

Molecular Characterization of Metastatic Endometrial Cancer by Mass Spectrometry



A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

By

Parul Mittal

B.Sc. (Biotechnology), M.Sc. (Biotechnology)

Adelaide Proteomics Centre
School of Biological Sciences
The University of Adelaide, South Australia
August 2017

This page intentionally left blank

Tables of Contents

Declaration	I
Acknowledgement	III
List of publications	V
Abbreviations	VI
Abstract	VII
Chapter 1: Introduction	1
1.1 Endometrial Cancer	3
1.2 Histopathological Diagnosis	5
1.2.1 Histological Subtypes	5
1.2.2 Tumour Grade	5
1.2.3 Staging	6
1.3 Genetic Abnormalities in Endometrial Cancer	8
1.4 Molecular Classification of Endometrial Cancer	9
1.5 Diagnostic, Prognostic and Predictive Biomarkers	10
1.5.1 Diagnostic Biomarkers	10
1.5.2 Prognostic and Predictive Biomarkers	12
1.6 Current Status of Molecular Biomarkers	15
1.7 Treatments for Endometrial Cancer	18
1.8 Metastasis and its Gene Protein Signature in Solid Tumours	19
1.9 Aims of the Thesis	20

Chapter 2: Published review article: Proteomics of Endometrial Cancer Diagnosis, Treatment and Prognosis	21
2.1 Summary	23
2.2 Statement of Authorship	25
2.3 Manuscript	27
Chapter 3: Materials and Methods	41
3.1 Novel Approaches for Biomarker Discovery in Tissue	43
3.1.1 Published Book Chapter: Applications of Mass Spectrometry Imaging to Cancer	45
3.1.1.1 Summary	47
3.1.1.2 Statement of Authorship	49
3.1.1.3 Manuscript	51
3.1.2 Data Dependent Acquisition/ Data Independent Acquisition	93
3.1.3 Immunohistochemistry	95
3.2 Tissue Microarray Design and Construction	96
Chapter 4: Published Research Article: Lymph Node Metastasis of Primary Endometrial Cancers: Associated Proteins Revealed by MALDI Imaging	99
4.1 Summary	101
4.2 Statement of Authorship	103
4.3 Manuscript	107
4.4 Supplementary File	115

Chapter 5: Published Research Article: Annexin A2 and Alpha Actinin 4 Expression Correlates with Metastatic Potential of Primary Endometrial Cancer	125
5.1 Summary	127
5.2 Statement of Authorship	131
5.3 Manuscript	135
5.4 Supplementary File	147
Chapter 6: Unpublished Work in Manuscript Format: Role of <i>N</i> -linked Glycans in Endometrial Cancer	149
6.1 Statement of Authorship	151
6.2 Manuscript	153
Chapter 7: Discussion	169
7.1 Main Findings and Other Considerations	171
7.2 Clinical Applications of MALDI-TOF	176
7.3 Role of Extracellular Matrix in Endometrial Cancer	178
7.4 Conclusion and Future Perspectives	179
7.5 Overall Summary	181
Appendix I: Classification of MALDI-MS Imaging Data of Tissue Microarrays using Canonical Correlation Analysis-Based Variable Selection	183
Appendix II: MALDI Mass Spectrometry Imaging Reveals Decreased CK5 Levels in Vulvar Squamous Cell Carcinomas Compared to the Precursor Lesion Differentiated Vulvar Intraepithelial Neoplasia	199
References	217

List of Tables and Figures

Figure 1: Schematic illustration of the female reproductive system

Figure 2: Schematic representation of different stages of endometrial cancer

Figure 3: Haematoxylin and eosin stained annotated tumour cores

Figure 4: Receiver operating characteristic (ROC) analysis of the different logistic regression models

Figure 5: Receiver operating characteristic (ROC) curves for the prediction of lymph node metastasis by several logistic regression models

Figure 6: Annotated haematoxylin and eosin stained endometrial cancer formalin fixed paraffin embedded tissue section

Figure 7: Tandem mass spectrometry fragmentation spectra

Figure 8: High mannose glycans are elevated in endometrial cancer

Figure 9: Hybrid/complex *N*-glycan structures predominately observed in the adjacent normal region compared to the tumour region

Table 1: *N*-glycan structures assigned by PGC-LC-ESI-ion trap MS/MS

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968. The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Parul Mittal

August 2017

This page intentionally left blank

Acknowledgements

Many people have helped me throughout my PhD. Some through hands-on help in the lab, some through inspiration and motivation and some by simply taking my mind off things for a while. Without them I would never have been able to finish my PhD. Some may go unmentioned here, but are nonetheless appreciated.

First and foremost, I would like to express my sincere gratitude to my principal supervisor, Prof Peter Hoffmann, for his guidance and advice during these years. He is highly thanked for giving me the opportunity to start my scientific career and do my PhD in his group. I am most grateful to my co supervisor, Dr Manuela Klingler-Hoffmann, for always being positive and seeing the possibilities instead of the obstacles. I admire you both as professional and accomplished role models for women in science. I would never have enjoyed my PhD so much without you. I owe my warmest thanks to my second co supervisor Prof Martin K Oehler, for providing patient samples and guidance during this project. Working with you has been very educational and a real privilege.

I'm extremely grateful to Georgia Arentz for sharing her knowledge, expertise and guidance during this project. I would like to thank Gurjeet Kaur, for her assistance annotating the tumour tissue sections. Chris Cursaro deserves my most sincere thanks for his invaluable technical support. I would also like to thank all the past and present members of the Adelaide Proteomic Centre, for their companionship and willingness to lend an ear when things weren't the best and for their words of encouragement. I wish to thank all my dear friends outside the world of science from the bottom of my heart for being there for me whenever needed. You all have been invaluable in taking my mind off work and putting things into perspective.

I would like to thank all the co-authors of the articles included in my thesis; Lyndal, James, Yee Leung, Colin, and Jonathan for their collaboration, their critical comments that subsequently improved the various papers. Special thanks to the patients and their families for their voluntary participation. Without them, these studies would never have been possible.

My deepest gratitude belongs to my grandparents who raised me with the knowledge that anything is possible if you work hard enough. I would like to thank my dear parents, brother and sister in law, who always encouraged me to try and achieve whatever I wanted and were interested not only in the destination but even much more in the journey. Thank you for raising

me with that attitude. My beautiful niece, Arya; your smiles enlighten my days. Special thanks to my in-laws, for your never ending support, interest and understanding during all these years.

Finally, I owe my dearest gratitude to my husband. You mean the world to me and this thesis would not have been possible without your love and belief in me. Thank you for all your listening, troubleshooting, interest, and most importantly allowing me to pursue my own personal goals. You have taught me patience that really was needed many times in the completion of this thesis.

Thank you all, it is because of you that I have made it this far.

Parul

List of Publications

This thesis is based on the following publications, which are referred to in text and attached as appendices:

- I. **Parul Mittal**, Manuela Klingler-Hoffmann, Georgia Arentz, Chao Zhang, Gurjeet Kaur, Martin K. Oehler, Peter Hoffmann, Proteomics of endometrial cancer diagnosis, treatment, and prognosis, *Proteomics Clinical Applications*, 2016; 217-229
- II. Georgia Arentz, **Parul Mittal**, Chao Zhang, Ying-Ying Ho, Matthew Briggs, Lyron Winderbaum, Manuela Klingler-Hoffmann, Peter Hoffmann, Chapter 2 - Application of Mass Spectrometry Imaging to cancer, in: Richard, R. D., Liam, A. M. (Eds.), *Advances in Cancer Research*, Academic Press 2017, Volume 134, 1st Edition, 27-66
- III. **Parul Mittal**, Manuela Klingler-Hoffmann, Georgia Arentz, Lyron Winderbaum, Noor Alia Lokman, Chao Zhang, Lyndal Anderson, James Scurry, Yee Leung, Colin JR Stewart, Jonathan Carter, Gurjeet Kaur, Martin K. Oehler, Peter Hoffmann, Lymph node metastasis of primary endometrial cancers: Associated proteins revealed by MALDI imaging, *Proteomics*, 2016, 16, 1793-1801
- IV. **Parul Mittal**, Manuela Klingler-Hoffmann, Georgia Arentz, Lyron Winderbaum, Gurjeet Kaur, Lyndal Anderson, James Scurry, Yee Leung, Colin JR Stewart, Jonathan Carter, Peter Hoffmann, Martin K. Oehler, Annexin A2 and alpha actinin 4 expression correlates with metastatic potential of primary endometrial cancer, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1865, 2017, 846-857
- V. Lyron Winderbaum, Inge Koch, **Parul Mittal**, Peter Hoffmann, Classification of MALDI-MS imaging data of tissue microarrays using Canonical Correlation Analysis based variable selection, *Proteomics*, 2016, 16, 1731-1735
- VI. Chao Zhang, Georgia Arentz, Lyron Winderbaum, Noor A Lokman, Manuela Klingler-Hoffmann, **Parul Mittal**, Christopher Carter, Martin K. Oehler, Peter Hoffmann, MALDI Mass Spectrometry Imaging reveals decreased CK5. Levels in vulvar squamous cell carcinomas compared to the precursor lesion differentiated vulvar intraepithelial neoplasia, *IJMS*, 2016, 17; 1088-1099

Abbreviations

α -Actin-2	Alpha-actin-2
ACTN4	Alpha-actinin-4
ANXA1	Annexin A1
ANXA1	Annexin A2
CHCA	α -cyano-4-hydroxycinnamic acid
CCA	Canonical Correlation Analysis
Da	Dalton
DDA	Data dependent analysis
DIA	Data independent analysis
EC	Endometrial cancer
ECM	Extracellular matrix
ESI	Electro-spray ionization
FFPE	Formalin fixed paraffin embedded
FIGO	International federation of gynaecology and obstetrics
H&E	Haematoxylin and eosin
HIAR	Heat induced citric acid antigen retrieval
IHC	Immunohistochemistry
ITO	Indium tin oxide
LC	Liquid chromatography
LMD	Laser capture micro-dissection
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LNM	Lymph Node Metastasis
MALDI	Matrix assisted laser desorption/ionization
MALDI MSI	Matrix assisted laser desorption/ionization mass spectrometry imaging
μ L	Microlitre
mL	Millilitre
mm	Millimetre
MS	Mass spectrometry
m/z	Mass to charge ratio
PEN	Polyethylene Naphthalate
PNGase F	peptide- <i>N</i> -glycosidase F
ppm	Parts per million
QTOF	Quadrupole time of flight
TMA	Tissue micro array

Abstract

One of the most reliable prognostic factors in endometrial cancer is the presence of lymph node metastasis. Clinicians presently face the challenge that radiological imaging and conventional surgical-pathological variables such as tumour size, depth of invasion and grade of disease are unreliable in determining if the endometrial cancer has metastasized. Although only 10% of endometrial cancer patients suffer from lymph node metastasis, the majority of them undergo lymphadenectomy, which can be associated with significant complications including lower extremity lymphedema. Based on the assumption that metastasis is mainly determined by the properties of the primary tumour and its interaction with the surrounding tissues, a tissue based proteomic approach combining two complementary methods, peptide matrix assisted laser desorption/ionisation mass spectrometry imaging (MALDI MSI) and liquid chromatography-tandem mass spectrometry (LC-MS/MS) was undertaken to identify molecular discriminators in primary endometrial cancers which correlate with lymph node metastasis.

In a discovery approach, MALDI MSI was carried out on two tissue micro arrays (TMA), containing a total of 43 patients. Upon data acquisition, a canonical correlation analysis (CCA) based method was applied to rank the acquired m/z values based on their power to discriminate the primary carcinomas with and without metastatic potential. The highly ranked m/z values were able to classify 38 out of 43 patients (88.4%) correctly. The top discriminative m/z values were identified using a combination of *in situ* sequencing and LC-MS/MS from digested tumour samples. The differential abundance of the two identified proteins, plectin and α -Actin-2 was further validated using data independent acquisition LC-MS/MS and immunohistochemistry.

In a targeted approach, we aimed to improve the prediction model for endometrial cancer metastasis preoperatively. From publically available data and published research, we compiled a list of 60 target proteins with the potential to display differential abundance between primary endometrial cancers with lymph node metastasis versus those without. Using data dependent acquisition LC-MS/MS, we were able to detect 23 of these proteins in an independent cohort of endometrial cancer patients. Using data independent acquisition LC-MS/MS, the differential abundance of 5 of those proteins was observed ($p < 0.05$). Upon validation by immunohistochemistry, our data indicates that annexin A2 is upregulated while annexin A1 and alpha actinin 4 were downregulated in primary endometrial cancers with lymph node metastasis versus those without. The results of this immunohistochemistry analysis were used

to generate a predictive model of endometrial cancer metastasis. Additionally the predictive model using highly ranked m/z values identified by MALDI MSI was generated and compared with other models containing the histopathological variables. However, when compared the MALDI MSI model showed significantly higher predictive accuracy than the model using immunohistochemistry data.

Our results showed that the highly ranked m/z values identified from MALDI MSI data serve as new independent prognostic information beyond the established risk factors. The developed molecular classification tool has the potential to predict which tumours have metastasized and which patients would therefore benefit from radical surgery while avoiding those who will not benefit from it and consequently decreasing the risk of post-surgical morbidity. In conclusion, these findings demonstrate a successful application of MALDI MSI for the identification of protein biomarkers of endometrial cancer metastasis.

INTRODUCTION

This page intentionally left blank

1 Introduction

Endometrial cancer is one of the leading causes of morbidity and mortality among women worldwide. Despite the disease being diagnosed at an early stage, the major challenge faced by clinicians is how to best predict the clinical course of disease and evaluate patient outcomes so that the most appropriate treatment can be identified. Predictive tissue markers for lymph node metastasis are therefore essential to determine the optimal treatment strategy and to reduce morbidity in endometrial cancer patients. Such molecular markers could provide insight into the tumour biology and the process that leads to metastasis and may immediately serve as diagnostic markers to guide surgeons. Inadequacies in our current system of endometrial cancer classification and risk stratification have prompted a call to action [1, 2], to identify new biologically informative tools.

1.1 Endometrial Cancer

Endometrial cancers (ECs) are the most common gynaecological malignancies in the developed countries and are the fourth most common cancer in women overall [3], increasing globally in both incidence and mortality [4]. Each year, EC develops in approximately 142,000 women worldwide, with an estimated 42,000 deaths from this cancer [5]. It is also one of the most common causes of death from gynaecological malignancy among women in Australia. It affects approximately 1/75 Australian women by the age of 75, resulting in around 1700 new cases and 230 deaths from the disease every year [6].

The uterus is the primary female reproductive organ with its major role being to nourish the developing foetus. The uterus is shaped like an upside-down pear (refer to Figure 1); the superior is called fundus, the middle is corpus, and inferior is the cervix. The wall of the uterus consists of three layers of tissue; the inner lining of the uterus is called endometrium. The middle thick layer of the uterus is the myometrium and is made of specialised muscle tissue. The outer layer which is covering the uterus is called the serosa.

During the reproductive years, the endometrium changes throughout the menstrual cycle in response to hormones produced by the ovaries. In normal menstrual cycles, the endometrium is shed with each menstrual period. However, in most cases of EC, the endometrium has not discharged and remains increased in thickness. This stage is called endometrial hyperplasia (pre-cancerous stage), where the cancerous cells can grow quickly and out of control. When EC spreads, cancer cells leave the uterus (primary tumour site) and get lodged in lymph nodes

(metastasis) primarily in the pelvis near the uterus. The cancerous cells can further spread to the para-aortic lymph nodes, which lie in front of the lumbar vertebrae near the aorta. At the time of diagnosis of patients with EC, 6–8% will have lymph node metastasis (LNM), 4–6% will have pelvic lymph node metastasis (with or without para-aortic) and only 2% will have para-aortic lymph node metastasis [7-9].

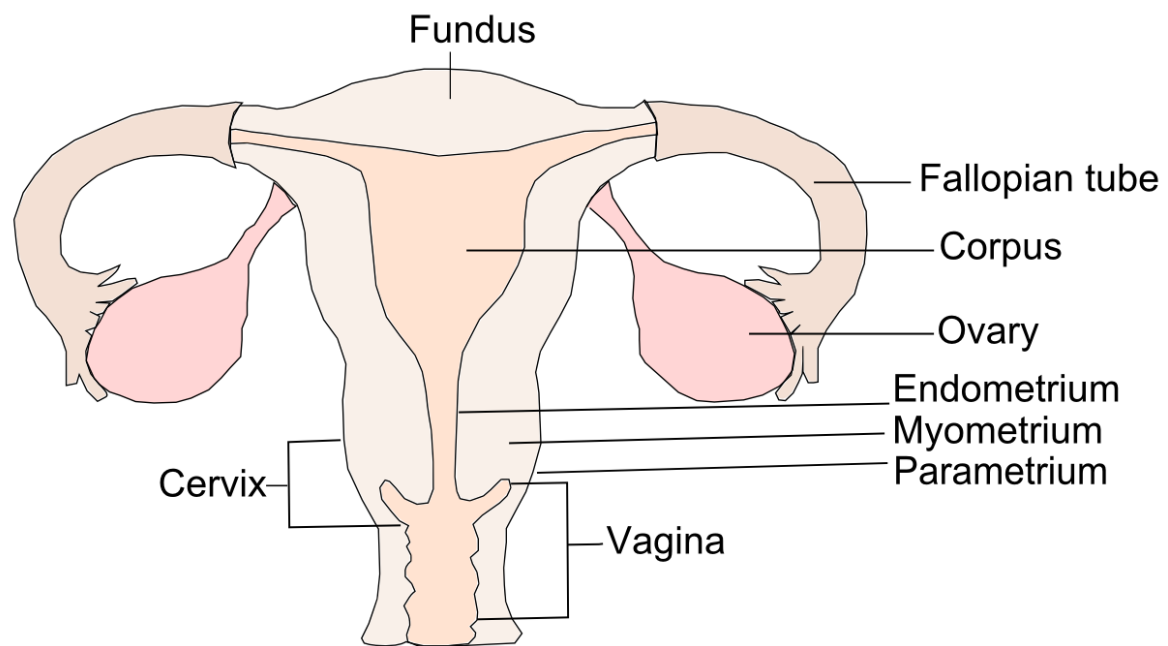


Figure1: Schematic illustration of the female reproductive system

1.2 Histopathological Diagnosis

1.2.1 Histological Subtypes

Prior to the 1980s, EC was broadly characterized as a single disease. Based on the specific molecular features and different gene expression profiles, Bokhman [10] first described two main clinicopathological variants of EC: type I and type II. Type I EC, commonly referred to as the endometrioid type, comprises 80-90% of all EC [11]. They are usually less aggressive, well differentiated tumours that develop in peri-menopausal women, arise from glandular cells and are associated with having an increased ratio of hormone estrogen compared to the hormone progesterone in the body. Risk factors include obesity and atypical endometrial hyperplasia. Due to the presence of early symptoms, the vast majority of type I carcinomas are diagnosed when the tumour is still confined to the uterus and patients are cured by surgery alone [12]. Type II EC are non-endometrioid carcinomas accounting for 10-20% of all EC [11]. These are very aggressive, poorly differentiated, and estrogen-independent tumours that generally arise in postmenopausal women. Type II EC includes serous carcinomas, clear cell carcinomas and carcinosarcomas. These tumours are often diagnosed at late stage with poor prognosis and surgery is commonly followed by adjuvant therapy. Women with such tumours are at high risk of relapse and metastatic disease.

1.2.2 Tumour Grade

Tumour grading depends on the extent of abnormalities of tumour cells and tumour tissue organization as viewed under a microscope. The grade of EC is evaluated by solid tumour growth (the ratio of cancerous tissue to normal tissue) and cancer cell differentiation. Based on this, type I adenocarcinomas are divided into three grades, 1 to 3, depending on how closely they resemble the phenotype of normal endometrium. If the cancer cells look very similar to normal endometrium and consist of well-formed glands, with no more than 5% solid non-squamous areas they are graded as grade 1, well differentiated and less aggressive. Grade 2 are moderately differentiated, consisting of 6-50% solid non-squamous areas. Grade 3 are poorly differentiated adenocarcinomas, where the cells look very abnormal with more than 50% solid non-squamous areas and are aggressive, grow and multiply quickly and are considered as high grade.

1.2.3 Staging

Stage refers to the extent of the disease, such as how much the tumour has grown and whether it has grown to other nearby structures such as the cervix, bladder or rectum; whether the cancer has spread to local/regional lymph nodes, or it has metastasized to other parts of the body. Historically, EC was staged only by physical examination and ultra-sound of the uterus. In 1988, the EC staging system changed from a clinical to a surgical system, and was revised in 2009 by the International Federation of Gynaecology and Obstetrics (FIGO). Surgical staging usually involves removal of the uterus, ovaries, fallopian tubes, adjacent lymph nodes, and part of the vagina.

According to the new revised 2009 FIGO system [13], EC is categorised into four stages: In stage I, cancer is found in the uterus only (refer to Figure 2a). Based on how far the cancer has spread, stage I is divided into three substages IA, IB and IC. (1) Stage IA, the tumour confined to the inner layer of cells of the uterus (endometrium). (2) Stage IB, tumour invades less than one half of the muscle wall of the uterus (myometrium). (3) Stage IC, tumour invades more than one half of the myometrium. In stage II, cancer has spread into connective tissue of the cervix, but has not spread outside the uterus (refer to Figure 2b). In stage III, cancer has spread beyond the uterus and cervix, but has not spread beyond the pelvis (refer to Figure 2c). Stage III is again divided into three substages IIIA, IIIB, and IIIC. (1) Stage IIIA, tumour spreads to the outer layer of the uterus and/or to the fallopian tubes, ovaries, and ligaments of the uterus. (2) Stage IIIB, where tumour spreads to the vagina and/or to the parametrium (connective tissue and fat next to the uterus). (3) Stage IIIC, tumour spreads to lymph nodes in the pelvis (IIIC1) and/or the aorta (IIIC2). In stage IV, tumour spreads beyond the pelvis into the bladder and/or bowel wall (stage IVA) and/or the abdomen and/or lymph nodes in the groin (stage IVB), (refer to Figure 2d).

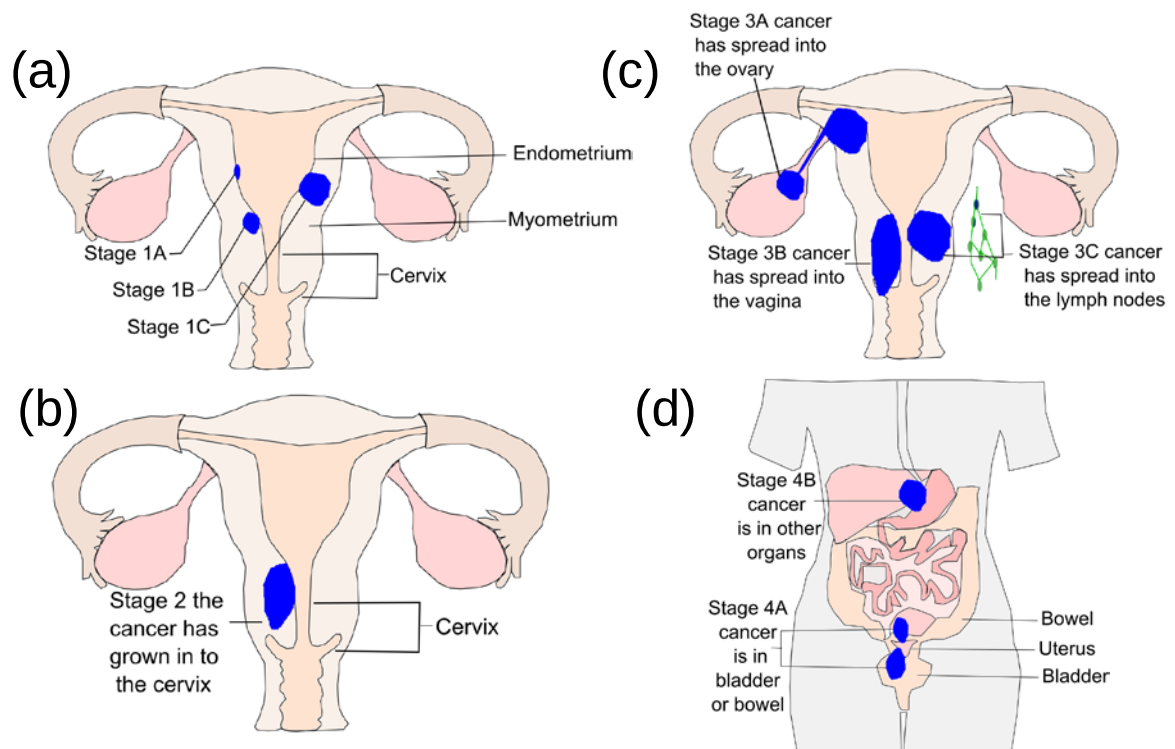


Figure 2: Schematic representation of different stages of endometrial cancer

1.3 Genetic Abnormalities in Endometrial Cancer

Over the past two or three decades, our understanding of the genetic alterations of endometrial cancer has increased considerably (reviewed in [14]). The recent development and implementation of next generation sequencing to comprehensively search for somatic mutations has resulted in a rapid and significant shift in our understanding of the molecular events underlying development of EC [15]. Complementary DNA (cDNA) analysis clearly demonstrated that the molecular genetic alterations involved in the development of type I differ from those of type II EC [16, 17]. The main molecular features of type I EC are microsatellite instability (MI) and mutations in the phosphatase and tensin homologue deleted on chromosome 10 (PTEN), phosphatidylinositol 3-kinase catalytic subunit (PIK3CA), Kirsten rat sarcoma viral oncogene homolog (K-ras) and β -catenin genes. Although the clinical and prognostic relevance of each of these alterations has not been fully elucidated, there is evidence showing that MI, and mutations in the PTEN and β -catenin genes might be associated with a favourable outcome. In contrast the main features of type II EC are p53 mutations, loss of heterozygosity (LOH) on several chromosomes, as well as other molecular alterations such as inactivation of p16 and E-cadherin, c-erbB2 amplification, cyclins D1 and E amplification, and alterations in genes involved in the regulation of the mitotic spindle checkpoint (STK-15) [18, 19]. Exome sequencing analyses show that the genes most frequently mutated in type I are PTEN (77%), PIK3CA (53%), PIK3R1 (37%), CTNNB1 (36%), ARID1A (35%), K-ras (24%), CTCF (20%), RPL22 (12%), p53 (11%), FGFR2 (11%), and ARID5B (11%). In contrast, the genes most frequently mutated in type II are p53 (90.7%), PIK3CA (41.9%), PPP2R1A (36.6%), FBXW7 (30.2%), CHD4 (16.3%), CSMD3 (11.6%), and COLA 11 (11.6%) [20]. However, it is important to notice that there is some degree of overlapping such as p53 which is mutated in > 90% of type II, but also in 11% of type I. A number of molecularly targeted therapeutics, assessing the potential clinical utility of these findings are in clinical trials. Among these agents are the mTOR inhibitors temsirolimus (CC1-779), everolimus (RAD001), and ridaforolimus (AP23573), however, there is currently no FDA-approved targeted therapy for EC (reviewed in [14]).

1.4 Molecular Classification of Endometrial Cancer

Genomic analysis of endometrial tumours has shown that approximately one-fifth of tumours that were classified as type I by histology are type II at the molecular level, revealing the limitations of this approach [21]. However, unambiguously classifying tumours is important because different histological subtypes have different clinical behaviours and concurrent treatment considerations (reviewed in [22]). Molecular classification of ECs is expanding our understanding and appears to correlate well with clinical outcome data, representing an important step forward [23]. In 2013, a new system for classification based on the molecular categories was identified by The Cancer Genome Atlas (TCGA), providing new insights into EC subtypes [20]. TCGA is a publicly funded project aiming to catalogue and discover major cancer-causing genome alterations in large cohorts of over 30 human tumours through large-scale genome sequencing and integrated multi-dimensional analyses [24]. Based on an integrated analysis of somatic mutation rates, frequency of copy number alterations, and microsatellite instability status in 373 EC patients, TCGA identified four distinct molecular subgroups of EC, which are defined as (i) DNA polymerase epsilon (POLE) ultramutated (7%), (ii) microsatellite instable hypermutated (28%), (iii) copy-number low (39%), and (iv) copy-number high (serous-like and mostly p53 mutant) (26%) [25]. TCGA suggested that EC patients harbouring POLE mutations had a less aggressive clinical course and improved progression-free survival when compared with the three other groups identified [20]. In the near future, one could envision that the molecular classification of EC might assist clinicians in guiding an optimal treatment for EC patients. However, sensitivity and specificity of the identified biomarkers remain to be validated before this can become reality. Additionally, future studies are required to identify specific molecular alterations which determine tumour behaviour and metastatic potential which can improve risk classification, treatment decisions, and identify targetable pathways in EC.

1.5 Diagnostic, Prognostic and Predictive Biomarkers

Biomarkers are indicators of relevant biological conditions and are defined as a “characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [26]. Accordingly, clinical and histological features of the cancer as well as the imaging and molecular alterations measurable in biological samples can all be regarded as biomarkers. Depending on their use, we can distinguish different types of biomarkers: diagnostic, prognostic and predictive biomarkers. Diagnostic markers are used to identify disease at early stage; prognostic biomarkers will indicate about the disease outcome and the survival of the patient independent of treatment; predictive biomarkers are related to the expected response of a patient to a treatment.

1.5.1 Diagnostic Biomarkers

Post-menopausal bleeding is an early symptom in up to 95% of woman with EC and most patients are therefore diagnosed at an early stage [27]. Other symptoms can include: pain or discomfort in the lower abdomen, pain during intercourse, loss of appetite and weight, tiredness or weakness, feeling or being sick, constipation, passing urine more often than usual, and pain in the back or legs.

The current diagnosis of EC is based on established conventional pathological markers, such as clinical stage, grade, depth of myometrium invasion and lymph-vascular space invasion (LVSI). To evaluate these markers, several non-invasive clinical imaging techniques such as transvaginal ultrasound (TVS), computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) scans are usually performed. Although, these techniques allow tumour lesion localization, they are unable to be utilized to make a sufficiently specific and secure diagnosis. Therefore, a sample of endometrial tissue must be removed from the lining of the endometrium (biopsy) and analysed by a pathologist. Upon cancer diagnosis, the surgical removal of the uterus is performed. Upon removal, a more thorough analysis can be performed and the tumour can be correctly staged and graded.

The clinical performance of a laboratory test can be described in terms of diagnostic accuracy, or the ability to correctly classify subjects into clinically relevant subgroups [28]. Diagnostic accuracy establishes how accurately the test discriminates between those with and without the disease and is determined by calculating the test’s sensitivity, specificity, likelihood ratio and

receiver operating characteristic (ROC) curve [29]. Diagnostic predictability establishes the ability of the test to predict the presence or absence of disease for a given test result and is determined by calculating the positive and negative predictive values. The former denotes the proportion of patients with true positive results (TPR) who have the disease, and the latter defines the proportion of patients with a false positive result (FPR) who do not have the disease [29]. For example, a test with 100% sensitivity and 95% specificity for EC is still not suitable to screen the general population because, if 5000 women were screened, the test would correctly identify the one true positive result, but would generate 250 (5% of 5000) false-positive results in the process. As an optimal standard, laboratory test development should strive for a 100% specific and 100% sensitive screening test for EC.

Imaging techniques, such as TVS, CT scan and MRI are usually used clinically to diagnose recurrent endometrial lesions, but they may be limited when used alone or for visualizing small disseminated lesions, small lymph node metastasis and postoperative/post radiation changes [30]. A study by Kim SH *et al.* compared the ability of TVS, CT scan and MRI to predict the depth of myometrium invasion and they found that the accuracy, sensitivity, and specificity of TVS, CT scan, and MRI were 69%/50%/81%, 61%/40%/75%, and 89%/90%/88%, respectively [31]. MRI has been shown to be superior to both TVS and CT scanning in determining whether EC has spread to the cervix or myometrium, with an overall staging accuracy between 85 and 93% [32]. The sensitivity and specificity of MRI in the detection of cervical involvement in EC is 72% and 93%, respectively [32]. Savelli *et al.* examined the accuracy of TVS and MRI in preoperative staging of EC. In predicting myometrial invasion, TVS had a sensitivity of 84%, a specificity of 83% and an accuracy of 84% [33]. In predicting cervical involvement in EC, TVS had a sensitivity of 93%, a specificity of 92% and an accuracy of 92% [33]. Accordingly, these tests are generally unreliable in diagnosis of the disease with and without metastasis. Despite the availability of the aforementioned tests, there are currently no recommended diagnostic tests available for routine screening for EC patients [34].

1.5.2 Prognostic and Predictive Biomarkers

Currently accepted morphology-based prognostic biomarkers in use today for EC patient management include the histological subtype, grade, stage of the disease, lymph node metastasis (LNM), depth of myometrial invasion, presence of lymph-vascular space invasion (LVSI) and cervical involvement [35]. Based on these risk factors, EC can be classified as low risk, intermediate risk and high risk for LNM. Low risk EC are FIGO stage IA with grade 1 or 2 of endometrioid type histology. High risk EC are FIGO stage IB of grade 3 or of non-endometrioid histology, or stage II or III EC. All other FIGO stage I EC are at intermediate risk for LNM. This group can be further divided into low and high intermediate risk subgroups [36]. However, it has been shown that up to 20% of the presumed low-risk tumours, will eventually recur, and roughly 50% of the patients with presumed high-risk tumours will be fully cured of their disease [37, 38]. Accordingly, there are no reliable prognostic biomarkers in routine clinical use for EC.

Histological subtype, according to traditional microscopy parameters, has consistently been proven to be an important predictor of survival, but also a determinant for the extent of the initial surgical procedure and subsequent use of adjuvant therapy. However, significant inter-observer variability exists, even amongst expert gynaecological pathologists, when subtyping and grading EC. FIGO stage is one of the strongest prognostic factors, with a significant reduction in 5-year survival rate correlating with higher stages. The 5-year survival rate for EC patients with localised disease is 96%. However, when the disease has spread regionally to nearby organs or lymph nodes, the 5-year survival drops to 67%. After the cancer has spread to distant organs, 5-year survival is 17%. Nonetheless, it has been reported that the preoperative grade is upgraded in 15-20% of cases in the final histopathology and up to 24% of patients have been identified to belong to a higher stage group upon pathological examination of the removed uterus [39]. Multiple studies have confirmed the positive correlation of LNM and LVSI [40]. Patients having LVSI are at greater risk of developing metastasis compared to those without and this could be used as a best marker for the prediction of LNM. Unfortunately, the information of the LVSI can currently only be gained after surgery.

Preoperative assessment of lymph node involvement represents a critical step for determining the extent of surgery in EC patients. Of patients with clinical stage I disease, lymph node metastasis (LNM) is the most important prognostic factor, as metastasis to regional lymph nodes changes the FIGO status from stage I to IIIC [41]. Metastasis to the pelvic nodes occur

in approximately 10% of EC cases and in 30% of these cases the para-aortic lymph nodes are also involved [42]. However, only 10% of EC patients suffer from metastasis, the majority undergoes lymphadenectomy (removals of the lymph nodes) that is associated with significant complications, such as lymphedema and nerve damage [43, 44]. Two large randomized trials demonstrate that lymphadenectomy does not improve survival compared with no lymphadenectomy in patients with early-stage EC [45, 46]. Therefore, in order to provide EC patients appropriate therapeutic treatment, it is important to identify biomarkers that can reliably predict the status of LNM in EC patients before surgery.

Over the last decade, various models have been proposed and investigated to predict LNM in EC patients, but none of these strategies have been widely accepted. Ideally, predictive biomarkers that could predict the status of LNM before surgery may allow surgeons to accurately select those patients who would benefit from complete surgical staging, protecting patients at low risk from possible complications resulting from the procedures.

Pollom *et al.* evaluated the pathological records and preoperative parameters of 296 EC patients and observed that LVSI, deep myometrial involvement, and cervical stromal invasion were significantly associated with LNM, whereas tumour size of ≥ 4 cm was borderline significant [47]. Using these factors, authors have constructed a predictive model for LNM, in which the absence of all four risk factors implied an irrelevant risk of LNM ($< 1\%$). Out of four, LVSI was strongly associated with LNM, and the authors suggested a preoperative evaluation of LVSI in all cases. Whilst information about the LVSI is clearly an important predictor of LNM, the downside is that such information can only be obtained after hysterectomy which usually also includes the removal of adjacent lymph nodes.

A recent study by Koskas *et al.* compared 3 preoperative and 7 postoperative models currently available to predict LNM in EC. The models were all applied to an independent cohort of 519 EC patients and the results show that the different models did not perform equally well [48]. One preoperative model, the Korean Gynecologic Oncology Group [KGOG] based on CA125 level and MRI loco-regional extension criteria [49] and three postoperative, a French nomogram based on pathological hysterectomy characteristics, age and race [50], Gynecologic Oncology Group (GOG), based on uterine pathology characteristics [51], and Mayo clinic based on pathological hysterectomy characteristics such as grade, LVSI, tumour diameter, myometrial and cervical invasion [52], performed well with AUC values of ≥ 0.75 and false negative rates of $< 5\%$. Overall, the above-mentioned studies found that postoperative models

based on final pathological characteristics predicted LNM better than the preoperative models based on clinical imaging techniques. This finding probably reflects the inaccuracy of currently available imaging techniques to assess the LNM status preoperatively. Therefore, a preoperative model that can guide the surgeons in deciding the status of LNM in EC is warranted.

1.6 Current Status of Molecular Biomarkers

Recent advances in molecular biology have shown that malignant tumours arise from alterations in genes controlling proliferation, apoptosis, angiogenesis, invasion, and metastasis [53]. Overall, in cancer, about a dozen predictive molecular biomarkers have demonstrated robust clinical utility and have already been incorporated into clinical algorithms. They include well known examples such as ERBB2 amplification for treatment with HER2-targeted therapies in breast cancer [54] and BCRABL gene fusion for treatment with imatinib in chronic myeloid and acute lymphoblastic leukaemia [55]. In contrast, EC does not have a molecular biomarker to guide treatment decisions [56]. Clinical symptoms and imaging are currently the main pillars to detect endometrial carcinoma progression or recurrence. Molecular biomarkers such as growth differentiation factor 15, cancer antigen 125 (CA125) and human epididymis protein 4 (HE4) have shown promise, but are not routinely applied clinically [57-60].

Serum cancer antigen 125 (CA125) is currently the best clinical marker for women with late-stage ovarian cancer. However, few studies have identified CA125 to be a useful marker in the early detection of EC [61-63]. Elevations of CA 125 were first described in EC patients with recurrent and advanced disease by Niloff *et al.* in 1984 [64]. Since then, many studies have confirmed serum CA125 concentrations in patients with EC are associated with deep myometrial invasion, extra uterine spread, positive peritoneal cytology, LNM, recurrence, advanced stages, and poor survival [65-68]. However, the sensitivity and specificity of CA125 may be too low for it to be used as an accurate diagnostic marker for EC [69]. Elevated serum CA125 levels have been detected in 11–43% of EC [70-74]. Hsieh *et al.* suggested a preoperative CA125 level greater than 40 U/ml should be considered as a criterion for full pelvic lymphadenectomy in the surgical staging of EC with the sensitivity and specificity of 77.8 and 81.0%, respectively [67]. Chao *et al.* determined the best cut-off value of serum CA125 level of 30 U/mL with sensitivity and specificity of 84.6% of 84.3% in predicting the status of LNM in EC [75]. However, it has been reported that only 10-20% of patients with EC and only 25% of patients with recurrent disease have elevated serum CA125 levels [76, 77]. For stage I disease, human epididymis protein 4 (HE4) yielded a 17% improvement in sensitivity compared with that of CA 125 [78]. In evaluating the relationship between patient survival and serum HE4, higher HE4 levels were associated with poor overall survival ($p = 0.02$) [79].

Other serum tumour markers have been studied for their potential use, both individually and in combination, to form a panel of proteins in patients with EC, including cancer antigen 15-3 (CA15-3), 19-9 (CA19-9), 72-4 (CA72-4), carcinoembryonic antigen (CEA), tumour marker OVX1 and macrophage colony-stimulating factor (M-CSF) [71, 80-83] with limited success. A study by Lo *et al.* has shown the increased levels of CA125, CA15.3 and CA19.9 were significantly associated with poor prognostic clinical parameters in EC patients [84]. Nagai *et al.* has established a one-step nucleic acid amplification (OSNA) assay using CK19 for detecting EC with LNM with a sensitivity of 93.3%, and specificity of 99.5% [85]. Mhawech-Fauceglia *et al.* investigated the role of U3 small nucleolar ribonucleoprotein protein (IMP3) in EC and found that IMP3 was the best predictor for type II compared with type I ($p < 0.001$) [86]. Voss *et al.* found that CD151 was an independent marker for survival and recurrence-free survival in EC [87].

Cocco *et al.* highlighted the serum amyloid A (SAA) as a biomarker for early detection of disease recurrence and the monitoring of EC therapy [88]. They observed that mean concentrations of SAA were significantly higher in individuals with grade 3 EC than individuals with grade 1 (252.6 vs 39.7 $\mu\text{g/ml}$; $p = 0.02$) or grade 2 EC (252.6 vs 47.9 $\mu\text{g/ml}$; $p = 0.02$) [88]. It was found that higher serum Midkine (MK) was associated with poor prognosis of EC patients ($p = 0.009$) [89]. Moore *et al.* reported HE4 and soluble mesothelin-related peptides were associated with low sensitivity in the detection of early-stage I and late-stage II-IV EC with 44.8% and 59.4%, respectively [78]. Recently, Farias-Eisner *et al.* developed a multiple logistic regression model (MLRM) using the values for ApoA-1 (apolipoprotein A-1), TF (transferrin), and TTR (transthyretin) to better detect EC [90]. This panel distinguished normal samples from early-stage EC with a sensitivity of 71% and a specificity of 88%. Additionally, the panel distinguished normal samples from late stage EC with a sensitivity of 82% and a specificity of 86% [90].

Hormone receptor status in particular has consistently been shown to be a relevant prognostic marker and might also influence the choice of treatment for metastatic disease. The overexpression of steroid receptors correlates with low-grade, non-endometrioid histology as well as a favourable outcome in many studies. Loss of oestrogen receptors and progesterone receptors in curettage specimens has been significantly associated with aggressive phenotypes and poor survival in EC patients [91]. Also, DNA ploidy has been shown to be an independent prognostic marker for EC. The true clinical value for this method in a routine diagnostic setting remains to be further validated [92]. There is evidence suggesting that the presence of

microsatellite instability (MI), or mutations of PTEN and CTNNB-1 (β -catenin) could indicate a good prognosis for EC, while the presence of p53 abnormalities or PIK3CA mutations could suggest aggressive progression of cancer. However, accumulated data regarding the prognostic significance of these alterations are still scarce.

1.7 Treatments for Endometrial Cancer

Improvements in histology and diagnostic imaging have simplified EC diagnosis, nonetheless it remains difficult to define individual risks accurately, and treatment is still chosen based on the conventional FIGO stage [93]. Due to the inconsistency of histotype and grade classification, the same women may receive different treatments. In Australia, the standard treatment for EC is surgery, including hysterectomy (complete removal of the uterus and cervix), bilateral salpingo-oophorectomy (removal of the fallopian tubes and both ovaries), and pelvic and para-aortic lymphadenectomy (complete removal of the lymph nodes). Despite only 10% of EC patients suffering from metastasis, all undergo radical treatment including the removal of lymph nodes (lymphadenectomy); a precautionary measure carried out due to our current inability to accurately stage the disease. It was observed that the surgical removal of the pelvic or para-aortic lymph nodes can be associated with a high morbidity and postoperative complications including deep vein thrombosis and lower extremity lymphedema, occurring in up to 38% of patients [94]. Lymphedema is the build-up of lymph fluid in the tissues just under the skin, resulting in swelling, tightness and discomfort in the affected part of the body. Although removals of the lymph nodes is the best way to identify LNM, its clinical importance remains controversial [95]. Based on the results shown in several studies, no survival benefit for routine lymphadenectomy in EC patients has been established [46]. It is believed that the risk of lymphadenectomy outweighs its benefits in low risk EC patients [96]. Therefore, many clinicians do not recommend the routine para-aortic lymphadenectomy, when the cancer is diagnosed at a very early stage and is found to be slow growing. On the contrary, the prognosis of the patients with LNM would be poor if LNM was not identified and these patients received only hysterectomy without lymphadenectomy or post-operative radiotherapy [97]. Therefore, to determine the optimal treatment for EC patients, use of biomarkers that can predict primary tumours of EC with and without LNM is warranted.

1.8 Metastasis and its Gene Protein Signature in Solid Tumours

Metastasis is a crucial step of malignant progression and remains the primary cause of death from solid cancers. It is a poorly understood process which involves complex angiogenic, proteolytic, cell adhesion, and immunologic events. Metastatic cancer cells proliferate, lose contact with neighbouring cells, migrate through the interstitial matrix, invade blood and lymph vessels, and grow out again in lymph nodes or distant organs. The metastatic cells possess specific properties to perform all these actions, which will be reflected in their proteome.

In 2003, Ramaswamy *et al.* demonstrated that a gene-expression program of metastasis is present in the bulk of a primary tumour, challenging the idea that metastasis originates from rare cells within the tumour which have acquired an exclusive ability to metastasize by random mutations [98]. Recently, it has also been shown that most metastasis result from synchronously disseminating cells from the primary tumour, supporting parallel rather than linear evolution [99]. A number of proteomics approaches have identified protein signatures from primary tumours to accurately predict LNM, overall survival and/or disease recurrence [100-105]. For example, an 11-protein signature for aggressive triple-negative breast cancer was developed and validated. Using this protein signature as guidance, more than 60% of patients could have avoided adjuvant chemotherapy compared with conventional prognostic guidelines [100].

1.9 Aims of the Thesis

The research presented in this thesis aimed to identify molecular discriminators which can distinguish between patients with and without metastatic disease based on the profiling of their primary endometrial carcinoma. More specifically, the study was designed to achieve the following aims:

Aim 1: To identify molecular tissue markers in primary cancers of the endometrium, which correlate with regional lymph node metastasis:

This research has been summarized, published and presented in Chapter 4.

Aim 2: To develop a preoperative predictive model for the identification of lymph node metastasis in endometrial cancer:

This research has been summarized, published and presented in Chapter 5.

Aim 3: To explore the feasibility and relevance of glycan imaging in endometrial cancer:

This research has been summarized and presented in a publication format in Chapter 6.

Proteomics of Endometrial Cancer Diagnosis, Treatment and Prognosis

This page intentionally left blank

2.1 Summary

The review article presented here provides a brief summary of proteomic technology applied in EC diagnosis, treatment, and prognosis. The first part of the review article focuses on recently identified biomarkers for EC, their importance in clinical use, as well as the proteomic methods used in their discovery. The second part highlights some of the cutting edge technologies that promise to contribute to a better understanding of EC by quantifying relative changes in the protein expression. Moreover as a proof of principle, we have also shown the ability of matrix assisted laser desorption and ionisation mass spectrometry imaging (MALDI MSI) in discriminating the regions of adjacent normal endometrium from cancerous areas.

This page intentionally left blank

2.2 Statement of Authorship

Title of Paper	Proteomics of endometrial cancer diagnosis, treatment and prognosis
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	P. Mittal, M. Klingler-Hoffmann, G. Arentz, C. Zhang, G. Kaur, M.K. Oehler, P. Hoffmann, Proteomics of endometrial cancer diagnosis, treatment, and prognosis, <u>Proteomics Clinical Applications</u> , 10 (2016); 217-229.

Principal Author

Name of Principal Author (Candidate)	Parul Mittal
Contribution to the Paper	Performed presented MALDI Imaging experiment, and wrote the first draft of the manuscript
Overall percentage (%)	80%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 24/11/16

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Manuela Klingler-Hoffmann
Contribution to the Paper	Supervised development of work, data interpretation and manuscript evaluation
Signature	Date 24/02/2017

Name of Co-Author	Georgia Arentz		
Contribution to the Paper	Data interpretation and manuscript evaluation		
Signature		Date	24 / 2 / 2017

Name of Co-Author	Chao Zhang		
Contribution to the Paper	Participated in experimental design		
Signature		Date	31 / 4 / 2017

Name of Co-Author	Gurjeet Kaur		
Contribution to the Paper	Annotated tissue samples and manuscript evaluation		
Signature		Date	25 / 11 / 16

Name of Co-Author	Martin K. Oehler		
Contribution to the Paper	Supervised development of work, data interpretation and manuscript evaluation		
Signature		Date	12/04/17

Name of Co-Author	Peter Hoffmann		
Contribution to the Paper	Supervised development of work, data interpretation, manuscript evaluation and acted as corresponding author		
Signature		Date	30 / 1 / 2017

REVIEW

Proteomics of endometrial cancer diagnosis, treatment, and prognosis

Parul Mittal^{1,2}, Manuela Klingler-Hoffmann^{1,2}, Georgia Arentz^{1,2}, Chao Zhang^{1,2}, Gurjeet Kaur³, Martin K. Oehler⁴ and Peter Hoffmann¹

¹ Adelaide Proteomics Centre, School of Biological Sciences, The University of Adelaide, Adelaide, Australia

² Institute for Photonics and Advanced Sensing (IPAS), The University of Adelaide, Adelaide, Australia

³ Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Pulau Pinang, Malaysia

⁴ Department of Gynaecological Oncology, Royal Adelaide Hospital, North Terrace, Adelaide, Australia

This review discusses the current status of proteomics technology in endometrial cancer diagnosis, treatment and prognosis. The first part of this review focuses on recently identified biomarkers for endometrial cancer, their importance in clinical use as well as the proteomic methods used in their discovery. The second part highlights some of the emerging mass spectrometry based proteomic technologies that promise to contribute to a better understanding of endometrial cancer by comparing the abundance of hundreds or thousands of proteins simultaneously.

Received: May 8, 2015

Revised: August 13, 2015

Accepted: November 2, 2015

Keywords:

Biomarker / Endometrial cancer

1 Introduction

Endometrial cancer (EC) is the most common malignant tumor of the female reproductive tract. According to the American Cancer Society, an estimated 52 630 new EC cases were diagnosed and 8590 patients died from the disease in 2014 in the United States. From 2006 to 2010, incidence rates of EC increased by 1.5% per year among women younger than 50 years and by 2.6% per year among women 50 years and older [1]. Based on the classification system introduced by Bokhman in 1983, EC is divided into two histological subtypes: endometrioid adenocarcinomas (Type I) and nonendometrioid (Type II) carcinomas [2]. Type I adenocarcinomas account for 90% of all EC. These cancers are usually low grade, diagnosed at an early stage and have a good prognosis. They are associated with oestrogen excess, obesity, and atypical hyperplasia. In contrast, type II nonendometrioid tumors are high grade tumors with more aggressive biological

behavior compared to type I disease. Clinically, type II cancers are often diagnosed at advanced stage when prognosis is poor. Type II tumors include serous and clear cell carcinomas, as well as carcinosarcomas.

Proteomics, more precisely the comparative quantitation of protein subsets, holds great promise in improving outcome of patients with EC as it provides unique tools for discovery of new biomarkers and therapeutic targets. To date, biomarkers have proven tremendous clinical value in early diagnosis [3, 4], categorizing different subtypes of malignancies [5, 6], and in monitoring patient's response to therapy [7, 8]. Regardless of the numerous proteomic studies that have contributed to the standardization of experimental protocols for digestion [9], separation [10], enrichment [11], identification [12], and quantification [13] of low abundant proteins by highly efficient mass spectrometric techniques, proteomic research is still restricted by both technology and bioinformatics tools. Many protein and peptide peaks have been reported to bear significant diagnostic [14], prognostic [15], or predictive value [16] for EC; however, the candidate biomarkers have not yet been validated for use in clinical patient care [17]. Some authors have speculated that this might be due to studies using a single proteomics approach, which is not sufficient to gain an in depth understanding of the protein function and does not eliminate false-negative and false-positive results [18]. Nevertheless, implementing "omics" integration approaches

Correspondence: Professor Peter Hoffmann, Adelaide Proteomics Centre, University of Adelaide, Gate 8 Victoria Drive, Adelaide 5005, South Australia

E-mail: peter.hoffmann@adelaide.edu.au

Fax: +61-08-0-8313-4362

Abbreviations: ANXA2, annexin A2; COX, cyclooxygenase; DDA, data-dependent acquisition; DIA, data-independent acquisition; EC, endometrial cancer; FFPE, formalin-fixed paraffin embedded; MSI, microsatellite instability; ROC, receiver operating characteristic

Colour Online: See the article online to view Figs. 1 and 4 in colour.

and the use of the rapidly emerging MS-based proteomic technologies will aid in the venture to elucidate protein markers and their function, ultimately providing more reliable, sensitive, and specific biomarkers for EC.

