Different localization patterns of anthocyanin species in the pericarp of black rice revealed by imaging mass spectrometry." PLoS One **7**(2): e31285.
Yost, R. A. and R. K. Boyd (1990). "Tandem mass spectrometry: quadrupole and hybrid instruments." Methods Enzymol **193**: 154-200.
Yost, R. A. and C. G. Enke (1977). "Selected ion fragmentation with a tandem quadrupole mass spectrometer." J Am Chem Soc **100**(7): 2274-75.
Zenobi, R. (2013). "Single-cell metabolomics: analytical and biological perspectives." Science **342**(6163): 1243259.
Zhang, S., C. K. Van Pelt, et al. (2003). "Quantitative determination of noncovalent binding interactions using automated nanoelectrospray mass spectrometry." Anal Chem **75**(13): 3010-8.
Zhang, Y., Z. Yuan, et al. (2011). "Coupling of liquid chromatography with mass spectrometry by desorption electrospray ionization (DESI)." Chem Commun (Camb) **47**(14): 4171-3.
Zhao, Y. Y. and R. C. Lin "UPLC-MS(E) application in disease biomarker discovery: the discoveries in proteomics to metabolomics." Chem Biol Interact **215**: 7-16.
Zhu, W., J. W. Smith, et al. (2009). "Mass spectrometry-based label-free quantitative proteomics." J Biomed Biotechnol **2010**: 840518.
Ziemiecki, A., R. G. Muller, et al. (1990). "Oncogenic activation of the human trk proto-oncogene by recombination with the ribosomal large subunit protein L7a." EMBO J **9**(1): 191-6.
Zimmerman, T. A., D. Debois, et al. (2011). "An analytical pipeline for MALDI in-source decay mass spectrometry imaging." Anal Chem **83**(15): 6090-7.
Zubarev, R. A., N. L. Kelleher, et al. (1998). "Electron capture dissociation of multiply charged protein cations. A nonergodic process." J Am Chem Soc **120**: 3265-66.
Zubarev, R. A. and A. Makarov (2013). "Orbitrap mass spectrometry." Anal Chem **85**(11): 5288-96.
Zybailov, B. L., L. Florens, et al. (2007). "Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors." Mol Biosyst **3**(5): 354-60.

ANNEX


Y:\data_MS\...\PALESA\20131203_01\Famos
12/3/2013 12:21:18 PM
Famos
RT: 0.00 - 89.99
NL:
2.47E7
Base Peak F:
FTMS + c NSI
Full ms
[300.00-
1600.00] MS
Famos
Relative Abundance
Time (min)
Famos #6393 RT: 56.26 AV: 1 NL: 2.82E6
F: FTMS + c NSI Full ms [300.00-1600.00]
714.64471
z=12
659.74957
z=13
779.61053
z=11
857.47101
z=10
952.63409
z=9
811.89960
z=?
1071.46289
z=8
612.69635
z=14
371.10110
z=1
939.74127
z=9
1030.29114
z=?
1224.38599
z=7
329.23206
z=1
462.14651
z=1
536.16547
z=?
971.07916
z=?
1132.51086
z=?
1208.09631
z=7
1287.15405
z=?
1428.28430
z=6
1495.84717
z=?
1585.42017
z=?
Relative Abundance
m/z

Supplementary Figure 1. Undigested proteins appearing at the full MS profile of the peak at retention time = 56.26 min of a 1.5-h chromatographic gradient run of an extract obtained using the FAMOS set-up, indicating incomplete trypsin digestion. Top shows the base peak chromatogram. Bottom shows the full ms, showing highly charged peaks, whose monoisotopic mass corresponds suspiciously to that of ubiquitin.



Supplementary Figure 2. Nodes sorted according to degree. Five hubs can be identified as they have the highest degree value in the entire network.

Function	Groups	Group Genes
Formation of Fibrin Clot (Clotting Cascade)	Group0	A2M, APOH, C1QBP, CAV1, F2, FGA, FGB, FGG, HRG, KNG1, PLG, SERPINC1,VTN
Cytoplasmic Ribosomal Proteins	Group1	APEX1, CALR, CANX, CAPG, CCT2, CCT8, EEF1G, EEF2, EIF4A1, GDI1, GNB2L1, HSP90AA1, LAMA5, MYO1C, PABPC1, PPP2CA, PPP2R1A, PRDX3, RAN, RPL12, RPL13, RPL14, RPL15, RPL18, RPL18A, RPL23, RPL23A, RPL31, RPL6, RPL7A, RPL8, RPLP2, RPN1, RPS12, RPS13, RPS18, RPS19, RPS27A, RPS4X, RPS8, RPSA, SEPT9, SORBS1, YWHAZ
Complement and coagulation cascades	Group2	A2M, ANXA1, APOA1, APOA2, APOH, C1QBP, C3, C4A, C4B, CALR, CAV1, CFB, CFH, CLU, COL18A1, F2, FBN1, FGA, FGB, FGG, FN1, GNAI2, GSTP1, HNRNPC, HRG, HSP90AA1, HSPG2, HYOU1, KNG1, LAMP2, LTF, PLG, PSAP, SERPINA1, SERPINC1, TLN1, TNC, VCL, VTN
Laminin interactions	Group3	COL18A1, COL3A1, COL6A2, DCN, FN1, HSPG2, LAMA4, LAMA5, LAMB2, LAMC1, NID1, NID2, TNC, VTN
regulation of nuclease activity	Group4	CALR, HYOU1, LMNA, NPM1, PDIA6, TLN1, TPP1, VCP
cytoplasmic membrane-bounded vesicle lumen	Group5	A2M, ANXA1, APOA1, APOA2, APOH, C1QBP, C3, C4A, C4B, CALR, CAV1, CFB, CFH, CLU, COL18A1, F2, FBN1, FGA, FGB, FGG, FN1, GNAI2, GSTP1, HNRNPC, HRG, HSP90AA1, HSPG2, HYOU1, KNG1, LAMP2, LTF, PLG, PSAP, SERPINA1, SERPINC1, TLN1, TNC, VCL, VTN
Valine, leucine and isoleucine degradation	Group6	ACAT1, ALDH6A1, ALDH7A1, DECR1, ECHS1, HADHA, MAOB, MCCC2, TPI1
Glucose metabolism	Group7	AKR1A1, ALDH7A1, FBP1, GOT2, GPI, MDH2, PPP2CA, PPP2R1A, RPS27A, TPI1

Supplementary Table 1. Top 8 groupings observed after GO pathway analysis using ClueGO.