In the first part of this review article, we will summarize the various MS-based proteomic approaches that have been used in EC studies. In the second part, we will introduce emerging proteomic technologies that may not only be used as discovery tools but also allow the implementation of MS in the diagnosis and prognosis of EC.

2 Clinical considerations of endometrial cancer

The clinical management of EC remains a challenge with patients presenting with a full spectrum of disease ranging from those with excellent prognosis and high curability to aggressive disease with poor outcome [19]. EC is staged according to the International Federation of Gynaecology and Obstetrics (FIGO) system (Table 1) [20]. The classification of EC into low and high risk disease is dependent on a number of parameters, only two of which are assessable pre-operatively [http://wiki.cancer.org.au/australiawiki/index.php?oldid=86682] after a biopsy has been obtained, namely the histological type and grade. However, Jacques et al. have shown a large percentage of cancer will be classified as a higher grade tumor after definitive surgery [21]. Therefore, the exact type, grade, and stage of EC can only be determined by surgery and subsequent histopathological assessment.

Table 1. Classification of carcinoma of the endometrium, International Federation of Gynaecology and Obstetrics (FIGO 2010), table modified from Creasman et al. [100]

FIGO stage	Tumor localization
Stage I	Tumor confined to the corpus uteri
IA	No or less than half myometrial invasion
IB	Invasion equal or more than half of the myometrium
Stage II	Tumor invades cervical stroma, but does not extend beyond the uterus
Stage III	Local and/or regional spread of the tumor
IIIA	Tumor invades the serosa of the corpus uteri and/or adnexa
IIIB	Vaginal and/or parametrial involvement
IIIC	Metastasis to pelvic and/or para-aortic lymph nodes
IIIC1	Positive pelvic nodes
IIIC2	Positive para-aortic lymph nodes with or without positive pelvic lymph nodes
Stage IV	Tumor invades bladder and/or bowel mucosa, and/or distant metastasis
IVA	Tumor invades bladder and/or bowel mucosa
IVB	Distant metastasis, including intraabdominal metastasis and/or inguinal lymph nodes

2.1 Diagnosis

EC is frequently diagnosed at an early stage, as it regularly presents with symptoms such as postmenopausal bleeding that usually develops early in the disease process. Nevertheless, the discovery of serum biomarkers for early detection of EC has become a high priority [22]. A number of serum biomarkers have been identified so far, with the most commonly used serum biomarker in gynecological oncology being Carbohydrate Antigen 125 (CA 125). In EC elevated serum CA 125 level have been detected in 11–43% of the cases and is shown to correlate with advanced stage and with the presence of extrauterine disease [23–25]. However, due to lack of specificity and sensitivity, CA 125 has limited significance in the diagnosis of EC. Human Epididymis protein 4 (HE4) has recently emerged as a promising biomarker for EC [26]. Brennan et al. highlighted the utility of serum HE4 using ELISA for preoperative risk stratification to identify high-risk patients within low-grade endometrioid EC patients who might benefit from lymphadenectomy [27]. For stage I EC, HE4 showed a 17% improvement in sensitivity when compared to CA 125 [28].

Hareyama et al. investigated the immunohistochemical CA72-4 expression in EC. The authors reported an elevated level of serum CA72-4 in 22–32% of the cases, that was associated with depth of myometrial invasion, adnexal metastasis, lymphovascular space invasion, and pelvic and para-aortic lymph node metastasis [29]. Konno et al. demonstrated the level of serum soluble Fas (sFas) is significantly higher in EC patients with advanced cancer when compared to localized cancer ($p < 0.0001$) [30]. A study by Sawada et al. showed raised serum levels of immunosuppressive acidic protein in 55–76% of EC cases where the level was found to increase with the stage of the disease [31]. Hakala et al. reported the serum level of macrophage colony-stimulating factor 1 (mcsf 1), correlated significantly with tumor grade and poor prognosis in 25–73% of EC cases [32,33]. A recent study by Kang-Wai Mu et al. reported differential levels of zinc alpha-2 glycoprotein, alpha 1-acid glycoprotein, and CD59 in the urine of EC patients when compared to urine from healthy controls using 2DIGE and O-Glycan binding lectin [34].

2.2 Treatment

EC is usually treated by surgery consisting of a hysterectomy, bilateral salpingo-oophorectomy, and pelvic lymph node dissection and has a very good prognosis when tumour mass is confined to the uterus [28]. In contrast, EC with metastasis to the lymph nodes has a high mortality rate. Lymph node metastasis is therefore a crucial factor in the prognosis and choice of treatment of EC patients. Clinicians presently face the challenge that conventional surgical-pathological variables (e.g. tumor size, depth of invasion, and grade of disease) and radiological imaging are unreliable in determining if a gynecological cancer has spread. Consequently, although

only 5% of patients suffer from metastatic disease, the majority undergo radical treatment including removal of the lymph nodes. Lymph node dissection, however, is associated with significant complications such as lower extremity lymphedema [35]. Predictive tissue markers for lymph node metastasis are therefore warranted to determine the optimal treatment strategy and to avoid morbidity in gynecological cancer.

2.3 Prognosis

The prognosis of EC depends on various factors including the histological subtype, grade, and stage of the disease (which is determined by depths of tumor invasion, lymph node metastasis, and spread of the disease to other organs) [36]. The depth of myometrial invasion and histological grade strongly correlate with the prevalence of lymph node metastasis and patient survival [37].

Various protein biomarkers have been described in EC that could be of benefit for the prediction of disease outcome. According to Cocco et al. serum amyloid A may represent a novel biomarker for EC to monitor disease recurrence and response to therapy [38]. Lo et al. demonstrated that increased levels of CA-125, CA 15-3, and CA 19-9 were significantly associated with poor prognostic clinical parameters [39]. Lambropoulou et al. evaluated the prognostic significance of survivin, c-erbB2, and cyclooxygenase (COX-2) levels in EC and stated that in a 10-year follow-up, patients with tumors expressing more of these three antigens had significantly lower survival rate than those with smaller expression score [40]. Yilmaz et al. have also shown higher nuclear expression of survivin in type I when compared to type II EC ($p = 0.040$); but no difference for cytoplasmic survivin and matrix metalloproteinase-2 expressions between type I and type II EC [41]. Zeimet et al., identified L1 cell adhesion molecule (L1CAM) as the best variable for predicting recurrence (sensitivity = 0.74; specificity = 0.91) and death (sensitivity = 0.77; specificity = 0.89) [42]. Although, the above-mentioned biomarkers are able to identify high risk patients with low grade EC, for clinical application additional biomarkers are required. New prognostic EC biomarkers will help distinguish patients who are at a low-risk of developing metastasis as compared to those who are at a high risk. Ideally, novel biomarkers will also be able to distinguish patients who have already developed metastasis. This would consequently allow patients to selectively undergo radical surgery while avoiding those who will not benefit from it and subsequently decreasing the risk of postsurgical morbidity.

3 Molecular genetics of endometrial cancer

Aside from the morphologic and clinical features, EC can be further distinguished into type I and type II EC on the basis of genetic alterations [43]. Type I and type II ECs are associ-

ated with mutations of independent gene sets [44]. Type I endometrial adenocarcinomas are characterized by mutations in the PTEN, K-ras, and β catenin, as well as DNA mismatch repair genes [35] while type II nonendometrioid EC frequently show aneuploidy and p53 mutations [35].

PTEN has been reported to be the most frequent genetic alteration of type I EC. PTEN, a tumor suppressor gene has been shown to be altered in up to 83% of endometrial adenocarcinomas and 55% in precancerous lesions [45]. Microsatellite instability (MSI) has been demonstrated in 20% of type I EC [46]. According to Bilbao et al. both PTEN mutations and MSI represent early events in endometrial carcinogenesis [47]. Other genetic alterations that occur in type I EC include mutations in K-ras [48] and β catenin [49]. PTEN, MSIS, and K-Ras mutations often coexist with each other, whereas mutations in β -catenin are usually observed alone [50].

The most common genetic alteration in type II EC is within p53, the tumor suppressor gene, with an occurrence of 93% [48, 51]. Other frequent genetic alteration in type II ECs are inactivation of the tumor suppressor gene p16 and overexpression of the oncogene, HER-2/neu [52].

4 Metastasis and protein biomarkers

Metastatic cancer cells proliferate, lose contact with neighboring cells, migrate through interstitial matrix, invade blood and lymph vessels, and colonize lymph nodes or distant organs [53]. Most cancer cells fail to undergo metastasis due to deficiency in one of the required steps like invasion, detachment or survival [54]. Discovery of EC biomarkers would significantly aid gynecological oncologists who currently face the challenge that radiological imaging and conventional surgical-pathological variables such as tumor size, depth of invasion, and grade of disease are unreliable in determining if an EC has metastasized.

The working model of metastasis implies that primary tumor cells acquire genetic alterations over time, which enables these cells to metastasize and form new solid tumors at distant sites [55]. It has been proposed that the gene expression program of metastasis is actually present in the bulk of primary tumor cells [56]. Ramaswamy et al. showed that some primary tumors are preconfigured to metastasize, and their susceptibility is detectable at the time of diagnosis. They defined a 17-gene signature pattern associated with metastasis, in which eight genes are upregulated, while nine genes are downregulated. They also concluded that metastasis is not dependent on only a single gene but on the complete 17 gene sequence.

Yi et al., have shown the expression of COX-2 to play an important role in metastasis of EC [57]. They identified the COX-2 inhibitor NS-398 to inhibit proliferation, viability, and invasion of the EC cell line RL95-2 [57]. Maxwell et al. confirmed annexin A2 (ANXA2) and peroxiredoxin were both being overexpressed in stage 1 EC when compared to normal endometrium [58].

5 MS-based proteomic approaches in endometrial cancer

The use of proteomic technologies are now enabling the identification and relative quantification of multiple proteins simultaneously from a single experiment to identify disease related and specific biomarkers. Proteomics in general deals with the large-scale determination of gene and cellular function directly at the protein level [59]. The proteome has also been defined as the protein complement expressed by a genome [60] but in reality, due to PTM and alternative splicing, it is estimated that about 22 000 protein coding genes [61] code for more than 500 000 proteins [62] in the human proteome.

In serum it is estimated that cancer derived proteins are 10 million times less abundant than common high abundant proteins secreted by normal cells. Therefore, it is challenging to correctly identify and quantify tumor-derived proteins from the whole serum proteome [63]. Alternatively, it is obvious that cancer derived proteins will be present in higher concentration in native tissue, organs, and their proximal fluids from where the tumor has originated as compared to the distant sites in which the tumor derived proteins may be secreted or leaked. Therefore, targeting those regional sites will dramatically increase the possibility of isolating and identifying tumor-specific biomarkers [64].

Proteomic analysis principally relies on MS for protein identification. The uses of MS-based techniques have made it possible to build comprehensive profiles of near complete proteomes, comparing the expressions of individual proteins that may serve as biomarkers. Generally, MS-based proteomic techniques can be classified as “gel based” and “gel free” (Fig. 1). In gel-based approaches such as 1D-PAGE, 2D-PAGE, and 2D-DIGE, proteins are separated on polyacrylamide gels via electrophoresis in one or two dimensions,

creating unique proteomic patterns based on the mass and/or charge/pH of the proteins.

Gel-free MS approaches can be either label based or label free. Label-based proteomic techniques are common for MS approaches where proteins are tagged either with isotopes such as ICAT, iTRAQ or labeled chemically such as SILAC. Labeling facilitates the quantification of proteins during MS analysis and is especially useful when acquiring data in the standard form of data-dependent acquisition (DDA), as labeling allows for the mixing and analysis of multiple samples at once. This alleviates the stochastic nature of DDA that can otherwise lead to sampling biases and poor reproducibility. Label-free techniques in the form of data-independent acquisition (DIA) help resolve the issue of sampling biases and eliminates any problems encountered due to artifacts from labeling that can interfere with the identity of the protein, leading to false conclusions [65,66]. DIA is addressed in more depth later in this article.

MS-based proteomics techniques have developed substantially over the past several years. Previously, protein analysis was reliant on 1D and 2D gel electrophoresis followed by sequencing; finally resulting in the identification of tens to hundreds of proteins. This limitation has been overcome with novel mass spectrometric approaches, which can identify many more proteins at once. Identification and quantification of proteins/peptides by MS can be done either by “bottom-up” or “top-down” approaches. The term “bottom-up” refers to the reconstruction of the protein sequence after proteolytic digestion, and thereby, identification and quantification of the peptide fragments using MS and appropriate databases [67]. The second approach “top-down,” refers to the identification of the whole proteins directly without enzymatic digestion [68].

To date, the most widely used proteomic technologies applied to the identification of the EC biomarkers includes

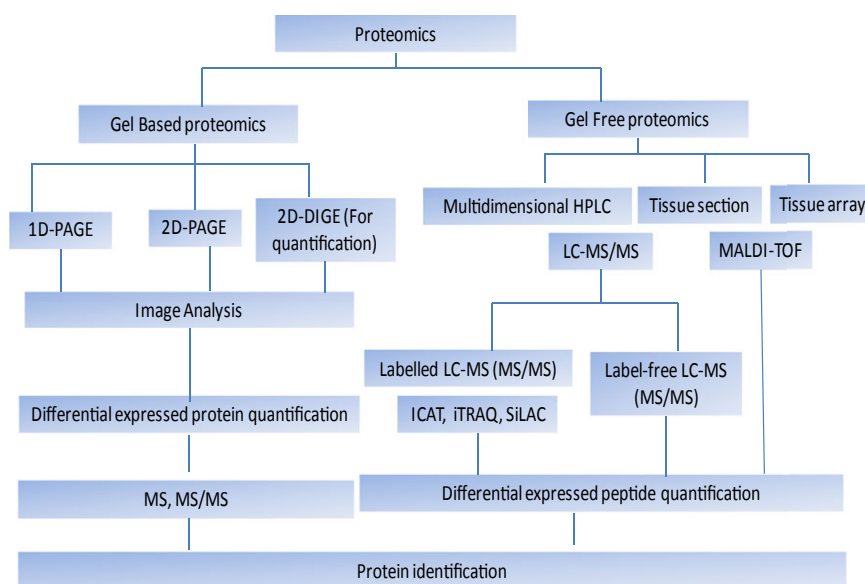


Figure 1. Proteomic approaches used in the EC biomarker discovery: Gel based and Gel free. In Gel-based proteomic approaches, proteins are separated on the basis of their charge/pH or mass by isoelectric focussing and electrophoresis. Gel-free techniques include MS-based technologies.

LC-MS. This approach is generally used with "bottom-up" methods. LC-MS has been used to identify and quantify differential change in protein abundance in healthy versus disease state. Identification of proteins by LC-MS/MS can be broken down into a few essential components. First, complex protein samples are digested with an enzyme (protease) into peptides, which are separated by a high resolution chromatographic technique coupled directly to a MS. Following chromatographic separation samples are sprayed into the MS while being ionized, a process called ESI. Here, the intact peptide mass is measured, followed by collision with an inert gas to induce fragmentation, upon which the mass of the fragment ions are measured, which is commonly referred as MS/MS or tandem MS. The acquired data is then processed to gain peptide sequences and protein identifications, and examined using mathematical algorithms, functions to detect and identify differences within the sample and/or between the samples. In traditional label-free approaches, peptide separation is compared across chromatographic profiles, and quantitation is determined either by spectral counting or by peak intensity during MS [69].

A recent study by Alconada et al. has identified and validated ANXA2 as a reliable biomarker of recurrent disease in EC by 2D DIGE followed by traditional LC-MS/MS. They demonstrated in vitro and with an in vivo mouse model that the increased expression of ANXA2 is associated with an improved ability of the cells to metastasize [70]. Monge et al. analyzed the role of the transcription factor ERM/ETV5 in myometrial invasion by 2D-DIGE to evaluate the differential expression of proteins in EC cell lines overexpressing ERM/ETV5 [71]. The authors have demonstrated that ERM/ETV5 acts by involving matrix metalloproteinase-2 to provide the migratory and invasive capabilities associated with the switch to myometrial infiltration in Hec-1A EC cell line [71]. They further characterized a role for ETV5 in a modulated response to oxidative stress associated with the promotion of invasion in EC [71]. Ihata et al. analyzed the amino acid concentrations in plasma samples by LC-MS and reported that the amino acid profile index is a potential new modality that could eventually play a significant role in the preoperative evaluation of EC and its screening [72].

The MALDI technique has also been widely used for the discovery of gynecological cancer biomarkers but in a more limited context for EC. Qiu et al. have reported that MALDI-TOF-MS is a high-throughput, sensitive, highly predictive, and rapid method for the early detection, diagnosis, and prognosis of EC and will be widely used in future clinical screening work [73]. In this technique, molecules including proteins, peptides, lipids and metabolites are first co-crystallized with large excess of suitable matrix (usually a weak organic acid) and then spotted on to a MALDI target plate. MALDI-TOF-MS is then performed, a process in which the laser transfers the energy to the aromatic matrix molecules and the matrix transfer a proton to the analyte, resulting in soft energy transfer without fragmentation. According to the m/z ratios, the ionized analytes are then separated in the TOF mass

analyzer and the detector transforms the incoming ions into an electric current that is proportional to their abundance [74]. In the case of protein analysis, the generated protein/peptide mass list can then be compared to protein databases that allow for the matching of peptide masses and identification of proteins, a process known as PMF. In addition using tandem MS, the parent ion can be further fragmented into daughter ions by the collision with an inert gas and peptide sequence is inferred by identifying the daughter ions. Casado-Vela et al. used an integrated approach combining three different but complementary strategies (in-solution digest followed by reverse phase HPLC-MS/MS, protein separation by SDS-PAGE followed by HPLC-MS/MS analysis, 2D-PAGE followed by MALDI-TOF/TOF analysis) to describe comprehensive proteomic analysis of EC fluid aspirate that led to the successful identification of 803 different proteins in the International Protein Index human database (v3.48) [75]. Therefore, EC biomarkers can be used for screening, estimating risk of disease, distinguishing benign from malignant, determining prognosis, and monitoring status of the disease, either to detect recurrence or response to therapy. Importantly, some biomarkers are only used in a specific setting, whereas others can serve more than one purpose [76].

6 Emerging techniques for the proteomic analysis of endometrial cancer

Traditional proteomic approaches as described above, continue to be the most widely applied technologies for protein identification; nonetheless, a major drawback of these proteomics techniques is that not every peptide from a complex biological sample can be detected. As a result, the proteomes will be under sampled. New techniques are emerging that offer innovative solutions to the analysis of low abundant proteins from complex biological samples [77]. Those proteomics technologies have the potential for direct clinical applications including the identification of novel biomarkers for the monitoring of therapy response, and disease progression, delivering on the great promise of personalized cancer medicine [78].

6.1 Data-independent acquisition

Protein identification by LC-MS is a straightforward technique, as the detection of only two unique peptides from a protein is considered sufficient. However, relative quantification by LC-MS is a more complex process and the failure to identify or detect a peptide does not necessarily confirm absence, as the peptides may simply be below the threshold of detection [79]. Traditional LC-MS or shotgun approaches are based on data-dependent acquisition (DDA), where peptides are selected for isolation and fragmentation based on their relative abundance within the sample. This makes the technique stochastic and can result in under sampling of the injected

analyte [80]. Abundance measurements can be inaccurate as the quantification of a peptide is highly influenced by coeluting peptides and their signal to noise ratios, meaning peptides of a higher abundance are quantified more accurately and more often than peptides of low abundance. We do however; know that many cancer-relevant proteins including signaling and regulatory proteins are typically expressed in low concentrations. As a result, the DDA approach tends to miss out on acquiring the most-valuable information [81]. Data-independent acquisition (DIA) aims to solve the problem of DDA by measuring all precursor ions within a defined mass range of ~25 Da, fragmenting and measuring all of the ions within that mass range [82]. Typically a 25 Da mass window is stepped across the entire mass range covered in traditional LC-MS/MS (~150–2200 Da) in cycle times compatible with the coupled chromatographic separation [82]. Vowinckel et al. shown the number of precisely quantifiable peptides can be increased up to five-fold when using DIA compared to DDA, with a CV less than 5% [83].

6.2 MALDI imaging

Another innovative technology that can be used for both biomarker discovery and validation is MALDI imaging (MALDI-MSI), which facilitates the direct analysis of protein distribution and abundance in thin tissue sections [84]. This technique has the potential to detect and characterize tumor margins and their environment in a spatial context [56]. MALDI-MSI is a valuable method for the identification of biomarkers and can complement histology, immunohistochemistry [85], and molecular pathology in various fields of histopathological diagnostics, especially with regard to identification and grading of tumors [86]. Whereas, MALDI-MSI is optimally suited to detect a large number of molecular features in a given tissue sample, the tissue microarray method enables expansion of this analysis to hundreds of tissue samples in a single experiment [87, 88]. In MALDI-MSI, proteins or in situ digested peptides are directly ionized from the surface of the tissue sample and a characteristic histological pattern is derived for hundreds or thousands of individual proteins/peptides simultaneously. The identification of proteins may be further investigate by PMF, a process of correlating the acquired mass spectra with the theoretical peptides generated from a given protein or proteins to find the closest match or by in situ MS/MS [89]. During in situ MS/MS, a desired peptide is isolated based on its m/z ratio and fragmented (i.e. breaking peptide bonds within the peptides), producing a series of fragment ions that are detected as a MS/MS spectrum. The fragmentation pattern is compared to the theoretical fragmentation pattern for every peptide in the proteome to find the closest match. In this way the sequence of the peptide ion is inferred from its fragmentation pattern [90]. Realistically in the case of MALDI-MSI, this form of MS/MS is limited to highly abundant ionizable peptides whereas the identifications of lower abundance peptides can

be made by matching back to fragmentation spectra acquired from LC-MS/MS using a tissue extract.

MALDI-MSI has been successfully used to classify HER2 receptor expression in breast cancer [91]. Meding et al. have shown the power of MALDI-MSI in classification of metastasis for different tumor entities [92]. In addition, their results indicated that even closely related entities such as primary tumor of the colon and its liver metastasis could be classified efficiently [92]. Quaa et al. detected several molecular signals associated with phenotypic features of oesophageal cancer cells, highlighting the potential of MALDI-MSI to identify new molecular markers with relevance to oesophageal cancer [87]. Another strength of MALDI-MSI is that it can be used in the discovery phase for the identification of tumor-derived proteins and then the verification of the presence of those particular proteins can be carried out in body fluid (blood, urine, or saliva), aiming to develop noninvasive diagnostic tests. Addona et al. developed a pipeline integrating proteomic technologies from the discovery to the verification stages of plasma biomarker identification and applied it to identify early biomarkers of cardiac injury [93]. Thus, combining the histopathology reports with MALDI-MSI data will help surgeons to make informed medical decisions.

In the similar approach, we performed MALDI imaging on EC formalin-fixed paraffin embedded (FFPE) tissue samples collected from the Royal Adelaide Hospital (RAH), South Australia. Ethics approval was granted by RAH Human ethics committee.

In this study, 6 μm FFPE tissue sections were sliced and mounted onto indium tin oxide-coated conductive glass slides (Bruker Daltonics, Bremen, Germany) and samples were prepared as described by Gustafsson et al. [12]. Briefly upon Citric acid antigen retrieval (CAAR), in situ tryptic digestion was performed using an Image prep station (Bruker Daltonics, Bremen, Germany) at 37°C for 2 h. CHCA matrix solution was prepared at 7 mg/mL in 50% ACN/0.2% TFA and was overlaid onto the tissue section using an Image Prep station. Following matrix deposition, analysis was performed on an UltrafleXtreme MALDI-TOF/TOF (Bruker Daltonics, Bremen, Germany) in positive reflectron mode with FlexControl V3.0.1 and FlexImaging V4.0.1 (Bruker Daltonics) software package. The data was acquired at 100 μm lateral resolution with a laser frequency of 2000 Hz over m/z 800–4500. Following MS acquisition, an overlapped MALDI spectral profile of tumor and healthy regions was generated by FlexAnalysis software (V4.0.1, Bruker Daltonics, Bremen, Germany) and as expected, a number of m/z values have been identified that are differentially expressed between healthy and tumor regions (Fig. 2).

In order to determine data-dependent visualization of tissue morphological regions, the raw data was loaded into SCiLS lab software (V2014b, Bruker Daltonics, Bremen, Germany). Data were processed by baseline removal and TIC normalization. Healthy and tumor regions were then grouped based on annotations done by a pathologist and discriminatory m/z values were calculated by ROC (receiver operating

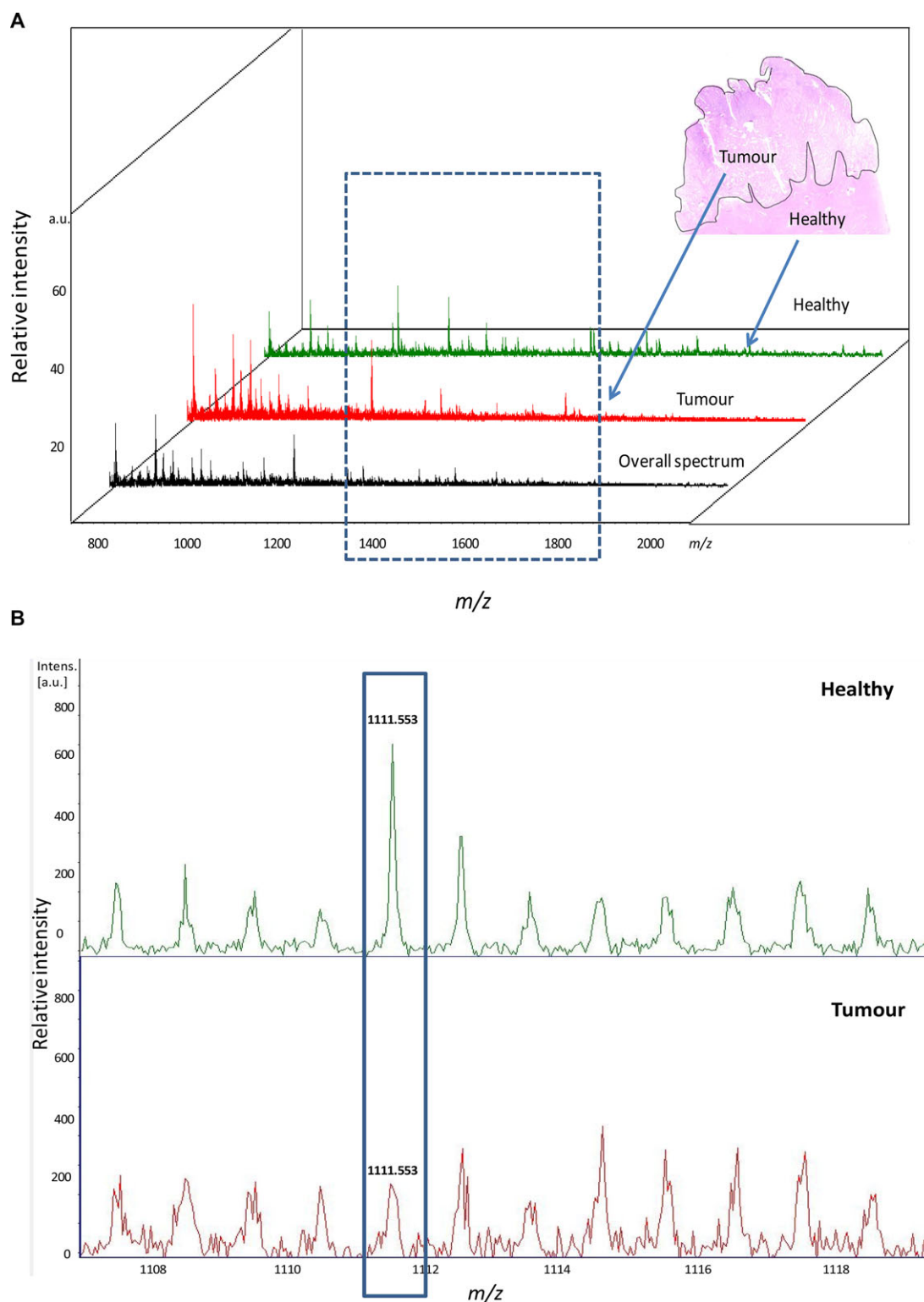


Figure 2. In this analysis in situ tryptic digestion was performed on 6 μm thick FFPE tissue. The matrix used was α -cyano-4-hydroxycinnamic acid and data were collected in positive reflectron mode using an UltrafleXtreme MALDI-TOF/TOF (Bruker Daltonics, Bremen, Germany). An overlapped MALDI spectral profile of tumor and healthy regions was generated by FlexAnalysis software (V4.0.1, Bruker Daltonics, Bremen, Germany). (A) The overlapped spectral profile show a multitude of differentially expressed m/z species between healthy and tumor regions simultaneously that allows a straightforward correlation of the expression pattern within the tissue morphology. (B) Comparative representative spectra of m/z 1111.553 Da $\pm$ 0.250 Da between healthy and tumor regions.

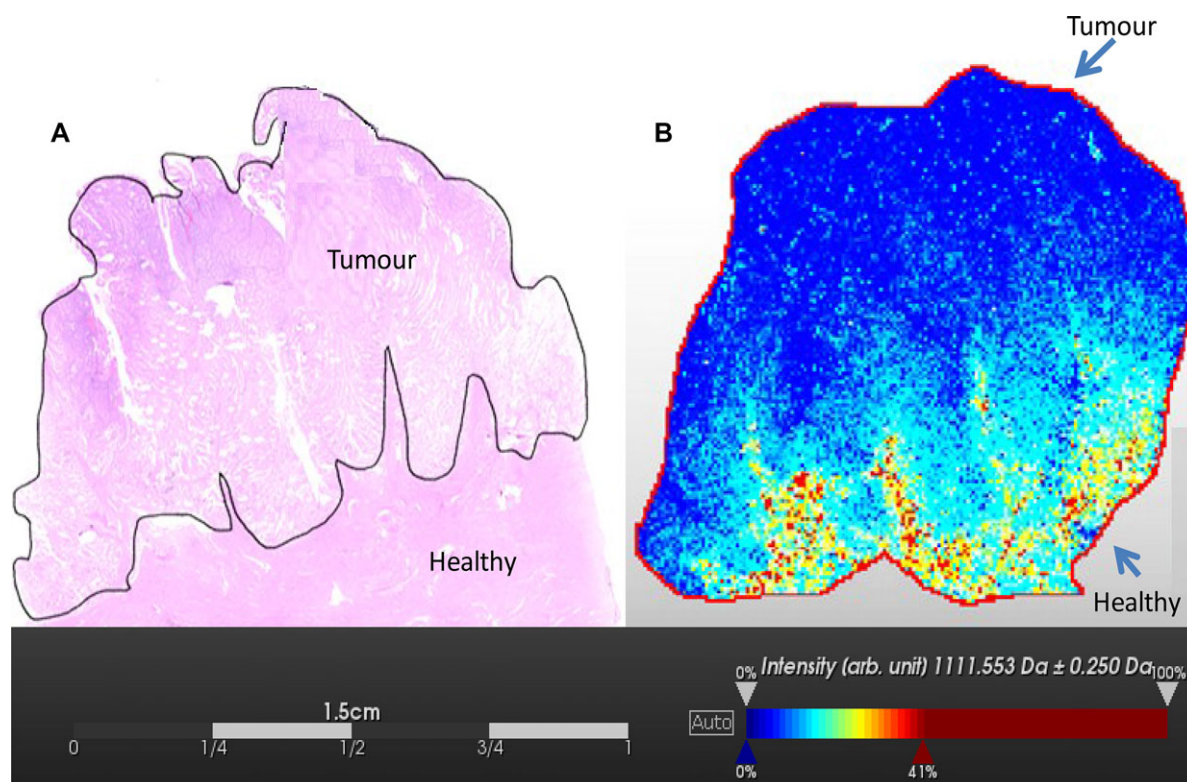


Figure 3. MALDI-MSI of an EC FFPE sample. The expression pattern of m/z 1111.553 Da $\pm$ 0.250 Da is visualized between healthy and tumor regions (B), when comparison is made with corresponding H&E stained annotated image (A). Visualization of difference in the intensities was performed in SCiLS lab software (V2014b, Bruker Daltonics, Bremen, Germany) with edge preserving image denoising and automatic hotspot removal applied. This m/z species is seen to be specifically downregulated in tumor region as compared to healthy region.

characteristic) curve. ROC curve compares sensitivity versus specificity across a range of m/z values by plotting a curve called an AUC. The AUC assumes values between 0 and 1 and expresses the discrimination power of the m/z signal. A perfect discrimination would yield an AUC value equal to 1 (abundant in group 1) or 0 (abundant in group 2) [94]. The AUC closer to 0.5 indicates that there is no discrimination between two sample groups. An example of a discriminatory m/z value (m/z 1111.553 Da $\pm$ 0.250 Da) is shown in Fig. 3. Here the spatial intensity of m/z 1111.553 Da $\pm$ 0.250 Da is clearly higher in the healthy tissue regions as compared to the tumor tissue region, shown also in the ROC curve (Fig. 4). The results from this study will provide crucial new tools to assist in the diagnosis and prognosis of EC, with the ultimate aim to prevent overtreatment of patients whose primary tumors do not have metastatic potential.

6.3 Targeted approach

Sensitive detection of low abundant proteins in complex samples has been typically achieved by ELISA that require antibodies specific to the protein(s) of interest. However, this poses a problem as development of antibodies is associated with high cost, long development lead times, and high failure

rates [95]. To overcome this problem, for the past few decades intensive research has been carried out aiming to develop more targeted strategies that have been designed specifically to analyze preselected peptides/proteins of interest within complex samples.

Unlike untargeted approaches, which aim to detect all proteins within a complex sample in an unbiased manner, multiple reaction monitoring (MRM) is a targeted approach, in which the protein(s) of interest are preselected for quantification and analyzed using highly sensitive triple quadrupole MS [96]. MRM-MS provides high sensitivity and accuracy needed in the discovery phase and offers high reproducibility necessary for clinical validation [97]. Due to the high specificity, sensitivity, and throughput of MRM, it can be used for the validation of a single protein of interest or a subset of proteins [98]. To ensure good quantification by MRM, three of the most intense ion transitions (precursor ion fragmentation to product ions) per protein are monitored. DeSouza et al. demonstrated that the MRM of iTRAQ labeled peptides enables absolute quantification of pyruvate kinase in cancerous and normal endometrial tissues [99]. Elizabeth et al. pioneered a new technique that couples the ability to perform MRM on a triple quadrupole mass spectrometer with a MALDI source. The authors have shown that this approach has the feasibility for the precise and accurate quantitation

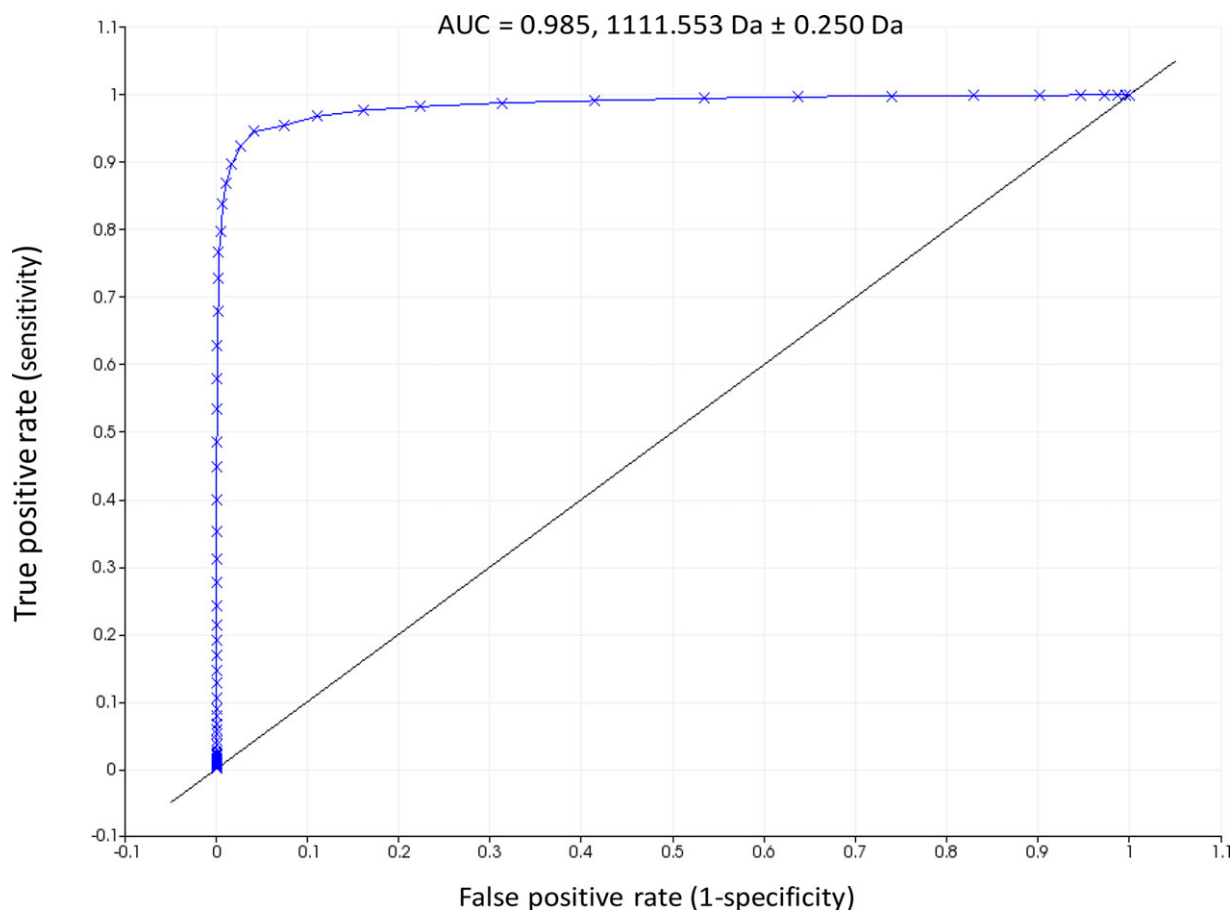


Figure 4. The ROC curve of m/z 1111.553 Da $\pm$ 0.250 Da (AUC = 0.985) for healthy versus tumor region.

of tissue protein concentrations over two orders of magnitude, while maintaining the spatial localization information for the proteins [100]. The greatest advantage of the MALDI MRM-based imaging technique is improved sensitivity and selectivity of the analysis, enabling measurement of the accurate protein concentration with the addition of stable isotope labeled internal standards. This technique combines the spatial information gained by traditional MALDI MSI with the accurate quantification achievable by MRM and may developed as an optimal strategy for the analysis of abundant peptide biomarkers.

7 Conclusion

Above-mentioned studies have discovered a number of candidate EC biomarkers, for example CA-125, CA 19-9, CA 15-3, HE4 for the diagnosis; and L1CAM, COX-2, Survivin, c-erb B2 for the prognosis of EC. Unfortunately due to their lack of sensitivity and specificity, none of those protein biomarkers are currently used in clinical practice. Recently, the major focus of biomarker research has shifted from single biomarker discovery to the multiparametric analysis of proteins, as the

proteomic pattern analysis can obtain exceptional results. As part of this review, we have introduced two emerging techniques with exceptional sensitivity and dynamic range. These techniques are refining our capability to identify and quantify relative changes in protein expression and will ultimately result in the discovery of new EC biomarkers panels.

Moreover, the integration of the data obtained from various “omics” approaches such as genomics, transcriptomics, and proteomics will help to reduce the false positives and false negatives obtained from single “omic” approaches and provide us with a number of potential protein targets. Taken together, this is an exciting time for EC biomarker research and we are hopeful that novel EC biomarker panels will perform well and will be used in the clinic in the future.

The authors have declared no conflict of interest.

8 References

- [1] American Cancer Society, *Cancer Facts and Figures 2014*. American Cancer Society, Atlanta 2014.
- [2] Bokhman, J. V., Two pathogenetic types of endometrial carcinoma. *Gynecol. Oncol.* 1983, 15, 10–17.

- [3] Toiyama, Y., Tanaka, K., Kitajima, T., Shimura, T. et al., Elevated serum angiopoietin-like protein 2 correlates with the metastatic properties of colorectal cancer: a serum biomarker for early diagnosis and recurrence. *Clin. Cancer Res.* 2014, 20, 6175–6186.
- [4] Bie, Y., Zhang, Z., Diagnostic value of serum HE4 in endometrial cancer: a meta-analysis. *World J. Surg. Oncol.* 2014, 12, 169.
- [5] Qendro, V., Lundgren, D. H., Rezaul, K., Mahony, F. et al., Large-scale proteomic characterization of melanoma expressed proteins reveals nestin and vimentin as biomarkers that can potentially distinguish melanoma subtypes. *J. Proteome Res.* 2014, 13, 5031–5040.
- [6] Le Faouder, J., Laouirem, S., Alexandrov, T., Ben-Harallah, S. et al., Tumoral heterogeneity of hepatic cholangiocarcinomas revealed by MALDI imaging mass spectrometry. *Proteomics* 2014, 14, 965–972.
- [7] Kralj, E., Žakelj, S., Trontelj, J., Pajič, T. et al., Monitoring of imatinib targeted delivery in human leukocytes. *Eur. J. Pharm. Sci.* 2013, 50, 123–129.
- [8] Ubhi, B. K., Cheng, K. K., Dong, J., Janowitz, T. et al., Targeted metabolomics identifies perturbations in amino acid metabolism that sub-classify patients with COPD. *Mol. Biosyst.* 2012, 8, 3125–3133.
- [9] Meding, S., Martin, K., Gustafsson, O. J., Eddes, J. S. et al., Tryptic peptide reference data sets for MALDI imaging mass spectrometry on formalin-fixed ovarian cancer tissues. *J. Proteome Res.* 2013, 12, 308–315.
- [10] Weiland, F., Zammit, C. M., Reith, F., Hoffmann, P., High resolution two-dimensional electrophoresis of native proteins. *Electrophoresis* 2014, 35, 1893–1902.
- [11] Danilchik, J., Sebela, M., Petrivalsky, M., Lenobel, R., A sensitive quantification of the peptide apidaecin 1 isoforms in single bee tissues using a weak cation exchange pre-separation and nanocapillary liquid chromatography coupled with mass spectrometry. *J. Chromatogr. A* 2014, 1374, 134–144.
- [12] Gustafsson, O. J., Eddes, J. S., Meding, S., McColl, S. R. et al., Matrix-assisted laser desorption/ionization imaging protocol for in situ characterization of tryptic peptide identity and distribution in formalin-fixed tissue. *Rapid Commun. Mass Spectrom.* 2013, 27, 655–670.
- [13] Li, J., Zhou, L., Wang, H., Yan, H. et al., A new sample preparation method for the absolute quantitation of a target proteome using O labeling combined with multiple reaction monitoring mass spectrometry. *Analyst* 2015, 140, 1281–1290.
- [14] Ueda, Y., Enomoto, T., Kimura, T., Miyatake, T. et al., Serum biomarkers for early detection of gynecologic cancers. *Cancers* 2010, 2, 1312–1327.
- [15] Engelsens, I. B., Akslen, L. A., Salvesen, H. B., Biologic markers in endometrial cancer treatment. *APMIS* 2009, 117, 693–707.
- [16] Dong, P., Kaneuchi, M., Konno, Y., Watari, H. et al., Emerging therapeutic biomarkers in endometrial cancer. *BioMed. Res. Intern.* 2013, 2013, 11.
- [17] Galvao, E. R., Martins, L. M., Ibiapina, J. O., Andrade, H. M., Monte, S. J., Breast cancer proteomics: a review for clinicians. *J. Cancer Res. Clin. Oncol.* 2011, 137, 915–925.
- [18] Ge, H., Walhout, A. J., Vidal, M., Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Gen.* 2003, 19, 551–560.
- [19] Zhang, Y., Wang, J., Controversies in the management of endometrial carcinoma. *Obst. Gynecol. Int.* 2010, 2010, 862908.
- [20] Haltia, U. M., Butzow, R., Leminen, A., Loukovaara, M., FIGO 1988 versus 2009 staging for endometrial carcinoma: a comparative study on prediction of survival and stage distribution according to histologic subtype. *J. Gynecol. Oncol.* 2014, 25, 30–35.
- [21] Jacques, S. M., Qureshi, F., Munkarah, A., Lawrence, W. D., Interinstitutional surgical pathology review in gynecologic oncology: I. Cancer in endometrial curettages and biopsies. *Int. J. Gynecolog. Pathol.* 1998, 17, 36–41.
- [22] Ueda, Y., Enomoto, T., Kimura, T., Miyatake, T. et al., Serum biomarkers for early detection of gynecologic cancers. *Cancers* 2010, 2, 1312–1327.
- [23] Powell, J. L., Hill, K. A., Shiro, B. C., Diehl, S. J., Gajewski, W. H., Preoperative serum CA-125 levels in treating endometrial cancer. *J. Reprod. Med.* 2005, 50, 585–590.
- [24] Jhang, H., Chuang, L., Visintainer, P., Ramaswamy, G., CA 125 levels in the preoperative assessment of advanced-stage uterine cancer. *Am. J. Obst. Gynecol.* 2003, 188, 1195–1197.
- [25] Scambia, G., Gadducci, A., Panici, P. B., Foti, E. et al., Combined use of CA 125 and CA 15-3 in patients with endometrial carcinoma. *Gynecol. Oncol.* 1994, 54, 292–297.
- [26] Brennan, D. J., Hackethal, A., Metcalf, A. M., Coward, J. et al., Serum HE4 as a prognostic marker in endometrial cancer—a population based study. *Gynecol. Oncol.* 2014, 132, 159–165.
- [27] Brennan, D. J., Hackethal, A., Metcalf, A. M., Coward, J. et al., Serum HE4 as a prognostic marker in endometrial cancer—a population based study. *Gynecol. Oncol.* 2014, 132, 159–165.
- [28] Moore, R. G., Brown, A. K., Miller, M. C., Badgwell, D. et al., Utility of a novel serum tumor biomarker HE4 in patients with endometrioid adenocarcinoma of the uterus. *Gynecol. Oncol.* 2008, 110, 196–201.
- [29] Hareyama, H., Sakuragi, N., Makinoda, S., Fujimoto, S., Serum and tissue measurements of CA72-4 in patients with endometrial carcinoma. *J. Clin. Pathol.* 1996, 49, 967–970.
- [30] Konno, R., Takano, T., Sato, S., Yajima, A., Serum soluble Fas level as a prognostic factor in patients with gynecological malignancies. *Clin. Cancer Res.* 2000, 6, 3576–3580.
- [31] Sawada, M., Okudaira, Y., Matsui, Y., Shimizu, Y., Immunosuppressive acidic protein in patients with gynecologic cancer. *Cancer* 1984, 54, 652–656.
- [32] Hakala, A., Kacinski, B. M., Stanley, E. R., Kohorn, E. I. et al., Macrophage colony-stimulating factor 1, a clinically useful tumor marker in endometrial adenocarcinoma: comparison

- with CA 125 and the aminoterminal propeptide of type III procollagen. *Am. J. Obst. Gynecol.* 1995, **173**, 112–119.
- [33] Suzuki, M., Ohwada, M., Sato, I., Nagatomo, M., Serum level of macrophage colony—stimulating factor as a marker for gynecologic malignancies. *Oncology* 1995, **52**, 128–133.
- [34] Mu, A. K., Lim, B. K., Hashim, O. H., Shuib, A. S., Detection of differential levels of proteins in the urine of patients with endometrial cancer: analysis using two-dimensional gel electrophoresis and o-glycan binding lectin. *Int. J. Mol. Sci.* 2012, **13**, 9489–9501.
- [35] Frederick, P. J., Straughn, J. M., Jr., The role of comprehensive surgical staging in patients with endometrial cancer. *Cancer Control Center* 2009, **16**, 23–29.
- [36] Larson, D. M., Connor, G. P., Broste, S. K., Krawisz, B. R., Johnson, K. K., Prognostic significance of gross myometrial invasion with endometrial cancer. *Obst. Gynecol.* 1996, **88**, 394–398.
- [37] Boronow, R. C., Morrow, C. P., Creasman, W. T., Disaia, P. J. et al., Surgical staging in endometrial cancer: clinical-pathologic findings of a prospective study. *Obst. Gynecol.* 1984, **63**, 825–832.
- [38] Cocco, E., Bellone, S., El-Sahwi, K., Cargnelutti, M. et al., Serum amyloid A: a novel biomarker for endometrial cancer. *Cancer* 2010, **116**, 843–851.
- [39] Lo, S. S., Cheng, D. K., Ng, T. Y., Wong, L. C., Ngan, H. Y., Prognostic significance of tumour markers in endometrial cancer. *Tumor Biol.* 1997, **18**, 241–249.
- [40] Lambropoulou, M., Papadopoulos, N., Tripsianis, G., Alexiadis, G. et al., Co-expression of survivin, c-erbB2, and cyclooxygenase-2 (COX-2): prognostic value and survival of endometrial cancer patients. *J. Cancer Res. Clin. Oncol.* 2010, **136**, 427–435.
- [41] Yilmaz, E., Koyuncuoglu, M., Gorken, I. B., Okyay, E. et al., Expression of matrix metalloproteinase-2 and survivin in endometrioid and nonendometrioid endometrial cancers and clinicopathologic significance. *J. Gynecol. Oncol.* 2011, **22**, 89–96.
- [42] Zeimet, A. G., Reimer, D., Huszar, M., Winterhoff, B. et al., L1CAM in early-stage type I endometrial cancer: results of a large multicenter evaluation. *J. Natl. Cancer Inst.* 2013, **105**, 1142–1150.
- [43] Bansal, N., Yendluri, V., Wenham, R. M., The molecular biology of endometrial cancers and the implications for pathogenesis, classification, and targeted therapies. *Cancer Control* 2009, **16**, 8–13.
- [44] Hecht, J. L., Mutter, G. L., Molecular and pathologic aspects of endometrial carcinogenesis. *J. Clin. Oncol.* 2006, **24**, 4783–4791.
- [45] Arango, H. A., Hoffman, M. S., Roberts, W. S., DeCesare, S. L. et al., Accuracy of lymph node palpation to determine need for lymphadenectomy in gynecologic malignancies. *Obst. Gynecol.* 2000, **95**, 553–556.
- [46] Bilbao-Sieyro, C., Ramirez, R., Rodriguez-Gonzalez, G., Falcon, O. et al., Microsatellite instability and ploidy status define three categories with distinctive prognostic impact in endometrioid endometrial cancer. *Oncotarget* 2014, **5**, 6206–6217.
- [47] Bilbao, C., Rodriguez, G., Ramirez, R., Falcon, O. et al., The relationship between microsatellite instability and PTEN gene mutations in endometrial cancer. *Int. J. Cancer* 2006, **119**, 563–570.
- [48] Lax, S. F., Kendall, B., Tashiro, H., Slebos, R. J., Hedrick, L., The frequency of p53, K-ras mutations, and microsatellite instability differs in uterine endometrioid and serous carcinoma: evidence of distinct molecular genetic pathways. *Cancer* 2000, **88**, 814–824.
- [49] Moreno-Bueno, G., Hardisson, D., Sánchez, C., Sarrió, D. et al., Abnormalities of the APC/b-catenin pathway in endometrial cancer. *Oncogene* 2002, **21**, 7981–7990.
- [50] Saegusa, M., Hashimura, M., Yoshida, T., Okayasu, I., β -Catenin mutations and aberrant nuclear expression during endometrial tumorigenesis. *Br. J. Cancer* 2001, **84**, 209–217.
- [51] Lluís Catasus, A. G., Miriam Cuatrecasas and Jaime Prat. Concomitant PI3K-AKT and p53 alterations in endometrial carcinomas are associated with poor prognosis. *Modern Pathol.* (2009). 2009, **22**, 522–529.
- [52] Doll, A., Abal, M., Rigau, M., Monge, M. et al., Novel molecular profiles of endometrial cancer—new light through old windows. *J. Steroid Biochem. Mol. Biol.* 2008, **108**, 221–229.
- [53] Takes, R. P., Baatenburg de Jong, R. J., Wijffels, K., Schuur, E. et al., Expression of genetic markers in lymph node metastases compared with their primary tumours in head and neck cancer. *J. Pathol.* 2001, **194**, 298–302.
- [54] René Bernards, Weinberg, R. A., A progression puzzle. 2002, **418**, 823.
- [55] Fidler, I. J., Kripke, M. L., Metastasis results from preexisting variant cells within a malignant tumor. *Science* 1977, **197**, 893–895.
- [56] Ramaswamy, S., Ross, K. N., Lander, E. S., Golub, T. R., A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* 2003, **33**, 49–54.
- [57] Yi, Z., Jingting, C., Yu, Z., Proteomics reveals protein profile changes in cyclooxygenase-2 inhibitor-treated endometrial cancer cells. *Int. J. Gynecol. Cancer* 2009, **19**, 326–333.
- [58] Maxwell, G. L., Hood, B. L., Day, R., Chandran, U. et al., Proteomic analysis of stage I endometrial cancer tissue: identification of proteins associated with oxidative processes and inflammation. *Gynecol. Oncol.* 2011, **121**, 586–594.
- [59] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, **422**, 198–207.
- [60] Hochstrasser, D. F., Proteome in perspective. *Clin. Chem. Lab. Med.* 1998, **36**, 825–836.
- [61] Pertea, M., Salzberg, S. L., Between a chicken and a grape: estimating the number of human genes. *Genome Biol.* 2010, **11**, 206.
- [62] Ewing, B., Green, P., Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* 2000, **25**, 232–234.
- [63] Ong, S.-E., Whole proteomes as internal standards in quantitative proteomics. *Genome Med.* 2010, **2**, 49.
- [64] Kulasingam, V., Diamandis, E. P., Tissue culture-based breast cancer biomarker discovery platform. *Int. J. Cancer* 2008, **123**, 2007–2012.

- [65] Grossmann, J., Roschitzki, B., Panse, C., Fortes, C. et al., Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *J. Proteomics* 2010, 73, 1740–1746.
- [66] Arentz, G., Weiland, F., Oehler, M. K., Hoffmann, P., State of the art of 2D DIGE. *Proteomics* 2015, 9(3–4), 277–288.
- [67] Tiss, A., Timms, J., Menon, U., Gammerman, A., Cramer, R., Proteomics approaches towards early detection and diagnosis of ovarian cancer. *J. Immunother. Cancer* 2014, 2(Suppl 1), O5.
- [68] Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E. et al., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011, 480, 254–258.
- [69] Liu, H., Sadygov, R. G., Yates, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 2004, 76, 4193–4201.
- [70] Alonso-Alconada, L., Santacana, M., Garcia-Sanz, P., Muinelo-Romay, L. et al., Annexin-A2 as predictor biomarker of recurrent disease in endometrial cancer. *Int. J. Cancer* 2014, 136, 1863–1873.
- [71] Monge, M., Colas, E., Doll, A., Gil-Moreno, A. et al., Proteomic approach to ETV5 during endometrial carcinoma invasion reveals a link to oxidative stress. *Carcinogenesis* 2009, 30, 1288–1297.
- [72] Ihata, Y., Miyagi, E., Numazaki, R., Muramatsu, T. et al., Amino acid profile index for early detection of endometrial cancer: verification as a novel diagnostic marker. *Int. J. Clin. Oncol.* 2014, 19, 364–372.
- [73] Qiu, F., Gao, Y. H., Jiang, C. G., Tian, Y. P., Zhang, X. J., Serum proteomic profile analysis for endometrial carcinoma detection with MALDI-TOF MS. *Arch. Med. Sci.* 2010, 6, 245–252.
- [74] Hillenkamp, F., Karas, M., Beavis, R. C., Chait, B. T., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* 1991, 63, 1193A–1203A.
- [75] Casado-Vela, J., Rodriguez-Suarez, E., Iloro, I., Ametzazurra, A. et al., Comprehensive proteomic analysis of human endometrial fluid aspirate. *J. Proteome Res.* 2009, 8, 4622–4632.
- [76] Henry, N. L., Hayes, D. F., Cancer biomarkers. *Mol. Oncol.* 2012, 6, 140–146.
- [77] Meehan, K. L., Rainczuk, A., Salamonsen, L. A., Stephens, A. N., Proteomics and the search for biomarkers of female reproductive diseases. *Reproduction* 2010, 140, 505–519.
- [78] Ong, S.-E., Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 2005, 1, 252–262.
- [79] Arentz, G., Weiland, F., Oehler, M. K., Hoffmann, P., State of the art of 2D DIGE. *Proteomics Clin. Appl.* 2014, 9, 277–288.
- [80] Wang, N., Li, L., Exploring the precursor ion exclusion feature of liquid chromatography–electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis. *Anal. Chem.* 2008, 80, 4696–4710.
- [81] Tate, S., Larsen, B., Bonner, R., Gingras, A.-C., Label-free quantitative proteomics trends for protein–protein interactions. *J. Proteomics* 2013, 81, 91–101.
- [82] Vowinkel, J., Capuano, F., Campbell, K., Deery, M. J. et al., The beauty of being (label)-free: sample preparation methods for SWATH-MS and next-generation targeted proteomics. *F1000Research* 2013, 2, 272.
- [83] Cornett, D. S., Reyzer, M. L., Chaurand, P., Caprioli, R. M., MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat. Meth.* 2007, 4, 828–833.
- [84] Bateson, H., Saleem, S., Loadman, P. M., Sutton, C. W., Use of matrix-assisted laser desorption/ionisation mass spectrometry in cancer research. *J. Pharmacol. Toxicol. Methods* 2011, 64, 197–206.
- [85] Kriegsmann, J., Kriegsmann, M., Casadonte, R., MALDI TOF imaging mass spectrometry in clinical pathology: a valuable tool for cancer diagnostics (review). *Int. J. Oncol.* 2014, 46, 893–906.
- [86] Quaas, A., Bahar, A. S., von Loga, K., Seddiqi, A. S. et al., MALDI imaging on large-scale tissue microarrays identifies molecular features associated with tumour phenotype in oesophageal cancer. *Histopathology* 2013, 63, 455–462.
- [87] Steurer, S., Borkowski, C., Odinga, S., Buchholz, M. et al., MALDI mass spectrometric imaging based identification of clinically relevant signals in prostate cancer using large-scale tissue microarrays. *Int. J. Cancer* 2013, 133, 920–928.
- [88] Breuer, E. K., Murph, M. M., The role of proteomics in the diagnosis and treatment of women's cancers: current trends in technology and future opportunities. *Int. J. Proteomics* 2011, 2011.
- [89] Wang, P., Whiteaker, J. R., Paulovich, A. G., The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biol. Ther.* 2009, 8, 1083–1094.
- [90] Rauser, S., Marquardt, C., Balluff, B., Deininger, S.-O. et al., Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. *J. Proteome Res.* 2010, 9, 1854–1863.
- [91] Meding, S., Nitsche, U., Balluff, B., Elsner, M. et al., Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging. *J. Proteome Res.* 2012, 11, 1996–2003.
- [92] Addona, T. A., Shi, X., Keshishian, H., Mani, D. R. et al., A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nat. Biotechnol.* 2011, 29, 635–643.
- [93] Klein, O., Nebrich, G., Trede, D., Becker, M. et al., SCiLS Lab 2D: Comparative Analysis for Uncovering Discriminative M/z-markers, 2015, SCiLS application note#5.
- [94] Shi, T., Fillmore, T. L., Sun, X., Zhao, R. et al., Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc. Natl. Acad. Sci. USA* 2012, 109, 15395–15400.
- [95] Huttenhain, R., Malmstrom, J., Picotti, P., Aebersold, R., Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr. Opin. Chem. Biol.* 2009, 13, 518–525.
- [96] Cohen Freue, G. V., Borchers, C. H., Multiple reaction monitoring (MRM): principles and application to coronary artery disease. *Circ. Cardiovasc. Genet.* 2012, 5, 378.

- [97] Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L., Park, B. K., Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J. Chromatogr. B, Analy. Technol. Biomed. Life Sci.* 2009, 877, 1229–1239.
- [98] DeSouza, L. V., Taylor, A. M., Li, W., Minkoff, M. S. et al., Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J. Proteome Res.* 2008, 7, 3525–3534.
- [99] Clemis, E. J., Smith, D. S., Camenzind, A. G., Danell, R. M. et al., Quantitation of spatially-localized proteins in tissue samples using MALDI-MRM imaging. *Anal. Chem.* 2012, 84, 3514–3522.
- [100] Creasman, W., Revised FIGO staging for carcinoma of the endometrium. *Int. J. Gynaecol. Obst.* 2009, 105, 109.

This page intentionally left blank

MATERIALS AND METHODS

This page intentionally left blank

3.1 Novel Approaches for Biomarker Discovery in Tissue

Over the past decade, major research has been undertaken to find gene alleles or mutations whose expression correlates with a given disease. The discovery of the complete sequence of the human genome has opened the door for proteomics by providing a sequence-based framework for mining proteomes [106, 107]. The proteome is a dynamic reflection of both genes and the environment and is thought to hold special promise for biomarker discovery. The human genome harbors 26000–31000 protein encoding genes [108]; whereas the total number of human protein products, including splice variants and essential post-translational modifications (PTMs), has been estimated to be close to one million [109]. Proteome alterations in disease may occur in many different ways that are not predictable from genomic analysis, and it is clear that a better understanding of these alterations will have a substantial impact in medicine [106]. As a matter of fact, most of today's licensed tests are protein-based assays and most drug targets are proteins [110, 111].

Mass spectrometry (MS) is a versatile and powerful tool for the analysis of proteins and peptides as well as other biological compounds such as lipids, metabolites, glycans, and other small molecules. The ability of MS to identify ever smaller amounts of protein from increasingly complex mixtures is a primary driving force in proteomics [112]. All mass spectrometers, regardless of type, ionization mode or performance characteristics, produce mass spectra, which plot the mass-to-charge ratio of the ions observed (x-axis) versus detected ion abundance (y axis). Subsequent interpretation of spectra, using characteristics such as isotope distribution, accurate mass and amino acid sequence information (in tandem MS (MS/MS) experiments), may allow portions of the spectrum to be labelled with a protein or peptide identity.

Currently available platforms for MS based proteomic discovery fall roughly under two categories: (1) pattern-based methods that focus on production of an MS-derived protein pattern such as MALDI and surface-enhanced laser desorption-ionization (SELDI) [113, 114], (2) “identity-based” methods that rely on proteolytic digestion of proteins to peptides with analysis by liquid chromatography (LC)-MS/MS [115]. The latter method suffers from the limitation of losing the structural information of the tissue, which is critical for clinical application, especially in heterogeneous carcinomas where different structural elements will express a unique proteome with subsequent unique cellular function. To reduce this morphological complexity while maintaining the cellular specificity, laser capture

microdissection (LMD) can be performed, which can isolate specific cell populations for analysis however, this method is time consuming but also might lead to sample loss due to incomplete homogenization. When compared, MALDI can be much more easily automated than LC-MS/MS and allows direct analysis of tissue sections without tissue homogenization. However, a major disadvantage for the analysis of complex samples appears to be the pronounced signal suppression in MALDI, which is observed to a lesser degree in LC-MS/MS [116].

3.1.1 Applications of Mass Spectrometry Imaging to Cancer

This page intentionally left blank

3.1.1.1 Summary

Matrix assisted laser desorption/ionisation mass spectrometry imaging (MALDI MSI) has received considerable attention in recent years. First described by Caprioli *et al.*, MALDI MSI has been developed to create a link between the molecular assessment of numerous molecules and morphological information about their spatial distribution [117]. The ability of this technique to compare the histology of a tissue to its corresponding spatially resolved mass spectrometric information makes it as a valuable tool to understand the molecular mechanism behind tumour metastasis.

The strength of this method compared to immunohistochemistry (a standardized technique employed in pathology laboratories globally), is that it is independent of prior knowledge of protein composition or availability of antibodies. Previous studies of several different cancers have also shown that MALDI MSI can acquire a more comprehensive proteomic picture than immunohistochemistry (IHC) [118, 119]. MALDI MSI is a non-biased, non-targeted approach which collects molecular and spatial information from fresh frozen or formalin fixed paraffin embedded (FFPE) tissue samples, the standard sample preparation procedure used in clinical pathology. This method requires very little tissue material; a single tissue section from an endoscopic biopsy is enough for analysis whilst retaining morphology, allowing the tissue to be stained post MALDI MSI acquisition. For MALDI MSI analysis, matrix is applied uniformly or in a discrete grid array across the tissue. A laser irradiates the matrix and ionizes the proteins, generating two dimensional data containing both the proteins detected and the tissue location. The sampling positions become ‘pixels’ that are compiled to generate a picture or image for each molecule detected. Each mass spectrum is measured at a pixel with a pair of spatial coordinates (x/y). Any given mass-to-charge (m/z) signal in the spectrum can be displayed with its relative intensity over the entire array using either freely or commercially available software such as SCiLS lab (GmbH, Bremen, Germany). The generated intensity images allow new insights into tumour tissues.

Multiple cancer research studies using MALDI MSI have demonstrated the strength of this technology for uncovering new markers that correlate with disease severity as well as prognosis and therapeutic response [105, 120, 121]. The presented book chapter gives a detailed overview of how MALDI MSI is used in the field of cancer biomarker discovery, the current developments and the challenges faced. This approach represents an exciting opportunity to

translate research discoveries from the laboratory to the clinic to improve patient health and survival.

3.1.1.2 Statement of Authorship

Title of Paper	Chapter 2 - Applications of Mass Spectrometry Imaging to Cancer
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	G. Arentz, <u>P. Mittal</u> , C. Zhang, Y-Y. Ho, M. Briggs, L. Winderbaum, M. Klingler-Hoffmann, P. Hoffmann, Chapter 2 - Applications of Mass Spectrometry Imaging to Cancer, in: Richard, R. D., Liam, A. M. (Eds.), <u>Advances in Cancer Research</u> , Academic Press 2017, Volume 134, 1st Edition, pp. 27-66.

Principal Author

Name of Co-Author (Candidate)	Parul Mittal		
Contribution to the Paper	Written "Protein Imaging in Cancer Research"		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	30/01/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Principal Author	Georgia Arentz		
Contribution to the Paper	Written "introduction" and manuscript evaluation		
Signature		Date	24/02/2017

Name of Co-Author	Chao Zhang		
Contribution to the Paper	Written "Practical Considerations for Peptide MSI"		
Signature		Date	3 / 4 / 2017

Name of Co-Author	Ying Ying Ho		
Contribution to the Paper	Written "Lipid Imaging in Cancer Research"		
Signature		Date	03.04.2017

Name of Co-Author	Matthew Briggs		
Contribution to the Paper	Written "Glycan Imaging in Cancer Research"		
Signature		Date	07/04/17

Name of Co-Author	Lyron Winderbaum		
Contribution to the Paper	Written "Data analysis"		
Signature		Date	30/1/2017

Name of Co-Author	Manuela Klingler-Hoffmann		
Contribution to the Paper	Written "Drug Imaging in Cancer Research" and manuscript evaluation		
Signature		Date	30/1/2017

Name of Co-Author	Peter Hoffmann		
Contribution to the Paper	Written "Conclusion", manuscript evaluation and acted as corresponding author		
Signature		Date	30/1/2017



Applications of Mass Spectrometry Imaging to Cancer

G. Arentz^{*,†}, P. Mittal^{*,†}, C. Zhang^{*,†}, Y.-Y. Ho^{*}, M. Briggs^{*,†,‡},
L. Winderbaum^{*}, M.K. Hoffmann^{*,†}, P. Hoffmann^{*,†,1}

^{*}Adelaide Proteomics Centre, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia

[†]Institute for Photonics and Advanced Sensing (IPAS), University of Adelaide, Adelaide, SA, Australia

[‡]ARC Centre for Nanoscale BioPhotonics (CNBP), University of Adelaide, Adelaide, SA, Australia

¹Corresponding author: e-mail address: peter.hoffmann@adelaide.edu.au

Contents

1. Introduction	28
1.1 The Advantages of MSI	29
1.2 The Basic Principles of MSI	30
2. Protein MSI in Cancer Research	33
2.1 Distinguishing Tissue Types by Peptide MSI	36
2.2 Determining Tumor Margins by Peptide MSI	37
2.3 Prediction of Metastasis by Peptide MSI	37
2.4 Analysing Chemoresponse by Peptide MSI	39
2.5 Identification of Diagnostic and Prognostic Markers by Peptide MSI	39
2.6 Characterisation of Intra- and Intertumor Variability by Peptide MSI	41
2.7 Practical Considerations for Proteolytic Peptide MSI: Sample Preparation	42
2.8 Practical Considerations for Peptide MSI: Spatial Resolution	43
2.9 Practical Considerations for Peptide MSI: Mass Analysers	44
2.10 Practical Considerations for Peptide MSI: Identification	44
3. Lipid MSI in Cancer Research	45
3.1 Profiling Lipids in Cancer by DESI-MSI	45
3.2 Profiling Lipids in Cancer by SIMS-MSI	46
4. Glycan MSI in Cancer Research	47
5. Drug Imaging in Cancer Research	50
5.1 MALDI-MSI on Tissue Sections	51
5.2 MALDI-MSI on Whole Body Sections	53
5.3 MALDI-MSI on 3D Tissue Cultures	53
6. Data Analysis	53
6.1 Spatial Information	53
6.2 Preprocessing: Peak Detection	55
6.3 Classification of FFPE-TMAs and the Importance of Dimension Reduction	56
7. Concluding Remarks	57
7.1 The Future of MSI: Molecular Pathology	57
References	57

Abstract

Pathologists play an essential role in the diagnosis and prognosis of benign and cancerous tumors. Clinicians provide tissue samples, for example, from a biopsy, which are then processed and thin sections are placed onto glass slides, followed by staining of the tissue with visible dyes. Upon processing and microscopic examination, a pathology report is provided, which relies on the pathologist's interpretation of the phenotypic presentation of the tissue. Targeted analysis of single proteins provide further insight and together with clinical data these results influence clinical decision making. Recent developments in mass spectrometry facilitate the collection of molecular information about such tissue specimens. These relatively new techniques generate label-free mass spectra across tissue sections providing nonbiased, nontargeted molecular information. At each pixel with spatial coordinates (x/y) a mass spectrum is acquired. The acquired mass spectrums can be visualized as intensity maps displaying the distribution of single m/z values of interest. Based on the sample preparation, proteins, peptides, lipids, small molecules, or glycans can be analyzed. The generated intensity maps/images allow new insights into tumor tissues. The technique has the ability to detect and characterize tumor cells and their environment in a spatial context and combined with histological staining, can be used to aid pathologists and clinicians in the diagnosis and management of cancer. Moreover, such data may help classify patients to aid therapy decisions and predict outcomes. The novel complementary mass spectrometry-based methods described in this chapter will contribute to the transformation of pathology services around the world.

ABBREVIATIONS

CHCA α -cyano-4-hydroxycinnamic acid
DESI desorption electrospray ionization
DHB 2,5-dihydroxybenzoic acid
FFPE formalin-fixed paraffin embedded
FT-ICR Fourier transformed ion cyclotron resonance
IHC immunohistochemistry
LNM lymph node metastasis
MALDI matrix-assisted laser desorption/ionization
MSI mass spectrometry imaging
PCA principle component analysis
SIMS secondary ion mass spectrometry
TOF time-of-flight



1. INTRODUCTION

In order to optimize the treatment outcomes for cancer patients and even provide a cure, several issues regarding the clinical management of disease need to be addressed. Based on pathological and clinical parameters

patients are classified to receive the best possible therapy. The accuracy of this process would be better if a more complete understanding of the molecular events involved in the development and progression of cancer was available. Currently, mass spectrometry imaging (MSI) is being extensively applied to the in situ molecular analysis of cancerous cells and tissues with the aim of identifying tumor margins, classifying primary tumor tissues with regards to chemoresponse and metastatic status, identifying diagnostic and prognostic markers, and the analysis of drug response rates and resistance. Elucidation of the spatial location and abundance of peptides, proteins, lipids, glycans, and drug metabolites in relation to cancerous tissues and cells has the potential to significantly improve the diagnosis, staging, and treatment of disease.

1.1 The Advantages of MSI

Interpatient heterogeneity significantly impacts the success of diagnostic and prognostic tests, because it influences the way patients respond to the same treatment. Moreover, proteomic analysis is impacted by intratumor heterogeneity, as tumors are complex structures comprised of a number of different cell types, such as epithelial or endothelial, stromal, vascular, and inflammatory cells. This is particularly important when considering concepts such as innate chemoresistance, metastasis, and stem-cell-like cancer cells, which are believed to consist of a small subset of cells within the bulk of the tumor tissue. Analysis methods that require homogenisation of a sample lose important information about the spatial location of the molecules being studied. Moreover, the information of a potential difference in protein abundance of a small number of cells might not be reflected in the average abundance of the tissue. Certain analysis methods, such as immunohistochemistry (IHC) or fluorescence microscopy, allow for the quantification of compounds within a spatial context. However, these methods are targeted, requiring prior knowledge about the target and the sample in question, and only a small number of molecules can be analyzed simultaneously.

The rapidly evolving technique of MSI allows for the in situ analysis of biological samples, combining classical mass spectrometry (MS) with histological tissue analysis. MSI has the capacity to determine the relative intensity and spatial distribution of several hundreds of compounds from cells and tissue while retaining important spatial information ([Schwamborn &](#)

Caprioli, 2010). In the context of cancer research this affords the technique the ability to characterize the molecular features of a sample while maintaining morphology and requiring no prior knowledge of the molecular expression profile. The type of molecules detected, or ions in the context of MS, is dependent on the sample preparation, instrumentation, and acquisition protocols used.

1.2 The Basic Principles of MSI

MS analysis includes the processes of analyte ionization, mass analysis, and detection (de Hoffmann & Stroobant, 2007). During ionization analyte molecules become electrically charged and liberated into the gas phase. Once ionized the mass of the charged molecules is determined by analysis in the mass spectrometer; the analyte ions are separated based on their mass-to-charge ratio (m/z) and detected, commonly by a microchannel plate detector. In terms of MSI, there are two basic forms: microprobe mode imaging and microscope mode imaging (Bodzon-Kulakowska & Suder, 2016; Klerk, Maarten Altelaar, Froesch, McDonnell, & Heeren, 2009; Luxembourg, Mize, McDonnell, & Heeren, 2004). In microprobe mode MSI, where information is collected sequentially from each pixel (pixel sizes typically 5–100 μm), each mass spectrum contains information about the m/z and relative intensity of the detected ions at each location. Following data acquisition, the collected information is reconstructed into a molecular image. In microscope mode MSI, spectral information is collected over a relatively large sample area (100–300 μm) simultaneously; in this form of MSI, the ions produced over the analyzed area are measured using a position-sensitive two-dimensional detection system. The information collected by both microprobe and microscope mode can be displayed as ion maps which show the distribution of the detected m/z values and their intensities, thereby characterizing cancer cells or tissues within their environment in a spatial context.

Following MSI the sample morphology is frequently retained and can be histochemically analyzed postdata acquisition (Crecelius et al., 2005; Schwamborn et al., 2007), making it feasible to directly compare MSI results with histological staining. These methods are complementary and have a high potential for aiding pathologists and clinicians in their diagnosis and management of disease (Rodrigo et al., 2014). As indicated earlier, MSI usually requires very little material and, for example,

a single tissue section from an endoscopic biopsy is enough to perform a successful MSI experiment.

For MSI, the ionization methods matrix-assisted laser desorption/ionization (MALDI), secondary ion mass spectrometry (SIMS), and desorption electrospray ionization (DESI) are popular, and are frequently coupled to time-of-flight (TOF) or Fourier transform type mass analyzers. A very brief description of the commonly applied techniques and their capabilities is provided in the following sections.

1.2.1 MALDI-TOF

MALDI-TOF is commonly used for the analysis of proteins, peptides, lipids, metabolites, and glycans (Cho et al., 2015; Signor & Boeri Erba, 2013). The application of MALDI-TOF to imaging MS was first introduced in 1997 (Caprioli, Farmer, & Gile, 1997) and the technique has gained significant popularity over the last two decades. For analysis by MALDI-TOF-MSI, a sample is placed on a conductive surface such as a stainless steel plate or an indium tin oxide-coated glass slide. Commonly fresh frozen or formalin-fixed paraffin-embedded (FFPE) tissue sections 4–10 μm thick are analyzed by MALDI-TOF-MSI, with FFPE tissues requiring additional preparation steps of deparaffinization and antigen retrieval (Ronci et al., 2008). Once the tissue is placed on the slide, depending on the type of analysis being performed, a digestion step may be carried out. For proteolytic peptide analysis, the most common enzyme utilized is trypsin, for glycan analysis a peptide-*N*-glycosidase F (PNGase F) digestion may be employed to release N-linked glycan structures from proteins (Gustafsson et al., 2015). The analysis of lipids, metabolites, endogenous peptides, and intact proteins directly from tissue does not require enzyme treatment. Prior to analysis a small organic acid, referred to as the matrix, is suspended in an acidified solvent buffer and deposited over the sample (Goodwin, 2012; Kaletas et al., 2009). During this process the matrix incorporates the analyte molecules, forming crystals. A focussed laser beam can then be applied to the surface in order to induce analyte ionization. The small organic molecules of the matrix mixture are designed to absorb the energy of the focussed laser beam resulting in an explosive desorption of the matrix and analyte crystals. The protons present in the acidified matrix solution also facilitate the analyte ionization. MALDI is a “soft” ionization technique, meaning many analytes may be analyzed without significant fragmentation.

Following ionization the ions are rapidly accelerated through a strong electric field in the ion source, the ions then enter the TOF tube or “drift” region, allowing separation based on velocity and thus m/z . The charged analyte molecules strike the detector and the number of events within a time period are recorded, providing intensity information. By comparing the measured information to calibration standards, the mass of the detected analyte ion can be calculated. The spatial resolution of MALDI-TOF imaging sits at around 20 μm and can accommodate a molecular mass range of over 100 kDa.

1.2.2 MALDI-FT-ICR

MALDI-FT-ICR is routinely applied to the imaging analysis of lipids, drugs, and metabolites. MALDI is carried out as described earlier with the Fourier transform ion cyclotron resonance (FT-ICR) acting as the mass analyser, where the m/z of analyte ions is determined by the cyclotron frequency of the ions within the fixed magnetic field of the instrument (Scigelova, Hornshaw, Giannakopoulos, & Makarov, 2011).

Following ionization the charged molecules are trapped in a Penning trap, which consists of a series of electrodes that use a homogeneous axial magnetic field to capture the molecules radially, and an inhomogeneous quadrupole electric field to confine the molecules axially. Within the Penning trap, a RF electric field is applied in the plane of ion cyclotron motion which excites the ions to move in a larger, synchronous, cyclotron radius. Once the excitation field is removed, the ions are left rotating at their cyclotron frequencies in phase. The ions induce a charge which is detected as an image current by electrode detectors in close proximity. The recorded signal consists of a superposition of sine waves; this information can be used to calculate the masses of the detected ions using a Fourier transform equation. In FT-ICR a superconducting magnet is used to generate a highly stable magnetic field resulting in very high levels of resolution.

1.2.3 SIMS-TOF

SIMS utilizes beams of primary monoatomic (e.g., Bi^+ , Au^+) or polyatomic (C_{60}^+ , Bi_{+3}^+ , Cs_{+n}^+) ions in a vacuum for ionization (Fletcher, Vickerman, & Winograd, 2011). During SIMS the sample surface is bombarded with a beam of primary ions, causing molecules on the surface of the sample to be released, a process referred to as sputtering. During sputtering energy from the primary ions is transferred to the analyte molecules. Around 1% of the sputtered analyte molecules possess an electric

charge; these ions can be detected by a mass analyzer, such as a TOF (as described earlier). The energy of the primary ion beam is typically high compared to the covalent bond energies of the analyte molecules, resulting in their fragmentation. This makes SIMS a “hard” ionization technique, although polyatomic ion beams offer softer ionization than monoatomic ion beams (Bodzon-Kulakowska & Suder, 2016). SIMS requires no matrix and offers higher depth and spatial resolution ($< 10\text{ }\mu\text{m}$) as compared to MALDI. However, MALDI is more sensitive as it has access to higher yields of analyte and is able to detect larger molecules (Fletcher, Lockyer, & Vickerman, 2011), with the highest detectable mass range of SIMS sitting at around 1 kDa.

1.2.4 DESI

DESI uses an electrospray source of highly charged aqueous spray droplets under ambient conditions to gently desorb and ionize analyte molecules from the sample surface (Takáts, Wiseman, Gologan, & Cooks, 2004). An advantage of DESI is that it allows for the analysis of a sample in its native state. In the ion source an electrospray emitter is used to generate charged microdroplets from a solvent, which create a thin film on the sample surface that dissolves the analyte molecules. Secondary microdroplets are formed from the kinetic impact of the primary droplets, which contain the highly charged solvent and dissolved analyte. These secondary microdroplets are ionized from the sample surface to the mass spectrometer inlet. A DESI source can be coupled to most standard electrospray mass spectrometers (Hsu & Dorrestein, 2015) where following ionization, the analytes are sent via an ion transfer tube to the mass analyser for detection. During ionization little analyte fragmentation occurs, hence DESI is considered a “soft” ionization technique with a spatial resolution of around $50\text{--}100\text{ }\mu\text{m}$ and an upper mass range detection limit of around 2 kDa.



2. PROTEIN MSI IN CANCER RESEARCH

Changes in protein function play a crucial role in the development of cancer. Such alterations can be caused by number of events including altered localization, posttranslational modifications (PTMs), and/or abundance levels. Furthermore, the biology of a tumor depends not only on the cancerous cells but also on their interaction with surrounding stroma, blood vessels, and the immune system (Kriegsmann, Kriegsmann, & Casadonte, 2015). The ability to spatially resolve different proteins while simultaneously

acquiring information concerning their relative abundance is particularly important in heterogeneous diseases such as cancer. Therefore, an approach combining high-resolution MS with in situ spatial analysis of the diseased area has the potential to directly identify and quantify proteins and peptides.

When MSI first entered the field of tissue-based research, it was used to analyze the spatial distribution of proteins (Stoeckli, Chaurand, Hallahan, & Caprioli, 2001). Following this ground-breaking research, different imaging methods have been established and applied to a wide range of biological problems. Among the several MSI techniques, MALDI-MSI is one of the most commonly used. MALDI-MSI has the capability of detecting a broad range of compounds present in a tissue section, provided the compound can be ionized and desorbed into gaseous phase during the MALDI process. High-resolution MS, such as MALDI-FT-ICR, has greater strength in the field of low-molecular weight compounds, while MALDI-TOF is used for the analysis of peptides and small proteins up to 25 kDa (Aichler & Walch, 2015). MALDI-MSI of thin tissue sections can result in the detection of over 500 individual protein signals within the mass range of 2–70 kDa that can be directly correlated with specific morphological regions of the tissue (Chaurand, Schwartz, & Caprioli, 2004).

Protocols have been developed for the preparation of tissue sections and matrix application that provide a high level of reproducibility (Gustafsson et al., 2013). The methodology behind protein MSI, via the analysis of their proteolytic peptides, is straightforward; a digestion enzyme is uniformly applied over the entire tissue surface either by a spotting or spray deposition method. Thereafter, using the same application method, internal calibrants and matrix are applied. Data acquisition is then carried out by a MS (Gustafsson et al., 2013), where analyte ionization, mass analysis, and detection occur. This results in the subsequent generation of mass spectrums at regular series of points across the tissue section. The mass spectra produced are then visualized as ion intensity maps; generating two-dimensional distribution maps of the detected mass spectra containing information about the location and the relative intensity of the detected peptides. This visualization is similar in nature to images generated by histological or IHC analysis, but with the advantage that multiple molecules can be examined within a single measurement (Aichler & Walch, 2015) (Fig. 1).

Three-dimensional MALDI-MS images can also be generated by analyzing a number of consecutively cut tissue sections, and following data acquisition, combining the results to generate three-dimensional maps (Andersson, Groseclose, Deutch, & Caprioli, 2008; Sinha et al., 2008).

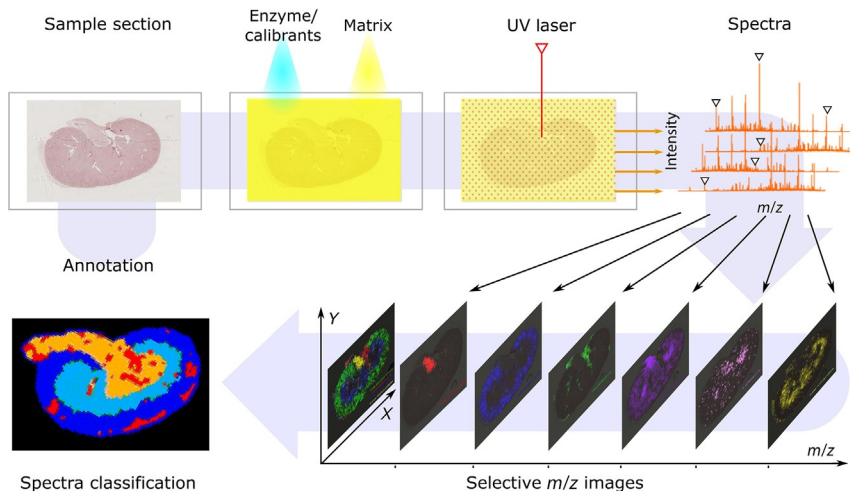


Fig. 1 Work flow of MALDI mass spectrometry imaging.

Recently, methods were developed for the integration of three-dimensional volume reconstructions of spatially resolved MALDI-MSI ion intensity images taken of whole mouse heads with corresponding *in vivo* data provided by magnetic resonance imaging (MRI) (Sinha et al., 2008). The added spatial dimension provides contextual information to the MALDI-MSI data that allow the interrogation of proteomic relationships within a tissue volume. The work displayed a correlation between the *in vivo* MRI and proteomic profiles obtained by MALDI-MSI, with both techniques able to distinguish tumor from healthy tissues (Sinha et al., 2008).

Until recently, MALDI-MSI had been carried out almost exclusively on fresh frozen samples as it was believed proteins were inaccessible from FFPE tissue due to the cross-linking caused by the formalin fixation. However, this limitation has been overcome and equal number of proteins can be identified from both fresh frozen and FFPE tissues (Diehl et al., 2015). The use of FFPE tissues in cancer research is a huge advantage, given that the samples are well preserved by the formalin fixation process and can be stored for long periods of time at room temperature. Another attractive feature is that the samples are often well documented as they have been processed by clinical pathology departments, meaning information on patient treatment response, disease progression, and other relevant clinical data have been cataloged (Gorzolka & Walch, 2014). Additionally, FFPE tissues are well suited for the construction of tissue microarrays (TMAs), where needle core biopsies from multiple tissue samples are assembled into a paraffin block. However,

fresh frozen TMAs have also been studied in a limited context (Fejzo & Slamon, 2010).

One of the main advantages of MALDI-MSI is the technique's ability to detect changes in disease-associated protein expression prior to histological transformation, making it a useful tool in the identification of early stage disease biomarkers (McDonnell et al., 2010). There have been number of cancer studies that have successfully applied peptide imaging to the discovery of potential biomarkers with the capacity to differentiate tumors of various subtypes, stages, and/or degree of metastasis.

2.1 Distinguishing Tissue Types by Peptide MSI

The ability of peptide MSI to distinguish specific types of tissue, such as regions of healthy, cancer, stroma, and vasculature, has been shown multiple times (Mittal, Klingler-Hoffmann, Arentz, Zhang, et al., 2016). Schwamborn et al. performed MSI on prostate cancer specimens and was able to classify the cancerous tissue from normal tissue with 85% sensitivity and 90% specificity, with an overall cross validation accuracy of 88% (Schwamborn et al., 2007). Moreover, MSI of prostate cancer TMAs on a large scale allowed for the immediate prioritization of MSI signals based on associations with clinic-pathological and molecular data (Steurer et al., 2013). A comparison of these signals with clinic-pathological features revealed statistical association with a favorable phenotype such as grade, stage, and relationship with prolonged time to prostate-specific antigen recurrence (Steurer et al., 2013). Peptide MSI experiments have been able to discriminate normal, preinvasive, and invasive lung tumor tissues with an accuracy of 90% (Rahman et al., 2011), to differentiate normal tissue from oral squamous cell carcinoma (Patel et al., 2009), to discriminate normal tissue from gastric cancer tissues (Kim et al., 2010), and have identified mitogen-activated protein kinase/extracellular signal-regulated kinase kinase 2 (MEKK2) as a marker with the capacity to accurately discern cancerous prostate from normal tissue (Cazares et al., 2009).

One of the most challenging aspects of neurochemistry is the detection of endogenous neuropeptides due to their low in vivo concentrations ranging from pico- to femto-molar levels, which makes it difficult to obtain enough material for quantitative analysis (Andersson, Andren, & Caprioli, 2010; Strand, 2003). MALDI-MSI helps overcome this issue as it allows the direct detection of neuropeptides in discrete regions of a brain sections, enabling the study of the physiological and disease-related

metabolic processing of neuropeptides (Andersson et al., 2010). Recently, Andersson et al. has been able to sequence and identify neuropeptides by in situ tandem mass spectrometry (MS/MS) directly off rat brain sections using a MALDI-QTOF mass spectrometer (Andersson et al., 2008).

Furthermore, peptide MALDI-MSI has been employed to differentiate benign regions from malignant in ovarian cancer (Lemaire, Ait Menguellet, et al., 2007) and in thyroid cancer (Pagni et al., 2015). Likewise, biomarkers have been identified that could distinguish Hodgkin's lymphoma from lymphadenitis (Schwamborn et al., 2010).

2.2 Determining Tumor Margins by Peptide MSI

One of the major aims of clinical oncology is to ensure a tumor has been completely removed during surgery to minimize the possibility of disease recurrence (Han et al., 2011). Defining tumor margins also plays a pivotal role in cancer staging, which subsequently influences the administration of treatment. Moreover, the molecular characterization of tumor margins will assist in understanding the process of invasion into surrounding tissues and may help in the clinical management of disease (Agar et al., 2010). MALDI-MSI has been proposed as a valuable tool for gaining information with regards to tumor margins and heterogeneity (Chaurand, Sanders, Jensen, & Caprioli, 2004). Peptide MSI has interestingly revealed histologically normal tissue adjacent to renal cell carcinoma regions, share many molecular characteristics with that of the tumor tissue, providing unprecedented insight into cancer development (Oppenheimer, Mi, Sanders, & Caprioli, 2010). The technique has also been used to identify protein biomarkers for the differentiation of delineated hepatocellular carcinoma from adjacent cirrhotic tissue (Le Faouder et al., 2011). Analysis of ovarian cancer interface zones, the regions between tumor and normal tissues, by MSI revealed a unique peptide profile from both the tumor and normal tissue, and detected plastin 2 and peroxiredoxin 1 as interface markers specific to ovarian cancer (Kang et al., 2010).

2.3 Prediction of Metastasis by Peptide MSI

Metastatic status is often the defining feature in determining if a patient will receive chemotherapy and is currently resolved by the removal and analysis of local lymph nodes, which frequently results in noncancer-related health complications. Identification of primary tumor markers of metastasis

would prevent the unnecessary removal of lymph nodes, thereby reducing patient suffering and associated treatment costs, and allow the introduction of more personalized cancer therapies (Casadonte et al., 2014). However, identification of metastasis based on analysis of the primary tumor tissues is challenging, especially when the primary carcinoma site is unknown (Pavlidis & Pentheroudakis, 2016). A statistical classification model based on peptide MSI data has been developed for determining the metastatic status of primary cancers of pancreatic origin and their metastasis to secondary breast cancer with an overall accuracy of 83% (Casadonte et al., 2014). In another study the HER2 status of gastric cancers could be predicted with 90% accuracy using peptide expression patterns originating from breast cancers (Balluff et al., 2010). These studies display the potential of peptide MSI in the profiling of cancers, independent of their site of origin.

The likelihood of lymph node metastasis (LNM) is one of the most important factors to consider while determining the appropriate treatment for a cancer patient. Several studies have applied MALDI-MSI to the prediction of LNM status. Yanagisawa et al. was able to classify primary lung tumors with and without LNM with an accuracy of 85% (Yanagisawa et al., 2003). Different peptide signatures have been identified with the ability to distinguish melanoma (LNM) from control lymph nodes (Hardesty, Kelley, Mi, Low, & Caprioli, 2011). Similarly, the overexpression of S100A10, thioredoxin, and S100A6, have been shown to be highly associated with LNM in papillary thyroid carcinoma (PTC) (Nipp et al., 2011), and it was suggested the three proteins may be used for risk stratification regarding metastatic potential in PTC. Other protein biomarkers have been suggested for the prediction of LNM in several cancers from peptide MSI experiments, such as S100A8 and S100A9 for gastric adenocarcinoma (Choi et al., 2012), COX7A2, TAGLN2, and S100-A10 for monitoring the Barrett's adenocarcinoma development as well as for predicting regional (LNM) and disease outcome (Elsner et al., 2012).

Recently, it has been shown that the LNM status of primary endometrial tumors could be predicted with an accuracy of 88% using a classification system based on tissue peptide markers detected by MALDI-MSI (Mittal, Klingler-Hoffmann, Arentz, Winderbaum, et al., 2016; Winderbaum, Koch, Mittal, & Hoffmann, 2016). In the study FFPE-TMAs comprised of duplicate tumor tissue cores from endometrial cancer patients diagnosed with and without LNM were analyzed by MALDI-TOF-MSI. An overview of the sample preparation and approach used for the proteomic marker

identification by MALDI-MSI-(top panel) and LC-MS/MS (bottom panel) is outlined in [Fig. 2](#).

2.4 Analysing Chemoresponse by Peptide MSI

Innate and acquired chemoresistance significantly impact the survival rates of late stage cancer patients. Administration of chemotherapy to patients with specific cancer types has traditionally been standardized, despite the knowledge that response can differ significantly among a cohort of patients due to genetic diversity and tumor heterogeneity. Identification of markers or tests that predict resistance to standard chemotherapy regimens would help save patients from damaging, ineffective treatments, and may help determine if an alternative treatment course may be more beneficial.

IHC is one of the most widely applied proteomic techniques in clinical practice. However, IHC is often unable to distinguish the different isoforms of the same protein that may play different role in the disease development. Recent studies using MSI have been able to distinguish two isoforms of the protein defensin. This is of particular interest given the 2 defensin protein isoforms respond differently in breast cancer tumors when exposed to chemotherapy ([Bauer et al., 2010](#)). The finding was discovered during the analysis of a breast cancer patient receiving neoadjuvant paclitaxel and radiation therapy, and was one of the first biomarker discovery studies to utilize MALDI-MSI ([Bauer et al., 2010](#)). Mitochondrial defects have also been found and linked to chemotherapy response in oesophageal adenocarcinoma patients using peptide MSI ([Aichler et al., 2013](#)), displaying the techniques potential to change classifications of tumors and chemoresponse.

2.5 Identification of Diagnostic and Prognostic Markers by Peptide MSI

A high proportion of cancer patients are diagnosed in the later stages of disease, when the cancer has already metastasized to distant lymph nodes or organs. Detecting cancer in the earlier stages of disease, prior to metastasis, would help increase survival rates and reduce the treatment burden for patients. In a genetically engineered mouse model, peptide MSI was used to detect several discriminative peptide masses that could distinguish between intraepithelial neoplasia, intraductal papillary mucinous neoplasm, and normal pancreatic tissue ([Grüner et al., 2012](#)). Using MALDI-MSI to analyze TMAs, significant differences between papillary and clear cell renal cell cancer were found ([Steurer, Seddiqi, et al., 2014](#)). Analysis of bladder

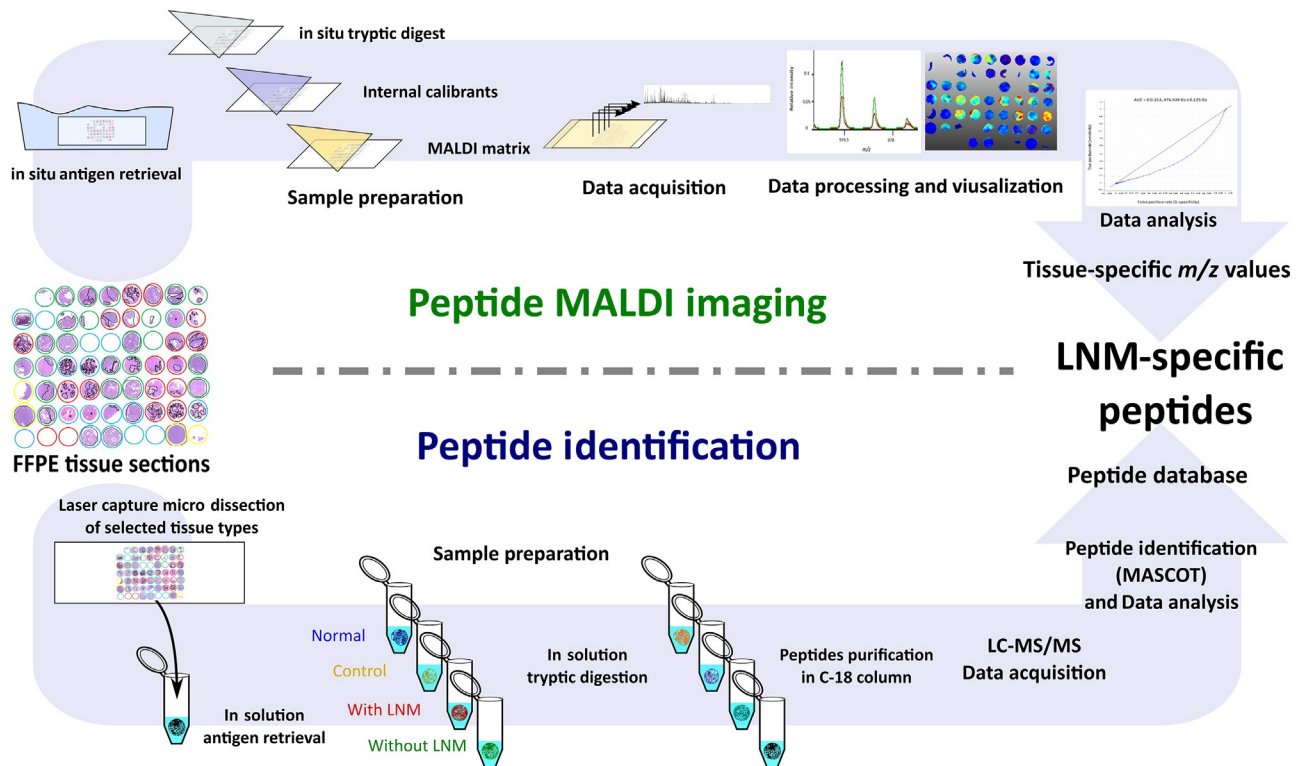


Fig. 2 Work flow for proteomic marker identification by peptide MALDI-MSI and label-free quantification.

cancer TMAs by MSI lead to the identification of multiple signals in the mass spectra associated with unfavorable tumor phenotype and clinical characteristics, such as tumor recurrence, progression, and patient survival time (Steurer, Singer, et al., 2014).

MALDI-MSI has also been used to identify a protein signature that can accurately determine the HER2 status in breast cancer with sensitivity of 83%, for specificity of 92%, and overall accuracy of 89% (Rauser et al., 2010). A seven-protein signature of novel tissue markers was identified as an independent indicator of unfavorable overall survival in intestinal type gastric cancer (Balluff, Rauser, et al., 2011). Other investigators identified histone 4 expression as specific to poorly differentiated gastric cancer tissues by peptide MSI (Morita et al., 2010). A comparison of genomically stable (diploid) and unstable (aneuploid) colon cancer tissues against normal mucosa by MSI identified thymosin beta 4 as an independent prognostic marker for colorectal cancer (Gemoll, Strohkamp, Schillo, Thorns, & Habermann, 2015). Analysis of lung cancer TMAs containing 112 needle core biopsies were used to generate a support vector machine model utilizing MSI information on 73 peptides that could classify adenocarcinomas with an accuracy of 97% and squamous cell carcinomas with an accuracy of 98% (Groseclose, Massion, Chaurand, & Caprioli, 2008).

2.6 Characterisation of Intra- and Intertumor Variability by Peptide MSI

Understanding the spatiotemporal patterns and dynamics of intratumor heterogeneity is crucial for clinical management and designing personalized targeted regimens for cancer patients (Renovanz & Kim, 2014). Balluff et al. showed for the first time that MALDI-MSI of tumor tissues can reveal microscopically indistinct tumor populations that might have adverse impacts in clinical outcome (Balluff et al., 2015). In the proof-of-principal study tumor-specific mass spectra were grouped using an advanced statistically clustering algorithm in order to obtain segmentation maps of molecularly distinct regions. The distinct regions were then statistically compared with patient clinical data to identify *phenotypic* tumor subpopulations (Balluff et al., 2015). This approach revealed that several of the detected tumor subpopulations are associated with different overall survival characteristics of gastric cancer patients ($p=0.025$) and the presence of regional metastasis in patients with breast cancer ($p=0.036$). This approach enables researchers to gain deeper insights into the underlying biological

process and changes of tumor subpopulations on a genetic, metabolic, and proteomic level, which may result in novel targeted therapies (Balluff et al., 2015).

The ability of MALDI-MSI to reveal inter- and intratumoral biomolecular heterogeneity of histologically contiguous tumors has been shown, specifically in the capacity to distinguish different types and grades of myxoid sarcomas (Willems et al., 2010). By applying hierarchical clustering to a gastric cancer MALDI-MSI dataset, it has been demonstrated the different histological regions could be distinguished solely on the basis of their protein and peptide expression profiles (Deininger, Ebert, Fütterer, Gerhard, & Röcken, 2008). Ion mobility separation combined with MALDI-MSI has been used to classify prostate cancer TMAs and has identified a number of proteins, including type I collagen and tumor necrosis factor receptor, directly from the tissue that could discriminate different tumor classes (Djidja et al., 2010). Low grade urothelial neoplasia could be separated from high grade urothelial neoplasia with an overall cross validation rate of 97.18% upon analysis of peptide MSI data (Oezdemir et al., 2012). Likewise, based on protein profiling, a grading system for meningioma was developed (Agar et al., 2010). Proteomic signatures associated with breast cancer-activated stromal tissues have been identified by MSI, and hierarchical clustering of the proteomic signals could be used to distinguish the activated intratumoral stroma from the quiescent extratumoral stroma (Dekker et al., 2014). High grade and low grade glioma can also be distinguished by MALDI-MSI (Chaurand et al., 2004).

2.7 Practical Considerations for Proteolytic Peptide MSI: Sample Preparation

MSI of proteolytic peptides in cancer is frequently performed on fresh frozen or FFPE tissues; each sample type has unique advantages and limitations with regards to MSI. The protein component of fresh tissues is easily accessible, especially when compared to FFPE tissues, due to a lack of chemical fixation. Hence, no retrieval steps are required prior to the proteolytic digestion of fresh frozen samples. A disadvantage of fresh frozen tissues, however, is that they require rapid freezing upon collection and storage at extremely low temperatures to inhibit endogenous enzymatic degradation (Fentz, Zornig, Juhl, & David, 2004). Fresh tissues also require solvent washing to remove lipids and other biomolecules that may compete for ionization and interrupt the detection of peptides. Washing steps must be performed carefully, as extensive washing may cause the loss of soluble proteins

(McDonnell et al., 2010; Seeley, Oppenheimer, Mi, Chaurand, & Caprioli, 2008). Two significant advantages of FFPE samples are that they are widely used in modern clinical practice, meaning very large sample banks have often been developed, and that the fixation process allows for storage of the samples at room temperature for indefinite periods of time without loss of morphological information (Wisztorski et al., 2007). The downfall is that formalin fixation results in the formation of methylene bridges between amino residues that causes protein cross-linking (Metz et al., 2004). This cross-linking makes the proteins inaccessible to proteolysis and must be reversed prior to any further preparation steps. Heat-induced antigen retrieval prior to in situ proteolytic digestion is a widely used method for reversing protein cross-links, which allows for the peptide profiling of FFPE tissues that is comparable to fresh frozen tissue analysis (Casadonte & Caprioli, 2011).

2.8 Practical Considerations for Peptide MSI: Spatial Resolution

Currently, MALDI is the most popular ionization method applied in the field of MSI. The success of analyte ionization in a MALDI experiment relies significantly on the matrix used; the type of matrix applied will impact on the mass range of the detectable ions, the intensity of detection, and the minimum spatial resolution achievable (Lemaire et al., 2006). For peptide MALDI-MSI commonly used matrices include α -cyano-4-hydroxycinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB) (van Hove, Erika, Smith, & Heeren, 2010). The homogeneity of matrix crystallization is a major limiting factor with regard to the spatial resolution that can be achieved by MALDI-MSI, as larger and nonuniform matrix crystals result in lower resolutions. DHB matrix is prone to producing large, needle-like crystals, making CHCA a more suitable matrix as it commonly forms small, homogeneous crystals resulting in higher image resolution (Schwartz, Reyzer, & Caprioli, 2003). The way in which the matrix is applied also has a significant impact on the spatial resolution, with the application of larger size droplets increasing the likelihood of analyte displacement. Instrumentation has been developed that allows matrix solutions to be sprayed or spotted onto the sample surface in a more homogeneous manner compared to manual application (Aerni, Cornett, & Caprioli, 2006). For example, using a matrix spray device CHCA can be applied to form uniform crystals that allow for spatial resolution peptide imaging of $<25\ \mu\text{m}$ (McDonnell et al., 2010; Schuerenberg & Deininger, 2010).

Laser specifications also have a significant impact on the spatial resolution achievable. As opposed to the limitations of matrix crystallization, laser beam resolutions can easily reach less than 10 μm , theoretically allowing for peptide imaging at the single cell level (Balluff, Schöne, Höfler, & Walch, 2011). However, reducing the size of the laser beam concurrently reduces the yield of ionized analytes. Higher frequency lasers and overlapping laser shots can counteract the issue of lower ion yields during detection (Jurchen, Rubakhin, & Sweedler, 2005). Thus, currently matrix is the limiting factor for the spatial resolution achievable in MALDI-MSI.

2.9 Practical Considerations for Peptide MSI: Mass Analysers

The type of mass analyser used for MSI will largely determine the mass resolution and mass range of the detectable ions. TOF mass analysers are commonly coupled to MALDI sources and have high acquisition speeds but typically lower mass resolution (Caprioli et al., 1997). Higher mass accuracy can be achieved through the use of internal or external calibrants. Other commonly used mass analysers include FT-ICR and Orbitraps; these instruments have significantly higher mass accuracy and resolution (Taban et al., 2007).

Internal or external calibrants can be used to improve the mass accuracy of acquired spectra during or after data acquisition (Gobom et al., 2002). Calibrants are often comprised of a small number of specific analytes with known masses that span a specific mass range. Within this mass range, the use of calibrants can significantly increase the mass accuracy of the detected analytes. However, the mass error of any analyte molecules that fall outside the mass range of the calibrants cannot be corrected (Coombes, Koomen, Baggerly, Morris, & Kobayashi, 2005).

2.10 Practical Considerations for Peptide MSI: Identification

Currently, MALDI-MSI enables the discrimination of heterogeneous tissue types, but does not allow for the direct identification of the peptide sequences. During data acquisition the intact mass, or the m/z , of the detected peptides is recorded. This allows peptide mass fingerprinting (PMF) to be performed, a process where the intact peptide masses are matched back to a database containing the *in silico* digested proteome or proteomes of specific species (Pappin, Hojrup, & Bleasby, 1993). Due to a lack of peptide sequence information, often PMF can provide a list of potential peptide identifications but is not able to provide definitive

identifications for those peptides with similar masses. However, the higher the mass accuracy of the instrument used the more likely it is a definitive identification will be gained (Horn, Peters, Klock, Meyers, & Brock, 2004). The use of complementary MS techniques, such as matching data to peptides sequenced by traditional liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) or by in situ tandem mass spectrometry (MALDI-MS/MS), can help clarify peptide sequence identifications (Lemaire, Desmons, et al., 2007). However, there can be significant difficulties associated with these options also. Multiple sequence matches are often encountered when matching MSI m/z values back to LC-ESI-MS/MS data obtained from the same sample. Generating high quality in situ MS/MS spectrums is often challenging due to the issue that acquisition of very complex peptide mass spectra directly off tissue results in poor fragmentation and high MS/MS complexity, both of which limit direct peptide sequencing from tissues (Gustafsson et al., 2012).



3. LIPID MSI IN CANCER RESEARCH

3.1 Profiling Lipids in Cancer by DESI-MSI

The soft ionization technique of DESI has been developed for the MSI of small molecules and lipids. DESI uses highly charged solvent droplets traveling at high velocities for analyte ionization in an ambient atmosphere. An advantage of DESI for the analysis of lipids and small molecules is that no sample preparation steps are required, unlike MALDI where matrix deposition is a necessity. A further advantage is that a DESI source can be coupled to most standard mass analysers, such as a FT-ICR, Orbitrap, Ion Trap, or a TOF.

In practice, DESI-MSI has been applied to the study of human cancers for the discrimination of tumor subtypes and grades, as well as for the identification of tumor margins (Calligaris et al., 2014; Eberlin et al., 2014). One particular application used DESI-MSI to generate lipid profiles for the classification of human brain tumors (Eberlin et al., 2012). This study was able to classify the subtype, grade, and tumor concentration of 36 human glioma samples via lipidomics imaging in agreement with the histopathology diagnoses (Eberlin et al., 2012). Lipidomic analysis by DESI-MSI has identified the phospholipids phosphatidylserine (PS), phosphatidylinositol (PI), and phosphatidylethanolamine as biomarkers for distinguishing healthy tissue

from benign and malignant tumor tissues (Abbassi-Ghadi et al., 2014; Dill et al., 2011, 2009; Eberlin et al., 2012).

Discriminatory lipid signatures between cancerous and normal breast tissue have been detected by DESI-MSI (Calligaris et al., 2014), with the delineation of tumor margins possible through the analysis of PI (18:0/20:4). PI (18:0/20:4) has been shown to be abundant at the tumor center and tumor edge, whereas it was absent or weak in normal tissues. This lipid sharply stratifies the tumor margin in agreement the histopathological staining. Using DESI-MSI, another study found PS (20:4/18:0) and PI (18:0/20:4) are significantly lower in concentration in primary tumor tissues and LNMs when compared to normal lymph node tissues (Abbassi-Ghadi et al., 2014). The same lipid, PI (18:0/20:4), was found to be a biomarker for distinguishing benign and malignant breast tumors (Yang et al., 2015). Collectively, the discussed research shows MSI lipid profiling of cancer tissues can be used for classifying tumor types, with high specificity in differentiating histopathologic grade, and in the mapping of tumor margins.

3.2 Profiling Lipids in Cancer by SIMS-MSI

A less popular ionization method for MSI in the lipidomics field is static SIMS. Generally, a static SIMS source is coupled to a TOF (or QTOF) instrument, and is particularly suitable for the analysis of intact molecules such as lipids and metabolites. It uses a beam of energetic primary ions with either monatomic source, such as Cs^+ and Ga^+ for limited mass range of low hundreds Daltons, or cluster ions such as SF_5^+ , Bi_3^+ , and Au_3^+ that have extended the mass range to a couple of thousands Daltons (Lanni, Rubakhin, & Sweedler, 2012). The potential of SIMS-TOF imaging for lipidomics has been shown in rat brain sections, a well-established model system for MSI (Benabdellah et al., 2010; Passarelli & Winograd, 2011; Sjoval, Lausmaa, & Johansson, 2004; Sjoval, Lausmaa, Nygren, Carlsson, & Malmberg, 2003). In the field of cancer lipidomics, it is frequently used in cellular imaging, where single cells or a population of cells are isolated after drug treatment for the characterisation of molecules on the cell surface. For example, lipid composition analysis of individual human breast cancer stem cells (CSCs) reported significantly lower expression levels of palmitoleic acids FA (16:1) as compared to nonstem cancer cells (NSCCs), and is thought to have been successfully characterized from complex clinical specimens using TOF-SIMS (Waki et al., 2014).



4. GLYCAN MSI IN CANCER RESEARCH

Glycosylation is the enzymatic process where oligosaccharides, polysaccharides, and carbohydrates (i.e., glycans) are attached to proteins, lipids, or other organic molecules (Ohtsubo & Marth, 2006). In recent years, the word “glycomics” has emerged to describe the study of glycan structures from an organism’s “glycome.” The mammalian glycome repertoire is comprised of 10 known monosaccharides. These monosaccharides include xylose (Xyl), glucose (Glc), galactose (Gal), mannose (Man), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), fucose (Fuc), glucuronic acid (GlcA), iduronic acid (IdoA), and *N*-acetylneuraminic acid (NeuAc or Neu5Ac), a sialic acid (Wuhrer, Deelder, & van der Burgt, 2011). Glycan complexity arises from variation in expression of specific glycosyltransferases, resulting in a diversity of monosaccharide arrangements. Furthermore, modifications such as sulfation, acetylation, or phosphorylation increase the complexity of glycosylation.

The major type of glycosylation is protein glycosylation, where glycans are attached to asparagine residues by their amine group (N-linked) or at serine/threonine residues by their hydroxyl group (O-linked) (North, Hitchen, Haslam, & Dell, 2009). Protein glycosylation is an important PTM which has relevance in many biological processes such as cell signaling, immune responses, extracellular interaction, and cell adhesion (Varki, 1993). In the tumor microenvironment, aberrant protein glycosylation such as the expression of truncated glycans have been well described in various cancers (Dube & Bertozzi, 2005). These structures may resemble glycomic “fingerprints” which discriminate between healthy and cancerous tissues or potentially discriminate between various cancer subtypes (Abbott et al., 2010).

One of the most common glycosylation changes in various cancers is an increase in branching of N-glycans, where large tetra-antennary structures are formed by the expression of *N*-acetylglucosaminyltransferase V (GnT-V) (Lau & Dennis, 2008). The presence of increased branching antennas creates additional sites for the addition of NeuAc by sialyltransferases in various cancers (Bull et al., 2013). Likewise, terminal modifications on glycan structures such as fucosylation on the cancer cell surface can also give rise to the presence of Lewis and sialyl Lewis antigens (sialyl Lea; [Neu5Ac α 2-3Gal β 1-3(Fuc α 1-4)GlcNAc β] and sialyl Lex; [Neu5Ac α 2-3Gal β 1-4(Fuc α 1-3)

GlcNAc β]) which have been shown to correlate with tumor progression and metastasis (Nakagoe et al., 2000). It is evident that aberrant glycosylation is indeed a key event in metastasis and therefore the discovery of cancer-specific protein glycosylation markers would be beneficial in the staging and treatment of patients (Drake et al., 2010).

As glycomics gains recognition in this postgenomics era, MS-based methodologies are becoming indispensable and routinely used for the reliable profiling of glycans from clinical samples. Glycan MSI is a recent MS-based development which allows the visualization of the tissue-specific spatial distribution of glycans. N-glycans have been the most comprehensively studied type of protein glycosylation, as they can be easily released from proteins by enzymatic cleavage using PNGase F (Jensen, Karlsson, Kolarich, & Packer, 2012). The majority of these studies have applied optimized N-glycan MALDI-MSI methods to FFPE tissues, with the most recent focus on various cancer tissues.

One of the first N-glycan MALDI-MSI methods to emerge was published in 2013 (Powers et al., 2013). This proof-of-principle paper described the measurement of N-glycans by MALDI-MSI across fresh frozen mouse brain tissue. The m/z values from the sum spectrum were assigned to N-glycans based on further MS structural analysis, for example, high-performance liquid chromatography (HPLC)-MS. The authors concluded that this new approach could lead to novel disease-related targets for biomarker and therapeutic applications. Subsequently, N-glycan MALDI-MSI has been successfully applied to the analysis of glioblastoma tumor xenografts compared to normal brain tissues (Toghi Eshghi et al., 2014), where it was found specific N-glycans associated with distinct tissue regions. N-glycan imaging of FFPE pancreatic and prostate cancers, and a human hepatocellular carcinomas in TMAs found tumor and nontumor regions could be distinguished in the TMA format, enabling high-throughput analysis (Powers et al., 2014).

The use of complimentary MS methods, such as liquid chromatography tandem MS (LC-MS/MS) with MALDI-TOF-MSI, have been developed to enable the mapping of tissue-specific N-glycans at high resolution (Gustafsson et al., 2015). The application of these techniques has resulted in several interesting findings. A panel of over 30 N-glycans has been detected from FFPE human hepatocellular carcinoma tissues (Powers, Holst, Wuhler, Mehta, & Drake, 2015), from which the N-glycans were extracted and the sialylated species were stabilized by an off-tissue ethylation reaction. N-glycan sialylation linkage-specificity was further found to

correlate with the different tissue types analyzed. A novel method for the analysis of N-glycan sialylated species (i.e., α 2–3 and α 2–6) by MALDI-MSI in colon carcinoma found sialylated N-glycans with α 2–3 linkage were observed in stroma, tumor, and necrotic regions, while sialylated N-glycans with α 2–6 linkage were observed in necrotic, collagen-rich, and red blood cell regions (Holst et al., 2016).

The analysis of FFPE stage III ovarian cancer tissues found high mannose structure (Man5 + Man3GlcNAc2) in the tumor regions, while triantennary complex structures (Hex3HexNAc3Deoxyhexose1 + Man3GlcNAc2) were observed in the stromal regions (Fig. 3). This observation was consistent between not only the patients investigated, but also between each N-glycan family. The tissue analyzed in the study had been annotated by a pathologist, and it was found specific glycan structures were differentially detected in specific tissue regions, such as tumor, stroma, and adipose tissue.

In summary, N-glycan MALDI-MSI has developed immensely in recent times, with improvements in areas such as sensitivity and mass accuracy,

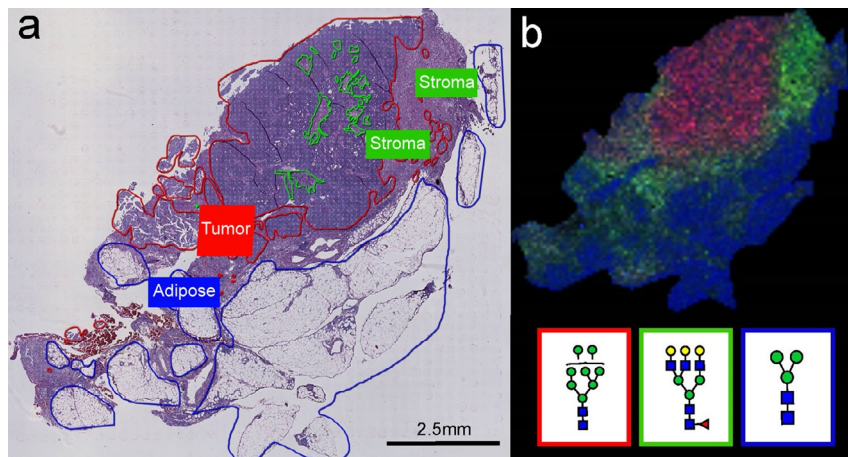


Fig. 3 Glycan analysis of ovarian cancer. The formalin-fixed paraffin-embedded stage III ovarian cancer section was treated with antigen retrieval prior to printing of 15 nL/spot dialyzed PNGase F with 250 μ m spacing. 2,5-DHB (20 mg/mL) was sprayed onto the sections and MS spectra were acquired by oversampling at 100 μ m intervals using a MALDI-TOF/TOF MS instrument. (A) Haematoxylin and eosin stain of the ovarian cancer section with tumor (red), stroma (green), and adipose (blue) regions annotated by a pathologist. (B) An ion intensity map of m/z 1743.7 (red), m/z 2174.9 (green), and m/z 933.3 (blue) on the ovarian cancer section. Yellow circle, galactose (Gal); green circle, mannose (Man); blue square, N-acetylglucosamine (GlcNAc); red triangle, fucose (Fuc).

compositional analysis by complementary techniques (i.e., LC-MS/MS and lectin histochemistry staining), and identification of linkage-specific sialylated species by in situ derivatization methods. However, further work in the field is required to properly answer clinical questions and discover cancer-specific markers. TMAs have not yet been fully utilized in N-glycan MSI experiments, but offer promise in providing researchers and clinicians with high-throughput data for analysis.



5. DRUG IMAGING IN CANCER RESEARCH

Quantification of drugs and their metabolites is crucial in oncology where drug distribution within the tumor tissue is thought to play a pivotal role in response to therapy and could partly explain the variable response rates observed among patients with similar tumor types (Minchinton & Tannock, 2006). Traditionally, autoradiography has been used to visualize the localization of a drug. However, this technique suffers from some significant limitations. One of them is that the synthesis of the radiolabeled drug is an expensive and time-consuming process. Second, the technique does not provide information on the molecular structure. Hence, metabolites either cannot be distinguished from the parent drug or are not detected if the radiolabel is lost in a metabolic process (Kertesz et al., 2008). The other clinically used noninvasive drug imaging techniques are positron emission tomography and MRI. These techniques can provide information on the drug distribution but suffers from the limitation of low spatial resolution. Due to the relatively simple sample preparation steps required for MSI the technique has rapidly emerged as an alternative to the conventional methods of analysis. MSI allows detailed distribution analysis of the parent drug and its metabolites in a single experiment and without any labeling of the targeted compound in contrast to conventional immunohistochemical methods. Such drug characterization in situ, by both spatial and temporal behaviors within tissue compartments, provides a new understanding of the dynamic processes impacting drug uptake and metabolism at the local sites targeted by a drug therapy (Kwon et al., 2015). Multiple MSI techniques including MALDI-MSI, SIMS, and nanostructure initiator mass spectrometry (NIMS) have been used to analyze the distribution of the drug and metabolites. However, MALDI-MSI is the most frequently used method for looking at anticancer drug distribution in tumor samples.

5.1 MALDI-MSI on Tissue Sections

MALDI-MSI was used for the first time in 2003 to directly analyze and image pharmaceutical compounds in intact tissues (Reyzer, Hsieh, Ng, Korfmacher, & Caprioli, 2003). From then, the technique has been applied to the imaging of a number of drugs. Prideaux et al. applied MALDI-MRM-MSI to visualize the distribution of the second line tuberculosis drug moxifloxacin at a range of time points after dosing in tuberculosis-infected rabbits, which was further validated by quantitative LC-MS/MS of lung and granuloma extracts from adjacent biopsies taken from the same animals (Prideaux et al., 2011). Drug distribution within the granulomas was observed to be inhomogeneous, and very low levels were observed in the caseum in comparison to the cellular granuloma regions (Prideaux et al., 2011). Fehniger et al. measured the occurrence of inhaled bronchodilator, ipratropium, within human bronchial biopsies obtained by fiber optic bronchoscopy shortly after dosing exposure and showed the drug is rapidly absorbed into the airway wall (Fehniger et al., 2011). Nilson et al. provided the first evidence that compounds administered by inhaled delivery at standard pharmacological dosage can be quantitatively detected by MALDI-MSI with accuracy and precision (Nilsson et al., 2010). The distribution of the inhaled drug tiotropium was tracked and quantified in the lungs of dosed rats and a concentration gradient (80 fmol–5 pmol) away from the central airways into the lung parenchyma and pleura was observed (Nilsson et al., 2010). Hsieh et al. visualized the spatial distribution of astemizole (withdrawn in most countries due to its side effects) and its primary metabolite in rat brain tissues and showed that the drug alone is likely to be responsible for the associated central nervous system side effects upon elevated exposure (Hsieh, Li, & Korfmacher, 2010). Using MALDI-MSI the specific distribution of unlabeled chloroquine was examined in the retinas of rats and was found to be similar to autoradiograms results reported previously (Yamada, Hidefumi, Shion, Oshikata, & Haramaki, 2011).

Atkinson et al. showed the distribution of the bioreductive anticancer drug AQ4N (banoxatrone), its active metabolite AQ4, and ATP in treated human tumor xenografts (Atkinson, Loadman, Sutton, Patterson, & Clench, 2007). The distribution of ATP was found similar to that of AQ4N, i.e., in regions of abundant ATP there was no evidence of conversion of AQ4N into AQ4, indicating that the cytotoxic metabolite AQ4 is confined to hypoxic regions of the tumor as intended (Atkinson et al., 2007). Using

MALDI-FT-ICR, Cornett et al. imaged the antitumor drug imatinib and two of its metabolites from a mouse brain glioma and showed the metabolites to be more abundant in tumor region compared to normal, which is consistent with other imaging studies of the drug (Cornett, Frappier, & Caprioli, 2008). In order to better understand the penetration of the anti-cancer drug oxaliplatin, MALDI-MSI of tissue sections from treated rat kidneys was performed and it was observed that the drug and its metabolites were localized exclusively in the kidney cortex, suggesting the drug did not penetrate deeply into the organ (Bouslimani, Bec, Glueckmann, Hirtz, & Larroque, 2010). Sugihara et al. demonstrated the first data on the localization of personalized medicine (vemurafenib) within tumor compartments of malignant melanoma (MM). In a proof-of-concept in vitro study the overexpression and localization of the drug in the MM was shown using MS fragment ion signatures (Sugihara et al., 2014). Using MALDI-MSI, it has been shown the drug irinotecan reaches tumors mainly through microvessels whereas the conversion of the drug into its active form SN-38 in tumors of colorectal cancer was poor compared to normal tissue (Buck et al., 2015). The distribution of tamoxifen in both ER-positive and ER-negative breast cancer tumor tissues as analyzed by MALDI-MSI detected the drug at significantly lower intensities in tumor cells compared with stroma in ER-negative samples (Végvári et al., 2016). Aikawa et al. employed a quantitative MSI method (a combination of MALDI-MSI and LC-MS/MS) in a preclinical mouse model to evaluate the intrabrain distribution of an anticancer drug (Aikawa et al., 2016). While no differences were observed between the mice for the drug concentration, diffuse drug distribution was found in the brain of the knockout mice vs wild mice (Aikawa et al., 2016).

To remove the background signals from matrix degradation observed in MALDI-MSI, the method nanoparticle-assisted laser desorption ionization (n-PALDI) was developed which utilises nanoparticles as matrixes. It was found the protocol could be used to investigate the distribution of anticancer agents in primary tumors and in metastasis, to ascertain whether resistance is related to inadequate drug penetration in poorly vascularized parts of the tumor (Morosi et al., 2013). Using the n-PALDI protocol, the different distribution of paclitaxel in tumor and normal tissues was visualized and related to the dosage-schedules and pathological features of the tumors (Morosi et al., 2013).

5.2 MALDI-MSI on Whole Body Sections

Rohner et al. first demonstrated the MALDI-MSI of drugs in whole body animal sections in 2005 (Rohner, Staab, & Stoeckli, 2005). Since then, a wide range of drugs have been imaged in whole body animal sections, allowing the label-free tracking of both endogenous and exogenous compounds with spatial resolution and molecular specificity (Khatib-Shahidi, Andersson, Herman, Gillespie, & Caprioli, 2006). MALDI-MSI has been used to simultaneously detect drugs and their individual metabolite distributions at various time points across whole body tissue sections (Khatib-Shahidi et al., 2006). Stoeckli et al. measured the distribution of a ^{14}C labeled compound (dosed intratracheally) in whole rat tissue sections using MALDI-MSI and whole body autoradiography (WBA) which displayed good quantitative agreement between both techniques (Stoeckli, Staab, & Schweitzer, 2007). Trim et al. compared the distribution of vinblastine within whole body sections using MALDI-MSI and WBA which displayed MALDI-MSI to be advantageous by separating the drug from an endogenous isobaric lipid (Trim et al., 2008).

5.3 MALDI-MSI on 3D Tissue Cultures

Liu et al. applied MALDI-MSI to the analysis of 3D spheroids in order to assess the distribution of pharmaceuticals and their metabolites. As a proof-of-concept study, MALDI-MSI was applied to in HCT 116 colon carcinoma multicellular spheroids to analyze the distribution of irinotecan, showing the time-dependent penetration of the drug and three metabolites (Liu, Weaver, & Hummon, 2013).



6. DATA ANALYSIS

6.1 Spatial Information

The main advantage of MSI over other methods is the access it provides to spatial information. Possibly the most common approach to leveraging the spatial information in MSI data is simply in displaying results as spatial plots. Fig. 4 shows the results of several analyses of such data. These analyses are blind to the spatial information in the data but their results have spatial patterns that can be easily seen when plotted. Plotting such results next to a stained image of the tissue allows for the molecular information in the MSI data

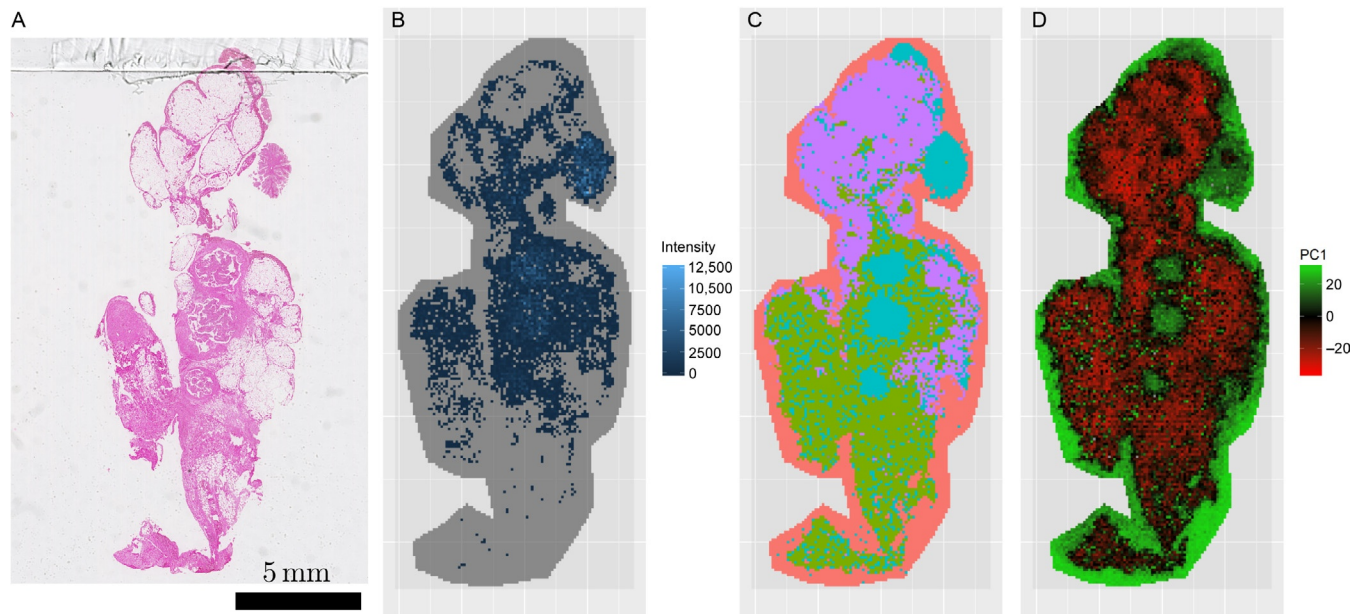


Fig. 4 Results of various analyses of tryptic peptide MALDI-MSI data collected from the same section of tissue. (A) H&E stained section of several primary serous ovarian tumors embedded in surrounding stromal tissue. (B) Intensity of a single peptide peak at $m/z = 1628.8$ confirmed as belong to heterogeneous nuclear ribonucleoprotein A1 which has been shown to be of interest in the past ([Chen, Zhang, & Manley, 2010](#); [Lee et al., 2010](#)). (C) Results of 4-means clustering on the log-intensities including all measured m/z values. (D) Results of the first principle component including all measured m/z values on the log-intensities.

to be correlated to the histology of the tissue, thereby allowing biologically relevant use of the spatial information. Such results can even be overlaid directly with stained images (Walch, Rauser, Deininger, & Höfler, 2008), and used to partially automate and aid in histological annotations in an unbiased manner (Cornett et al., 2006). A popular alternative to the k-means clustering shown in Fig. 4 is the so-called semisupervised hierarchical clustering approach applied by Deininger et al. (2008). These semisupervised methods allow the user to manipulate the clustering results until they agree with the observed histology. Jones et al. (2011) demonstrated how multiple approaches to multivariate analysis, such as k-means and principle component analysis (PCA) (Fig. 4), could be combined in an automated manner to produce more reliable partitioning of spatial regions based on the multivariate MSI data.

Spatial information can also be applied to the partitioning of data in a histology driven fashion. To use Fig. 4 as an example, one could use the histology shown in the H&E stain to distinguish different tissue types such as tumor and surrounding stroma, and then by matching the image to the acquired region, partition the spectra into groups representing specific tissue. These groups of spectra could then be compared in an attempt to detect molecular differences between tissue types—this is also a very common approach used in the analysis of MSI data. Using this technique Oppenheimer et al. (2010) demonstrated that histologically healthy tissues within a margin of tumor tissues share molecular characteristics, and hypothesized that this phenomena could be involved in tumor recurrence postresection. Such information could be used by surgeons to identify an appropriate margin to excise the tissue surrounding a tumor.

6.2 Preprocessing: Peak Detection

MSI data have a uniquely complicated structure due to the spatial information it contains, and additionally can be computationally challenging due to the sheer amount of information recorded during acquisition. The data therefore require novel approaches to analysis when compared to other, more traditional forms of mass spectrometry (Jones, Deininger, Hogendoorn, Deelder, & McDonnell, 2012). Furthermore, the appropriate approach will vary depending on the application. One approach to addressing the quantity of data is to perform some, preferably computationally fast, data reduction early in the data analysis workflow and then work

with the reduced data. This allows for analyses to be performed quickly, but also means that all further analyses will be affected by the choices made in the data reduction step. The most simplistic approach to data reduction is centroiding—essentially reducing spectra to their local maxima. Centroiding is perhaps the most simplistic approach to peak detection. Essentially centroiding involves reducing the acquired spectra to their local maxima, allowing peaks in the spectra to be identified, information relevant to these peaks to be stored and the remainder of the “noise” to be removed. Yang, He, and Yu (2009) provide a good overview of peak detection in the context of MALDI-MS. Wavelet approaches to peak detection, such as those suggested by Du, Kibbe, and Lin (2006), Lange, Gropl, Reinert, Kohlbacher, and Hildebrandt (2006), show some promise but have not yet made it into mainstream software packages, and so do not currently see much use. Regardless, peak detection prior to further analysis is a useful tool as it can reduce the volume of data by as much as three orders of magnitude. However, care must be taken when interpreting final results; it is important to be aware that all analyses will be affected by any peak detection steps and that a quality control analysis of the data following peak detection should be implemented to ensure its performance.

6.3 Classification of FFPE-TMAs and the Importance of Dimension Reduction

The application of MSI to FFPE-TMAs for classification of various diagnostic and prognostic clinical factors is a particularly promising field of research. The ability to use FFPE tissue allows access to large archives including tissue samples and associated patient meta data (Hood et al., 2005). The use of TMAs with MSI facilitates the collection of label-free and untargeted mass data from large numbers of patients (Groseclose et al., 2008), which in turn allows for classification problems to be considered. One of the key obstacles to the classification of such data is that, due to the untargeted nature of MSI, the resulting data are of high-dimensionality. The natural approach to address the dimensionality of such data is to perform some dimension reduction prior to classification. Mascini et al. (2015) proposed the use PCA for dimension reduction prior to classification. PCA is a classical multivariate approach that preserves the maximum amount of variability in the dimension-reduced data. Winderbaum et al. (2016) proposed using a cross correlation analysis (CCA) to rank MSI variables based on their contribution to the highest multivariate correlation to the clinical variable and to reduce the dimension of the data by restricting to only the most highly ranked MSI

variables. Either way, dimension reduction prior to classification is crucial. In order to apply these ideas in clinical practice, large cohort studies and/or clinical trials would need to be carried out to validate the performance of these MS-based classifiers. To date most studies have been exploratory, using small samples of patients as a proof of principle, with a study by [Steurer et al. \(2013\)](#) being a notable exception. Dependent on large scale validation studies, classification of MSI data could potentially be used in the individualisation of treatments, predicting factors such as LNM or potentially even chemotherapy response.



7. CONCLUDING REMARKS

7.1 The Future of MSI: Molecular Pathology

For over 100 years clinical pathology has relied upon a description of morphology by trained specialists for tumor classification and grading. The information that can be gained from such an analysis is limited, however, and is often purely descriptive. During the last decade mass spectrometers have found their way from physics and biochemistry research laboratories into clinical pathology laboratories, bringing together researchers from a wide spectrum of disciplines, such as physics, biochemistry, medicine, pathology, and mathematics. A key feature of MSI is the ability to discern molecular signatures of disease with the additional bonus that the features can be directly related back to the tissue morphology. In the context of clinical oncology, the application of MSI has the potential to shift traditional pathology from descriptive morphological analyses to detailed molecular analyses. As such, the introduction of MSI equipment into pathology laboratories will extend their capabilities and revolutionize management and treatment of cancer.

REFERENCES

- Abbassi-Ghadi, N., Veselkov, K., Kumar, S., Huang, J., Jones, E., Strittmatter, N., et al. (2014). Discrimination of lymph node metastases using desorption electrospray ionisation-mass spectrometry imaging. *Chemical Communications (Cambridge, England)*, 50(28), 3661–3664.
- Abbott, K. L., Lim, J. M., Wells, L., Benigno, B. B., McDonald, J. F., & Pierce, M. (2010). Identification of candidate biomarkers with cancer-specific glycosylation in the tissue and serum of endometrioid ovarian cancer patients by glycoproteomic analysis. *Proteomics*, 10(3), 470–481.
- Aerni, H.-R., Cornett, D. S., & Caprioli, R. M. (2006). Automated acoustic matrix deposition for MALDI sample preparation. *Analytical Chemistry*, 78(3), 827–834.

- Agar, N. Y., Malcolm, J. G., Mohan, V., Yang, H. W., Johnson, M. D., Tannenbaum, A., et al. (2010). Imaging of meningioma progression by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Analytical Chemistry*, 82(7), 2621–2625.
- Aichler, M., Elsner, M., Ludyga, N., Feuchtinger, A., Zangen, V., Maier, S. K., et al. (2013). Clinical response to chemotherapy in oesophageal adenocarcinoma patients is linked to defects in mitochondria. *The Journal of Pathology*, 230(4), 410–419.
- Aichler, M., & Walch, A. (2015). MALDI Imaging mass spectrometry: Current frontiers and perspectives in pathology research and practice. *Laboratory Investigation*, 95(4), 422–431.
- Aikawa, H., Hayashi, M., Ryu, S., Yamashita, M., Ohtsuka, N., Nishidate, M., et al. (2016). Visualizing spatial distribution of alectinib in murine brain using quantitative mass spectrometry imaging. *Scientific Reports*, 6, 23749.
- Andersson, M., Andren, P., & Caprioli, R. M. (2010). MALDI imaging and profiling mass spectrometry in neuroproteomics. In O. Alzate (Ed.), *Neuroproteomics*. Boca Raton, FL: CRC Press/Taylor & Francis.
- Andersson, M., Groseclose, M. R., Deutch, A. Y., & Caprioli, R. M. (2008). Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. *Nature Methods*, 5(1), 101–108.
- Atkinson, S. J., Loadman, P. M., Sutton, C., Patterson, L. H., & Clench, M. R. (2007). Examination of the distribution of the bioreductive drug AQ4N and its active metabolite AQ4 in solid tumours by imaging matrix-assisted laser desorption/ionisation mass spectrometry. *Rapid Communications in Mass Spectrometry*, 21(7), 1271–1276.
- Balluff, B., Elsner, M., Kowarsch, A., Rausser, S., Meding, S., Schuhmacher, C., et al. (2010). Classification of HER2/neu status in gastric cancer using a breast-cancer derived proteome classifier. *Journal of Proteome Research*, 9(12), 6317–6322.
- Balluff, B., Frese, C. K., Maier, S. K., Schone, C., Kuster, B., Schmitt, M., et al. (2015). De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry. *The Journal of Pathology*, 235(1), 3–13.
- Balluff, B., Rausser, S., Meding, S., Elsner, M., Schone, C., Feuchtinger, A., et al. (2011). MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer. *The American Journal of Pathology*, 179(6), 2720–2729.
- Balluff, B., Schöne, C., Höfler, H., & Walch, A. (2011). MALDI imaging mass spectrometry for direct tissue analysis: Technological advancements and recent applications. *Histochemistry and Cell Biology*, 136(3), 227–244.
- Bauer, J. A., Chakravarthy, A. B., Rosenbluth, J. M., Mi, D., Seeley, E. H., De Matos Granja-Ingram, N., et al. (2010). Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. *Clinical Cancer Research*, 16(2), 681–690.
- Benabdellah, F., Seyer, A., Quinton, L., Touboul, D., Brunelle, A., & Laprevote, O. (2010). Mass spectrometry imaging of rat brain sections: Nanomolar sensitivity with MALDI versus nanometer resolution by TOF-SIMS. *Analytical and Bioanalytical Chemistry*, 396(1), 151–162.
- Bodzon-Kulakowska, A., & Suder, P. (2016). Imaging mass spectrometry: Instrumentation, applications, and combination with other visualization techniques. *Mass Spectrometry Reviews*, 35(1), 147–169.
- Bouslimani, A., Bec, N., Glueckmann, M., Hirtz, C., & Larroque, C. (2010). Matrix-assisted laser desorption/ionization imaging mass spectrometry of oxaliplatin derivatives in heated intraoperative chemotherapy (HIPEC)-like treated rat kidney. *Rapid Communications in Mass Spectrometry*, 24(4), 415–421.
- Buck, A., Halbritter, S., Spath, C., Feuchtinger, A., Aichler, M., Zitzelsberger, H., et al. (2015). Distribution and quantification of irinotecan and its active metabolite SN-38

- in colon cancer murine model systems using MALDI MSI. *Analytical and Bioanalytical Chemistry*, 407(8), 2107–2116.
- Büll, C., Boltje, T. J., Wassink, M., de Graaf, A. M., van Delft, F. L., den Brok, M. H., et al. (2013). Targeting aberrant sialylation in cancer cells using a fluorinated sialic acid analog impairs adhesion, migration, and in vivo tumor growth. *Molecular Cancer Therapeutics*, 12(10), 1935–1946.
- Calligaris, D., Caragacianu, D., Liu, X., Norton, I., Thompson, C. J., Richardson, A. L., et al. (2014). Application of desorption electrospray ionization mass spectrometry imaging in breast cancer margin analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(42), 15184–15189.
- Caprioli, R. M., Farmer, T. B., & Gile, J. (1997). Molecular imaging of biological samples: Localization of peptides and proteins using MALDI-TOF MS. *Analytical Chemistry*, 69(23), 4751–4760.
- Casadonte, R., & Caprioli, R. M. (2011). Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry. *Nature Protocols*, 6(11), 1695–1709.
- Casadonte, R., Kriegsmann, M., Zweynert, F., Friedrich, K., Bretton, G., Otto, M., et al. (2014). Imaging mass spectrometry to discriminate breast from pancreatic cancer metastasis in formalin-fixed paraffin-embedded tissues. *Proteomics*, 14(7–8), 956–964.
- Cazares, L. H., Troyer, D., Mendrinos, S., Lance, R. A., Nyalwidhe, J. O., Beydoun, H. A., et al. (2009). Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase 2 discriminates cancer from uninvolved prostate tissue. *Clinical Cancer Research*, 15(17), 5541–5551.
- Chaurand, P., Sanders, M. E., Jensen, R. A., & Caprioli, R. M. (2004). Proteomics in diagnostic pathology: Profiling and imaging proteins directly in tissue sections. *The American Journal of Pathology*, 165(4), 1057–1068.
- Chaurand, P., Schwartz, S. A., & Caprioli, R. M. (2004). Assessing protein patterns in disease using imaging mass spectrometry. *Journal of Proteome Research*, 3(2), 245–252.
- Chen, M., Zhang, J., & Manley, J. L. (2010). Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA. *Cancer Research*, 70(22), 8977–8980.
- Cho, Y. T., Su, H., Wu, W. J., Wu, D. C., Hou, M. F., Kuo, C. H., et al. (2015). Biomarker characterization by MALDI-TOF/MS. *Advances in Clinical Chemistry*, 69, 209–254.
- Choi, J. H., Shin, N. R., Moon, H. J., Kwon, C. H., Kim, G. H., Song, G. A., et al. (2012). Identification of S100A8 and S100A9 as negative regulators for lymph node metastasis of gastric adenocarcinoma. *Histology and Histopathology*, 27(11), 1439–1448.
- Coombes, K. R., Koomen, J. M., Baggerly, K. A., Morris, J. S., & Kobayashi, R. (2005). Understanding the characteristics of mass spectrometry data through the use of simulation. *Cancer Informatics*, 1, 41–52.
- Cornett, D. S., Frappier, S. L., & Caprioli, R. M. (2008). MALDI-FTICR imaging mass spectrometry of drugs and metabolites in tissue. *Analytical Chemistry*, 80(14), 5648–5653.
- Cornett, D. S., Mobley, J. A., Dias, E. C., Andersson, M., Arteaga, C. L., Sanders, M. E., et al. (2006). A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. *Molecular & Cellular Proteomics*, 5(10), 1975–1983.
- Crecelius, A. C., Cornett, D. S., Caprioli, R. M., Williams, B., Dawant, B. M., & Bodenheimer, B. (2005). Three-dimensional visualization of protein expression in mouse brain structures using imaging mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 16(7), 1093–1099.
- de Hoffmann, E., & Stroobant, V. (2007). *Mass spectrometry: Principles and applications* (3rd ed.). England: Wiley.

- Deininger, S.-O., Ebert, M. P., Fütterer, A., Gerhard, M., & Röcken, C. (2008). MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *Journal of Proteome Research*, 7(12), 5230–5236.
- Dekker, T. J. A., Balluff, B. D., Jones, E. A., Schöne, C. D., Schmitt, M., Aubele, M., et al. (2014). Multicenter matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI MSI) identifies proteomic differences in breast-cancer-associated stroma. *Journal of Proteome Research*, 13(11), 4730–4738.
- Diehl, H. C., Beine, B., Elm, J., Trede, D., Ahrens, M., Eisenacher, M., et al. (2015). The challenge of on-tissue digestion for MALDI MSI—A comparison of different protocols to improve imaging experiments. *Analytical and Bioanalytical Chemistry*, 407(8), 2223–2243.
- Dill, A. L., Eberlin, L. S., Costa, A. B., Zheng, C., Ifa, D. R., Cheng, L., et al. (2011). Multivariate statistical identification of human bladder carcinomas using ambient ionization imaging mass spectrometry. *Chemistry*, 17(10), 2897–2902.
- Dill, A. L., Ifa, D. R., Manicke, N. E., Costa, A. B., Ramos-Vara, J. A., Knapp, D. W., et al. (2009). Lipid profiles of canine invasive transitional cell carcinoma of the urinary bladder and adjacent normal tissue by desorption electrospray ionization imaging mass spectrometry. *Analytical Chemistry*, 81(21), 8758–8764.
- Djidja, M.-C., Claude, E., Snel, M. F., Francese, S., Scriven, P., Carolan, V., et al. (2010). Novel molecular tumour classification using MALDI-mass spectrometry imaging of tissue micro-array. *Analytical and Bioanalytical Chemistry*, 397(2), 587–601.
- Drake, P. M., Cho, W., Li, B., Prakobphol, A., Johansen, E., Anderson, N. L., et al. (2010). Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clinical Chemistry*, 56(2), 223–236.
- Du, P., Kibbe, W. A., & Lin, S. M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17), 2059–2065.
- Dube, D. H., & Bertozzi, C. R. (2005). Glycans in cancer and inflammation—Potential for therapeutics and diagnostics. *Nature Reviews. Drug Discovery*, 4(6), 477–488.
- Eberlin, L. S., Norton, I., Dill, A. L., Golby, A. J., Ligon, K. L., Santagata, S., et al. (2012). Classifying human brain tumors by lipid imaging with mass spectrometry. *Cancer Research*, 72(3), 645–654.
- Eberlin, L. S., Tibshirani, R. J., Zhang, J., Longacre, T. A., Berry, G. J., Bingham, D. B., et al. (2014). Molecular assessment of surgical-resection margins of gastric cancer by mass-spectrometric imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 111(7), 2436–2441.
- Elsner, M., Rauser, S., Maier, S., Schöne, C., Balluff, B., Meding, S., et al. (2012). MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma. *Journal of Proteomics*, 75(15), 4693–4704.
- Fehniger, T. E., Vegvari, A., Rezeli, M., Prikk, K., Ross, P., Dahlback, M., et al. (2011). Direct demonstration of tissue uptake of an inhaled drug: Proof-of-principle study using matrix-assisted laser desorption ionization mass spectrometry imaging. *Analytical Chemistry*, 83(21), 8329–8336.
- Fejzo, M. S., & Slamon, D. J. (2010). Tissue microarrays from frozen tissues—OCT technique. *Methods in Molecular Biology*, 664, 73–80.
- Fentz, J. S., Zornig, C., Juhl, H. H., & David, K. A. (2004). Tissue ischemia time affects gene and protein expression patterns within minutes following surgical tumor excision. *Biotechniques*, 36(6), 1030–1037.
- Fletcher, J. S., Lockyer, N. P., & Vickerman, J. C. (2011). Developments in molecular SIMS depth profiling and 3D imaging of biological systems using polyatomic primary ions. *Mass Spectrometry Reviews*, 30(1), 142–174.

- Fletcher, J. S., Vickerman, J. C., & Winograd, N. (2011). Label free biochemical 2D and 3D imaging using secondary ion mass spectrometry. *Current Opinion in Chemical Biology*, 15(5), 733–740.
- Gemoll, T., Strohkamp, S., Schillo, K., Thorns, C., & Habermann, J. K. (2015). MALDI-imaging reveals thymosin beta-4 as an independent prognostic marker for colorectal cancer. *Oncotarget*, 6(41), 43869–43880.
- Gobom, J., Mueller, M., Egelhofer, V., Theiss, D., Lehrach, H., & Nordhoff, E. (2002). A calibration method that simplifies and improves accurate determination of peptide molecular masses by MALDI-TOF MS. *Analytical Chemistry*, 74(15), 3915–3923.
- Goodwin, R. J. (2012). Sample preparation for mass spectrometry imaging: Small mistakes can lead to big consequences. *Journal of Proteomics*, 75(16), 4893–4911.
- Gorzolka, K., & Walch, A. (2014). MALDI mass spectrometry imaging of formalin-fixed paraffin-embedded tissues in clinical research. *Histology and Histopathology*, 29(11), 1365–1376.
- Groseclose, M. R., Massion, P. P., Chaurand, P., & Caprioli, R. M. (2008). High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. *Proteomics*, 8(18), 3715–3724.
- Grüner, B. M., Hahne, H., Mazur, P. K., Trajkovic-Arsic, M., Maier, S., Esposito, I., et al. (2012). MALDI imaging mass spectrometry for in situ proteomic analysis of preneoplastic lesions in pancreatic cancer. *PLoS One*, 7(6), e39424.
- Gustafsson, O. J., Briggs, M. T., Condina, M. R., Winderbaum, L. J., Pelzing, M., McColl, S. R., et al. (2015). MALDI imaging mass spectrometry of N-linked glycans on formalin-fixed paraffin-embedded murine kidney. *Analytical and Bioanalytical Chemistry*, 407(8), 2127–2139.
- Gustafsson, J. O., Edde, J. S., Meding, S., Koudelka, T., Oehler, M. K., McColl, S. R., et al. (2012). Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS. *Journal of Proteomics*, 75(16), 5093–5105.
- Gustafsson, O. J., Edde, J. S., Meding, S., McColl, S. R., Oehler, M. K., & Hoffmann, P. (2013). Matrix-assisted laser desorption/ionization imaging protocol for in situ characterization of tryptic peptide identity and distribution in formalin-fixed tissue. *Rapid Communications in Mass Spectrometry*, 27(6), 655–670.
- Han, E. C., Lee, Y.-S., Liao, W.-S., Liu, Y.-C., Liao, H.-Y., & Jeng, L.-B. (2011). Direct tissue analysis by MALDI-TOF mass spectrometry in human hepatocellular carcinoma. *Clinica Chimica Acta*, 412(3–4), 230–239.
- Hardesty, W. M., Kelley, M. C., Mi, D., Low, R. L., & Caprioli, R. M. (2011). Protein signatures for survival and recurrence in metastatic melanoma. *Journal of Proteomics*, 74(7), 1002–1014.
- Holst, S., Heijs, B., de Haan, N., van Zeijl, R. J., Briare-de Bruijn, I. H., van Pelt, G. W., et al. (2016). Linkage-specific in situ sialic acid derivatization for N-glycan mass spectrometry imaging of formalin-fixed paraffin-embedded tissues. *Analytical Chemistry*, 88(11), 5904–5913.
- Hood, B. L., Darfler, M. M., Guiel, T. G., Furusato, B., Lucas, D. A., Ringeisen, B. R., et al. (2005). Proteomic analysis of formalin-fixed prostate cancer tissue. *Molecular & Cellular Proteomics*, 4(11), 1741–1753.
- Horn, D. M., Peters, E. C., Klock, H., Meyers, A., & Brock, A. (2004). Improved protein identification using automated high mass measurement accuracy MALDI FT-ICR MS peptide mass fingerprinting. *International Journal of Mass Spectrometry*, 238(2), 189–196.
- Hsieh, Y., Li, F., & Korfmacher, W. A. (2010). Mapping pharmaceuticals in rat brain sections using MALDI imaging mass spectrometry. *Methods in Molecular Biology*, 656, 147–158.
- Hsu, C. C., & Dorrestein, P. C. (2015). Visualizing life with ambient mass spectrometry. *Current Opinion in Biotechnology*, 31, 24–34.

- Jensen, P. H., Karlsson, N. G., Kolarich, D., & Packer, N. H. (2012). Structural analysis of N- and O-glycans released from glycoproteins. *Nature Protocols*, 7(7), 1299–1310.
- Jones, E. A., Deininger, S.-O., Hogendoorn, P. C., Deelder, A. M., & McDonnell, L. A. (2012). Imaging mass spectrometry statistical analysis. *Journal of Proteomics*, 75(16), 4962–4989.
- Jones, E. A., van Remoortere, A., van Zeijl, R. J. M., Hogendoorn, P. C. W., Bovée, J. V. M. G., Deelder, A. M., et al. (2011). Multiple statistical analysis techniques corroborate intratumor heterogeneity in imaging mass spectrometry datasets of myxofibrosarcoma. *PLoS One*, 6(9), e24913.
- Jurchen, J. C., Rubakhin, S. S., & Sweedler, J. V. (2005). MALDI-MS imaging of features smaller than the size of the laser beam. *Journal of the American Society for Mass Spectrometry*, 16(10), 1654–1659.
- Kaletas, B. K., van der Wiel, I. M., Stauber, J., Guzel, C., Kros, J. M., Luiders, T. M., et al. (2009). Sample preparation issues for tissue imaging by imaging MS. *Proteomics*, 9(10), 2622–2633.
- Kang, S., Shim, H. S., Lee, J. S., Kim, D. S., Kim, H. Y., Hong, S. H., et al. (2010). Molecular proteomics imaging of tumor interfaces by mass spectrometry. *Journal of Proteome Research*, 9(2), 1157–1164.
- Kertesz, V., Berkel, V., Gary, J., Vavrek, M., Koeplinger, K. A., Schneider, B. B., et al. (2008). Comparison of drug distribution images from whole-body thin tissue sections obtained using desorption electrospray ionization tandem mass spectrometry and autoradiography. *Analytical Chemistry*, 80(13), 5168–5177.
- Khatib-Shahidi, S., Andersson, M., Herman, J. L., Gillespie, T. A., & Caprioli, R. M. (2006). Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. *Analytical Chemistry*, 78(18), 6448–6456.
- Kim, H. K., Reyzer, M. L., Choi, I. J., Kim, C. G., Kim, H. S., Oshima, A., et al. (2010). Gastric cancer-specific protein profile identified using endoscopic biopsy samples via MALDI mass spectrometry. *Journal of Proteome Research*, 9(8), 4123–4130.
- Klerk, L. A., Maarten Altelaar, A. F., Froesch, M., McDonnell, L. A., & Heeren, R. M. A. (2009). Fast and automated large-area imaging MALDI mass spectrometry in microprobe and microscope mode. *International Journal of Mass Spectrometry*, 285(1–2), 19–25.
- Kriegsmann, J., Kriegsmann, M., & Casadonte, R. (2015). MALDI TOF imaging mass spectrometry in clinical pathology: A valuable tool for cancer diagnostics (review). *International Journal of Oncology*, 46(3), 893–906.
- Kwon, H. J., Kim, Y., Sugihara, Y., Baldetorp, B., Welinder, C., Watanabe, K., et al. (2015). Drug compound characterization by mass spectrometry imaging in cancer tissue. *Archives of Pharmacological Research*, 38(9), 1718–1727.
- Lange, E., Gropl, C., Reinert, K., Kohlbacher, O., & Hildebrandt, A. (2006). High-accuracy peak picking of proteomics data using wavelet techniques. *Pacific Symposium on Biocomputing*, 11, 243–254. Paper presented at the Pacific Symposium on Biocomputing.
- Lanni, E. J., Rubakhin, S. S., & Sweedler, J. V. (2012). Mass spectrometry imaging and profiling of single cells. *Journal of Proteomics*, 75(16), 5036–5051.
- Lau, K. S., & Dennis, J. W. (2008). N-Glycans in cancer progression. *Glycobiology*, 18(10), 750–760.
- Le Faouder, J., Laouirem, S., Chapelle, M., Albuquerque, M., Belghiti, J., Degos, F., et al. (2011). Imaging mass spectrometry provides fingerprints for distinguishing hepatocellular carcinoma from cirrhosis. *Journal of Proteome Research*, 10(8), 3755–3765.
- Lee, D. H., Chung, K., Song, J.-A., Kim, T.-h., Kang, H., Huh, J. H., et al. (2010). Proteomic identification of paclitaxel-resistance associated hnRNP A2 and GDI 2 proteins in human ovarian cancer cells. *Journal of Proteome Research*, 9(11), 5668–5676.
- Lemaire, R., Ait Menguellet, S., Stauber, J., Marchaudon, V., Lucot, J.-P., Collinet, P., 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. *Journal of Proteome Research*, 6(11), 4127–4134.

- Lemaire, R., Desmons, A., Tabet, J. C., Day, R., Salzter, M., & Fournier, I. (2007). Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections. *Journal of Proteome Research*, 6(4), 1295–1305.
- Lemaire, R., Tabet, J. C., Ducoroy, P., Hendra, J. B., Salzter, M., & Fournier, I. (2006). Solid ionic matrices for direct tissue analysis and MALDI imaging. *Analytical Chemistry*, 78(3), 809–819.
- Liu, X., Weaver, E. M., & Hummon, A. B. (2013). Evaluation of therapeutics in three-dimensional cell culture systems by MALDI imaging mass spectrometry. *Analytical Chemistry*, 85(13), 6295–6302.
- Luxembourg, S. L., Mize, T. H., McDonnell, L. A., & Heeren, R. M. (2004). High-spatial resolution mass spectrometric imaging of peptide and protein distributions on a surface. *Analytical Chemistry*, 76(18), 5339–5344.
- Mascini, N. E., Eijkel, G. B., ter Brugge, P., Jonkers, J., Wesseling, J., & Heeren, R. M. A. (2015). The use of mass spectrometry imaging to predict treatment response of patient-derived xenograft models of triple-negative breast cancer. *Journal of Proteome Research*, 14(2), 1069–1075.
- McDonnell, L. A., Corthals, G. L., Willems, S. M., van Remoortere, A., van Zeijl, R. J. M., & Deelder, A. M. (2010). Peptide and protein imaging mass spectrometry in cancer research. *Journal of Proteomics*, 73(10), 1921–1944.
- Metz, B., Kersten, G. F. A., Hoogerhout, P., Brugghe, H. F., Timmermans, H. A. M., De Jong, A. D., et al. (2004). Identification of formaldehyde-induced modifications in proteins reactions with model peptides. *Journal of Biological Chemistry*, 279(8), 6235–6243.
- Minchinton, A. I., & Tannock, I. F. (2006). Drug penetration in solid tumours. *Nature Reviews. Cancer*, 6(8), 583–592. <http://dx.doi.org/10.1038/nrc1893>.
- Mittal, P., Klingler-Hoffmann, M., Arentz, G., Winderbaum, L., Lokman, N. A., Zhang, C., et al. (2016). Maldi imaging of primary endometrial cancers reveals proteins associated with lymph node metastasis. *Proteomics*, 16, 1793–1801.
- Mittal, P., Klingler-Hoffmann, M., Arentz, G., Zhang, C., Kaur, G., Oehler, M. K., et al. (2016). Proteomics of endometrial cancer diagnosis, treatment, and prognosis. *Proteomics. Clinical Applications*, 10(3), 217–229.
- Morita, Y., Ikegami, K., Goto-Inoue, N., Hayasaka, T., Zaima, N., Tanaka, H., et al. (2010). Imaging mass spectrometry of gastric carcinoma in formalin-fixed paraffin-embedded tissue microarray. *Cancer Science*, 101(1), 267–273.
- Morosi, L., Spinelli, P., Zucchetti, M., Pretto, F., Carrà, A., D'Incalci, M., et al. (2013). Determination of paclitaxel distribution in solid tumors by nano-particle assisted laser desorption/ionization mass spectrometry imaging. *PLoS One*, 8(8), e72532.
- Nakagoe, T., Fukushima, K., Nanashima, A., Sawai, T., Tsuji, T., Jibiki, M., et al. (2000). Expression of Lewis(a), sialyl Lewis(a), Lewis(x) and sialyl Lewis(x) antigens as prognostic factors in patients with colorectal cancer. *Canadian Journal of Gastroenterology*, 14(9), 753–760.
- Nilsson, A., Fehniger, T. E., Gustavsson, L., Andersson, M., Kenne, K., Marko-Varga, G., et al. (2010). Fine mapping the spatial distribution and concentration of unlabeled drugs within tissue micro-compartments using imaging mass spectrometry. *PLoS One*, 5(7), e11411.
- Nipp, M., Elsner, M., Balluff, B., Meding, S., Sarioglu, H., Ueffing, M., et al. (2011). S100-A10, thioredoxin, and S100-A6 as biomarkers of papillary thyroid carcinoma with lymph node metastasis identified by MALDI imaging. *Journal of Molecular Medicine*, 90(2), 163–174.
- North, S. J., Hitchen, P. G., Haslam, S. M., & Dell, A. (2009). Mass spectrometry in the analysis of N-linked and O-linked glycans. *Current Opinion in Structural Biology*, 19(5), 498–506.
- Oezdemir, R. F., Gaisa, N. T., Lindemann-Docter, K., Gostek, S., Weiskirchen, R., Ahrens, M., et al. (2012). Proteomic tissue profiling for the improvement of

- grading of noninvasive papillary urothelial neoplasia. *Clinical Biochemistry*, 45(1–2), 7–11.
- Ohtsubo, K., & Marth, J. D. (2006). Glycosylation in cellular mechanisms of health and disease. *Cell*, 126(5), 855–867.
- Oppenheimer, S. R., Mi, D., Sanders, M. E., & Caprioli, R. M. (2010). Molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma. *Journal of Proteome Research*, 9(5), 2182–2190.
- Pagni, F., Mainini, V., Garancini, M., Bono, F., Vanzati, A., Giardini, V., et al. (2015). Proteomics for the diagnosis of thyroid lesions: Preliminary report. *Cytopathology*, 26(5), 318–324.
- Pappin, D. J. C., Hojrup, P., & Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, 3(6), 327–332.
- Passarelli, M. K., & Winograd, N. (2011). Lipid imaging with time-of-flight secondary ion mass spectrometry (ToF-SIMS). *Biochimica et Biophysica Acta*, 1811(11), 976–990.
- Patel, S. A., Barnes, A., Loftus, N., Martin, R., Sloan, P., Thakker, N., et al. (2009). Imaging mass spectrometry using chemical inkjet printing reveals differential protein expression in human oral squamous cell carcinoma. *Analyst*, 134(2), 301–307. <http://dx.doi.org/10.1039/B812533C>.
- Pavlidis, N., & Pentheroudakis, G. (2016). Cancer of unknown primary site. *The Lancet*, 379(9824), 1428–1435.
- Powers, T. W., Holst, S., Wuhrer, M., Mehta, A. S., & Drake, R. R. (2015). Two-dimensional N-glycan distribution mapping of hepatocellular carcinoma tissues by MALDI-imaging mass spectrometry. *Biomolecules*, 4, 2554–2572.
- Powers, T. W., Jones, E. E., Betesh, L. R., Romano, P. R., Gao, P., Copland, J. A., et al. (2013). Matrix assisted laser desorption ionization imaging mass spectrometry workflow for spatial profiling analysis of N-linked glycan expression in tissues. *Analytical Chemistry*, 85(20), 9799–9806.
- Powers, T. W., Neely, B. A., Shao, Y., Tang, H., Troyer, D. A., Mehta, A. S., et al. (2014). MALDI imaging mass spectrometry profiling of N-glycans in formalin-fixed paraffin embedded clinical tissue blocks and tissue microarrays. *PLoS One*, 9(9), e106255.
- Prideaux, B., Dartois, V., Staab, D., Weiner, D. M., Goh, A., Via, L. E., et al. (2011). High-sensitivity MALDI-MRM-MS imaging of moxifloxacin distribution in tuberculosis-infected rabbit lungs and granulomatous lesions. *Analytical Chemistry*, 83(6), 2112–2118.
- Rahman, S. M., Gonzalez, A. L., Li, M., Seeley, E. H., Zimmerman, L. J., Zhang, X. J., et al. (2011). Lung cancer diagnosis from proteomic analysis of preinvasive lesions. *Cancer Research*, 71(8), 3009–3017.
- Rauser, S., Marquardt, C., Balluff, B., Deininger, S.-O., Albers, C., Belau, E., et al. (2010). Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. *Journal of Proteome Research*, 9(4), 1854–1863.
- Renovanz, M., & Kim, E. L. (2014). Intratumoral heterogeneity, its contribution to therapy resistance and methodological caveats to assessment. *Frontiers in Oncology*, 4, 142.
- Reyzer, M. L., Hsieh, Y., Ng, K., Korfmacher, W. A., & Caprioli, R. M. (2003). Direct analysis of drug candidates in tissue by matrix-assisted laser desorption/ionization mass spectrometry. *Journal of Mass Spectrometry*, 38(10), 1081–1092.
- Rodrigo, M. A., Zitka, O., Krizkova, S., Moulick, A., Adam, V., & Kizek, R. (2014). MALDI-TOF MS as evolving cancer diagnostic tool: A review. *Journal of Pharmaceutical and Biomedical Analysis*, 95, 245–255.
- Rohner, T. C., Staab, D., & Stoeckli, M. (2005). MALDI mass spectrometric imaging of biological tissue sections. *Mechanisms of Ageing and Development*, 126(1), 177–185.

- Ronci, M., Bonanno, E., Colantoni, A., Pieroni, L., Di Ilio, C., Spagnoli, L. G., et al. (2008). Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: Application to MALDI-TOF imaging MS investigations. *Proteomics*, 8(18), 3702–3714.
- Schuerenberg, M., & Deininger, S.-O. (2010). Matrix application with ImagePrep. *Imaging mass spectrometry* (pp. 87–91). New York, USA: Springer.
- Schwamborn, K., & Caprioli, R. M. (2010). Molecular imaging by mass spectrometry—Looking beyond classical histology. *Nature Reviews. Cancer*, 10(9), 639–646. <http://dx.doi.org/10.1038/nrc2917>.
- Schwamborn, K., Krieg, R. C., Jirak, P., Ott, G., Knuchel, R., Rosenwald, A., et al. (2010). Application of MALDI imaging for the diagnosis of classical Hodgkin lymphoma. *Journal of Cancer Research and Clinical Oncology*, 136(11), 1651–1655.
- Schwamborn, K., Krieg, R. C., Reska, M., Jakse, G., Knuchel, R., & Wellmann, A. (2007). Identifying prostate carcinoma by MALDI-Imaging. *International Journal of Molecular Medicine*, 20(2), 155–159.
- Schwartz, S. A., Reyzer, M. L., & Caprioli, R. M. (2003). Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: Practical aspects of sample preparation. *Journal of Mass Spectrometry*, 38(7), 699–708.
- Scigelova, M., Hornshaw, M., Giannakopoulos, A., & Makarov, A. (2011). Fourier transform mass spectrometry. *Molecular & Cellular Proteomics*, 10(7). M111.009431.
- Seeley, E. H., Oppenheimer, S. R., Mi, D., Chaurand, P., & Caprioli, R. M. (2008). Enhancement of protein sensitivity for MALDI imaging mass spectrometry after chemical treatment of tissue sections. *Journal of the American Society for Mass Spectrometry*, 19(8), 1069–1077.
- Signor, L., & Boeri Erba, E. (2013). Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometric analysis of intact proteins larger than 100 kDa. *Journal of Visualized Experiments*, 79, e506351.
- Sinha, T. K., Khatib-Shahidi, S., Yankeelov, T. E., Mapara, K., Ehtesham, M., Cornett, D. S., et al. (2008). Integrating spatially resolved three-dimensional MALDI IMS with in vivo magnetic resonance imaging. *Nature Methods*, 5(1), 57–59. <http://dx.doi.org/10.1038/nmeth1147>.
- Sjovall, P., Lausmaa, J., & Johansson, B. (2004). Mass spectrometric imaging of lipids in brain tissue. *Analytical Chemistry*, 76(15), 4271–4278.
- Sjovall, P., Lausmaa, J., Nygren, H., Carlsson, L., & Malmberg, P. (2003). Imaging of membrane lipids in single cells by imprint-imaging time-of-flight secondary ion mass spectrometry. *Analytical Chemistry*, 75(14), 3429–3434.
- Steurer, S., Borkowski, C., Odinga, S., Buchholz, M., Koop, C., Huland, H., et al. (2013). MALDI mass spectrometric imaging based identification of clinically relevant signals in prostate cancer using large-scale tissue microarrays. *International Journal of Cancer*, 133(4), 920–928.
- Steurer, S., Seddiqi, A. S., Singer, J. M., Bahar, A. S., Eichelberg, C., Rink, M., et al. (2014). MALDI imaging on tissue microarrays identifies molecular features associated with renal cell cancer phenotype. *Anticancer Research*, 34(5), 2255–2261.
- Steurer, S., Singer, J. M., Rink, M., Chun, F., Dahlem, R., Simon, R., et al. (2014). MALDI imaging-based identification of prognostically relevant signals in bladder cancer using large-scale tissue microarrays. *Urologic Oncology*, 32(8), 1225–1233.
- Stoeckli, M., Chaurand, P., Hallahan, D. E., & Caprioli, R. M. (2001). Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. *Nature Medicine*, 7(4), 493–496.
- Stoeckli, M., Staab, D., & Schweitzer, A. (2007). Compound and metabolite distribution measured by MALDI mass spectrometric imaging in whole-body tissue sections. *International Journal of Mass Spectrometry*, 260(2–3), 195–202.
- Strand, F. L. (2003). Neuropeptides: General characteristics and neuropharmaceutical potential in treating CNS disorders. *Progress in Drug Research*, 61, 1–37.

- Sugihara, Y., Végvári, Á., Welinder, C., Jönsson, G., Ingvar, C., Lundgren, L., et al. (2014). A new look at drugs targeting malignant melanoma—An application for mass spectrometry imaging. *Proteomics*, 14(17–18), 1963–1970.
- Taban, I. M., Altelaar, A. F., van der Burgt, Y. E., McDonnell, L. A., Heeren, R. M., Fuchser, J., et al. (2007). Imaging of peptides in the rat brain using MALDI-FTICR mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 18(1), 145–151.
- Takáts, Z., Wiseman, J. M., Gologan, B., & Cooks, R. G. (2004). Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science*, 306(5695), 471–473.
- Toghi Eshghi, S., Yang, S., Wang, X., Shah, P., Li, X., & Zhang, H. (2014). Imaging of N-linked glycans from formalin-fixed paraffin-embedded tissue sections using MALDI mass spectrometry. *ACS Chemical Biology*, 9(9), 2149–2156.
- Trim, P. J., Henson, C. M., Avery, J. L., McEwen, A., Snel, M. F., Claude, E., et al. (2008). Matrix-assisted laser desorption/ionization-ion mobility separation-mass spectrometry imaging of vinblastine in whole body tissue sections. *Analytical Chemistry*, 80(22), 8628–8634.
- van Hove, A., Erika, R., Smith, D. F., & Heeren, R. M. A. (2010). A concise review of mass spectrometry imaging. *Journal of Chromatography. A*, 1217(25), 3946–3954.
- Varki, A. (1993). Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology*, 3(2), 97–130.
- Végvári, Á., Shavkunov, A. S., Fehniger, T. E., Grabau, D., Niméus, E., & Marko-Varga, G. (2016). Localization of tamoxifen in human breast cancer tumors by MALDI mass spectrometry imaging. *Clinical and Translational Medicine*, 5(1), 1–8.
- Waki, M., Ide, Y., Ishizaki, I., Nagata, Y., Masaki, N., Sugiyama, E., et al. (2014). Single-cell time-of-flight secondary ion mass spectrometry reveals that human breast cancer stem cells have significantly lower content of palmitoleic acid compared to their counterpart non-stem cancer cells. *Biochimie*, 107(Pt. A), 73–77.
- Walch, A., Rauser, S., Deininger, S.-O., & Höfler, H. (2008). MALDI imaging mass spectrometry for direct tissue analysis: A new frontier for molecular histology. *Histochemistry and Cell Biology*, 130(3), 421–434.
- Willems, S. M., van Remoortere, A., van Zeijl, R., Deelder, A. M., McDonnell, L. A., & Hogendoorn, P. C. W. (2010). Imaging mass spectrometry of myxoid sarcomas identifies proteins and lipids specific to tumour type and grade, and reveals biochemical intra-tumour heterogeneity. *The Journal of Pathology*, 222(4), 400–409.
- Winderbaum, L., Koch, I., Mittal, P., & Hoffmann, P. (2016). Classification of MALDI-MS imaging data of tissue microarrays using canonical correlation analysis based variable selection. *Proteomics*, 16, 1731–1735.
- Wisztorski, M., Lemaire, R., Stauber, J., Menguelet, S. A., Croix, D., Mathé, O. J., et al. (2007). New developments in MALDI imaging for pathology proteomic studies. *Current Pharmaceutical Design*, 13(32), 3317–3324.
- Wuhrer, M., Deelder, A. M., & van der Burgt, Y. E. (2011). Mass spectrometric glycan rearrangements. *Mass Spectrometry Reviews*, 30(4), 664–680.
- Yamada, Y., Hidefumi, K., Shion, H., Oshikata, M., & Haramaki, Y. (2011). Distribution of chloroquine in ocular tissue of pigmented rat using matrix-assisted laser desorption/ionization imaging quadrupole time-of-flight tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 25(11), 1600–1608.
- Yanagisawa, K., Yu, S., Xu, B. J., Massion, P. P., Larsen, P. H., White, B. C., et al. (2003). Proteomic patterns of tumour subsets in non-small-cell lung cancer. *The Lancet*, 362(9382), 433–439.
- Yang, L., Cui, X., Zhang, N., Li, M., Bai, Y., Han, X., et al. (2015). Comprehensive lipid profiling of plasma in patients with benign breast tumor and breast cancer reveals novel biomarkers. *Analytical and Bioanalytical Chemistry*, 407(17), 5065–5077.
- Yang, C., He, Z., & Yu, W. (2009). Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics*, 10(1), 4.

3.1.2 Data Dependent Acquisition/ Data Independent Acquisition

Even though MALDI MSI is highly sensitive and allows researchers to map the spatial distribution of a large number of analytes simultaneously, several challenges still remain. One major limitation of this technique is the lack of direct protein identification and therefore an additional step is required to overcome this issue [122]. Three strategies based on mass spectrometry methods allow protein identification: top-down and bottom-up experiments directly on the tissue sections (*in situ*) by MALDI-MS/MS, and indirect identification using tissue lysate by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Bottom-up experiments on tissue sections consist of the tryptic digestion of the proteins followed by MALDI MSI with MS/MS analysis. While top-down is used to identify the intact proteins using a mass spectrometer, without performing any enzymatic digestion. However, the protein identifications directly from tissues using MS/MS experiments for peptides or proteins is hampered by the low sensitivity caused by ion-suppression effects due to the complex molecular composition of tissues [122]. Secondly, each protein analysed by MALDI MSI results in a singly charged state ion, and ion activation by MALDI fails to produce fragmentation and identification [123]. To overcome this limitation, the protein identification can be completed by matching MALDI MSI data with data generated from LC-MS/MS.

For proteomics, two main LC-MS/MS methods have been developed: data dependent acquisition (DDA) and data independent acquisition (DIA). They have in common that the proteins are digested into peptides by an enzyme and are then separated using high-performance liquid chromatography (HPLC). They differ in the mass spectrometric method used. The events occurring in both DDA and DIA: (a) ionization: peptides elute into the mass spectrometer from the HPLC and are ionized; (b) MS1 scan: the abundance and mass-to-charge ratios (m/z) of all ions eluting at a given time are measured; and (c) MS2 scan: some or all detected ions are fragmented, and the abundances and m/z 's of the fragments are measured and recorded [124]. The methods vary in how ions are selected and measured in the MS2 scan.

For the last 15 years, the most widely accepted method for biomarker discovery is shotgun proteomics, where the instrument is operated in data dependent acquisition (DDA) mode. In DDA, a subset of the most abundant ions reaching the mass spectrometer detector during an MS1 scan are individually isolated and fragmented in sequential MS2 scans, and each MS2 scan can be analysed with a database search algorithm [125, 126]. DDA typically yields thousands of protein identifications. Although extremely powerful, DDA suffers from several

fundamental limitations. If in a single MS1 scan, too many peptide species co-elute and appear, then DDA stochastically samples only the most abundant peptides and misses the rest [127]. Thus, it remains a challenge to reproducibly quantify especially low-abundance peptides. To overcome this issue, recent improvements in mass spectrometer design and bioinformatics algorithms have resulted in the rediscovery and development of another sampling method called data independent acquisition (DIA) [128].

In DIA MS2 scan mode, the instrument fragments all of the precursors generated from a sample that are within a predetermined mass-to-charge ratio (m/z) and retention-time range. Usually, the instrument cycles through the precursor ion m/z range in segments of specified width, at each cycle producing a highly multiplexed fragment-ion spectrum. However, because of the increased complexity of these MS/MS spectra, the link between the fragment ions and the precursors from which they originate is lost, complicating the proper interpretation of DIA data. One approach to solve this issue is to compare the DIA spectra to sets of annotated and refined peptide-MS2 spectrum matches from DDA experiments called spectral libraries that show accurate, empirically determined fragmentation patterns for each peptide in the library [129]. This can be performed using a free, open-source Skyline software tool [130], which can be downloaded at <http://skyline.maccosslab.org/>. Here, the spectral library was constructed using a DDA experiment from 12 EC patient tissue sections (n=6 with LNM and n=6 without LNM). Using this spectral library, DIA was used for label free quantification in a targeted and partially untargeted fashion.

3.1.3 Immunohistochemistry

The principle of immunohistochemistry (IHC) has existed since the 1930s, but it was not until 1941 that the first IHC study was reported [131]. From then, IHC has become a crucial technique that serves as a diagnostic and prognostic method for identification of disease markers in tissue samples [132]. Staining of various marker proteins located either in the cell nuclei, cytoplasm or membrane are often considered as pathological determinants for classifying and grading diseases. For identifying the presence and the extent of expression of such proteins, qualitative assessments using IHC is commonly performed.

The basic principle of IHC is studying the abundance of a protein of interest based on the differential staining intensity of the biopsy tissue samples with antibodies specific to the protein of interest. In the IHC method, visualization of the antibody-antigen reaction is accomplished by the use of a secondary antibody conjugated to an enzyme, such as peroxidase. The processed sample slides are generally observed under a light microscope by a trained pathologist to assign a score based on the visual parameters set. One major issue in determining the standard by this approach is the amount of variability due to visual perception on a haematoxylin counterstained tissue section. To overcome this, we have used an open source plugin named IHC Profiler [132], which is compatible with ImageJ software to quantify staining of IHC images. IHC Profiler, an automated quantitation tool requires minimal supervision for analysis of both cytoplasmic and nuclear protein markers. IHC Profiler provides automated scoring along with a detailed histogram profile of the image on the basis of the overall DAB (3, 3'-diaminobenzidine) stain intensity obtained through colour deconvolution. Similar to the standard clinical grading procedure, the histogram profile obtained by the IHC Profiler is divided into 4 intensities: high positive, positive, low positive and negative. These four intensities were equally divided on the pixel colour intensity bar. Further statistical analysis can be performed using Graph Pad Prism 6 software (Graph Pad Software, La Jolla California USA).

3.2 Tissue Microarray Design and Construction

In order to detect robust cancer biomarkers, a high number of patient samples has to be investigated. The large collection of archived formalin fixed paraffin embedded (FFPE) tissue samples provides an excellent source for these studies. However, collecting individual tissue samples for each patient in a cohort and analysing them one at a time is not only slow and time consuming, but increases the likelihood of introducing discrepancies from tissue to tissue because of the technical variability. To address this issue, tissue micro arrays (TMAs) have been constructed in-house. Furthermore, the MALDI MSI measurement of a whole single tissue section, shown in Figure 3 of the published review article took 12 hours to finish and generated ~8900 spectra from the tumour region and the surrounding tissue at 100 μM resolution, from which only a limited number of spectra are selected from the tumour region for processing; meaning much of the acquired data is not utilized. The MALDI MSI analysis of TMAs, shown in Figure 1 of the published Chapter 4 took 16 hours to finish and ~34,000 spectra were generated from 29 patient samples at 60 μM resolution, analysed in duplicate and thereby significantly improving throughput. Using MALDI MSI, Steurer *et al.* performed the largest investigation of peptide patterns on TMAs on bladder cancer ($n = 697$) [133], renal cell cancer ($n = 789$) [134], and prostate cancer ($n = 1044$) [135]. This high throughput MALDI MSI analysis on TMAs was also used to detect peptide patterns in several cancers including pancreatic and breast cancer [136], gastric [137], non-small-cell lung cancer [138], ovarian [139], and oesophageal cancer [140].

The tissue microarray array (TMA) of Kononen *et al.* [141] is an extension of an idea originally developed by Battifora [142] and consists of an array of cylindrical cores of paraffin-embedded tissue that are removed from donor paraffin blocks (obtained from surgical pathology, autopsy, or research material) and transferred into a recipient paraffin block. For the TMA assembly in the present study, 1.5 mm tissue cores were excised from regions of primary tumour and arranged vertically in an empty paraffin block using a tissue arrayer (3DHistech TMA Master, SciTech, Victoria, Australia). Two tissue cores from tumour-dense areas were taken from each patient. The layouts of both TMAs were randomized to avoid any unforeseen bias that could be introduced by core location within the TMAs. Initially 57 tumour samples were used in the construction of the TMAs, however following careful review of the patient cohort, 14 mixed carcinomas were excluded from the study, resulting in a final cohort of 43 patients with only type I endometrioid adenocarcinomas ($n=16$ with LNM and $n=27$ without LNM). Tissue

sections 6 μm thick were cut from both TMA's and mounted onto indium tin oxide (ITO) (Bruker Daltonics, Bremen, Germany) conductive glass slides for MALDI MSI analysis. Sections were also placed on plain glass slides and polyethylene naphthalate (PEN) membrane slides, and H&E stained for pathology annotation, IHC and laser capture microdissection (LMD) respectively. Upon construction of the TMA, each tissue core was annotated again by a pathologist. The tumour cores with LNM are circled in red, without LNM are circled in green, control in yellow and the cores excluded from the study are circled in blue. Annotated H&E stained images of TMA1 and TMA2 are shown in Figure 3.

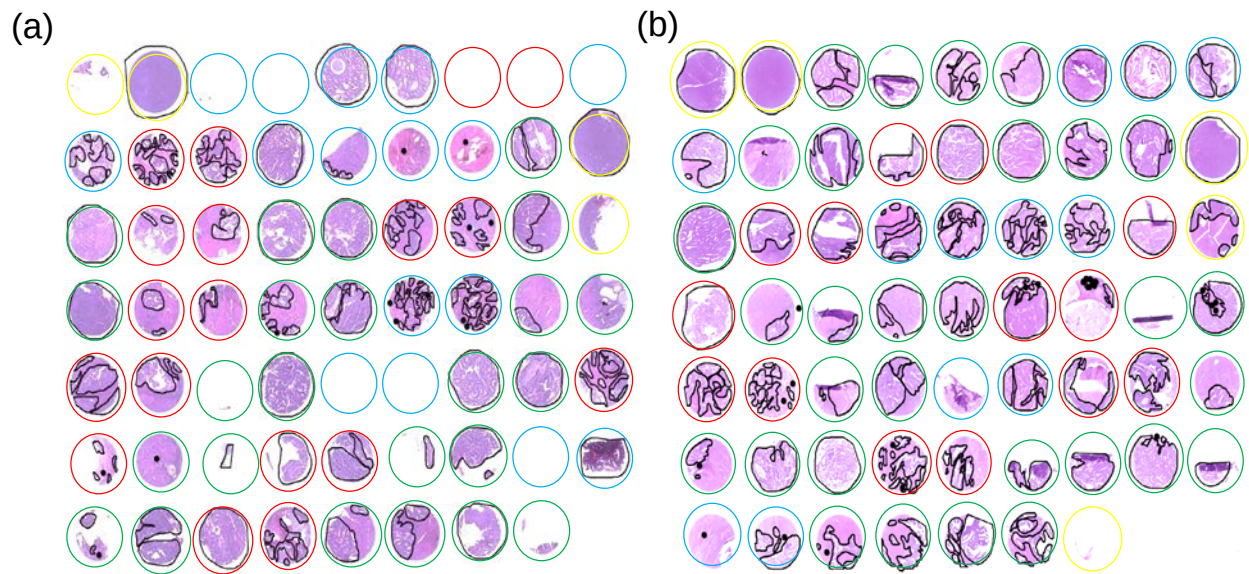


Figure 3: Haematoxylin and eosin stained annotated tumour cores (a) TMA1 (b) TMA2. Tumour cores with LNM are circled red; tumours without LNM are circled green, yellow circles indicate controls and blue circles indicate cores which have been excluded from the analysis. All tumour regions within the tumour cores are annotated in black.

This page intentionally left blank

Lymph Node Metastasis of Primary Endometrial Cancers: Associated Proteins Revealed by MALDI Imaging

This page intentionally left blank

4.1 Summary

In spite of the high and increasing incidence of endometrial cancer, our current models for prediction of lymph node metastasis in EC are suboptimal. Therefore, to determine the optimal treatment strategy and to reduce morbidity in EC, molecular biomarkers are needed. Such markers may immediately serve as diagnostic markers to guide surgeons. In the current study, a novel classification model was used to predict the lymph node status in primary tumours of EC.

For this, MALDI MSI was carried out on two TMA (TMA1 and TMA2), containing a total of 86 primary endometrial carcinoma tissue cores from 43 patients with (n=16) and without (n=27) LNM. MALDI MSI data has a high number of dimensions, making it difficult to process and interpret. Therefore, a Canonical Correlation Analysis (CCA) based method was applied to rank the intensities of the acquired m/z values based on their power to discriminate the primary carcinomas with metastatic potential from those without. In doing so, the most “important” m/z values were identified and only these were used during classification, thus reducing the dimensions of the data. By doing this, a classification accuracy of 38 out of 43 patients (88.4%) was achieved (statistical analysis of this chapter is a part of Lyrone Winderbaum’s thesis and partially published as a technical note [143], for details see appendix 1).

The top 20 m/z values considered to have the highest level of classification potential as determined by the CCA were identified using a combination of *in situ* MALDI MS/MS and LC-MS/MS. The spatial distribution of the discriminative m/z values were visualized in the form of ion intensity maps using SCiLS lab software. The intensity maps represent the intensity of the particular m/z values within the tissue section and it ranges from blue (lowest) to red (highest). The comparisons between the intensities were plotted as receiver operating characteristic (ROC) curves. ROC curves compare sensitivity (i.e. the true-positive rate) versus specificity (i.e. the false-positive rate) across a range of m/z values by plotting a curve called an AUC (area under the ROC curve). The AUC assumes values between 0 and 1 and expresses the discrimination power of the m/z signal. A perfect discrimination would yield an AUC value equal to 1 (abundant in group 1) or 0 (abundant in group 2) [144]. The AUC closer to 0.5 indicates that there is no discrimination between two sample groups.

The ion intensity maps displaying the spatial localization and relative intensity for m/z 967.42 (peptide belongs to protein plectin) and 976.42 (peptide belongs to protein α -Actin-2) across

the TMA1 are provided in Section 4.3, Figure 1 and 2. Although m/z 967.42 was included in the list of the top 20 list, SCiLS analysis only revealed a slight difference between the ion intensity between tumours with and without LNM with AUC of 0.396. However, the SCiLS analysis of m/z 976.42 revealed a difference between the ion intensity map between tumours with and without LNM with AUC of 0.313.

In order to verify the MALDI MSI results, DIA LC-MS/MS was carried out on laser capture micro dissected (LCM) digested healthy and primary tumour tissue (n=4 with LNM and n=4 without LNM), which had not been included in the TMA analysis. Prior to the comparison of primary tumours with and without LNM, the expression of each peptide in tumour tissue was normalised in comparison to the expression of the same peptide in corresponding normal tissue for each patient. A trend of decreased abundance in the primary tumours with LNM was observed for both plectin and α -Actin-2 (Section 4.3, Figure 3).

The differential abundance of plectin and α -Actin-2 was further validated across the TMAs by immunohistochemistry. As indicated from SCiLS analysis described above, IHC staining of plectin does not distinguishes EC patients with and without LNM (Section 4.3, Figure 4). However, significant difference in the staining intensity of α -Actin-2 was observed. The staining intensity of α -Actin-2 was reduced in primary tumours with LNM compared to those without (Section 4.3, Figure 5).

In summary, we were able to predict the LNM status of EC patients from primary tumour cores with an accuracy of 88.4% [143]. Additionally, we were able to identify and validate α -Actin-2 and plectin as potential discriminators for LNM in EC [145]. Our results show that lymph node status can be predicted with a relatively high accuracy in women within the patient cohort used.

4.2 Statement of Authorship

Title of Paper	Lymph node metastasis of primary endometrial cancers: Associated proteins revealed by MALDI imaging
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	P. Mittal, M. Klingler-Hoffmann, G. Arentz, L. Winderbaum, N.A. Lokman, C. Zhang, L. Anderson, J. Scurry, Y. Leung, C.J.R. Stewart, J. Carter, G. Kaur, M.K. Oehler, P. Hoffmann, Lymph node metastasis of primary endometrial cancers: Associated proteins revealed by MALDI imaging, <u>Proteomics</u> , 2016, 16, 1793-1801.

Principal Author

Name of Principal Author (Candidate)	Parul Mittal		
Contribution to the Paper	Performed analysis on all samples, interpreted data and wrote the manuscript		
Overall percentage (%)	40%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	15/08/2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- the candidate's stated contribution to the publication is accurate (as detailed above);
- permission is granted for the candidate to include the publication in the thesis; and
- the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Manuela Klingler-Hoffmann		
Contribution to the Paper	Equally contributed author, supervised development of work, data interpretation and manuscript evaluation		
Signature		Date	30/01/2017

Name of Co-Author	Georgia Arentz		
Contribution to the Paper	Data interpretation and manuscript evaluation		
Signature		Date	24/2/2017

Name of Co-Author	Lyron Winderbaun		
Contribution to the Paper	Data interpretation and manuscript evaluation		
Signature		Date	30/1/2017

Name of Co-Author	Noor A Lokman		
Contribution to the Paper	Data interpretation and manuscript evaluation		
Signature		Date	7/4/2017

Name of Co-Author	Chao Zhang		
Contribution to the Paper	Participated in experimental design		
Signature		Date	3/4/2017

Name of Co-Author	Lyndal Anderson		
Contribution to the Paper	Provided tissue samples, ethics approval and manuscript evaluation		
Signature		Date	23/11/16

Name of Co-Author	JAMES SCURRY		
Contribution to the Paper	Provided tissue samples and manuscript evaluation		
Signature		Date	23.08.2016

Name of Co-Author	Yee Leung		
Contribution to the Paper	Provided tissue samples and manuscript evaluation		
Signature	Digitally signed by Yee Leung DN: cn=Yee Leung, o=The University of Western Australia, ou=SWIH, email=yee.leung@uwa.edu.au, c=AU Date: 2016.08.15 12:41:56 +08'00'	Date	15 August 2016

Name of Co-Author	COLIN JR STEWART		
Contribution to the Paper	Provided tissue samples and manuscript evaluation		
Signature		Date	15/8/16

Name of Co-Author	JONATHAN CARTER		
Contribution to the Paper	Provided tissue samples and manuscript evaluation		
Signature		Date	15/8/16

Name of Co-Author	Martin K. Oehler		
Contribution to the Paper	Supervised development of work, data interpretation and manuscript evaluation		
Signature		Date	12/04/17

Name of Co-Author	Peter Hoffmann		
Contribution to the Paper	Supervised development of work, data interpretation, manuscript evaluation and acted as corresponding author		
Signature		Date	30/1/2017

This page intentionally left blank

RESEARCH ARTICLE

Lymph node metastasis of primary endometrial cancers: Associated proteins revealed by MALDI imaging

Parul Mittal^{1,2*}, Manuela Klingler-Hoffmann^{1,2*}, Georgia Arentz^{1,2}, Lyron Winderbaum^{1,2}, Noor A Lokman^{1,8}, Chao Zhang^{1,2}, Lyndal Anderson³, James Scurry⁴, Yee Leung⁵, Colin JR Stewart⁵, Jonathan Carter⁶, Gurjeet Kaur⁷, Martin K. Oehler^{8,9} and Peter Hoffmann^{1,2**}

¹ Adelaide Proteomics Centre, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

² Institute for Photonics and Advanced Sensing (IPAS), University of Adelaide, Adelaide, South Australia, Australia

³ Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia

⁴ Faculty of Health and Medicine, University of New South Wales, Callaghan, New South Wales, Australia

⁵ School of Women's and Infants' Health, University of Western Australia, Crawley, Western Australia, Australia

⁶ Department of Gynaecological Oncology, Chris O'Brien Lifehouse, Camperdown, New South Wales, Australia

⁷ Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Pulau Pinang, Malaysia

⁸ Discipline of Obstetrics and Gynaecology, Research Centre for Reproductive Health, School of Medicine, Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia

⁹ Department of Gynaecological Oncology, Royal Adelaide Hospital, Adelaide, South Australia, Australia

Metastasis is a crucial step of malignant progression and is the primary cause of death from endometrial cancer. However, clinicians presently face the challenge that conventional surgical-pathological variables, such as tumour size, depth of myometrial invasion, histological grade, lymphovascular space invasion or radiological imaging are unable to predict with accuracy if the primary tumour has metastasized. In the current retrospective study, we have used primary tumour samples of endometrial cancer patients diagnosed with ($n = 16$) and without ($n = 27$) lymph node metastasis to identify potential discriminators. Using peptide matrix assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI), we have identified m/z values which can classify 88% of all tumours correctly. The top discriminative m/z values were identified using a combination of *in situ* sequencing and LC-MS/MS from digested tumour samples. Two of the proteins identified, plectin and α -Actin-2, were used for validation studies using LC-MS/MS data independent analysis (DIA) and immunohistochemistry. In summary, MALDI-MSI has the potential to identify discriminators of metastasis using primary tumour samples.

Received: November 17, 2015

Revised: February 12, 2016

Accepted: April 5, 2016

Keywords:

Biomarker / Biomedicine / Endometrial cancer / Lymph node metastasis / Proteomics



Additional supporting information may be found in the online version of this article at the publisher's web-site

Correspondence: Professor Peter Hoffmann, Adelaide Proteomics Centre, University of Adelaide, Gate 8 Victoria Drive, Adelaide, South Australia, 5005

E-mail: peter.hoffmann@adelaide.edu.au

Fax: +61 (08) 0 8313 4362

Abbreviations: α -Actin-2, alpha-actin-2; CCA, canonical correlation analysis; EC, endometrial cancer; FFPE, formalin fixed paraffin embedded; LC-MS/MS, liquid chromatography mass spectrometry; LDA, linear discriminant analysis; LNM, lymph node

1 Introduction

Endometrial cancer (EC) is the most frequent malignant tumour occurring in the female reproductive system. Based on the histology and clinical practice EC is divided into

metastasis; L^{OO}, leave one out; MALDI-MSI, matrix assisted laser desorption/ionisation mass spectrometry imaging

*These authors contributed equally to the work.

**To whom all correspondence and requests for reprints should be addressed.

Significance of the study

Endometrial cancer is the most common gynaecological malignancy in Australia with 2256 diagnosed cases in 2010 and 381 associated deaths in 2011. The presence or absence of lymph node metastasis is the most important prognostic factor in early stage I endometrial cancer. Of the patients diagnosed with stage I disease, around 10% will have pelvic lymph node metastases (LNM). Despite the small percentage of patients who suffer from metastasis the majority undergo radical treatment including the removal of lymph nodes; a precautionary measure carried out due to our current inability to accurately stage the disease. Lymph node removal is

associated with significant complications including lower extremity lymphoedema, occurring in up to 38% of patients. A classification system based around predictive tissue markers of metastasis is therefore essential to determine the optimal treatment strategy for endometrial cancer patients and to reduce disease morbidity. In this study we show that data acquired from the MALDI imaging of primary endometrial carcinomas can be used to successfully predict the presence or absence of LNM with an overall accuracy of 88.4%. The development of such a classification method shows the diagnostic potential of MALDI imaging in determining the metastatic potential of primary carcinomas.

two subgroups: low-grade endometrioid adenocarcinomas (Type I) and high-grade endometrioid and non-endometrioid (Type II) carcinomas [1]. Type I EC accounts for 65% of all EC cases. These cancers are usually low grade, associated with oestrogen excess, obesity and atypical endometrial hyperplasia. Due to presence of early symptoms, the vast majority of type I carcinomas are diagnosed when the tumour is still confined to the uterus and these are cured by surgery in most cases resulting in a 5-year survival rate above 70% [2]. However, patients with recurrent EC have a 5 year survival rate below 40% despite intervention with chemotherapy and/or radiotherapy [2].

Currently accepted prognostic [prognostic = survival] factors for EC include the histological subtype, grade and International Federation of Gynecology and Obstetrics (FIGO) stage of the disease [3]. Deep (>50%) myometrial invasion and high histological grade are associated with the presence of lymph node metastasis (LNM) and adverse prognosis in EC [4]. Based on pathology findings, Milam et al. categorised EC patients as low risk for LNM if the tumour size was ≤ 2 cm, well or moderately differentiated, and depth of myometrial invasion was $\leq 50\%$. Tumours that did not meet all three criteria were considered at high risk for LNM [5]. A study by Jacques et al. showed that a large percentage of EC will be misclassified before surgery [6]. Although only 15% of EC patients have or develop metastasis, the majority undergo radical treatment including removal of lymph nodes. This procedure is associated with significant complications including lower extremity lymphoedema, deep vein thrombosis and vascular or nerve injury [7].

Metastasis is a complex process in which tumour cells from the primary neoplasm acquire the ability to survive detachment, intravascular circulation, and implant and proliferate at a secondary site [8]. Recent studies in a variety of cancers have shown that metastatic potential of a primary tumour can be determined by genomic and proteomic analysis [9–12]. Moreover, in recent years a number of proteomic approaches have identified primary tumour signatures that accurately predict the presence of LNM, overall survival or disease recurrence

[13–18]. Therefore, we hypothesised that the metastatic potential of a primary EC could be reflected at the proteome level and could be determined through the identification of molecular discriminators using MALDI mass spectrometry imaging (MALDI-MSI).

MALDI-MSI allows for the *in situ* characterisation of tissue sections, enabling the relative quantification and spatial expression profiling of thousands of peptides within and between tissues [19]. This technique allows a comparison of tissue histology with corresponding spatially resolved mass spectrometric information. The identification of differentially expressed m/z values cannot be achieved from a standard MALDI-MSI experiment alone [20] and moreover, single m/z values can match several peptide masses from the corresponding LC-MS/MS data of the tissue extract, even when using relative high mass accuracy in the MALDI-MSI experiment using internal calibrants. Although, *in situ* MS/MS data are often poor and don't reveal direct identification; combining the two methods and matching m/z values of the intact peptides and y and b ions from both fragmentation patterns provides identification in most cases.

Recently, we have shown the capacity of MALDI-MSI to discriminate regions of healthy endometrial tissue from tumour [21]. Here we present an analysis of primary EC specimens with ($n = 16$) and without LNM ($n = 27$) by MALDI-MSI. Upon data acquisition, a Canonical Correlation Analysis (CCA) based method was applied to rank the intensities of the acquired MALDI m/z values based upon their power to discriminate the primary carcinomas with metastasis from those without. This ranking was used to reduce the dimension of the data to the top ranked m/z values prior to classification by linear discriminant analysis (LDA), and the performance of this classification was judged by leave one out (LOO) cross validation (for details see Winderbaum et al. [22]).

The top m/z values were targeted for identification using the complementary techniques of *in situ* MALDI MS/MS and matching to peptide sequences obtained from traditional nano-flow liquid chromatography electrospray ionisation tandem mass spectrometry (nanoLC-ESI-MS/MS). The

differential expression of plectin and α -Actin-2 between the primary carcinomas with and without LNM was further validated using label-free quantitative LC-MS/MS and immunohistochemistry. In summary we have identified α -Actin-2 as a potential discriminator of increased risk of LNM in EC.

2 Materials and methods

2.1 Sample collection and tissue specimens

Formalin-fixed paraffin-embedded (FFPE) tissue samples were retrieved from the archives of the Institute of Medical and Veterinary Science, Adelaide, South Australia, Royal Prince Alfred Hospital, Sydney, New South Wales, John Hunter Hospital, Newcastle, New South Wales and King Edward Memorial Hospital, Perth, Western Australia. The study was approved by the ethics committees of the different institutions. The histo-morphological and clinical information for the patients is provided in Supporting Information Table 1.

2.2 Tissue microarray construction

TMA's were constructed as previously described [23]. Primary tumour sections were annotated by a pathologist and two cores representing tumour centre were used to construct two TMA's. However, after TMA construction, all medical records were carefully reviewed by a clinician again and samples with mixed carcinoma were excluded from the study. This led to overall exclusion of 14 samples from 57 patient samples and the final cohort was comprised of a total of 43 patients ($n = 16$ with LNM and $n = 27$ without LNM), which were assembled in two TMA's (named TMA1 and TMA2). Serial 6 μ m sections were mounted onto indium titanium oxide (ITO) conductive glass slides for MALDI-MSI analysis (Bruker Daltonics, Bremen, Germany). Sections were also placed on plain glass slides and Polyethylene Naphthalate (PEN) membrane slides (MicroDissect, Herborn, Germany), and haematoxylin and eosin (H&E) stained for pathology annotation and laser microdissection (LMD), respectively.

2.3 Sample preparation for FFPE MALDI-MSI

MALDI-MSI was carried out as previously described [24] in duplicate on consecutively cut TMA sections. Briefly, the tissues were deparaffinised, subjected to heat induced citric acid antigen retrieval (HIAR) (10 mM citric acid, pH = 6), followed by digestion with trypsin gold (Promega, Madison, WI) at 37°C for 2 h. Internal calibrants and α -cyano-4-hydroxycinnamic acid (CHCA) matrix were overlaid using an ImagePrep station [25].

2.4 TMA analysis by MALDI-MSI

MALDI imaging of the TMA's was carried out using an ultra-flexTreme MALDI-TOF/TOF MS (Bruker Daltonics, Bremen, Germany) with flexControl v3.0.1 and flexImaging v4.0.1 soft-

ware (Bruker Daltonics) in positive reflectron mode over a detection range of m/z 800–4000 Da. MS spectra was acquired in a raster based grid with a centre to centre resolution of 60 μ m. Technical replicates of each TMA were measured. After data acquisition, the matrix was removed with 70% ethanol, the TMA cores were H&E stained, and digitally scanned using a Nanozoomer (Hamamatsu Photonics, Shimadzu, Japan) and images were obtained using imaging software (NDP scan software v2.2, Hamamatsu Photonics). To align the MS data with the tissue histology, the H&E scanned cores were co-registered with the MALDI-MSI results and annotated using the flexImaging software. Tissue regions containing only areas of primary tumour were selected and the spectra lists for these regions were exported as .XML files.

2.5 MALDI-MSI canonical correlation analysis (CCA)

A detailed description of the CCA method can be found in Winderbaum et al. [22]. Briefly, arbitrarily located bins with a width of 0.25 Da were used to discretise m/z domains in order to group peaks. The intensity values of the peaks in each of these defined peak groups (m/z bin) were log-transformed and averaged across the annotated tumour areas for each patient. These averages were assembled into a data matrix with columns representing each analysed patient and rows corresponding to the m/z bins. The rows of this matrix were then ranked using the developed CCA based method for their ability to distinguish between primary carcinomas with and without LNM. A dimension reduced submatrix consisting of the top ranked rows was then analysed using linear discriminant analysis (LDA) in order to predict the LNM status of the 43 patients. All analyses were replicated in parallel using two alternate (shifted) bin locations, resulting in three analyses in total. A majority rule was used to combine the data.

2.6 Data analysis using SCiLS lab

For data dependent visualisation of tissue morphological regions, raw data was uploaded into the SCiLS lab software (v2015a, GmbH, Bremen, Germany). Here the data was pre-processed including top hat baseline removal and total ion count (TIC) normalisation, and peak alignment and picking was performed [26]. The spatial expression profiles of the m/z values found to have discriminative power in the CCA were visualised in the form of ion intensity maps, and receiver operating characteristic (ROC) curves of these m/z values comparing the intensities of the tumours with and without LNM was generated.

2.7 Identification of m/z values by in-situ MALDI MS/MS and nanoLC-ESI-MS/MS

In order to gain peptide identifications for the m/z values of interest, *in situ* MS/MS was performed directly from the tissue used in the MALDI-MSI analysis and searched using Mascot (Version 2.3.02) as previously described [24]. For

matching back to peptide sequences obtained by data dependent acquisition, nanoLC-ESI-MS/MS, primary tumour regions of interest were collected from the TMA cores using LMD, subjected to HIAR, and digested with trypsin [23]. NanoLC-ESI-MS/MS was performed using an Ultimate 3000 RSLC system (Thermo-Fisher Scientific) coupled to an Impact IITM QTOF mass spectrometer (Bruker Daltonics) via an Advance CaptiveSpray source (Bruker Daltonics). Peptide samples were pre-concentrated onto a C18 trapping column (Acclaim PepMap100 C18 75 $\mu\text{m} \times 20\text{ mm}$, Thermo-Fisher Scientific) at a flow rate of 5 $\mu\text{L}/\text{min}$ in 2% ACN 0.1% TFA for 10 min. Peptide separation was performed using a 75 μm ID C18 column (Acclaim PepMap100 C18 75 $\mu\text{m} \times 50\text{ cm}$, Thermo-Fisher Scientific) at a flow rate of 0.2 $\mu\text{L}/\text{min}$ using a linear gradient from 5 to 45% B (A: 5% ACN 0.1% FA, B: 80% ACN 0.1% FA) over 130 min, followed by a 20 min wash with 90% B, and a 20 min equilibration with 5% A. MS scans were acquired in the mass range of 300 to 2200 m/z in a data-dependent fashion using Bruker's Shotgun Instant ExpertiseTM method. Singly charged precursor ions were excluded from acquisition. Collision energy ranged from 23 to 65% as determined by the m/z of the precursor ion.

Acquired spectra were subjected to peak detection, deconvolution, and re-calibration according to a lock mass using Compass DataAnalysis for OTOF (Version 1.7, Bruker Daltonics). Processed spectra were then exported to Mascot generic format and submitted to Mascot (Version 2.3.02) for identification. Search parameters were as follows; SwissProt Homo sapiens database search, trypsin digestion with up to two missed cleavages, variable modification of oxidation of methionine, MS mass tolerance of 40 ppm and a MS/MS mass tolerance of 0.2 Da. In Mascot an ion score cut off of 20 with a peptide significance threshold of ≤ 0.05 was used, which corresponds to a false discovery rate (peptide level) of $< 2\%$. Matching between the MALDI-MSI and nanoLC-ESI-MS/MS was done by comparing the experimental m/z values of the nanoLC-ESI-MS/MS sequenced peptides to the m/z values from the compiled peak bins.

2.8 Quantification of peptides by data independent acquisition (DIA) nano LC-ESI-MS/MS results

DIA nano LC-ESI-MS/MS was performed on LMD normal and cancer tissues from four patients with, and four patients without LNM. Nano LC was performed as described above using an Ultimate 3000 RSLC system coupled to an Impact IITM QTOF mass spectrometer. The Impact IITM QTOF acquired data using Bruker's Middle Band CIDTM method where a mass range of m/z 375 to 1206 is scanned in 26 Da increments with increasing collision energies of 20 to 36. Data were analysed in the Skyline software against a spectral library generated from the previous nano LC-ESI-MS/MS experiments [27]. The peptide and transition settings during analysis were as follows; trypsin was specified as the cleavage enzyme with a maximum of one missed cleavage, precursor charge states 2 and 3, ion charges 1 and 2, ion types y and b from ion 3 to 6,

ion match tolerance 0.1 m/z , a MS/MS filtering DIA isolation scheme from m/z 400–1206 (26 Da windows), a resolution 10 000, and only scans within 5 min retention time window of spectral library MS/MS identification used. Summed area intensities for the analysed peptides were calculated from y and b ions 3 to 6 (starting from ion 3). For each peptide analysed, the relative intensity in the tumour tissue was normalised to the relative intensity of the normal tissue from each patient.

2.9 Immunohistochemistry (IHC)

For the analysis of plectin and α -Actin-2 by IHC, 6 μm TMA sections were analysed as previously described [28]. Briefly, the tissue sections were dewaxed, rehydrated with xylene and ethanol and subjected to microwave antigen retrieval for 10 min at 100°C (Sixth Sense, Whirlpool, VIC, Australia) in 10 mM citric acid buffer pH = 6. TMA sections were incubated overnight at 4°C with either α -Actin-2 (1/500, rabbit polyclonal, ProteinTech, Chicago, USA) or plectin (1/250, rabbit monoclonal, Abcam, MA, USA) in blocking buffer (5% goat serum), followed by incubation with biotinylated anti-rabbit immunoglobulin (1/400, Dako, NSW, Australia) and streptavidin-HRP (1/500, Dako). Immunoreactivity was detected using diaminobenzidine (DAB)/H₂O₂ (Sigma Aldrich) substrate and counterstaining with haematoxylin (Sigma Aldrich). TMA slides were digitally scanned using a Nanozoomer and images were obtained using NDP view imaging software. Analysis was carried out in IHC Profiler-Image J [29]. For each tissue core, three representative photomicrographic images at 40x magnification were analysed.

3 Results and discussion

3.1 MALDI-MSI

MALDI-MSI was carried out on two TMA (TMA1 and TMA2), two replicates per patient were used resulting in a total of 86 primary endometrial carcinoma tissue cores from 43 patients with ($n = 16$) and without ($n = 27$) LNM. Peak groups were generated from the MALDI-MSI data and then ranked using a CCA based method for their ability to distinguish between the primary carcinomas with and without LNM. A list of the top m/z bins (peak groups) with the capacity to differentiate the primary cancer types is shown in Supporting Information Table 2. Reducing the data to these m/z values, and using LDA to discriminate between primary carcinomas with and without LNM, a classification accuracy of 38 out of 43 patients (88.4%) was achieved by LOO cross validation (for details see Winderbaum et al. [22]).

3.2 Identification of discriminative m/z values

The top discriminating m/z bins of 0.25 Da were centred at: 802.42, 857.42, 915.42, 941.42, 944.42, 967.42, 975.42, 976.42, 1027.67, 1032.67, 1115.42, 1138.67, 1157.67, 1161.67,

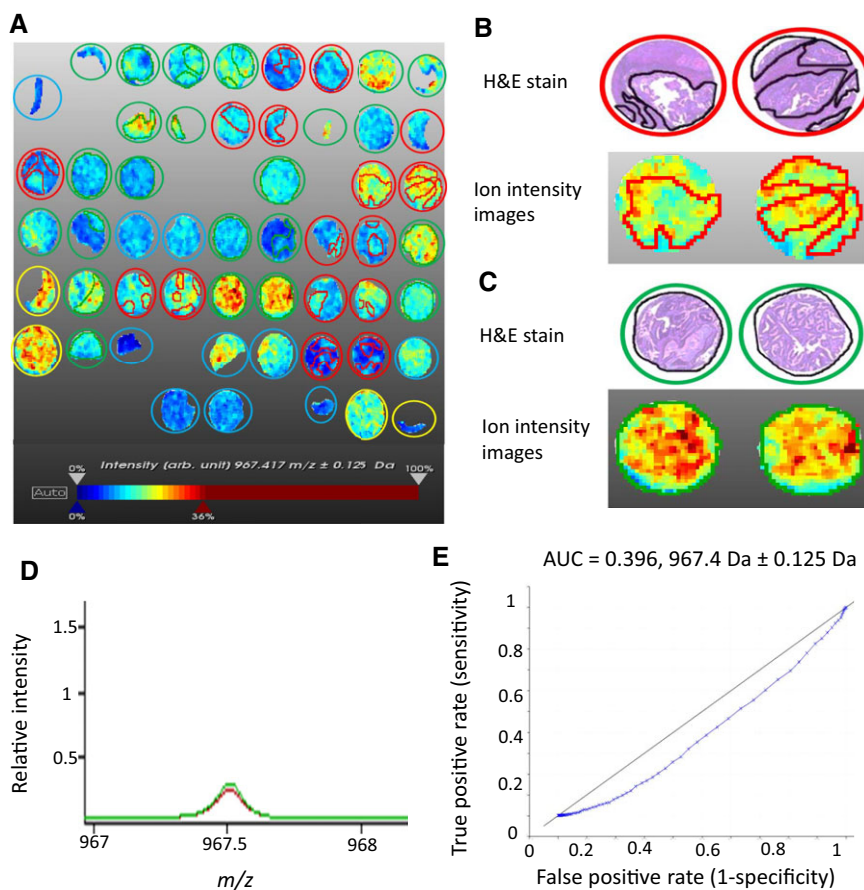


Figure 1. Representative MALDI-MSI images for m/z 967.42 $\pm$ 0.125 Da. (A) Overview of one TMA slide, MALDI-MSI image of (intensity range from blue (lowest) to red (highest)). The samples belonging to different groups are indicated by different coloured circles: control (yellow), with LNM (red), without LNM (green) and mixed carcinoma/not included in the study (blue). The tumour regions (red) within the samples have been annotated by a pathologist. (B) Magnification of two cancer tissue spots with LNM showing H&E stain and the ion intensity images of m/z 967.42. (C) Magnification of two cancer tissue spots without LNM showing H&E stain and the ion intensity images of m/z 967.42. (D) MALDI-MSI spectra displaying the mean spectrum from regions with LNM (red) and without LNM (green). (E) ROC curve with AUC of 0.396.

1167.67, 1198.67, 1242.67, 1406.67 and 1612.92 (in order of size). Potential peptide identifications for the top m/z bins, as ranked by the CCA, were compiled by matching back to sequences obtained by data dependent acquisition nanoLC-ESI-MS/MS (Supporting Information Table 2). Of the 20 m/z bins, 3 had no matches back to the tandem MS data, with the remaining 17 having 2 or more sequence matches. In order to confirm peptide identifications, the m/z values were targeted for in situ MS/MS directly off the tissue, from which 2 peptides could be verified; m/z 1198.67 AVFPSIVGRPR (α -Actin-2), and m/z 976.42 AGFAGDDAPR (α -Actin-2).

α -Actin-2 was targeted for further analysis given 3 of the top m/z bins matched to α -Actin-2 peptides. Moreover, the m/z value of 1501.42 was proteotypic for α -Actin-2 (data not shown). The m/z 967.42 matching to plectin was selected given only two possible nano LC-ESI-MS/MS sequence matches were obtained for this m/z value, and of these two matches the relative abundance of the plectin related peptide was significantly greater in the nano LC-ESI-MS/MS data than the alternative candidate and matched the high abundance in the MALDI-MSI data (data not shown).

Figures 1 and 2 show the ion intensity images for m/z 967.42 and 976.42 across the TMA. Tumour cores with LNM are circled red; tumours without LNM are circled green, yellow circles indicate controls and blue circles indicate cores

which have been excluded from the analysis. All tumour regions within the tumour cores are annotated in the corresponding colour. Magnification of two replicate cores from one representative patient with LNM (Figs. 1B, 2B) and without LNM (Figs. 1C, 2C) reveals the expression of the m/z value within the tissue cores (H&E stain top panel, ion intensity images bottom panel). Although m/z 967.42 was included in the list of the top 20 list, SCiLS analysis only revealed a slight difference between the ion intensity between tumours with and without LNM (Fig. 1). However, the SCiLS analysis of m/z 976.42 revealed a difference between the ion intensity map between tumours with and without LNM (Fig. 2).

3.3 Validation by DIA

The MALDI-MSI results for the α -Actin-2 and plectin peptides were verified by DIA nano LC-ESI-MS/MS. Analysis was carried out on LMD on normal and primary tumour tissue from 4 patients with LNM and 4 without LNM, who had not been included in the TMA analysis. DIA allows the differential quantification of the isobaric peaks by matching the retention time, as the unique fragment ions of the two species generate two different MS/MS chromatograms [30]. Therefore, retention time was used when matching back the

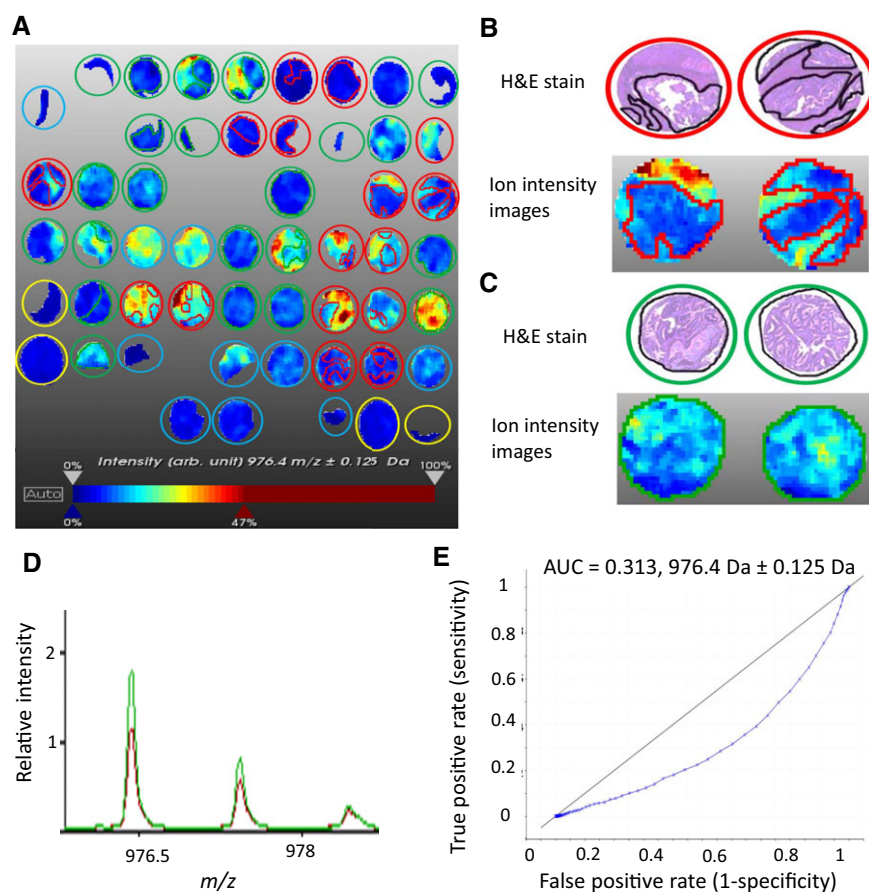


Figure 2. Representative MALDI-MSI images for m/z 976.42 $\pm$ 0.125 Da. (A) Overview of one TMA slide, MALDI-MSI image of (intensity range from blue (lowest) to red (highest)). The samples belonging to different groups are indicated by different coloured circles: control (yellow), with LNM (red), without LNM (green) and mixed carcinoma/not included in the study (blue). The tumour regions (red) within the samples have been annotated by a pathologist. (B) Magnification of two cancer tissue spots with LNM showing H&E stain and the ion intensity images of m/z 976.42. (C) Magnification of two cancer tissue spots without LNM showing H&E stain and the ion intensity images m/z 976.42. (D) MALDI-MSI spectra displaying the mean spectrum from regions with LNM (red) and without LNM (green). (E) ROC curve with AUC of 0.313.

data to the spectral library for DIA. The area intensities of the peptides matching back to the α -Actin-2 and plectin were summed for both tumour and normal tissues. The summed area intensity of the tumour was then normalised to that of the paired normal tissues. The normalised tumours were then compared with and without LNM using an unpaired T-Test in GraphPad Prism. A trend of increased expression

in the primary tumours without LNM was observed for both plectin and α -Actin-2 peptides (Fig. 3).

3.4 Validation by immunohistochemistry

The spatial expression profile of plectin was verified across the patient cohort by immunohistochemistry (Fig. 4A).

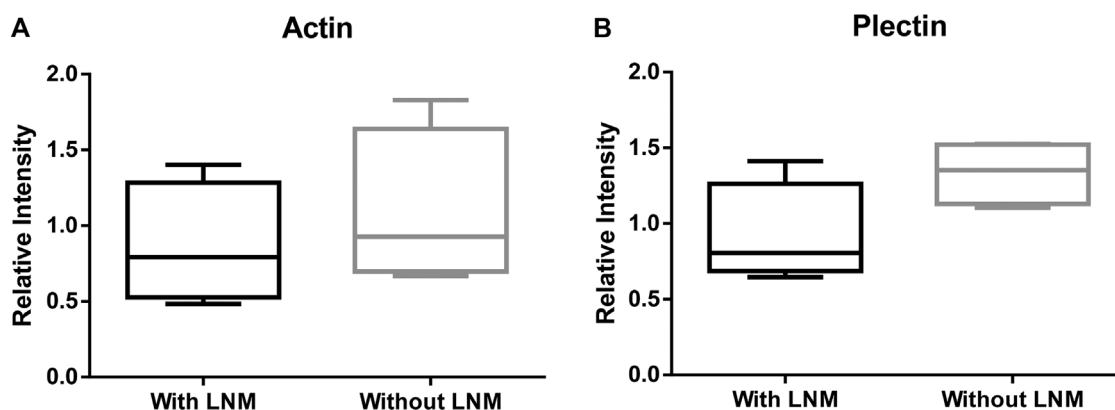


Figure 3. DIA analysis of tumour sections with ($n = 4$) and without ($n = 4$) LNM. (A) The relative abundance of plectin was analysed in comparison to normal tissue set to 1.0. (B) The relative abundance of α -Actin-2 was analysed in comparison to normal tissue set to 1.0. The error bar indicates the standard deviation.

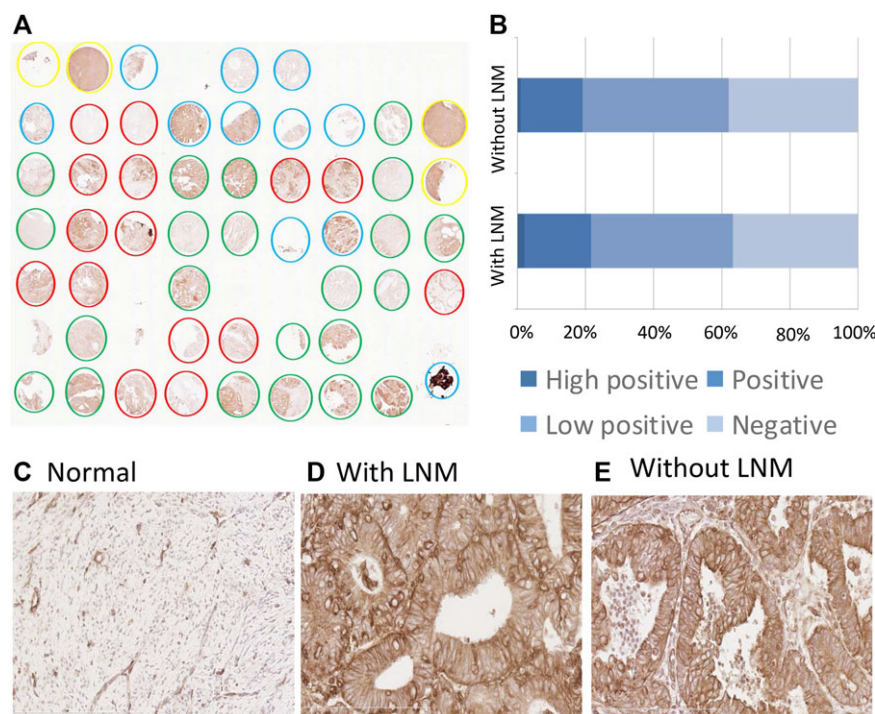


Figure 4. Immunohistochemical staining of plectin (A) 6 µm serial section of TMA 1 was used for immunohistochemistry (IHC). The different tissue types are encircled control (yellow), with LNM (red), without LNM (green) and mixed carcinoma/not included in the study (blue). (B) Quantitative analysis was performed using IHC profiler-Image J. For each tissue section, three representative photo-micrographic images at 40x magnification were used and each image was assigned a score of high positive, positive, low positive and negative staining. Shown is the average staining intensity distribution of all analysed images. Representative image of plectin immunostaining of normal tissue (C) tumour with LNM (D) without LNM (E).

Quantitative analysis of immunostaining was performed using IHC Profiler-Image J [29] and as expected this indicated no difference between tumours with and without LNM; shown is the average staining intensity across all tumour cores (Fig. 4B). Plectin scarcely stained normal tissue including stroma (Fig. 4C); the staining of tumours cells was strong (Figs. 4D–E). In summary plectin IHC can highlight tumour cells, but staining intensity does not distinguish cases with and without LNM.

This is in contrast to data presented in the human proteome atlas (Version 14, updated 2015-10-16), where a medium intensity stain of plectin was detected in healthy endometrium and absent or low staining was usually detected in EC with one of three antibodies. However, the other antibodies showed different staining patterns, making a precise interpretation of the data difficult.

It is known that tumour cell motility is required for invasion and metastasis [31, 32]. Plectin, has been found to be important in cytoskeletal network organisation [33]. A number of studies have shown that increased levels of plectin correlate with migration and invasion [33–36]. However, our data show an increase in plectin staining in EC tumour cells unrelated of their metastatic potential. Furthermore, we have identified a number of peptides from the protein plectin which showed potential to discriminate between tumours with and without LNM. One representative *m/z* 967.4 (amino acid 1045–1052) is shown in Fig. 1 and has been confirmed by DIA (Fig. 3). However, the immunohistochemistry analysis failed to identify a difference in staining intensity of tumours with LNM when compared to tumours without LNM. A recent study has identified S1047 in plectin as a potential phosphorylation

site offering one possible explanation of the discrepancy of results [37].

The spatial expression profile and differential expression of α -Actin-2 was verified across the patient cohort by immunohistochemistry (Fig. 5A). Quantitative analysis of staining was performed using IHC Profiler-ImageJ and a significant difference in the negative staining (1.8 fold, $p < 0.05$) between tumours without and with LNM was observed; shown is the average staining intensity across all tumour cores (Fig. 5B). α -Actin-2 stained normal tissue (Fig. 5C), while the staining of tumours with LNM was reduced (Fig. 5D) when compared to tumour without LNM (Fig. 5E). In summary α -Actin-2 IHC staining intensity has the potential to distinguish between tumours with and without LNM.

This is in agreement with the data presented in the human proteome atlas (Version 14, updated 2015-10-16), where a medium intensity stain of α -Actin-2 was detected in healthy endometrium and absent or low staining was detected in EC with four different antibodies.

Cytoskeletal proteins facilitate the biological modes of cells: migration, cell division, differentiation and cell death. It is therefore not surprising that these proteins are frequently identified in comparative proteomic studies. α -Actin-2, the human aortic smooth muscle actin gene, is one of six different actin isoforms which have been identified and has been described to facilitate migration of cells. A number of studies have shown that increased expression of α -Actin-2 leads prevents cellular motility [38, 39]. Accordingly, decreased expression of α -Actin-2 has been shown to contribute to the metastatic potential of basal cell carcinoma [40]. Our data indicate a down-regulation of α -Actin-2 in tumours with LNM

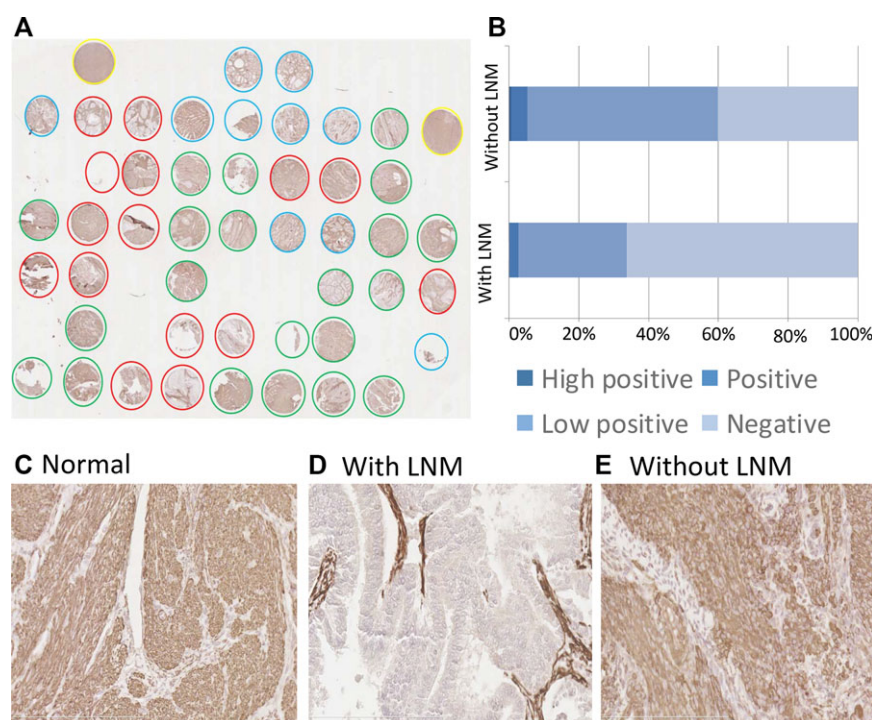


Figure 5. Immunohistochemical staining of α -Actin-2 (A) 6 μ m serial section of TMA 1 was used for immunohistochemistry (IHC). The different tissue types are encircled control (yellow), with LNM (red), without LNM (green) and mixed carcinoma/not included in the study (blue). (B) Quantitative analysis was performed using IHC profiler-Image J. For each tissue section, three representative photomicrographic images at 40x magnification were used and each image was assigned a score of high positive, positive, low positive and negative staining. Shown is the average staining intensity distribution of all analysed images. Representative image of α -Actin-2 immunostaining of normal tissue (C), tumour with LNM (D) and without LNM (E).

and therefore α -Actin-2 may have potential as a biomarker for EC metastasis.

4 Concluding remarks

In summary, MALDI-MSI has the potential to identify the markers of tumour metastasis by providing spatial intensity of proteins/peptides that might be associated with different tissue types and facilitate developing disease. Using MALDI-MSI data, we found a number of m/z values that could predict the status of LNM with an overall accuracy of 88.4%. Additionally, the m/z values were identified as α -Actin-2 and plectin via in situ MS/MS (Supporting Information Fig. 1) and label free quantification (Supporting Information Table 2). Furthermore, DIA (Supporting Information Figs. 2–6) and immunohistochemistry was used for relative quantification and validation. The role of α -Actin-2 and plectin in metastasis has already been described previously and could be useful as potential biomarkers for distinguishing EC with and without LNM.

5 References

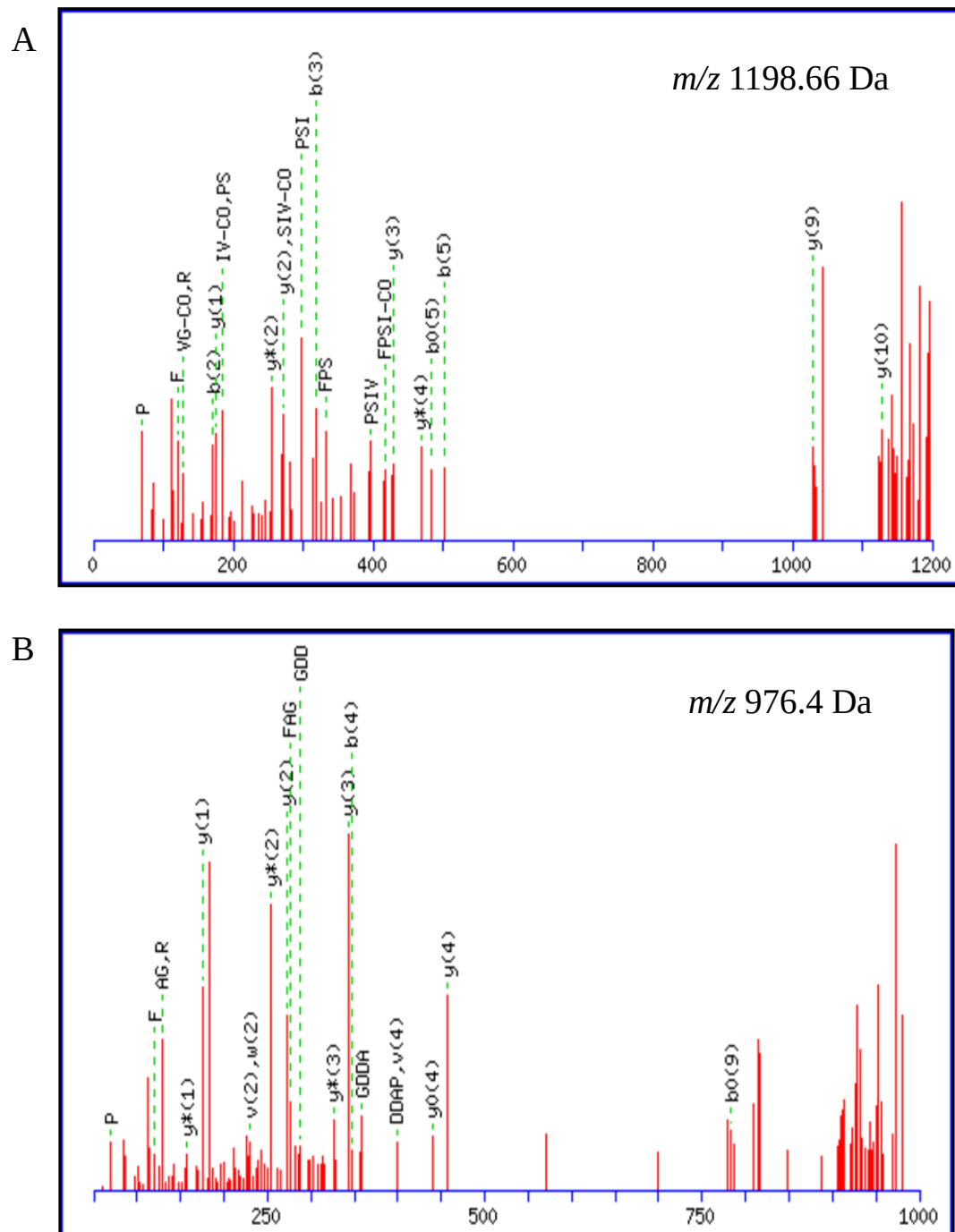
- [1] Bokhman, J. V., Two pathogenetic types of endometrial carcinoma. *Gynecol. Oncol.* 1983, 15, 10–17.
- [2] American Cancer Society 2015.
- [3] Larson, D. M., Connor, G. P., Broste, S. K., Krawisz, B. R., Johnson, K. K., Prognostic significance of gross myometrial invasion with endometrial cancer. *Obstet. Gynecol.* 1996, 88, 394–398.
- [4] Boronow, R. C., Morrow, C. P., Creasman, W. T., Disaia, P. J. et al., Surgical staging in endometrial cancer: clinical-pathologic findings of a prospective study. *Obstet. Gynecol.* 1984, 63, 825–832.
- [5] Milam, M. R., Java, J., Walker, J. L., Metzinger, D. S. et al., Nodal metastasis risk in endometrioid endometrial cancer. *Obstet. Gynecol.* 2012, 119, 286–292.
- [6] Jacques, S. M., Qureshi, F., Munkarah, A., Lawrence, W. D., Interinstitutional surgical pathology review in gynecologic oncology: I. Cancer in endometrial curettings and biopsies. *Int. J. Gynecol. Pathol.* 1998, 17, 36–41.
- [7] Creasman, W. T., Odicino, F., Maisonneuve, P., Quinn, M. A. et al., Carcinoma of the Corpus Uteri. *Int. J. Gynecol. Obstet.* 2006, 95, Supplement 1, S105–S143.
- [8] Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P. et al., *Molecular Cell Biology*, W. H. Freeman, New York 2000.
- [9] Meding, S., Nitsche, U., Balluff, B., Elsner, M. et al., Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging. *J. Proteome Res.* 2012, 11, 1996–2003.
- [10] Meding, S., Balluff, B., Elsner, M., Schöne, C. et al., Tissue-based proteomics reveals FXD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer. *J. Pathol.* 2012, 228, 459–470.
- [11] Hristov, A. C., Cope, L., Reyes, M. D., Singh, M. et al., HMGA2 protein expression correlates with lymph node metastasis and increased tumor grade in pancreatic ductal adenocarcinoma. *Mod. Pathol.* 2009, 22, 43–49.
- [12] Ji, F., Jin, X., Jiao, C. H., Xu, Q. W. et al., FAT10 level in human gastric cancer and its relation with mutant p53 level, lymph node metastasis and TNM staging. *World J. Gastroenterol.* 2009, 15, 2228–2233.

- [13] Liu, N. Q., Stingl, C., Look, M. P., Smid, M. et al., Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer. *J. Natl. Cancer Inst.* 2014, **106**, djt376.
- [14] Zamaruddin, S. N., Saleh, A., Yang, Y. H., Hamid, S. et al., Four-protein signature accurately predicts lymph node metastasis and survival in oral squamous cell carcinoma. *Hum. Pathol.* 2013, **44**, 417–426.
- [15] Pavlou, M. P., Dimitromanolakis, A., Martinez-Morillo, E., Smid, M. et al., Integrating meta-analysis of microarray data and targeted proteomics for biomarker identification: application in breast cancer. *J. Proteome Res.* 2014, **13**, 2897–2909.
- [16] Cao, H. H., Zhang, S. Y., Shen, J. H., Wu, Z. Y. et al., A three-protein signature and clinical outcome in esophageal squamous cell carcinoma. *Oncotarget* 2014, **6**, 5435–5448.
- [17] Shipitsin, M., Small, C., Giladi, E., Siddiqui, S. et al., Automated quantitative multiplex immunofluorescence in situ imaging identifies phospho-S6 and phospho-PRAS40 as predictive protein biomarkers for prostate cancer lethality. *Proteome Sci.* 2014, **12**, 40.
- [18] Mascini, N. E., Eijkel, G. B., Ter Brugge, P., Jonkers, J. et al., The use of mass spectrometry imaging to predict treatment response of patient-derived xenograft models of triple-negative breast cancer. *J. Proteome Res.* 2015, **14**, 1069–1075.
- [19] Cole, L. M., Bluff, J. E., Carolan, V. A., Paley, M. N. et al., MALDI-MSI and label-free LC-ESI-MS/MS shotgun proteomics to investigate protein induction in a murine fibrosarcoma model following treatment with a vascular disrupting agent. *Proteomics* 2014, **14**, 890–903.
- [20] Diehl, H. C., Beine, B., Elm, J., Trede, D. et al., The challenge of on-tissue digestion for MALDI MSI - a comparison of different protocols to improve imaging experiments. *Anal. Bioanal. Chem.* 2015, **407**, 2223–2243.
- [21] Mittal, P., Klingler-Hoffmann, M., Arentz, G., Zhang, C. et al., Proteomics of endometrial cancer diagnosis, treatment and prognosis. *Proteomics Clinical applications* 2015, **10**, 217–229.
- [22] Winderbaum, L., Koch, I., Mittal, P., Hoffmann, P., Canonical correlation analysis based variable ranking for the classification of MALDI-MS imaging data of tissue microarrays. *Proteomics* 2016, doi: 10.1002/pmic.201500451.
- [23] Meding, S., Martin, K., Gustafsson, O. J. R., Eddes, J. S. et al., Tryptic peptide reference data sets for MALDI imaging mass spectrometry on formalin-fixed ovarian cancer tissues. *J. Proteome Res.* 2012, **12**, 308–315.
- [24] Gustafsson, O. J., Eddes, J. S., Meding, S., McColl, S. R. et al., Matrix-assisted laser desorption/ionization imaging protocol for in situ characterization of tryptic peptide identity and distribution in formalin-fixed tissue. *Rapid Commun. Mass Spectrom.* 2013, **27**, 655–670.
- [25] Gustafsson, J. O. R., Eddes, J. S., Meding, S., Koudelka, T. et al., Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS. *J. Proteomics* 2012, **75**, 5093–5105.
- [26] Alexandrov, T., Kobarg, J. H., Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics* 2011, **27**, i230–i238.
- [27] Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y., MacCoss, M. J., Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nat. Protocols* 2015, **10**, 887–903.
- [28] Ween, M. P., Lokman, N. A., Hoffmann, P., Rodgers, R. J. et al., Transforming growth factor-beta-induced protein secreted by peritoneal cells increases the metastatic potential of ovarian cancer cells. *Int. J. Cancer* 2011, **128**, 1570–1584.
- [29] Varghese, F., Bukhari, A. B., Malhotra, R., De, A., IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS one* 2014, **9**, e96801.
- [30] Sidoli, S., Simithy, J., Karch, K. R., Kulej, K., Garcia, B. A., Low resolution data-independent acquisition in an LTQ-orbitrap allows for simplified and fully untargeted analysis of histone modifications. *Anal. Chem.* 2015, **87**, 11448–11454.
- [31] Liotta, L. A., Stracke, M. L., Aznavoorian, S. A., Beckner, M. E., Schiffmann, E., Tumor cell motility. *Semin. Cancer Biol.* 1991, **2**, 111–114.
- [32] Stracke, M. L., Aznavoorian, S. A., Beckner, M. E., Liotta, L. A., Schiffmann, E., Cell motility, a principal requirement for metastasis. *Exs* 1991, **59**, 147–162.
- [33] Rikardsen, O. G., Magnussen, S. N., Svineng, G., Hadler-Olsen, E. et al., Plectin as a prognostic marker in non-metastatic oral squamous cell carcinoma. *BMC Oral Health* 2015, **15**, 1–8.
- [34] Katada, K., Tomonaga, T., Satoh, M., Matsushita, K. et al., Plectin promotes migration and invasion of cancer cells and is a novel prognostic marker for head and neck squamous cell carcinoma. *J. Proteomics* 2012, **75**, 1803–1815.
- [35] Bausch, D., Thomas, S., Mino-Kenudson, M., Fernández-del, C. C. et al., Plectin-1 as a novel biomarker for pancreatic cancer. *Clin. Cancer Res.* 2011, **17**, 302–309.
- [36] Bausch, D., Mino-Kenudson, M., Fernández-del Castillo, C., Warshaw, A. et al., Plectin-1 is a Biomarker of Malignant Pancreatic Intraductal Papillary Mucinous Neoplasms. *J. Gastrointest. Surg.* 2009, **13**, 1948–1954.
- [37] Bian, Y., Song, C., Cheng, K., Dong, M. et al., An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J. Proteomics* 2014, **96**, 253–262.
- [38] Beha, G., Sarli, G., Brunetti, B., Sassi, F. et al., Morphology of the myoepithelial cell: immunohistochemical characterization from resting to motile phase. *Scientific World Journal* 2012, 252034.
- [39] Rønnevig-Jessen, L., Petersen, O. W., A function for filamentous alpha-smooth muscle actin: retardation of motility in fibroblasts. *J. Cell Biol.* 1996, **134**, 67–80.
- [40] Uzquiano, M. C., Prieto, V. G., Nash, J. W., Ivan, D. S. et al., Metastatic basal cell carcinoma exhibits reduced actin expression. *Mod. Pathol.* 2008, **21**, 540–543.

This page intentionally left blank

4.4 Supplementary File

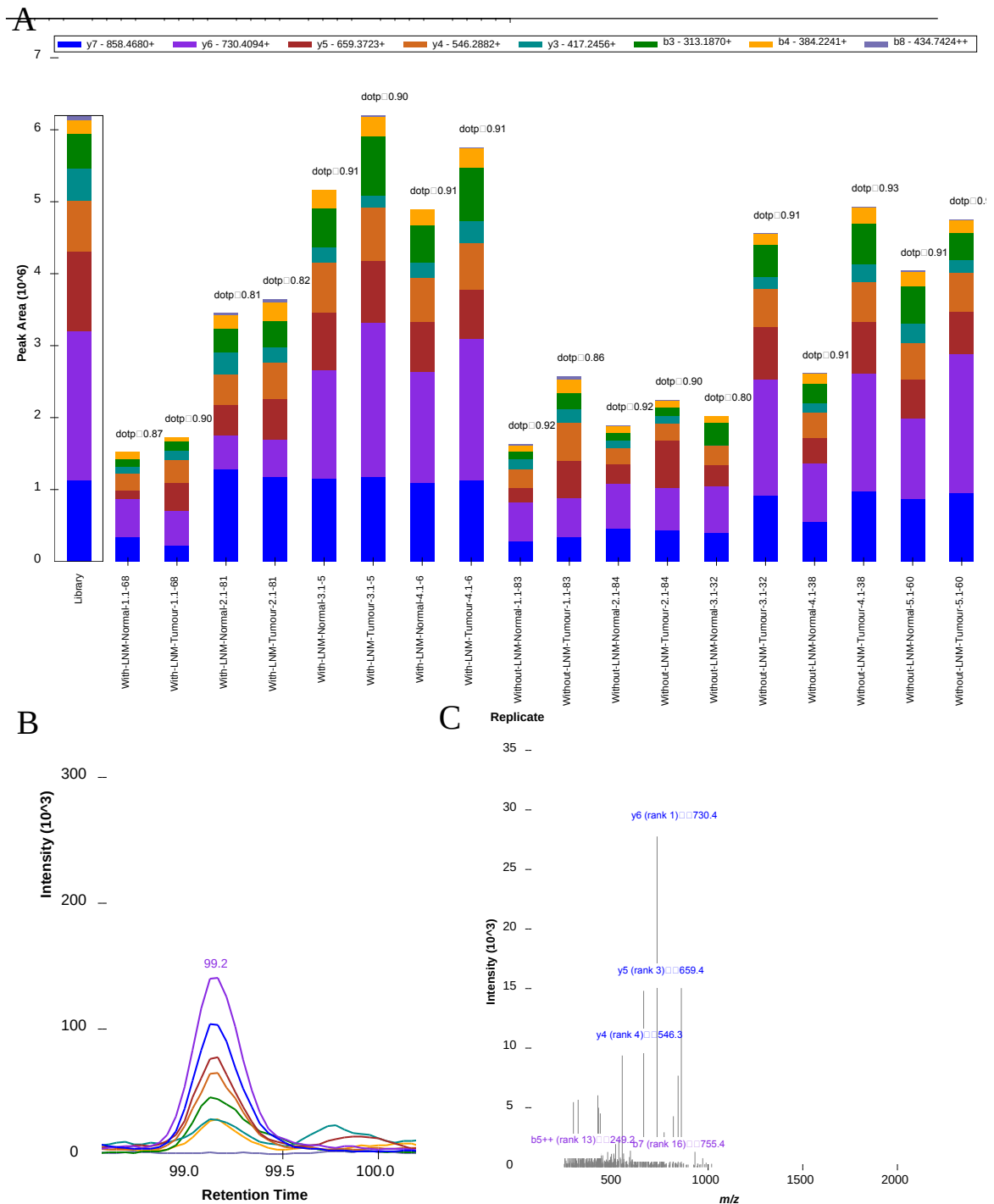
Supplementary Figure 1



Supplementary Figure 1: Tryptic peptide identified by *in situ* MS/MS as α -Actin-2, Aortic smooth muscle (ACTA_Human). 6 μ m FFPE EC TMA was analysed by tryptic peptide MALDI-MSI followed by *in situ* MS/MS on an ultrafleXtreme MALDI-TOF/TOF instrument. Selected m/z values from Supplementary Table 2 were fragmented *in situ* to generate a list of daughter ions to identify the peptide sequence (A) m/z 1198.66 Da AVFPSIVGRPR (B) m/z 976.4 Da AGFAGDDAPR

Supplementary Figure 2

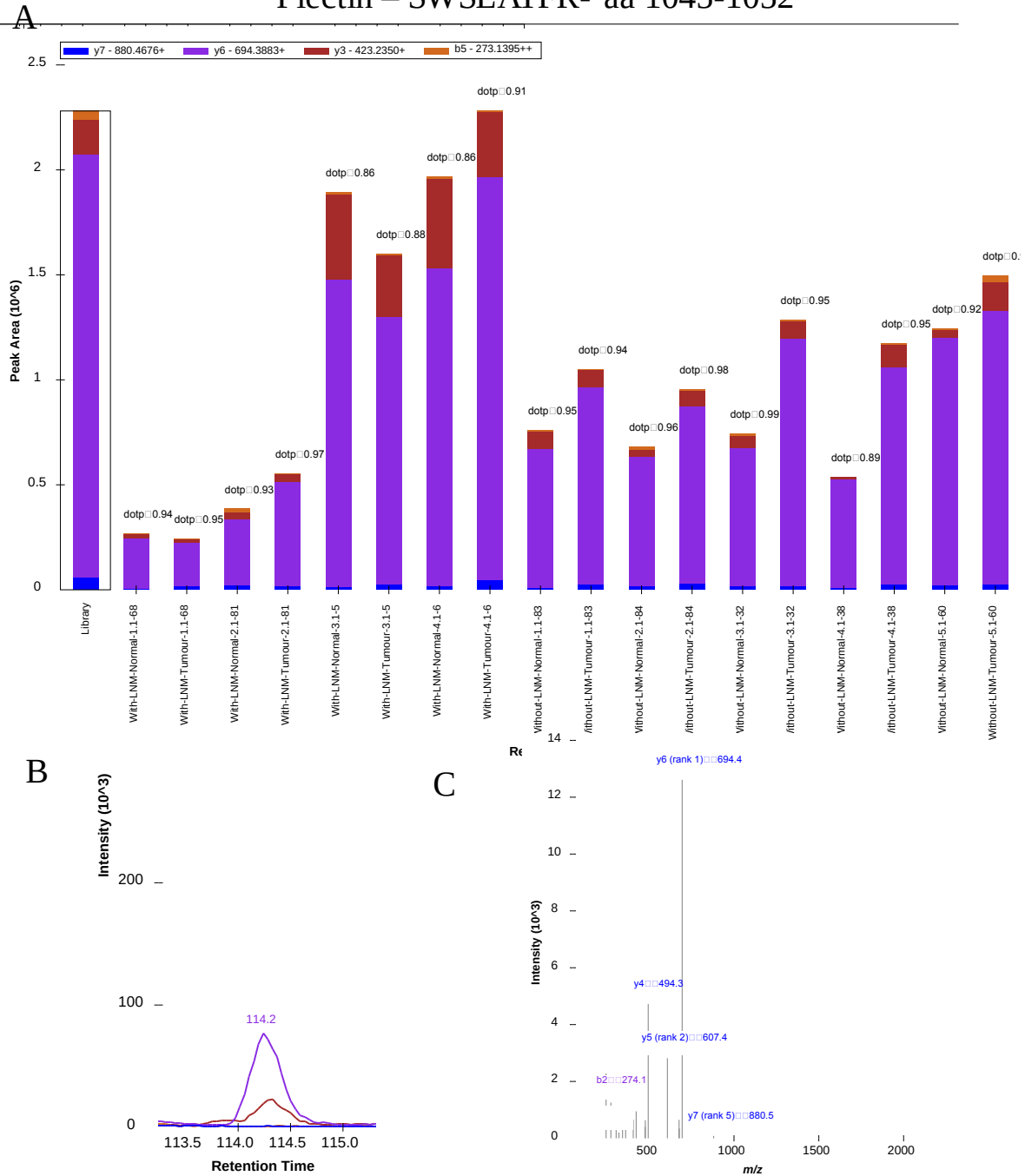
Plectin – ALQALEELR- aa 1670-1679



Supplementary Figure 2: DIA analysis confirmed proteins associated with LNM. The relative abundance of plectin peptide ALQALEELR was analysed by DIA of with (n=4) and without (n=4) LNM tumour sections in comparison to their paired normal tissue. (A) Peak area intensity (B) Peptide transition (C) Spectral Library

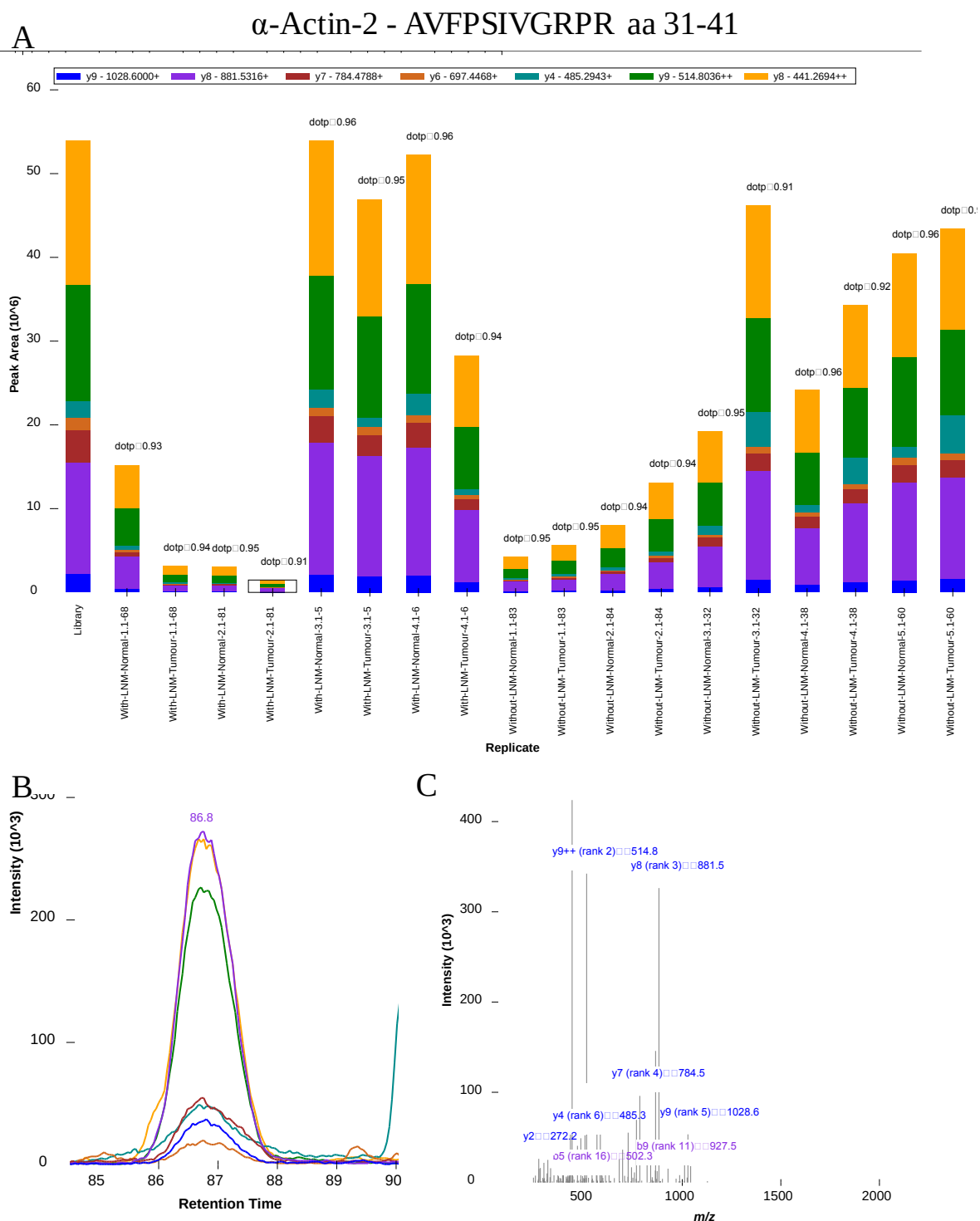
Supplementary Figure 3

Plectin – SWSLATFR- aa 1045-1052



Supplementary Figure 3: DIA analysis confirmed proteins associated with LNM. The relative abundance of plectin peptide SWSLATFR was analysed by DIA of with (n=4) and without (n=4) LNM tumour sections in comparison to their paired normal tissue. (A) Peak area intensity (B) Peptide transition (C) Spectral Library

Supplementary Figure 4

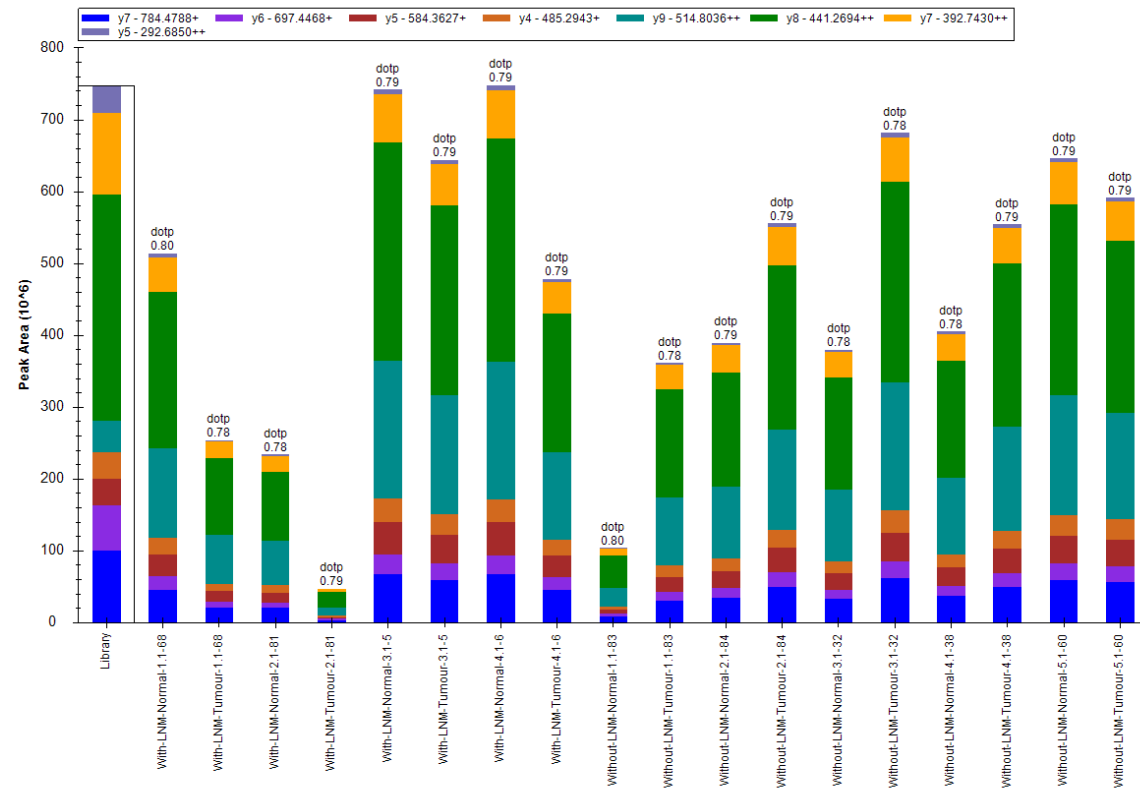


Supplementary Figure 4: DIA analysis confirmed proteins associated with LNM. The relative abundance of α -Actin-2 peptide AVFPSIVGRPR was analysed by DIA of with (n=4) and without (n=4) LNM tumour sections in comparison to their paired normal tissue.(A) Peak area intensity (B) Peptide transition (C) Spectral Library

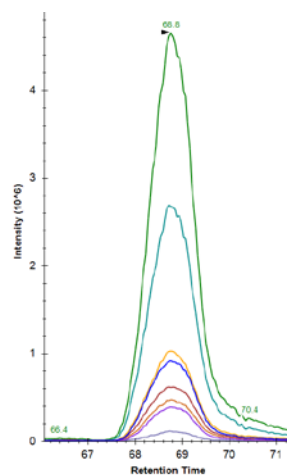
Supplementary Figure 5

α -Actin-2 – AVFPSIVGRPR - aa 31-41

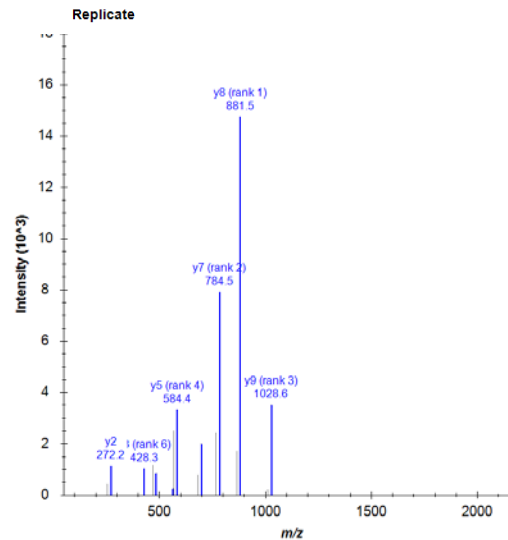
A



B



C

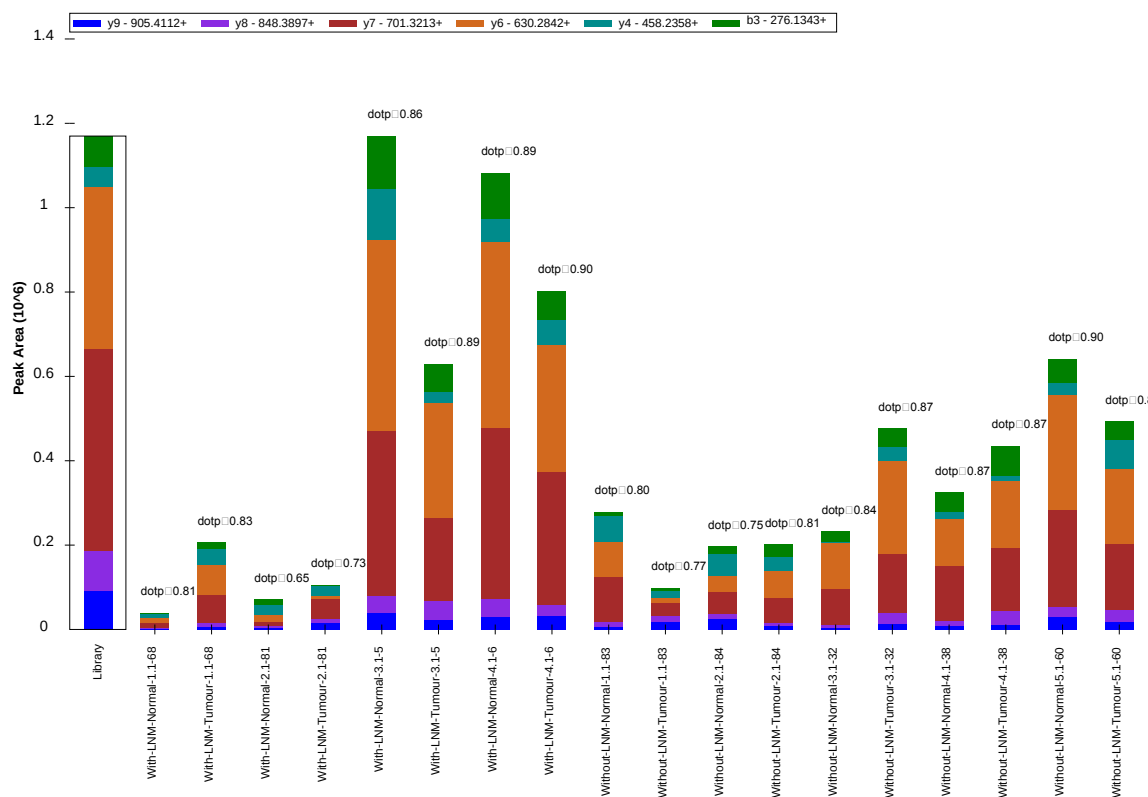


Supplementary Figure 5: DIA analysis confirmed proteins associated with LNM. The relative abundance of α -Actin-2 peptide AVFPSIVGRPR was analysed by DIA of with (n=4) and without (n=4) LNM tumour sections in comparison to their paired normal tissue.(A) Peak area intensity (B) Peptide transition (C) Spectral Library

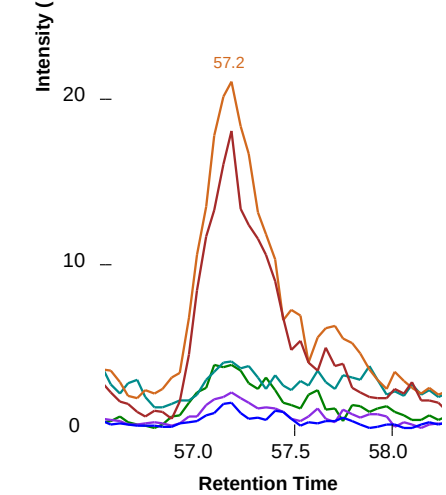
Supplementary Figure 6

α -Actin-2 - AGFAGDDAPR- aa 21-30

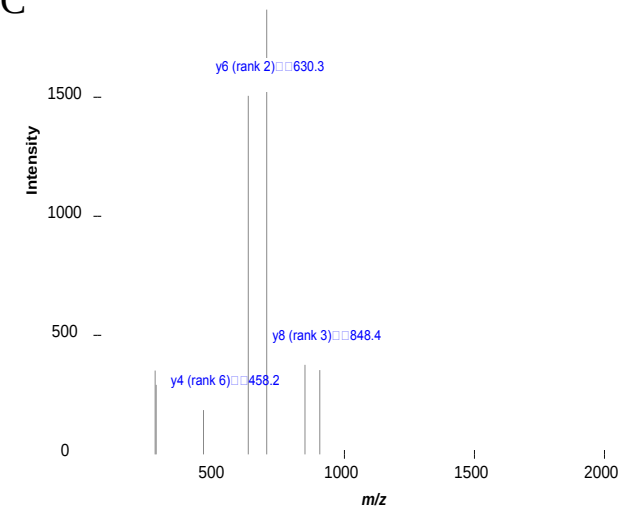
A



B



C



Supplementary Figure 6: DIA analysis confirmed proteins associated with LNM. The relative abundance of α -Actin-2 peptide AGFAGDDAPR was analysed by DIA of with (n=4) and without (n=4) LNM tumour sections in comparison to their paired normal tissue. (A) Peak area intensity (B) Peptide transition (C) Spectral Library

Supplementary Table 1: Clinicopathological characteristics of endometrial cancer patient samples used in the study

Histology	No. of patients
Histological stage (FIGO 2009)	
Stage I	27
Stage II	0
Stage III	16
Histological grade	
G1	20
G2	14
G3	9
Distance to Serosa	
$\leq 5\text{mm}$	16
$> 5\text{mm}$	26
Information not available	1
Myometrium invasion (MI)	
$\text{MI} \leq 5\text{mm}$	16
$\text{MI} > 5\text{mm}$	25
Information not available	2
Lymph Vascular Space Invasion (LVSI)	
Present	21
Absent	22
Lymph node metastasis	
Present	16
Absent	27

Supplementary Table 2: All possible nano-LC-ESI-MS/MS peptide matches to the top *m/z* peak groups as ranked by the CCA

Ranked m/z values by CCA ¹	Protein Accession Name	Peptide experimental m/z	Peptide experimental charge	Peptide experimental MR	Peptide calculated MR	Peptide score, Mascot	Peptide expected score, Mascot	Peptide sequence
967.417	PLEC_HUMAN	484.256	2	966.4974	966.4923	29.76	0.05	SWSLATFR
	NXN_HUMAN	484.3001	2	966.5856	966.5862	33.38	0.0067	EVIAGPLLR
1157.667	RS3_HUMAN	579.309	2	1156.6034	1156.5951	51.77	0.00014	IMLPWDPTGK
	SPTN1_HUMAN	386.5497	3	1156.6272	1156.6029	33	0.011	LAQFVEHWK
	CPNE1_HUMAN	579.2889	2	1156.5632	1156.5553	30.4	0.014	SDPFLEFFR
	PABP1_HUMAN	579.3409	2	1156.6673	1156.6604	29.37	0.017	FSPAGPILSIR
1198.667	ACTB_HUMAN	599.8602	2	1197.7058	1197.6982	77.62	1.70E-07	AVFPSIVGRPR
	ACTC_HUMAN	599.8602	2	1197.7058	1197.6982	77.62	1.70E-07	AVFPSIVGRPR
	ACTA_HUMAN	599.8602	2	1197.7058	1197.6982	77.62	1.70E-07	AVFPSIVGRPR
	RO60_HUMAN	599.8472	2	1197.6799	1197.6717	75.09	4.60E-07	LGLENAEALIR
	H14_HUMAN	599.8398	2	1197.665	1197.6605	62.3	9.50E-06	ASGPPVSELITK
802.417	no match	no match	no match	no match	no match	no match	no match	no match
975.417	EF1A1_HUMAN	488.2835	2	974.5524	974.5437	47.72	0.00028	LPLQDVYK
	ERF3A_HUMAN	488.2838	2	974.5531	974.58	46.91	0.00034	LPIVDKYK
	ATPB_HUMAN	488.2882	2	974.5618	974.5549	36.87	0.0029	IGLFGGAGVGK
	RL11_HUMAN	488.283	2	974.5515	974.5437	26.46	0.038	YDGILPGK
1242.667	HS90B_HUMAN	621.8602	2	1241.7059	1241.6979	81.32	1.20E-07	ADLINNLGTIAK
	HS90A_HUMAN	621.8602	2	1241.7059	1241.6979	81.32	1.20E-07	ADLINNLGTIAK
	H90B2_HUMAN	621.8602	2	1241.7059	1241.6979	81.32	1.20E-07	ADLINNLGTIAK
	RL22_HUMAN	621.846	2	1241.6774	1241.6728	73.35	9.70E-07	AGNLGGGVVTIER
	GRP75_HUMAN	621.8458	2	1241.677	1241.6728	65.65	5.70E-06	DAGQISGLNLVLR
	K2C7_HUMAN	621.8281	2	1241.6416	1241.6364	57.4	3.90E-05	GQLEALQVDGGR
	TERA_HUMAN	621.8246	2	1241.6346	1241.6252	51.76	0.00013	EVDIGIPDATGR
	2AAA_HUMAN	621.8785	2	1241.7424	1241.7343	50.48	6.40E-05	LSTIALALGVER
	SRP14_HUMAN	621.8211	2	1241.6277	1241.6227	50.45	0.00018	FQMAYSNLLR
	TLN1_HUMAN	414.8787	3	1241.6142	1241.604	28.44	0.027	FLPSELRDEH
1161.667	ACTB_HUMAN	581.3166	2	1160.6187	1160.6111	70.28	2.40E-06	EITALAPSTMK
	ACTC_HUMAN	581.3166	2	1160.6187	1160.6111	70.28	2.40E-06	EITALAPSTMK
	ACTA_HUMAN	581.3166	2	1160.6187	1160.6111	70.28	2.40E-06	EITALAPSTMK
	CUL5_HUMAN	581.3374	2	1160.6602	1160.6594	40.75	0.0026	TLWSLVAFPK
	RET4_HUMAN	581.28	2	1160.5455	1160.5325	35.46	0.0051	FSGTWYAMAK
1167.667	BCLF1_HUMAN	389.9103	3	1166.7092	1166.7023	66.54	8.00E-07	LLASTLVHSVK
	ASNS_HUMAN	584.3244	2	1166.6342	1166.6335	46.38	0.00078	ELYLFDVLR
	RS4X_HUMAN	389.8951	3	1166.6636	1166.656	44.25	0.00041	GNKPWISLPR
	HNRPM_HUMAN	584.299	2	1166.5834	1166.576	34.1	0.0086	NLPFDFTWK
	SBP1_HUMAN	584.3419	2	1166.6691	1166.6659	33.34	0.01	HEIVQTLSLK
1612.917	LV302_HUMAN	806.9034	2	1611.7923	1611.7853	127.92	3.40E-12	FSGSNSGNTATLTISR
	APOA1_HUMAN	806.9034	2	1611.7922	1611.7781	110.2	2.00E-10	LLDNWDSVTSTFSK
	NMT1_HUMAN	806.8977	2	1611.7808	1611.7603	74.14	7.70E-07	GFDVFNALDLMENK
	CLH1_HUMAN	806.9307	2	1611.8468	1611.8396	53.3	8.80E-05	ESYVETELIFALAK
1032.667	RS9_HUMAN	516.7995	2	1031.5845	1031.5764	59.12	2.50E-05	LFEGNALLR
	SYFB_HUMAN	516.8008	2	1031.5871	1031.5764	30.18	0.018	DLLFQALGR
	H31T_HUMAN	344.8695	3	1031.5868	1031.5876	26.64	0.04	YRPGTVALLR
1027.667	HXX1_HUMAN	514.3199	2	1026.6253	1026.6186	44.56	0.00061	GAALITAVGVVR
	PGM1_HUMAN	514.2806	2	1026.5466	1026.5386	39.81	0.0018	SIFDFSALK
	KAIN_HUMAN	514.2785	2	1026.5424	1026.5386	34	0.017	LGFTDLFSK
941.417	UBP2L_HUMAN	471.3083	2	940.6021	940.5957	59.13	1.30E-05	IDLAVLLGK
	CH60_HUMAN	314.5459	3	940.6159	940.6069	36.71	0.00051	IGIEIKR
	TXD17_HUMAN	471.3096	2	940.6046	940.5957	35.48	0.0012	VTAVPTLLK
	ACOC_HUMAN	471.2506	2	940.4867	940.4807	32.82	0.019	LLFFWNSK
	GSHB_HUMAN	471.2964	2	940.5782	940.5705	31.12	0.0055	ALAEGVLLR
	SCMC1_HUMAN	471.2448	2	940.475	940.4589	30.66	0.031	MNIFGGFR
	CPNE3_HUMAN	471.2799	2	940.5453	940.5342	25.92	0.03	SPLGEVAIR
976.417	In situ identified	In situ identified	In situ identified	In situ identified	In situ identified	In situ identified	In situ identified	AGFAGDDAPR
1406.667	HXX1_HUMAN	703.8682	2	1405.7218	1405.7123	100.28	2.20E-09	LSDETLIDIMTR
	ENOA_HUMAN	703.8667	2	1405.7188	1405.7089	80.57	2.00E-07	GNPTVEVDLFTSK
	SRPR_HUMAN	469.5788	3	1405.7145	1405.7023	71.03	1.90E-06	DAAGIAMEIAAFAR
	ENOG_HUMAN	703.8667	2	1405.7188	1405.7089	62.59	1.30E-05	GNPTVEVDLYTAK
	SAMP_HUMAN	703.8407	2	1405.6669	1405.6626	53.79	7.90E-05	AYSLFSYNTQGR
1138.667	CO6A3_HUMAN	569.8042	2	1137.5938	1137.5852	62.12	1.40E-05	SSIMAFAGNK
	G6PD_HUMAN	380.2176	3	1137.6309	1137.6182	61.92	3.50E-05	LKLEDFAR
	EF2_HUMAN	569.7671	2	1137.5197	1137.5091	32.49	0.0077	YEWDDVAEAR

1115.417	AL1A1_HUMAN	372.564	3	1114.6701	1114.6597	63.25	3.10E-06	ILDIESGKK
	PGS2_HUMAN	558.3289	2	1114.6432	1114.6387	62.77	7.50E-06	VSPGAFPLVK
	TMED4_HUMAN	372.5551	3	1114.6435	1114.6346	52.94	7.20E-05	DKLTEQLR
	TOP2B_HUMAN	558.3203	2	1114.626	1114.6169	50.68	0.00038	DLIQLVQR
	AK1A1_HUMAN	558.3176	2	1114.6207	1114.6209	50.34	0.0002	MPLIGLGTWK
	ALDR_HUMAN	558.3176	2	1114.6207	1114.6209	50.34	0.0002	MPILGLGTWK
	EIF3E_HUMAN	558.2876	2	1114.5606	1114.5369	43.46	0.0022	MLFDYLADK
	TENX_HUMAN	558.3062	2	1114.5979	1114.5983	30.26	0.041	SVTVTPLDAGR
944.417	H2A1B_HUMAN	472.7713	2	943.5279	943.524	60.32	2.50E-05	AGLQFPVGR
	H2A1D_HUMAN	472.7713	2	943.5279	943.524	60.32	2.50E-05	AGLQFPVGR
	H2A1A_HUMAN	472.7713	2	943.5279	943.524	60.32	2.50E-05	AGLQFPVGR
	RBMX_HUMAN	472.8001	2	943.5856	943.5776	50.79	4.50E-05	IVEVLLMK
	H2AV_HUMAN	472.7736	2	943.5326	943.524	47.17	0.00053	AGLQFPVGR
	WFS1_HUMAN	472.7905	2	943.5665	943.5702	26.76	0.022	TLTDLLR
857.417	no match	no match	no match	no match	no match	no match	no match	no match
1905.917	HSPB1_HUMAN	953.5054	2	1904.9963	1904.9843	129.4	1.70E-12	LATQSNEITIPVTFESR
	IGHG2_HUMAN	635.9761	3	1904.9064	1904.8866	79.15	2.00E-07	TTPPMLSDSGSFFLYSK
	SBP1_HUMAN	953.4614	2	1904.9082	1904.9004	69.19	2.00E-06	NTGTEAPDYLATVDVDPK
915.417	H2AY_HUMAN	458.2781	2	914.5416	914.5338	61.18	1.80E-05	AGVIFPVGR
	H2AW_HUMAN	458.2781	2	914.5416	914.5338	61.18	1.80E-05	AGVIFPVGR
	ZO1_HUMAN	458.2778	2	914.541	914.5338	44.67	0.0011	LGSWLAIR
	COCA1_HUMAN	458.279	2	914.5435	914.5437	37.8	0.0054	IGVLITD GK

¹ m/z value in the middle of the 0.25 Da wide bins

This page intentionally left blank

Annexin A2 and Alpha Actinin 4 Expression
Correlates with Metastatic Potential of Primary
Endometrial Cancer

This page intentionally left blank

5.1 Summary

Multiple studies have been performed over the past few decades with the aim to identify protein biomarkers for lymph node metastasis in EC, preoperatively. However, none of these biomarkers are currently in clinical use. The only FDA approved EC biomarker is CA125, which has low sensitivity and selectivity. For the current study we compiled a list of previously proposed biomarkers, derived from both genomics and proteomics studies, in order to validate their differential abundance at the protein level. We also aimed to develop a predictive model using IHC of suggested biomarkers for comparison to a predictive model generated from MALDI MSI data in Chapter 4.

To assemble a list of potential markers for EC metastasis, we analysed data from The Cancer Genome Atlas (TCGA), which was established in 2005 to catalogue data for high throughput genome analysis and more recently proteomics analysis. At the time of analyses, the databank contained 514 EC patients, with RNA sequencing data available for 333 patients (n=54 with LNM and n=279 without LNM). From the analysis, we have chosen eleven differentially expressed genes: six most significant up- and five down-regulated genes to add to our target list. Following an extensive literature review the list has grown to 60 target proteins, which might be differentially expressed in primary tumours with and without LNM (Section 5.4, Supplementary Table 1). Using DDA LC-MS/MS and whole tissue sections from 10 patients (n=5 with LNM and n=5 without LNM), we were able to detect 23/60 proteins (Section 5.3, Table 1) and confirmed a significant difference in abundance of four proteins by DIA LC-MS/MS: annexin A2 (ANXA2), receptor tyrosine protein kinase erb-2 (ERBB2), epidermal growth factor receptor (EGFR), and α actinin 4 (ACTN4) (Section 5.3, Figure 1). Annexin A1 (ANXA1) was also chosen because of its previously defined role in cancer metastasis [146]. The differential abundance of EGFR and Erb2 in EC has been described before [147] and we therefore did not include those proteins in our validation approach.

Following relative quantification analysis, the spatial distribution of ANXA1, ANXA2 and ACTN4 peptides was visualised using peptide MALDI MSI in whole tissue sections (n=5 with LNM and n=5 without LNM) (Section 5.3, Figure 2) as well as tissue microarrays containing samples from a total of 43 patients (with LNM n=16 and without LNM n=27) (Section 5.4, Supplementary Figure 1-3). The intensity of the ANXA2 peptide (m/z 1542.83 Da) appears to be higher in tumours with LNM as compared to those without (Section 5.3, Figure 2b). In contrast, the intensity of ACTN4 peptide (m/z 1429.76 Da) and ANXA1 peptide (m/z 1099.29

Da) appears to be lower in with LNM compared to those without LNM (Section 5.3, Figure 2c-d). The three ROC curves display a reasonable capacity to discriminate between patients with and without LNM (Section 5.3, Figure 2b-d).

The proteins identified were further validated across TMAs by IHC. This allowed the simultaneous analysis of 43 patients, of which 39 had not been analysed by relative quantification DIA LC-MS/MS. An increase in immunostaining intensity of ANXA2 ($p = 0.0875$; Section 5.3, Figure 3a) and significant decrease in immunostaining intensity of ACTN4 ($p = 0.0019$; Section 5.3, Figure 3b) and ANXA1 ($p = 0.0026$; Section 5.3, Figure 3c) were observed in tumours with LNM versus those without LNM. These results correlated strongly with the results obtained from the DIA LC-MS/MS and MALDI MSI. In summary, our data indicate that ANXA2 is upregulated, whereas ANXA1 and ACTN4 are downregulated in primary tumours with LNM compared to those without.

Additionally, a predictive model for EC metastasis was generated using the clinicopathologic features, MALDI MSI data acquired in Chapter 4 and IHC data obtained above. From the 43 patients analysed, we have 35 patients for which all the following clinical variables have been recorded: age, tumour grade, tumour size, tumour extent, and LVSI. Fitting a logistic regression model for LNM based on all these clinical variables and iteratively removing the least significant of them resulted in only one clinical variable remaining in the model as significant: LVSI (AUC 0.69) (Section 5.3, Figure 4a), and this agrees with a recently published result [40]. Given the apparent importance of LVSI in the prediction of LNM status, we include it in all the other models considered.

Models using the IHC data for annexin A2, annexin A1, and α actinin 4 were also considered. The percentage of the staining scores, low positive, positive, and high positive immunostaining, were combined for a total positive percentage staining for each antibody and was included with LVSI. The same model was repeated and involved iteratively removing the least significant until a model with only LVSI and ACTN4 in which both are significant at the 5% level was reached (Section 5.3, Figure 4b). Similarly, the same process was repeated but without LVSI to see how the IHC data perform on their own, and these results differed (shown in Figure 4). The IHC data indicated that adding the data to LVSI slightly improved the predictive accuracy. However, IHC data alone did not outperform LVSI.

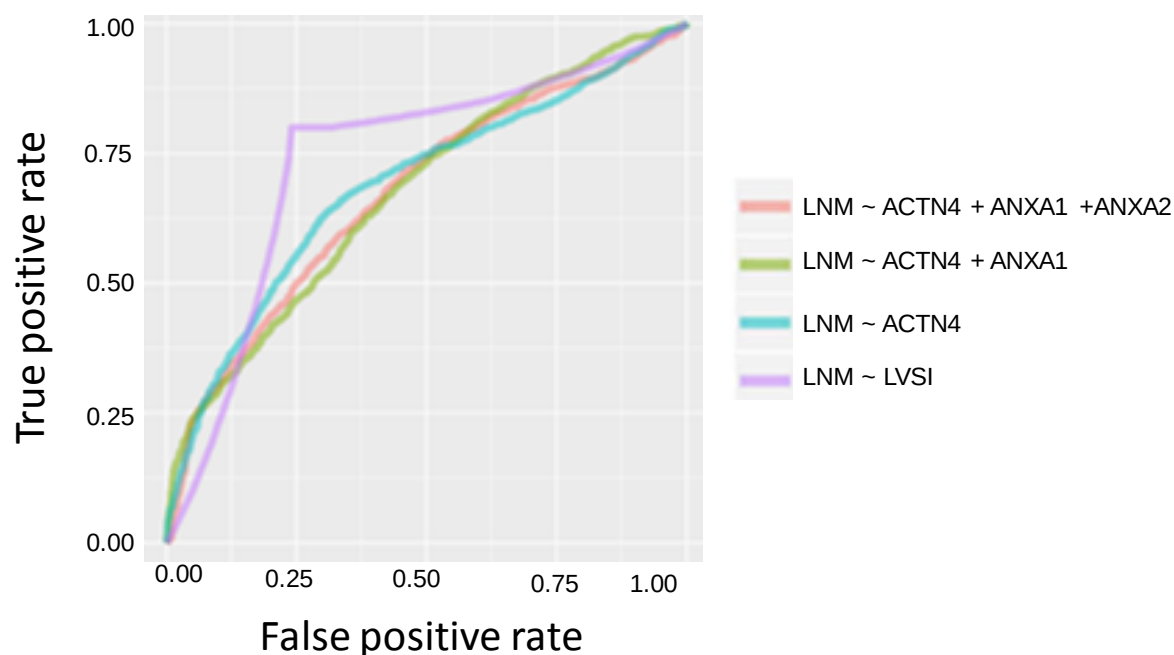


Figure 4: Receiver operating characteristic (ROC) analysis of the different logistic regression models. ROC analysis of the true positive rate (sensitivity) and false positive rate (specificity) for the prediction of lymph node metastasis of the immunohistochemistry analysis alone.

The prediction performance could be improved by adding highly ranked MALDI MSI variables. Given this, a model with LVSI and the top 10 MALDI MSI variables as ranked by CCA was considered, resulting in a model with LVSI and three of the top 10 MALDI MSI variables, ranked third, fifth and tenth by CCA as significant (AUC 0.86). The same process was repeated but without LVSI to see how the MALDI MSI variables performed on their own, and these results differed. The highly ranked MALDI MSI variables alone performed better than LVSI, shown in Figure 5.

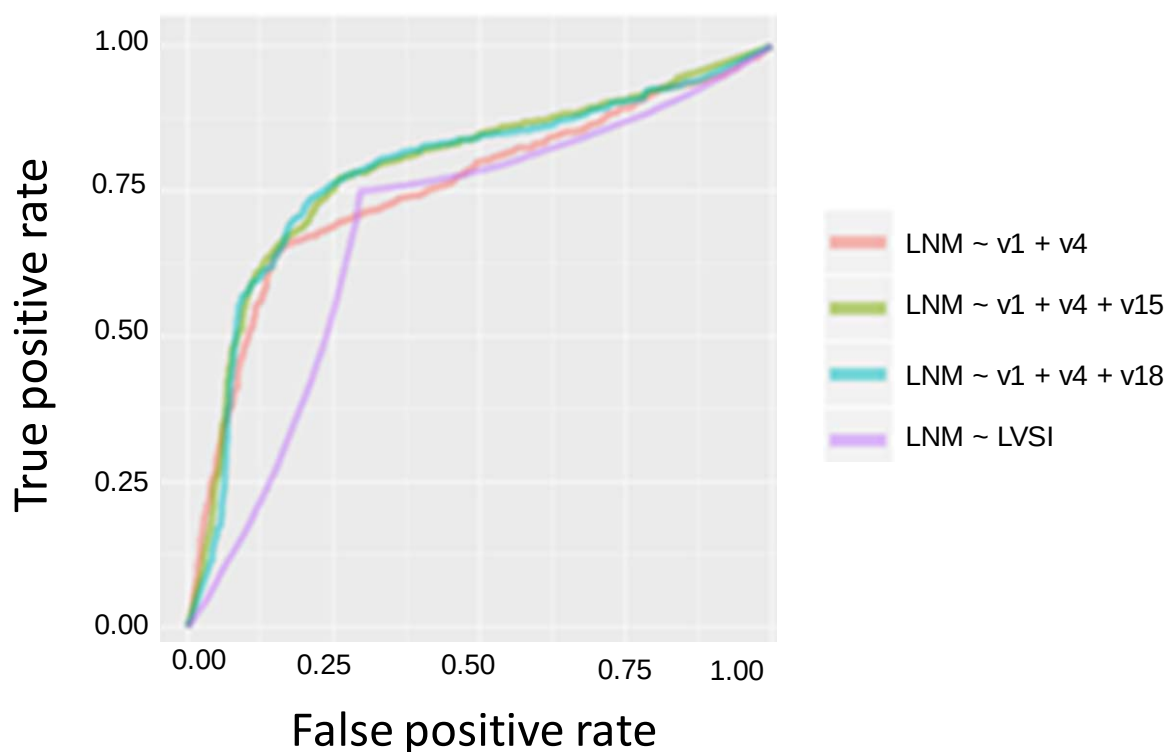


Figure 5: Receiver operating characteristic (ROC) curves for the prediction of LNM by several logistic regression models. ROC analysis of the true positive rate (sensitivity) and false positive rate (specificity) for the prediction of lymph node metastasis of the m/z values from the highly ranked MALDI MSI variables alone.

In summary, logistic regression analysis of our patient cohort indicated a correlation between LVSI and LNM, which has been shown previously [40]. A comparison of ROC curves from different multivariate regression models indicate that the multivariate logistic regression model of MALDI MSI data alone has the potential to predict LNM from primary tumours, and moreover potentially perform better than models containing immunohistochemistry data. Our study indicates that the molecular classification of EC can be achieved using MALDI MSI data, and provides independent prognostic information beyond the established risk factors. Furthermore a combination of LVSI and MALDI MSI data outperforms LVSI alone (Section 5.3, Figure 4b) and other predictive models in the literature [48]. The developed molecular classification tool has the potential to be used routinely in guiding treatment for individuals with EC.

5.2 Statement of Authorship

Title of Paper	Annexin A2 and alpha actinin 4 expression correlates with metastatic potential of primary endometrial cancer
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	<u>P. Mittal</u> , M. Klingler-Hoffmann, G. Arentz, L. Winderbaum, G. Kaur, L. Anderson, J. Scurry, Y. Leung, C.J.R. Stewart, J. Carter, P. Hoffmann, M.K. Oehler, Annexin A2 and alpha actinin 4 expression correlates with metastatic potential of primary endometrial cancer, <u>Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics</u> .

Principal Author

Name of Principal Author (Candidate)	Parul Mittal		
Contribution to the Paper	Performed analysis on all samples, interpreted data and wrote the first draft of the manuscript		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23/11/2016

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- the candidate's stated contribution to the publication is accurate (as detailed above);
- permission is granted for the candidate to include the publication in the thesis; and
- the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Manuela Klingler-Hoffmann		
Contribution to the Paper	Supervised development of work, data interpretation and manuscript evaluation		
Signature		Date	30/10/2017

Name of Co-Author	Georgia Arentz		
Contribution to the Paper	Data interpretation and manuscript evaluation		
Signature		Date	24/2/2017

Name of Co-Author	Lyron Winderbaun		
Contribution to the Paper	Data interpretation and manuscript evaluation		
Signature		Date	30/1/2017

Name of Co-Author	Gurjeet Kaur		
Contribution to the Paper	Annotated tissue samples and manuscript evaluation		
Signature		Date	25/11/16

Name of Co-Author	Lyndal Anderson		
Contribution to the Paper	Provided tissue samples, ethics approval and manuscript evaluation		
Signature		Date	23/11/16

Name of Co-Author	JAMES SCURRY		
Contribution to the Paper	Provided tissue samples, ethics approval and manuscript evaluation		
Signature		Date	24 NOVEMBER 2016

Name of Co-Author	Y Leung		
Contribution to the Paper	Provided tissue samples, ethics approval and manuscript evaluation		
Signature	Digitally signed by Yee Leung DN: cn=Yee Leung, o=The University of Western Australia, ou=SWIH, email=yee.leung@uwa.edu.au, c=AU Date: 2016.11.24 10:22:36 +08'00'	Date	24 November 2016

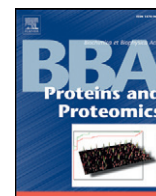
Name of Co-Author	COLIN STEWART		
Contribution to the Paper	Provided tissue samples, ethics approval and manuscript evaluation		
Signature		Date	24/11/16

Name of Co-Author	JONATHAN CARTER		
Contribution to the Paper	Provided tissue samples, ethics approval and manuscript evaluation		
Signature		Date	23/11/16

Name of Co-Author	Peter Hoffmann		
Contribution to the Paper	Supervised development of work, data interpretation, manuscript evaluation and acted as corresponding author		
Signature		Date	30/11/2017

Name of Co-Author	Martin K. Oehler		
Contribution to the Paper	Supervised development of work, data interpretation, manuscript evaluation and acted as corresponding author		
Signature		Date	12/04/17

This page intentionally left blank



Annexin A2 and alpha actinin 4 expression correlates with metastatic potential of primary endometrial cancer☆

Parul Mittal^{a,b}, Manuela Klingler-Hoffmann^{a,b}, Georgia Arentz^{a,b}, Lyron Winderbaum^{a,b}, Gurjeet Kaur^c, Lyndal Anderson^d, James Scurry^e, Yee Leung^f, Colin JR Stewart^f, Jonathan Carter^g, Peter Hoffmann^{a,b,*}, Martin K. Oehler^{h,i,**}

^a Adelaide Proteomics Centre, School of Biological Sciences, The University of Adelaide, SA 5005

^b Institute for Photonics and Advanced Sensing (IPAS), The University of Adelaide, SA 5005

^c Institute for Research in Molecular Medicine, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia

^d Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia

^e Faculty of Health and Medicine, University of New South Wales, Callaghan, New South Wales, Australia

^f School of Women's and Infants' Health, University of Western Australia, Crawley, Western Australia, Australia

^g Department of Gynaecological Oncology, Chris O'Brien Lifehouse, Camperdown, New South Wales, Australia

^h Discipline of Obstetrics and Gynaecology, School of Paediatrics and Reproductive Health, Research Centre for Reproductive Health, Robinson Institute, The University of Adelaide, SA 5005

ⁱ Department of Gynaecological Oncology, Royal Adelaide Hospital, Adelaide, SA 5005, Australia

ARTICLE INFO

Article history:

Received 15 June 2016

Received in revised form 30 September 2016

Accepted 16 October 2016

Available online 23 October 2016

Keywords:

Endometrial cancer

Metastasis

Annexin A2

α actinin 4

MALDI MSI

Label free LC-MS/MS

ABSTRACT

The prediction of lymph node metastasis using clinic-pathological data and molecular information from endometrial cancers lacks accuracy and is therefore currently not routinely used in patient management. Consequently, although only a small percentage of patients with endometrial cancers suffer from metastasis, the majority undergo radical surgery including removal of pelvic lymph nodes. Upon analysis of publically available data and published research, we compiled a list of 60 proteins having the potential to display differential abundance between primary endometrial cancers with versus those without lymph node metastasis. Using data dependent acquisition LC-ESI-MS/MS we were able to detect 23 of these proteins in endometrial cancers, and using data independent LC-ESI-MS/MS the differential abundance of five of those proteins was observed. The localization of the differentially expressed proteins, was visualized using peptide MALDI MSI in whole tissue sections as well as tissue microarrays of 43 patients. The proteins identified were further validated by immunohistochemistry. Our data indicate that annexin A2 protein level is upregulated, whereas annexin A1 and α actinin 4 expression are downregulated in tumours with lymph node metastasis compared to those without lymphatic spread. Moreover, our analysis confirmed the potential of these markers, to be included in a statistical model for prediction of lymph node metastasis. The predictive model using highly ranked *m/z* values identified by MALDI MSI showed significantly higher predictive accuracy than the model using immunohistochemistry data. In summary, using publicly available data and complementary proteomics approaches, we were able to improve the prediction model for lymph node metastasis in EC.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Endometrial cancer (EC) is the second most common gynaecological cancer and a significant contributor to gynaecological related mortality in the world [1]. Despite the disease being diagnosed at an early stage,

clinicians are faced with the challenge of predicting the stage of the disease so that the best surgical approach can be selected. Consequently, although only a small percentage of patients with EC suffer from metastasis the majority will undergo pelvic lymphadenectomy for diagnostic, prognostic and therapeutic reasons. This procedure is associated with

Abbreviations: EC, Endometrial cancer; LNM, Lymph Node Metastasis; FFPE, Formalin Fixed Paraffin Embedded; MALDI MSI, Matrix Assisted Laser Desorption/Ionisation Mass Spectrometry Imaging; LC-MS/MS, Liquid Chromatography Mass Spectrometry; IHC, Immunohistochemistry; ANXA2, Annexin A2; ANXA1, Annexin A1; ACTN4, α actinin 4; ERBB2, Receptor tyrosine protein kinase erb-2; EGFR, Epidermal growth factor receptor.

☆ This article is part of a Special Issue entitled: MALDI Imaging, edited by Dr. Corinna Henkel and Prof. Peter Hoffmann.

* Correspondence to: P. Hoffmann, Adelaide Proteomics Centre, University of Adelaide, Gate 8 Victoria Drive, Adelaide, SA, 5005, Australia.

** Correspondence to: M.K. Oehler, Department of Gynaecological Oncology, Royal Adelaide Hospital, North Terrace, Adelaide, SA 5000, Australia.

E-mail addresses: peter.hoffmann@adelaide.edu.au (P. Hoffmann), martin.oehler@adelaide.edu.au (M.K. Oehler).

significant morbidity including lower extremity lymphoedema [2]. Predictive tissue markers for LNM could prevent unnecessary surgery, thereby directly reducing EC morbidity. Primary pathological characteristics such as histological type, myometrial invasion, cervical stromal invasion, lympho-vascular space invasion (LVSI), tumour size, and tumour grade have been investigated for their association with LNM [3]. However, only lympho-vascular space invasion has been shown to correlate with LNM, but is unable to reliably predict metastatic lymph node spread [4].

Various models aiming to predict lymph node metastasis in EC have been presented. A recent study combined 10 individually published models for predicting LNM in EC to an independent cohort of 519 patients in order to evaluate and compare the predictive capacity of each model [5]. Of the models analysed, one preoperative model [6] and three postoperative models [7–9] performed well with AUC values of ≥ 0.75 and false negative rates of $< 5\%$. In general, the study found postoperative models based on pathological characteristics outperformed preoperative models based on clinical imaging techniques such as MRI. Advances in technology are also starting to result in the identification of biomarkers with better sensitivity and specificity. Recently, we used a MALDI MSI approach to analyse primary endometrial tumours and were able to correctly classify 88% of 43 patients [10].

Besides pathological characteristics, other parameters can be used as part of a predictive model, for example a model combining serum CA125 levels with the analysis of progesterone receptor (PR) and proliferation marker Ki67 by immunohistochemistry achieved an AUC of 0.82 [11].

Moreover, genetic markers of LNM, including p53, oestrogen and progesterone receptors, and DNA ploidy, show potential to form part of a prediction model [12]. Biomarkers identified by analysis of mRNA abundance either by microarray or RNA sequencing show great promise, but do not necessarily correlate with protein expression levels [13]. This discrepancy is relevant given many common pathology procedures such as immunohistochemistry (IHC) rely on antibodies which specifically measure the abundance of a protein within a sample. Mass spectrometry (MS) based proteomic profiling is therefore a promising alternative for biomarker discovery. Complementary MS techniques, such as liquid chromatography coupled electrospray ionisation MS/MS (LC-ESI-MS/MS) used with MALDI mass spectrometry imaging (MALDI MSI), can not only identify and quantify relative changes in protein abundance, they can also determine the spatial distribution of proteins and peptides of interest, which may correlate with disease severity.

MALDI MSI is a label free MS technique which can visualize the spatial intensity distribution of hundreds of peptides in a single tissue section whilst maintaining the sample morphology. Previously, we have shown the capacity of this technique in discriminating regions of normal endometrium from cancer [14] and moreover were able to predict the LNM status of EC patients from primary tumour cores with an accuracy of 88% [15]. Additionally, we were able to identify and validate α -Actin-2 and plectin as potential discriminators for LNM in EC [10].

For over 100 years clinical pathology has relied upon the description of tissue morphology by trained specialists for tumour classification and grading. The use of specific antibodies and immunohistochemistry (IHC) has provided targeted molecular information of the primary tumour, such as ErbB2 status for breast cancer treatment. In contrast, MALDI MSI provides non-targeted molecular information of any tissue. Previous studies in several different cancers have shown that MALDI MSI can acquire a more comprehensive proteomic picture than immunohistochemistry [16–18], and as such, MALDI MSI is slowly being introduced into commercial pathology laboratories around the world. The images generated by the technique allow new insights into tumour tissues as MALDI MSI allows for the detection and characterisation of tumour cells and their environment in a spatial context. The introduction of MALDI MSI equipment into pathology laboratories has the potential to extend current diagnostic and prognostic capabilities, revolutionising

clinical pathology. However, the full extent of this potential is yet to be demonstrated.

Given the multitude of studies that have identified potential biomarkers of EC metastasis, the aim of this study was to generate a predictive model of EC metastasis using previously identified biomarkers following their validation at the protein level, and compare the performance of the biomarker model to a predictive model generated from MALDI MSI data. To achieve this, a list of target proteins was compiled from a combination of literature searches and analysis of EC data extracted from The Cancer Genome Atlas (TCGA). Overall, a list of 60 potential biomarkers was compiled for comparison in EC with metastasis compared to those without metastasis. From this list, the biomarkers detectable at the protein level were characterised using the techniques of data dependent acquisition (DDA) LC-ESI-MS/MS, data independent acquisition (DIA) LC-ESI-MS, and MALDI MSI. Following this analysis 3 proteins were selected for validation by IHC: annexin A2, annexin A1, and α actinin 4. The results of this IHC analysis were used to generate a predictive model of EC metastasis. IHC was chosen as the method for validation and model generation given its routine use in clinical pathology laboratories. Separately, a predictive model of metastasis was generated from the MALDI MSI analysis of tissue microarrays containing tumours from patients with or without metastasis. When compared the MALDI MSI model showed a higher prognostic accuracy than the IHC model.

2. Material and Methods

2.1. Sample collection and Tissue specimens

Fifty-three tissue EC samples were retrieved from the archives of the Institute of Medical and Veterinary Science, Adelaide, South Australia, Royal Prince Alfred Hospital, Sydney, New South Wales, John Hunter Hospital, Newcastle, New South Wales and King Edward Memorial Hospital, Perth, Western Australia. All specimens were formalin fixed and paraffin embedded (FFPE), and tumours were classified according to the World Health Organisation, and staged according to the International Federation of Gynaecology and Obstetrics (FIGO). The study was approved by the ethics committees of the respective hospitals. None of the patients had received pre-operative chemotherapy and/or hormone therapy. The detailed histo-morphological and clinical characteristics were described in our previous study [10]. Haematoxylin and Eosin (H&E) stained tissue sections were annotated by a pathologist. EC patient samples in which tumour metastasized to pelvic lymph nodes were categorized as “with LNM” and the primary tumours which did not metastasize to lymph nodes were categorized as “without LNM”. Of the 53 patient samples used in the study, 43 were used in the construction and analysis of the Tissue Microarrays (TMAs), and 10 were used for LC-ESI-MS/MS and in whole tissue MALDI MSI analysis.

2.2. Tissue microarray design and construction

Forty-three EC samples were randomly divided into two equal groups for assembly of tissue microarrays (TMAs); a training set used to identify molecular discriminators (referred to as TMA1), and a test set for validation (referred to as TMA2). Of the 43 samples, 16 were positive, with the remaining 27 negative for LNM, respectively.

For TMA assembly, 1.5 mm tumour tissue cores were excised and arranged vertically in an empty paraffin block using a tissue arrayer (3DHitech TMA Master, SciTech, Victoria, Australia). Two tissue cores from tumour dense areas were taken from each patient. The layouts of both TMAs were randomized to avoid bias caused by core location within the TMAs (data not shown). Six μ m thick tissue sections were cut from both TMA's and mounted onto indium tin oxide (ITO) (Bruker Daltonics, Bremen, Germany) conductive glass slides for MALDI MSI analysis. Sections were also placed on plain glass slides

and polyethylene naphthalate (PEN) membrane slides, and H&E stained for pathology annotation and laser capture microdissection (LMD).

2.3. TCGA analysis

RNA sequencing data was extracted from The Cancer Genome Atlas (TCGA) EC database (<http://cancergenome.nih.gov/>) and analysed using edgeR [19]. Out of 514 EC patients, RNA sequencing data was available for 333 patients ($n = 54$ with LNM and $n = 279$ without LNM). Using the exact test proposed by Robinson *et al.* [20] for a mean difference between negative-binomial distributions, 627 of the 20,532 genes measured had P -values $< 2.5 \times 10^{-6}$ (0.05 with Bonferroni correction) for a mean difference between the RNA sequencing counts of with LNM and without LNM groups. Many of these significant P -values were caused by single outlying values and some had extremely small total counts, so a series of heuristics and bootstrap samples [21] were used to acquire a more stringent list of genes. Specifically, 100 bootstrap samples were generated and both the exact test P -value and log fold change were calculated for each bootstrap sample. Nineteen genes had log fold change in the same direction in all 100 bootstrap samples with P -values $< 2.5 \times 10^{-6}$ in at least 75 bootstrap samples, as well as having an original P -value $< 2.5 \times 10^{-10}$, and a median count of at least 10 (to filter out extremely low count measurements). Of these 19, 11 differentially expressed genes were selected to be included in the target list: six up regulated genes: C4BPB|725, OVGP1|5016, RNF183|138,065, C16orf89|146,556, SLC25A35|399,512, MUC6|4588 and 5 down regulated genes ALPK2|115,701, FBXO17|115,290, DLX4|1748, ADRA1B|147, and DNER|92,737. Notably, similar expression for RNF183|138,065, MUC6|4588 and DNER|92,737 was identified in recently published work [22]. Upon further literature research, we compiled a list of 60 target proteins with the potential to discriminate primary EC with or without metastasis (Supplementary table 1).

2.4. Peptide identifications by DDA nano-LC-ESI-MS/MS

The FFPE tissue blocks ($n = 5$ with LNM and $n = 5$ without LNM) for DDA nano-LC-ESI-MS/MS, not included in the TMAs were sectioned at 6 μm thickness, water bath mounted onto PEN membrane slides, deparaffinised and digested as previously described [10]. Briefly, the tissue sections were deparaffinised by heating at 60 °C for 1 h, stripped of paraffin wax using xylene, and rinsed in 100% ethanol, followed by haematoxylin staining. The regions of interest were excised by laser capture micro-dissection (LMD) (Leica Microsystems, Wetzlar, Germany) into 20 μL of 10 mM citric acid buffer (pH = 6) and incubated for 45 min at 98 °C. Samples were buffered with 10 mM of NH_4HCO_3 and digested as previously described [96] with 100 ng of trypsin gold (Promega, Madison, WI) for 2 h at 37 °C. Nano-LC-ESI-MS/MS was performed using an Ultimate 3000 RSLC system (Thermo-Fisher Scientific) coupled to an Impact II™ QTOF mass spectrometer (Bruker Daltonics) via an Advance CaptiveSpray source (Bruker Daltonics). Peptide samples were pre-concentrated onto a C18 trapping column (Acclaim PepMap100 C18 75 $\mu\text{m} \times 20$ mm, Thermo-Fisher Scientific) at a flow rate of 5 $\mu\text{L}/\text{min}$ in 2% ACN 0.1% TFA for 10 min. Peptide separation was performed using a 75 μm ID C18 column (Acclaim PepMap100 C18 75 $\mu\text{m} \times 50$ cm, Thermo-Fisher Scientific) at a flow rate of 0.2 $\mu\text{L}/\text{min}$ using a linear gradient from 5 to 45% B (A: 5% ACN 0.1% FA, B: 80% ACN 0.1% FA) over 130 min, followed by a 20 min wash with 90% B, and a 20 min equilibration with 5% A. MS scans were acquired in the mass range of 300 to 2200 m/z in a data-dependent fashion using Bruker's Shotgun Instant Expertise™ method. This method uses IDAS (intensity dependent acquisition speed) to adapt the speed of acquisition depending on the intensity of precursor ions (fixed cycle time), and RT² (RealTime Re-Think) to exclude previously selected precursor ions from undergoing re-fragmentation unless the chromatographic peak intensity of the ion has increased by a factor of 5. Singly charged precursor ions were excluded from

acquisition. Collision energy ranged from 23% to 65% as determined by the m/z of the precursor ion.

Acquired spectra were subjected to peak detection, de-convolution, and re-calibration according to a lock mass using Compass DataAnalysis for OTOF (Version 1.7, Bruker Daltonics). Processed spectra were then exported to Mascot generic format and submitted to Mascot (Version 2.3.02) for identification. Search parameters were as follows; SwissProt *Homo sapiens* database search, trypsin digestion with up to 2 missed cleavages, variable modification of oxidation of methionine, MS mass tolerance of 40 ppm, MS/MS mass tolerance of 0.2 Da, ion score cut off of 20, peptide expected score cut off ≤ 0.05 . In order to avoid false identifications only proteins with a minimum of 2 unique peptides were selected for further analysis.

2.5. Relative Quantification by DIA nano-LC-ESI-MS/MS

The FFPE tissue blocks ($n = 5$ with LNM and $n = 5$ without LNM) for DIA nano-LC-ESI-MS were prepared as described above. Nano-LC was performed as previously described using an Ultimate 3000 RSLC system coupled to an Impact HD™ QTOF mass spectrometer set to acquire data using Bruker's Middle Band CID™ method. This data independent acquisition (DIA) method scans a mass range of m/z 375 to 1206 in 26 Da increments and CID is performed with increasing collision energies of 20 to 36. The acquired data was analysed using Skyline (Version 3.1.0.7382) [97] against a spectral library generated from the search results of the data dependent acquisition experiment. Relative quantification analysis of the differentially expressed proteins was carried out using the spectral library peptides listed in Table 1 and Supplementary Table 2. The Skyline peptide and transition settings were as follows: Trypsin was specified as the cleavage enzyme with a maximum of 1 missed cleavage, precursor charge states 2 and 3, ion charges 1 and 2, ion types y and b from ion 2 to ion 6, an ion match tolerance 0.1 m/z , a MS/MS filtering DIA isolation scheme from m/z 375 to 1206 (26 Da windows), retention time window of 6 min, and a resolution of 10,000. For each protein, the area intensities for the analysed peptides was summed for each sample. The summed area intensities for each protein was then compared between the primary EC with or without LNM. For all results the standard error of the mean and P -values using un-paired t -tests were calculated using GraphPad Prism 6 v008.

2.6. Sample preparation for FFPE MALDI MSI

FFPE tissue sections for MALDI MSI, whole tissue sections ($n = 5$ with and $n = 5$ without LNM) as well as tissue microarrays containing a total of 43 patients, were prepared using the same protocol as previously described [98]. Briefly, the tissue sections were deparaffinised by heating at 60 °C for 1 h, stripped of paraffin wax using xylene, and rinsed in 100% ethanol. The tissue sections were subjected to heat induced citric acid antigen retrieval (HAIR) (10 mM citric acid, pH = 6) followed by digestion with trypsin gold (Promega, Madison, WI) using an ImagePrep station (Bruker Daltonics, Bremen, Germany) followed by incubation in a humidity chamber at 37 °C for 2 h. Internal calibrants [Angiotensin I (0.4 pmol/ μL), [Glu¹]-Fibrinopeptide B (0.4 pmol/ μL), Dynorphin A (1.8 pmol/ μL) and ACTH clip_{1–24} (1.6 pmol/ μL)] and α -cyano-4-hydroxycinnamic acid (CHCA) (7 mg/mL in 50% ACN/ 0.2% TFA) matrix was overlaid onto the tissue using an ImagePrep station [99]. Data was acquired using an ultrafleXtreme MALDI-TOF/TOF MS (Bruker Daltonics) with flexControl v3.0.1 and flexImaging v4.0.1 software in positive reflectron mode with detection range of m/z 800–4000 Da. MS spectra were acquired in a grid with a centre to centre resolution of 60 μm for TMAs and 100 μm for whole tissue sections. Two sections of each TMA were analysed.

Following data acquisition, the ITO slides were incubated in 70% ethanol to remove the matrix, H&E stained, and scanned using Nanozoomer (Hamamatsu, Shimadzu). To align the MS data with

Table 1

Relative abundance of the differentially expressed proteins.

Protein	Accession no	Tryptic peptides observed	Peptide mass (Da)	P-value	Expression observed in the current study (Fold change)	Expression observed as a metastatic protein	Reference
ANXA2_human	P07355	LSLEGDHSTPPSAYGSVK	1844.996	0.0002	2.0 Up	Up	[34–36]
		GVDEVTIVNLTNR	1542.753				
		DIISDTSGDFR	1225.277				
ERBB2_human	P04626	LPASPETHLDMLR	1479.714	0.0003	2.7 Down	Down	[55]
		VLGSGAFGTVYK	1198.384				
		EIPDLLEK	956.104				
RBM5_human	P52756	QQLIPELVR	1095.307	0.0004	4.1 Down	Down	[37]
YBOX_human	P67809	SVGDGETVEFDVVEGEK	1795.875	0.0008	2.2 Down	Independently associated	[22,94,95]
EGFR_human	P00533	NYDLNFLK	999.131	0.0025	1.8 Down	Independently associated	[22,49–51]
		NLQEILHGAVR	1249.435				
ACTN4_human	O43707	ALDFIASK	864.009	0.0074	2.9 Down	Up	[25–27]
		DGLAFNALIHR	1226.401				
		GYEEWLLNEIR	1421.572				
		DYETATLSDIK	1255.344				
		QFASQANVVGWPVIQTK	1774.009				
		VGWEQLLTIIAR	1386.614				
		TINEVENQILTR	1429.593				
		GISQEQMQEFR	1352.485				
		NFTAEELR	1092.217				
HPT_human	P00738	VTSIQDWVQK	1203.360	0.0625	1.7 Up	Up	[70–73]
ANXA1_human	P04083	GGPGSAVSPYPTFNPSDVAALHK	2356.577	0.0928	2.2 Down	Down	[30–33]
		GVDEATHIDILTK	1387.593				
		TPAQFDADLR	1262.342				
		DITSDTSGDFR	1213.223				
		NNAQRQQIK	1099.594				
		GTDVNVFNTILTR	1550.732				
ENPL_human	P14625	EEEAIQLDGLNASQIR		0.1579	1.7 Down	Up	[52–54]

the histological features the H&E scanned images were co-registered to the MALDI MSI results via the flexImaging software.

2.7. Data analysis using SCiLS lab

To determine the spatial expression profiles of the differentially expressed proteins, the raw data was loaded into SCiLS lab software (SCiLS GmbH, Bremen, Germany). Here, the raw data was pre-processed using the pre-processing pipeline including TopHat baseline removal, normalization to the total ion count (TIC), and peak alignment and peak picking was performed. Tissue specific regions of interest (ROI) (i.e. normal epithelium, region of primary EC with and without LNM) were created based on annotations completed by a pathologist. *m/z* values [M-H] of the differentially expressed proteins, as determined by the DIA nano-LC-ESI-MS analysis, were targeted for analysis using the SCiLS lab software. The spatial distribution of the peptides were visualised in the form of ion intensity maps. In all data sets, interval width was set to ± 0.125 Da.

2.8. Immunohistochemistry (IHC)

Six μ m consecutive TMA sections were used for immunohistochemistry (IHC). Tissue sections were dewaxed and rehydrated with xylene and ethanol. The endogenous peroxidase activity was quenched with 1% hydrogen peroxidase in PBS, followed by antigen retrieval using 10 mM citric acid (pH 6), at 100 °C for 10 min in a microwave (Sixth Sense, Whirlpool). To avoid nonspecific binding, tissue sections were blocked in 5% goat serum (Sigma Aldrich) for 30 min and incubated overnight at 4 °C with mouse monoclonal antibody annexin A2 (diluted 1:500, BD Biosciences), annexin A1 (diluted 1:5000, BD Biosciences) and α actinin 4 (diluted 1:1000, Sigma Aldrich) in blocking buffer (5% Goat serum). The tissues sections were subsequently incubated at room temperature for 1 h with biotinylated anti-mouse immunoglobulins (1/400, Dako, Australia) for annexin A2 and annexin A1 and anti-rabbit immunoglobulins for α actinin 4, followed by streptavidin horseradish peroxidase (1/500, Dako, Australia). Before counterstaining with haematoxylin, immunoreactivity was visualised using diaminobenzidine (DAB)/H₂O₂ substrate (Sigma Aldrich). The

slides were scanned using Nanozoomer. Appropriate positive and negative controls were included in the study. Quantitative analysis of protein expression was performed using the digital image analysis software 'ImageJ', coupled with a colour de-convolution plugin 'IHC profiler'. For each tissue section three representative photomicrographic images at 40 $\times$ magnification were obtained from the primary tumours with or without LNM. The colour de-convolution leads to the formation of three images resulting in the separation of DAB colour from haematoxylin in an automated manner [100]. From the IHC analysis of the TMAs, data from a total of 40 patients was obtained; initially tissue cores from 43 patients were present on the slides, however 3 cores were lost during sample processing.

2.9. Data analysis of MALDI-Imaging TMA data

Acquired spectra were recalibrated using the internal calibrants and pre-processed using flexAnalysis and flexImaging software (V4.0.1 Bruker Daltonics). During the pre-processing the Snap Algorithm with a signal to noise ratio of 2 was used for peak detection after Gaussian smoothing and a baseline subtraction using the TopHat method. Resulting peaks were grouped by *m/z* using the density based clustering algorithm DBSCAN* as described in Section 3 of [101], with a grouping tolerance $\varepsilon = 0.025$ and a minimum number of points of 150, generating 1254 groups. These peak groups then formed the variables in the following analysis, with patients representing the observations. Canonical Correlation Analysis (CCA) was used to rank the variables (peak groups) according to their contribution towards the highest multivariate correlation with LNM, as described in [15].

2.10. Logistic Regression Models

A number of logistic regression models for LNM were considered in three main contexts using clinical information (age, tumour grade, tumour size, percent invasion, and lympho-vascular space invasion (LVSI)), IHC results, and MALDI MSI results. In the first two contexts, model fit was approached by first fitting the largest possible additive model and then iteratively removing the least significant independent

variables from the model until all remaining independent variables had coefficients that were significantly non-zero at the 5% level as determined by individual *t*-tests. In the third context, MALDI MSI results, the full additive model could not be fit as it was highly overparameterised by the number of peakgroups in the data. As such, iterative addition of independent variables to the model, starting from the smallest model, was considered. In order to measure the predictive capabilities of these models, two-fold cross-validation was used with a stratified re-sampling scheme – patients with and without LNM were each randomly partitioned into two groups, and two subsets of the data thereby produced with equal ratio of with to without LNM patients. The models were then fit to one subset of the data and used to predict the LNM status of the other subset. This process was then repeated in the reverse direction – fitting the model to the second subset and predicting for the first. The random partitioning was then repeated 50 times and the results combined to produce a single smooth ROC curve for each model. The two directions for each of the 50 resamples resulted in 100 individual versions of each classification model being trained and tested. ROC curves for each of these individual 100 variations are also shown in order to provide a measure of the variability in the resampling step.

3. Results

3.1. DIA MS

Traditional shotgun proteomics (also known as data dependent acquisition) has been the standard approach in research to profile protein content in biological samples with the aim to identify protein candidates with potential diagnostic, prognostic, and/or therapeutic value. In practice, this approach requires significant resources and time and has limited reproducibility due to the stochastic nature of the acquisition protocol [102]. In contrast, targeted proteomic approaches such as data independent acquisition might offer higher sensitivity but require a predefined list of potential target proteins to analyse. To assemble a list of potential markers for EC metastasis, we combined potential biomarker candidates identified by literature review with 11 additional genes found to be differentially expressed in the TCGA data as described in Section 2.3, resulting in a list of 60 target proteins in total (Supplementary Table 1).

FASTA files for the proteins of interest were downloaded from UniProt (<http://www.uniprot.org/>) and imported into the Skyline software (Version 3.1.0.7382). Out of 60 target proteins, 23 were identified using DDA nano-LC-ESI-MS/MS analysis from ten EC tissue samples as summarized in Supplementary Table 2. The relative abundance of these 23 proteins was further analysed in EC with ($n = 5$) and without ($n = 5$) LNM by DIA nano-LC-ESI-MS/MS. Of the 23 proteins analysed, 4 had both at least 2 unique peptides identified and a significant difference in abundance between the EC tissues with or without metastasis: annexin A2 (ANXA2), receptor tyrosine protein kinase erb-2 (ERBB2), epidermal growth factor receptor (EGFR), and α actinin 4 (ACTN4) (Fig. 1). A fifth protein, annexin A1 (ANXA1), was selected from the list for further analysis despite not being significantly different between ECs with or without LNM ($P = 0.093$). This selection was made due to the well-known role of ANXA1 in the metastasis of other cancers.

3.2. MALDI MSI

The DIA analysis revealed the relative abundance of the proteins of interest in the homogenised tumour tissues. To reveal the spatial intensity distribution of the differentially expressed peptides, we performed MALDI MSI using single sections from 10 patients as well as tissue microarrays containing samples from 43 patients. Upon imaging the data of a randomly chosen EC patient sample with LNM and one randomly chosen EC patient sample without

LNM were imported together into the SCIls lab software. Fig. 2a shows the H&E stain of one representative pair of tumour sections with annotated tumour and normal tissues. Fig. 2b–d shows representative ion intensity maps for the m/z values [M–H] with the highest capacity to discriminate between the tumours with and without LNM, that match back to peptides from the proteins of interest as identified by nano-LC-ESI-MS/MS. All identified peptides from the proteins of interest were analysed in single sections and TMAs using the SCIls software, and the majority of peptides detected in the MSI data for each of the proteins of interest correlated in terms of trend abundance trend differences between the with and without LNM groups (Supplementary Figs. 1–2). The intensity of the ANXA2 peptide (m/z 1542.83 Da) appears to be higher in tumours with LNM as compared to those without (Fig. 2b). In contrast, the intensity of ACTN4 peptide (m/z 1429.76 Da) and ANXA1 peptide (m/z 1099.29 Da) appears to be lower in with LNM compared to those without LNM (Fig. 2c–d). The MALDI MSI data correlated well with the DIA data and moreover allowed for a visual inspection of the distribution of the peptides of interest within the single sections, containing normal tissue and tumour. The three ROC curves display a reasonable capacity to discriminate patients with and without LNM (Fig. 2b–d).

3.3. Validation by immunohistochemistry

The expression of EGFR and Erb2 in EC was described before [103] and we therefore did not include those proteins in our validation approach. The differential abundance of ANXA2, ACTN4 and ANXA1 within the EC patient cohort was verified by the IHC analysis of the TMAs. In total, tumour cores from 40 EC patients were analysed as some tissue cores were lost during tissue processing. Quantitative analysis was performed using ImageJ software coupled with IHC profiler [100]. Three representative images from each tissue core were captured, and the staining was assessed by calculating the percentage of the high positive, positive, low positive and negative staining. The staining intensities were summed for both groups and statistical analysis was performed using an unpaired *T*-Test in Graphpad prism. An increase in immunostaining intensity of ANXA2 ($P = 0.0875$; Fig. 3a) and significant decrease in immunostaining intensity of ACTN4 ($P = 0.0019$; Fig. 3b) and ANXA1 ($P = 0.0026$; Fig. 3c) were observed in tumours with LNM versus those without LNM. These results correlated strongly with the results obtained from the DIA and MALDI MSI.

3.4. Logistic Regression Models

The clinical parameters of age, tumour grade, tumour size, percent invasion, and LVSI were considered in building logistic regression model for LNM in the EC patients where data for all of the above clinic-pathological variables was available ($n = 35$). Fitting a logistic regression model for LNM based on all the clinic-pathological variables and iteratively removing the least significant of them resulted in only one clinical variable remaining: LVSI, and this is in accordance with recently published results [4]. Fig. 4a shows ROC curves produced by the two-fold cross validation and stratified resampling described in Section 2.10 for each of the models involved in this iterative reduction process. Given the apparent importance of LVSI in the prediction of LNM status, we include it as a covariate in all the other models we considered.

The IHC data for ANXA1, ANXA2 and ACTN4 were then considered. The percentage of the staining scores, low positive, positive, and high positive were combined for a total positive percentage for each protein, and the same iterative reduction scheme for model selection was used, beginning with the largest additive model including LVSI as well as all three IHC variables, with the resulting parsimonious model being that with only LVSI and ACTN4. For considering the MALDI MSI a similar iterative model selection scheme was used, however instead of starting

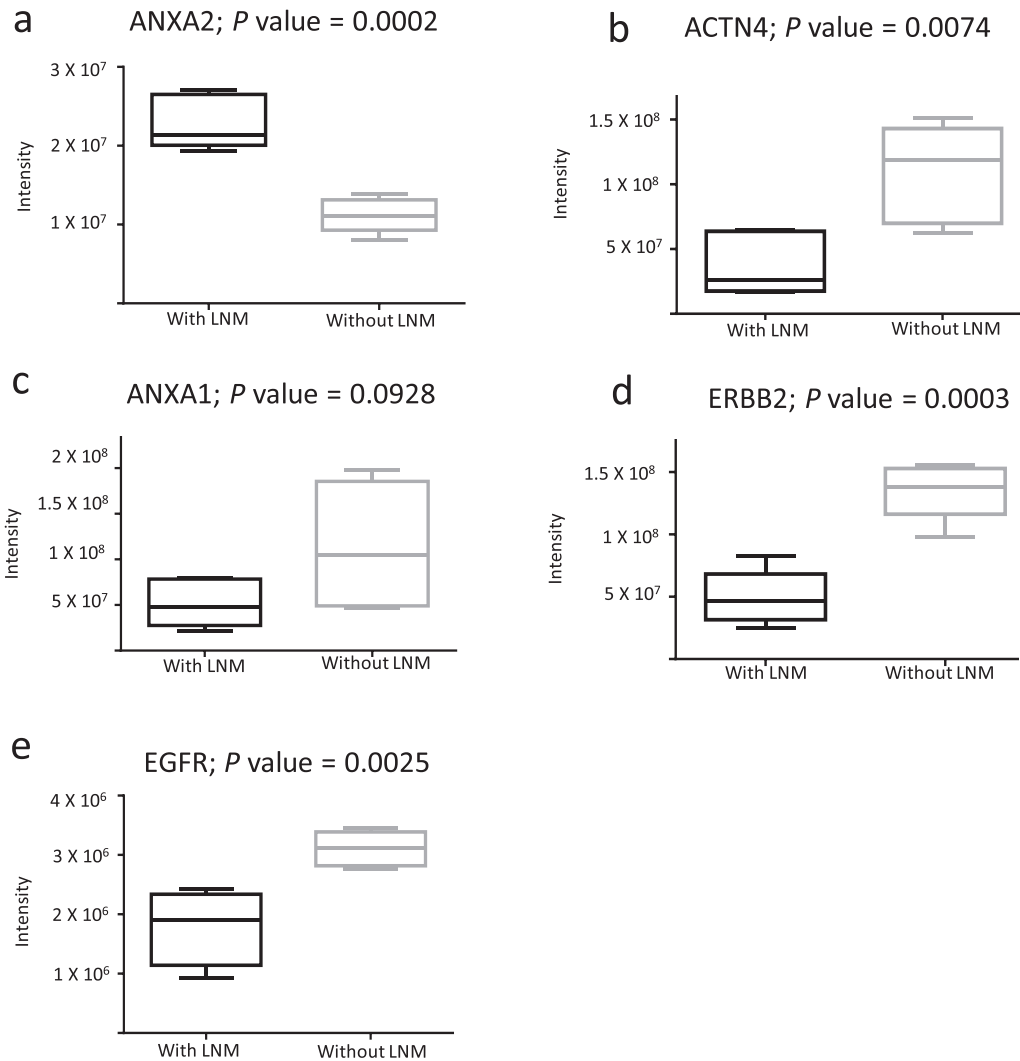


Fig. 1. DIA analysis identifies potential discriminators of lymph node metastasis in endometrial cancer: DIA analysis was performed on sections from tumours with ($n = 5$) or without ($n = 5$) lymph node metastasis. **(a)** The relative abundance of annexin A2 was compared between with and without LNM, based on 3 tryptic peptides and their specific fragments. **(b)** The relative abundance of α actinin 4 was compared between with and without LNM, based on 9 tryptic peptides and their specific fragments. **(c)** The relative abundance of annexin A1 was compared between with and without LNM, based on 6 tryptic peptides and their specific fragments. **(d)** The relative abundance of receptor tyrosine protein kinase erb-2 was compared between with and without LNM, based on 3 tryptic peptides and their specific fragments. **(e)** The relative abundance of epidermal growth factor receptor was compared between with and without LNM, based on 2 tryptic peptides and their specific fragments. The data shown as the means $\pm$ S.D.

with the full additive model (which is impossible due to over parameterisation), the data was iterated in the opposite direction – beginning with the model that includes only LVSI and iteratively adding the most significant of the top 10 MALDI MSI variables as ranked by CCA. This resulted in a model including LVSI and three of the top 10 MALDI MSI variables – the 3rd, 5th, and 10th highest ranked variables, specifically. The CCA ranking of the MALDI MSI variables is described in Section 2.9 [15] and the top 10 MSI variables (peak groups) are shown in Table 2. Fig. 4b shows ROC curves produced by the two-fold cross validation and stratified resampling described in Section 2.10 for each of the resulting models: LVSI and ACTN4, and LVSI with added MALDI MSI variables, as well as the model with LVSI alone for a comparative reference.

4. Discussion

The identification of sensitive and specific tumour biomarkers for prediction of LNM would improve treatment related morbidity of EC treatment. A significant body of work has been completed in this area, hence the first aim of this study was to compile a list of previously

proposed biomarkers, derived from both genomics and proteomics data, in order to validate the differential abundance of those biomarkers at the protein level in a single cohort of metastatic positive and negative primary EC samples. The second aim of the study was to generate a predictive model of metastasis using IHC of suggested biomarkers for comparison to a predictive model generated from MALDI MSI data.

Following a review of the literature and publicly available data a list of 60 markers was compiled, of which 23 were detected by DDA LC-MS/MS. A large number of the biomarkers assessed in this study originated from genomics data and it is well documented that mRNA and protein abundance levels do not always correlate [13]. This is a possible explanation as to why not all of the proposed markers were detectable at the protein level, and of the ones that were, why some were not found to be differentially expressed between LNM positive and negative cohorts. Despite the large number of potential markers originating from genomics data, in this study we chose to look at protein expression levels and verify our findings using an antibody based technique given the results could be directly translated into clinical practice via IHC, a standardised technique employed in pathology laboratories globally.

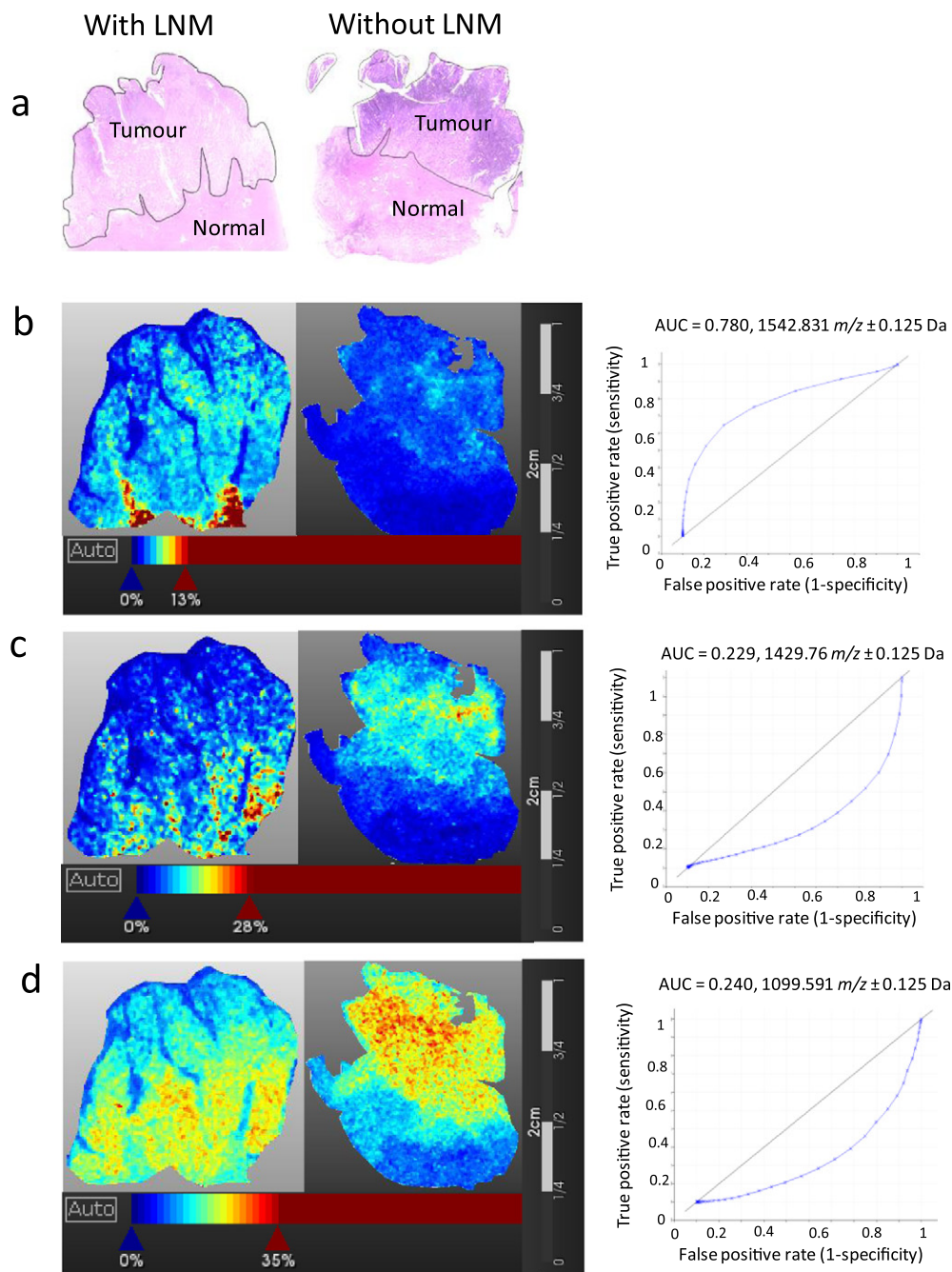


Fig. 2. Representative spatial ion intensity maps of potential discriminators: Tryptic peptide MALDI-MSI was performed on EC samples of patients diagnosed with ($n = 5$) or without ($n = 5$) lymph node metastasis. MALDI-MSI data acquired was used to generate the ion intensity maps using SCiLS lab software. Randomly one patient from each group was chosen to show the spatial intensity of the significant m/z values. **(a)** The annotated H&E stained sections are shown. The tumour region within the section was annotated by a pathologist (in black colour). **(b)** Representative MALDI-MSI images for $1542.831\ m/z \pm 0.125\ Da$, tryptic peptide belongs to annexin A2. **(c)** Representative MALDI-MSI images for $1429.76\ m/z \pm 0.125\ Da$, tryptic peptide belongs to α actinin 4. **(d)** Representative MALDI-MSI images for $1099.591\ m/z \pm 0.125\ Da$, tryptic peptide belongs to annexin A1. Automatic hotspot removal and weak denoising was applied. Scale bars are 2 cm and intensity scale for each ion are included, intensity ranges from blue (lowest) to red highest. Corresponding ROC curves indicating the sensitivity and selectivity are included on the right (**b-d**).

Of the 23 markers that were detectable at the protein level by mass spectrometry, 4 were both identified by at least two unique peptides, and found to be differentially abundant in the primary EC with and without LNM: ANXA2, ACTN4, ErBb2, and EGFR. All of these proteins have been previously implicated in metastasis. A fifth protein, ANXA1, was also chosen for further analysis despite not being significantly different between the sample cohorts, given the reported role ANXA1 plays in the metastatic process.

ANXA2, ACTN4 and ANXA1 were selected for further analysis by IHC of the TMAs, which revealed ACTN4 and ANXA1 to be significantly differentially abundant between ECs with or without LNM. When this IHC data was used to generate a logistic regression model for LNM in conjunction with LVSI, the iterative model selection scheme described in [Section 2.10](#) removed ANXA1 and ANXA2 from the model, and resulting in the model with only ACTN4 and LVSI. A similar iterative model selection scheme based on the top 10 MALDI MSI variables as

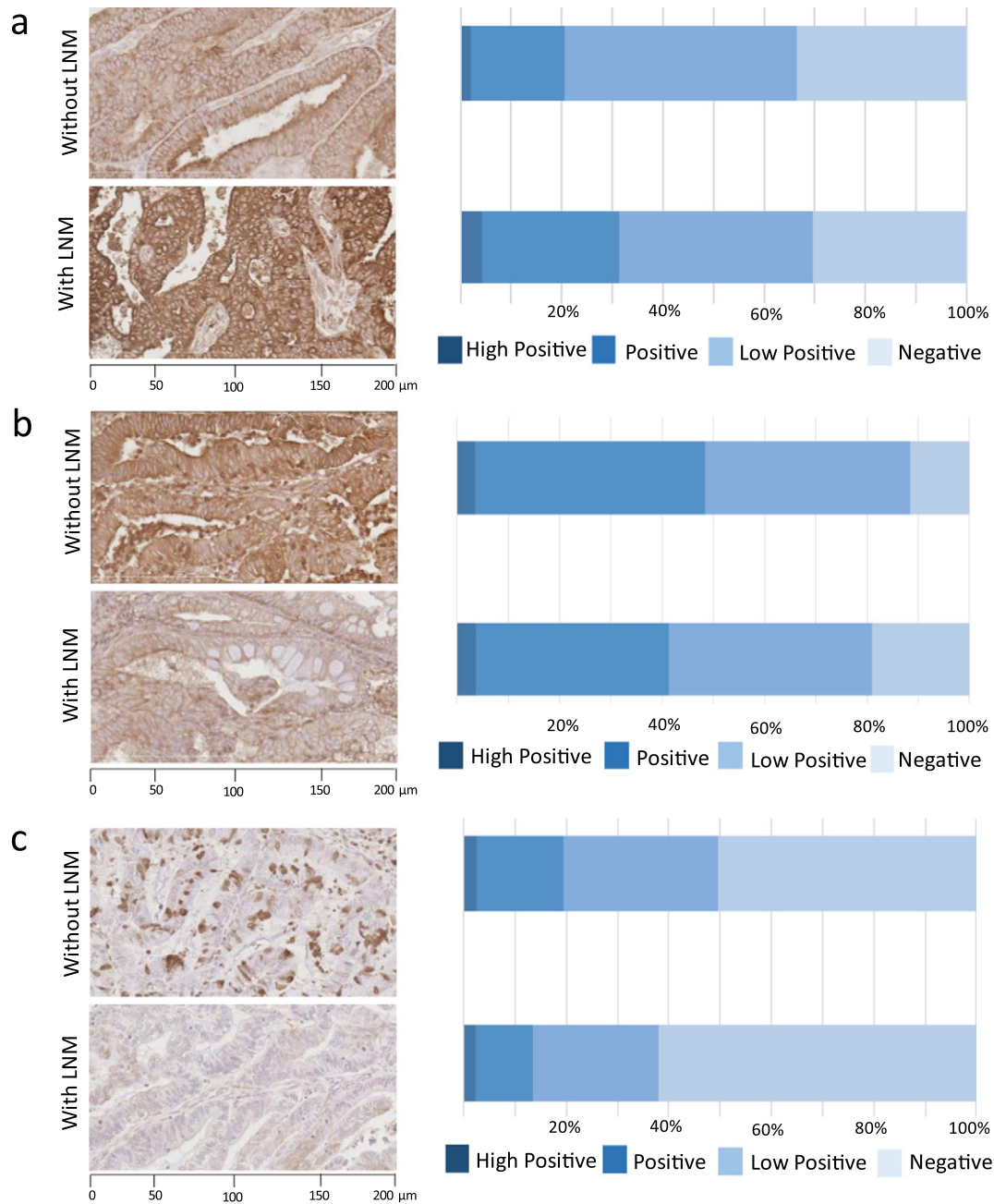


Fig. 3. Immunohistochemistry confirmed proteins associated with lymph node metastasis in endometrial cancer: 6 μ m serial sections of TMA1 were used for immunohistochemistry. Quantitative analysis was performed using IHC profiler-Image J. for each tissue core, three representative photo-micrographic images at 40 $\times$ magnification were used and each image was assigned a score of high positive, positive, low positive and negative staining. Shown is the average staining intensity distribution of all analysed images. **(a)** Representative image of immunostaining of annexin A2 on tumour with LNM and without LNM is shown. $P = 0.0875$ **(b)** Representative image of immunostaining of α actinin 4 on tumour with LNM and without LNM is shown. $P = 0.0019$ **(c)** Representative image of immunostaining of annexin A1 on tumour with LNM and without LNM is shown. $P = 0.0026$.

ranked by the CCA analysis described in Section 2.9 [15] resulted in a model with 3 MALDI MSI variables and LVSI being selected. Both the IHC (ACTN4) and MALDI MSI variables, when combined with LVSI, outperformed LVSI alone. However, the MALDI MSI model outperformed the ACTN4 IHC model.

The MALDI MSI models for the prediction of LNM as presented here is an interesting alternative to models based on more traditional proteomics approaches: here nothing is known about the identity of the top 10 variables as determined by the CCA, other than information regarding their mass range. This is in stark contrast to classical proteomics experiments where a large amount of information about each potential protein candidate must be confirmed before its use in such a prediction model. Most potential protein biomarkers are subjected

to a number of exclusion or inclusion criteria such as the number of unique peptides identified, peak intensity thresholds, identifiable modifications, mass tolerance, and/or enzyme specificity, which may result in a loss of potentially important markers. The MALDI MSI models described in the present study use data that has not been subjected to any downstream exclusion or inclusion, thereby avoiding any inherent bias introduced thereby. Furthermore we have demonstrated that these 'naïve' models can predict LNM from primary tumours, and as previously indicated [15,104], have the potential to perform better than comparable models using IHC data of validated protein candidates.

When fitting prediction models, it is important to consider measures of predictive accuracy as well as model-fit, as is demonstrated in Fig. 4a where a sequence of increasingly complex models based on clinical

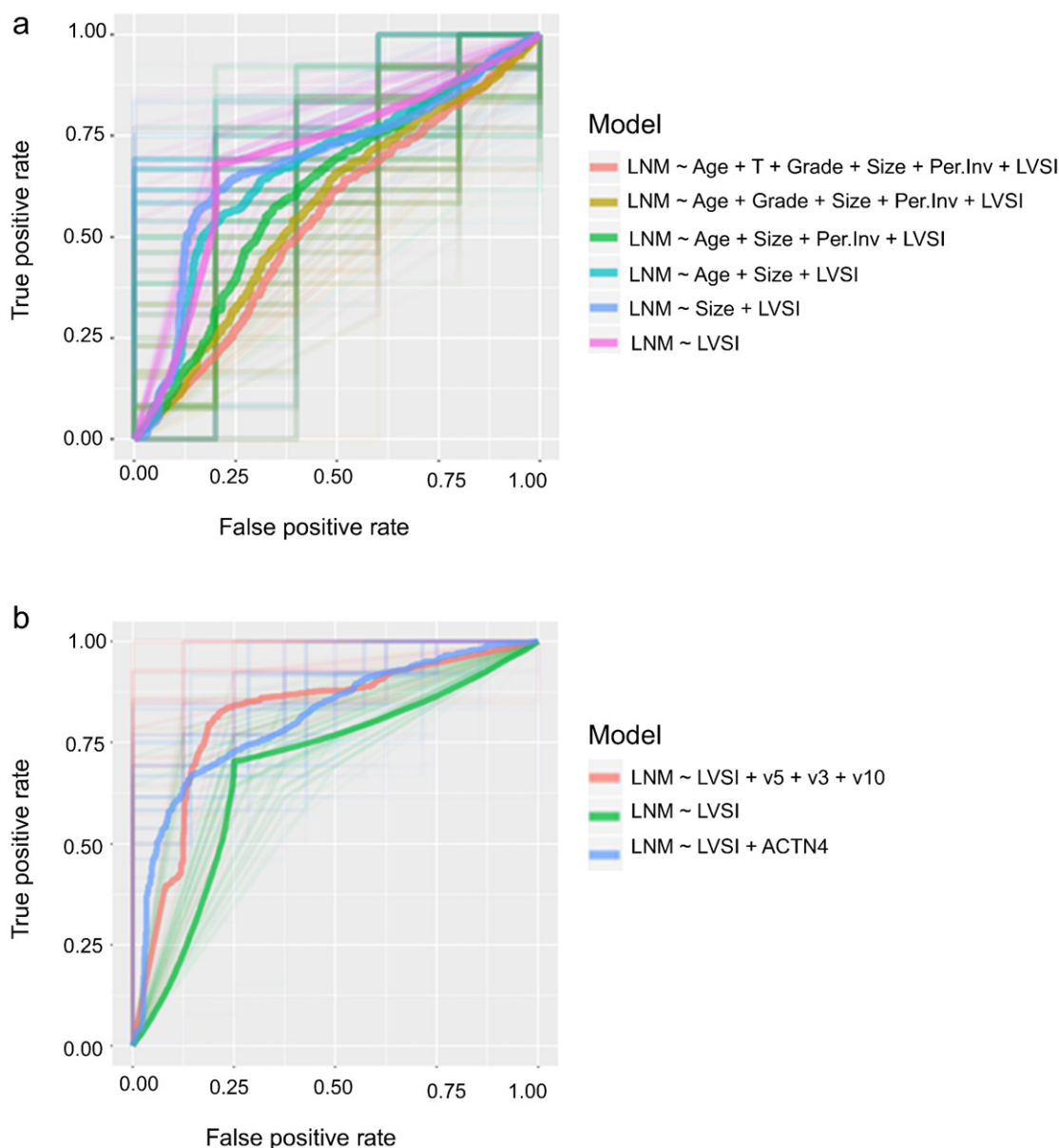


Fig. 4. Receiver operating characteristic (ROC) curves for the prediction of LNM status by several logistic regression models. The ROC curves shown are the result of the 2-fold cross validation and stratified resampling scheme described in Section 2.10. The combined ROC curves are shown as opaque lines, while the 100 ROC curves produced by the two directions of cross validation in each of the 50 partitions of the data are shown as transparent lines in order to provide a measure of the variability in the resampling. Results from a number of different models are shown: **(a)** compares the predictive performance of models using clinical variables: age, tumour grade (Grade), tumour size (Ts), percent invasion (Per. Inv.), and lympho-vascular space invasion (LVSI) – shown are the models involved in the iterative model selection scheme that resulted in the model with only LVSI being chosen, as described in Section 2.10, and **(b)** compares the model with LVSI alone, to the final model selected in each of the two contexts we consider: IHC (LVSI and ACTN4), and MALDI MSI (LVSI and the fifth, third, and tenth most highly CCA-ranked MALDI MSI variables – the CCA variable ranking is described in Section 2.9 [15]).

covariates are considered. The most complex models, including the most co-variables and naturally achieving the best model fit, also tend to perform badly in prediction due to overfitting effects. The trend seen in Fig. 4a reflects this, with the simplest model, containing only LVSI, performing best. The correlation between LVSI and metastasis has been previously shown [4], and our results support this. Whilst information about the LVSI is clearly an important predictor of metastasis the downside is that such information can only be obtained following surgical removal of the uterus which usually also includes the removal of the lymph nodes. Therefore, a preoperative model that can guide the surgeons in deciding the status of LNM in EC is warranted. The over-arching aim of this project was therefore to develop a preoperative model for the prediction of LNM EC biopsies taken from patients at diagnosis and prior to their hysterectomy and staging (lymph node removal) procedure.

Metastasis is a complex process in which the tumour cells are characterized by uncontrolled growth, invasion to surrounding tissues, and metastatic spread to distant sites. A critical step of invasion and metastasis is the detachment of cancer cells in the primary tumour, which is mainly controlled by adherent junction proteins such as α -catenin, β -catenin, vinculin, α -actinin, and actin. ACTN4 is a cytoplasmic protein belonging to the family of α -actinin proteins, which are actin binding proteins that play an important role in the formation of integrin cell adhesion which bind cells together. Enhanced expression of ACTN4 has recently been shown to increase cell motility and invasion [27]. In contrast, a recent study by Slater *et al.* [60] showed ACTN4 expression was down-regulated in EC compared to healthy tissue. Our study seems to be the first indicating low expression of ACTN4 in primary tumours with LNM compared to those without LNM. This highlights the complexity of cancer progression which may result in the

Table 2

The top 10 ranked MSI variables (peak groups).

Rank	Min. m/z^a	Max. m/z^b	Range ^c	Peaks ^d	AWM m/z^e
1	1905.908	1906.129	0.221	20,587	1906.005
2	967.402	967.636	0.234	46,719	967.5151
3	959.364	959.643	0.279	65,329	959.5133
4	1242.582	1242.754	0.172	13,503	1242.684
5	1027.449	1027.81	0.361	74,234	1027.549
6	1043.469	1043.62	0.151	5517	1043.538
7	1161.504	1161.647	0.143	38,841	1161.568
8	1157.474	1157.775	0.301	42,360	1157.635
9	1612.851	1613.064	0.213	65,684	1612.951
10	802.19	802.518	0.328	61,437	802.3933

^a The minimum m/z of the peak group.^b The maximum m/z of the peak group.^c The range in Da of the peak group.^d Number of peaks in the peak group.^e Abundance weighted mean m/z of the peak group.

same protein having significantly altered abundance during different stages of cancer development.

Annexin A1 (ANXA1) is a member of the calcium binding annexin protein family and is known suppressor of cell proliferation and mediator of apoptosis [105]. The tumour suppressor role of ANXA1 has been confirmed in many cancers including those of the prostate [106], breast [107], stomach [108,109] and oesophagus [32,110]. In this study the expression of ANXA1 was decreased in the tumours with LNM compared to those without. This finding is consistent with recent data from gastric cancer [109,111], where it was shown the expression of ANXA1 was decreased in LNM compared to primary gastric tumours.

Annexin A2 (ANXA2) is a peripheral membrane protein also belonging to the annexin family of calcium binding proteins. ANXA2 is known to mediate cellular process that influence cancer proliferation such as migration [112,113], invasion [114,115], and adhesion [116] in ovarian [117], pancreatic [118], breast [119], and prostate [120] cancers. The observed increased abundance of ANXA2 in ECs with LNM is consistent with previously published studies, implicating ANXA2 as an independent risk factor of EC [34,35].

ErB2 (also known as HER2) and EGFR are members of the epidermal growth factor receptor (EGFR) family and their overexpression has been shown to play a critical role in the pathogenesis of breast [121], prostate [122], bladder [123], ovarian [124] and gastric cancer [125]. However, neither HER2 overexpression nor HER2 amplification appeared to be associated with progression free survival or overall survival of EC patients [126]. Here, we describe low HER2 protein expression and low EGFR levels in EC with LNM compared to those without LNM; indicating an additional role for the tyrosine kinase receptors in cancer metastasis.

To the best of our knowledge this is the first time the performance of different models using pathological parameters, MALDI MSI data and/or IHC data have been directly compared. Our results show that MALDI MSI has the potential to help improve the clinical management of EC which will potentially result in better outcomes for patients.

Conflict of interest statement

The authors have declared no conflict of interest.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgements

Funding: This research was supported by NCRIS, Australia. MKH is a recipient of a Barbara Kidman Fellowship, University of Adelaide.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbapap.2016.10.010>.

References

- [1] C. de Haydu, J.D. Black, C.L. Schwab, D.P. English, A.D. Santin, An update on the current pharmacotherapy for endometrial cancer, *Expert. Opin. Pharmacother.* 17 (2016) 489–499.
- [2] Y. Todo, R. Yamamoto, S. Minobe, Y. Suzuki, U. Takeshi, M. Nakatani, Y. Aoyagi, Y. Ohba, K. Okamoto, H. Kato, Risk factors for postoperative lower-extremity lymphedema in endometrial cancer survivors who had treatment including lymphadenectomy, *Gynecol. Oncol.* 119 (2010) 60–64.
- [3] J.J. Chee, T.H. Ho, E.H. Tay, J.J. Low, K.L. Yam, Endometrioid adenocarcinoma of the uterus: surgico-pathological correlations and role of pelvic lymphadenectomy, *Ann. Acad. Med. Singap.* 32 (2003) 670–675.
- [4] R.W. Holloway, S. Gupta, N.M. Stavitski, X. Zhu, E.L. Takimoto, A. Gubbi, G.E. Bigsby, L.A. Brudie, J.E. Kendrick, S. Ahmad, Sentinel lymph node mapping with staging lymphadenectomy for patients with endometrial cancer increases the detection of metastasis, *Gynecol. Oncol.* 141 (2016) 206–210.
- [5] M. Koskas, M. Fournier, A. Vanderstraeten, F. Walker, D. Timmerman, I. Vergote, F. Amant, Evaluation of models to predict lymph node metastasis in endometrial cancer: A multicentre study, *Eur. J. Cancer* 61 (2016) 52–60.
- [6] S. Kang, W.D. Kang, H.H. Chung, D.H. Jeong, S.S. Seo, J.M. Lee, J.K. Lee, J.W. Kim, S.M. Kim, S.Y. Park, K.T. Kim, Preoperative identification of a low-risk group for lymph node metastasis in endometrial cancer: a Korean gynecologic oncology group study, *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 30 (2012) 1329–1334.
- [7] S. Bendifallah, A.S. Genin, I. Naoura, N. Chabbert Buffet, F. Clavel Chapelon, B. Haddad, D. Luton, E. Darai, R. Rouzier, M. Koskas, A nomogram for predicting lymph node metastasis of presumed stage I and II endometrial cancer, *Am. J. Obstet. Gynecol.* 207 (2012) 197.e191–197.e198.
- [8] W.T. Creasman, C.P. Morrow, B.N. Bundy, H.D. Homesley, J.E. Graham, P.B. Heller, Surgical pathologic spread patterns of endometrial cancer. A Gynecologic Oncology Group Study, *Cancer* 60 (1987) 2035–2041.
- [9] M.M. AlHilli, K.C. Podratz, S.C. Dowdy, J.N. Bakkum-Gamez, A.L. Weaver, M.E. McGree, G.L. Keeney, W.A. Cliby, A. Mariani, Risk-scoring system for the individualized prediction of lymphatic dissemination in patients with endometrioid endometrial cancer, *Gynecol. Oncol.* 131 (2013) 103–108.
- [10] P. Mittal, M. Klingler-Hoffmann, G. Arentz, L. Winderbaum, N.A. Lokman, C. Zhang, L. Anderson, J. Scurry, Y. Leung, C.J.R. Stewart, J. Carter, G. Kaur, M.K. Oehler, P. Hoffmann, Lymph node metastasis of primary endometrial cancers: Associated proteins revealed by MALDI imaging, *Proteomics* (2016) (n/a–n/a).
- [11] B. Yang, B. Shan, X. Xue, H. Wang, W. Shan, C. Ning, Q. Zhou, X. Chen, X. Luo, Predicting Lymph Node Metastasis in Endometrial Cancer Using Serum CA125 Combined with Immunohistochemical Markers PR and Ki67, and a Comparison with Other Prediction Models, *PLoS One* 11 (2016), e0155145.
- [12] H.M. Werner, H.B. Salvesen, Current status of molecular biomarkers in endometrial cancer, *Curr. Oncol. Rep.* 16 (2014) 403.
- [13] Y. Liu, A. Beyer, R. Aebbersold, On the Dependency of Cellular Protein Levels on mRNA Abundance, *Cell* 165 (2016) 535–550.
- [14] P. Mittal, M. Klingler-Hoffmann, G. Arentz, C. Zhang, G. Kaur, M.K. Oehler, P. Hoffmann, Proteomics of endometrial cancer diagnosis, treatment, and prognosis, *Proteomics Clin. Appl.* 10 (2016) 217–229.
- [15] L.J. Winderbaum, I. Koch, P. Mittal, P. Hoffmann, Classification of MALDI-MS imaging data of tissue microarrays using canonical correlation analysis based variable selection, *Proteomics* 16 (2016) 1793–1801.
- [16] J.O. Gustafsson, M.K. Oehler, A. Ruszkiewicz, S.R. McColl, P. Hoffmann, MALDI Imaging Mass Spectrometry (MALDI-IMS)-Application of Spatial Proteomics for Ovarian Cancer Classification and Diagnosis, *Int. J. Mol. Sci.* 12 (2011) 773–794.
- [17] B. Balluff, S. Rauser, S. Meding, M. Elsner, C. Schone, A. Feuchtinger, C. Schuhmacher, A. Novotny, U. Jutting, G. Maccarrone, H. Sarioglu, M. Ueffing, H. Braselmann, H. Zitzelsberger, R.M. Schmid, H. Hofler, M.P. Ebert, A. Walch, MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer, *Am. J. Pathol.* 179 (2011) 2720–2729.
- [18] G. Marko-Varga, T.E. Fehniger, M. Rezeli, B. Dome, T. Laurell, A. Vegvari, Drug localization in different lung cancer phenotypes by MALDI mass spectrometry imaging, *J. Proteome* 74 (2011) 982–992.
- [19] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics* 26 (2010) 139–140.
- [20] M.D. Robinson, G.K. Smyth, Small-sample estimation of negative binomial dispersion, with applications to SAGE data, *Biostatistics* 9 (2008) 321–332.
- [21] B. Efron, R.J. Tibshirani, An introduction to the bootstrap, CRC press, 1994.
- [22] N. Bou Zgheib, D.C. Marchion, S.H. Bush, P.L. Judson, R.M. Wenham, S.M. Apte, J.M. Lancaster, J. Gonzalez-Bosquet, Molecular determinants for lymph node metastasis in clinically early-stage endometrial cancer, *Oncol. Lett.* 11 (2016) 323–329.
- [25] S. Yamamoto, H. Tsuda, K. Honda, T. Kita, M. Takano, S. Tamai, J. Inazawa, T. Yamada, O. Matsubara, Actinin-4 expression in ovarian cancer: a novel prognostic indicator independent of clinical stage and histological type, *Mod. Pathol.* 20 (2007) 1278–1285.

- [26] K. Honda, T. Yamada, Y. Hayashida, M. Idogawa, S. Sato, F. Hasegawa, Y. Ino, M. Ono, S. Hirohashi, Actinin-4 increases cell motility and promotes lymph node metastasis of colorectal cancer, *Gastroenterology* 128 (2005) 51–62.
- [27] K. Honda, T. Yamada, R. Endo, Y. Ino, M. Gotoh, H. Tsuda, Y. Yamada, H. Chiba, S. Hirohashi, Actinin-4, a novel actin-bundling protein associated with cell motility and cancer invasion, *J. Cell Biol.* 140 (1998) 1383–1393.
- [30] L.Y. Xue, L.H. Teng, S.M. Zou, L.Q. Ren, S. Zheng, W. Luo, R. Bi, N. Lu, Expression of annexin I in different histological types of carcinomas, *Zhonghua Zhong Liu Za Zhi* 29 (2007) 444–448.
- [31] J. Da, X. Meng, P. Wang, Z. Yang, Y. Zhu, Significance on expressions of Annexin-I and its correlative gene proteins in endometrial hyperplasia, atypical hyperplasia and endometrial carcinoma, *Zhonghua Bing Li Xue Za Zhi* 30 (2001) 256–259.
- [32] R. Luthra, R.R. Singh, M.G. Luthra, Y.X. Li, C. Hannah, A.M. Romans, B.A. Barkoh, S.S. Chen, J. Ensor, D.M. Maru, R.R. Broaddus, A. Rashid, C.T. Albarracín, MicroRNA-196a targets annexin A1: a microRNA-mediated mechanism of annexin A1 downregulation in cancers, *Oncogene* 27 (2008) 6667–6678.
- [33] S.N. Voisin, O. Krakovska, A. Matta, L.V. DeSouza, A.D. Romaschin, T.J. Colgan, K.W. Siu, Identification of novel molecular targets for endometrial cancer using a drill-down LC-MS/MS approach with iTRAQ, *PLoS One* 6 (2011), e16352.
- [34] L. Deng, Y. Gao, X. Li, M. Cai, H. Wang, H. Zhuang, M. Tan, S. Liu, Y. Hao, B. Lin, Expression and clinical significance of annexin A2 and human epididymis protein 4 in endometrial carcinoma, *J. Exp. Clin. Cancer Res.* 34 (2015) 96.
- [35] L. Alonso-Alconada, M. Santacana, P. Garcia-Sanz, L. Muñelo-Romay, E. Colas, C. Mirantes, M. Monge, J. Cueva, E. Oliva, R.A. Soslow, M.A. Lopez, J. Palacios, J. Prat, J. Vallis, C. Krakstad, H. Salvesen, A. Gil-Moreno, R. Lopez-Lopez, X. Dolcet, G. Moreno-Bueno, J. Reventos, X. Matias-Guiu, M. Abal, Annexin-A2 as predictor biomarker of recurrent disease in endometrial cancer, *Int. J. Cancer* 136 (2015) 1863–1873.
- [36] S. Zhou, T. Yi, R. Liu, C. Bian, X. Qi, X. He, K. Wang, J. Li, X. Zhao, C. Huang, Y. Wei, Proteomics identification of annexin A2 as a key mediator in the metastasis and proangiogenesis of endometrial cells in human adenomyosis, *Mol. Cell. Probes* 11 (2012) (M112 017988).
- [37] S. Ramaswamy, K.N. Ross, E.S. Lander, T.R. Golub, A molecular signature of metastasis in primary solid tumors, *Nat. Genet.* 33 (2003) 49–54.
- [49] H. Niikura, H. Sasano, G. Matsunaga, K. Watanabe, K. Ito, S. Sato, A. Yajima, Prognostic value of epidermal growth factor receptor expression in endometrioid endometrial carcinoma, *Hum. Pathol.* 26 (1995) 892–896.
- [50] G. Scambia, P. Benedetti Panici, G. Ferrandina, F. Battaglia, M. Distefano, G. D'Andrea, R. De Vincenzo, F. Maneschi, F.O. Ranelletti, S. Mancuso, Significance of epidermal growth factor receptor expression in primary human endometrial cancer, *Int. J. Cancer* 56 (1994) 26–30.
- [51] L. Albitar, G. Pickett, M. Morgan, J.A. Wilken, N.J. Mahle, K.K. Leslie, EGFR isoforms and gene regulation in human endometrial cancer cells, *Mol. Cancer* 9 (2010) 166.
- [52] G.N. Chen, Y. Ma, Z.L. Yang, Expression of GRP78 and GRP94 in the liver tissues and their clinicopathological significance in children with hepatoblastoma, *Zhongguo Dang Dai Er Ke Za Zhi* 12 (2010) 634–636.
- [53] T.R. Cawthorn, J.C. Moreno, M. Dharsee, D. Tran-Thanh, S. Ackloo, P.H. Zhu, G. Sardana, J. Chen, P. Kupchak, L.M. Jacks, N.A. Miller, B.J. Youngson, V. Iakovlev, C.J. Guidos, K.A. Vallis, K.R. Evans, D. McCready, W.L. Leong, S.J. Done, Proteomic analyses reveal high expression of decorin and endoplasmin (HSP90B1) are associated with breast cancer metastasis and decreased survival, *PLoS One* 7 (2012), e30992.
- [54] X. Xiong, Y. Wang, C. Liu, Q. Lu, T. Liu, G. Chen, H. Rao, S. Luo, Heat shock protein 90beta stabilizes focal adhesion kinase and enhances cell migration and invasion in breast cancer cells, *Exp. Cell Res.* 326 (2014) 78–89.
- [55] H. Sakamoto, K. Kinoshita, Y. Nakayama, T. Takami, K. Ohtani, S. Izuka, K. Satoh, Gene expression in primary, metastatic and recurrent lesions of endometrial cancer, *Gynecol. Obstet. Investig.* 48 (1999) 215–220.
- [60] M. Slater, M. Cooper, C.R. Murphy, The cytoskeletal proteins alpha-actinin, Ezrin, and talin are De-expressed in endometriosis and endometrioid carcinoma compared with normal uterine epithelium, *Appl. Immunohistochem. Mol. Morphol.* 15 (2007) 170–174.
- [70] Y.D. Chai, L. Zhang, Y. Yang, T. Su, P. Charugundla, J. Ai, D. Messadi, D.T. Wong, S. Hu, Discovery of potential serum protein biomarkers for lymph node metastasis in oral cancer, *Head Neck* 38 (2016) 118–125.
- [71] U. Tabassum, O. Reddy, G. Mukherjee, Elevated serum haptoglobin is associated with clinical outcome in triple-negative breast cancer patients, *Asian Pac. J. Cancer Prev.* 13 (2012) 4541–4544.
- [72] L. Sun, S. Hu, L. Yu, C. Guo, L. Sun, Z. Yang, J. Qi, Y. Ran, Serum Haptoglobin as a novel molecular biomarker predicting colorectal cancer hepatic metastasis, *Int. J. Cancer* (2016).
- [73] H. Nabli, E. Tuller, K.L. Sharpe-Timms, Haptoglobin expression in endometrioid adenocarcinoma of the uterus, *Reprod. Sci.* 17 (2010) 47–55.
- [94] T. Ohga, K. Koike, M. Ono, Y. Makino, Y. Itagaki, M. Tanimoto, M. Kuwano, K. Kohno, Role of the human Y box-binding protein YB-1 in cellular sensitivity to the DNA-damaging agents cisplatin, mitomycin C, and ultraviolet light, *Cancer Res.* 56 (1996) 4224–4228.
- [95] T. Ise, G. Nagatani, T. Imamura, K. Kato, H. Takano, M. Nomoto, H. Izumi, H. Ohmori, T. Okamoto, T. Ohga, T. Uchiyumi, M. Kuwano, K. Kohno, Transcription factor Y-box binding protein 1 binds preferentially to cisplatin-modified DNA and interacts with proliferating cell nuclear antigen, *Cancer Res.* 59 (1999) 342–346.
- [96] S. Meding, K. Martin, O.J.R. Gustafsson, J.S. Eddes, S. Hack, M.K. Oehler, P. Hoffmann, Tryptic Peptide Reference Data Sets for MALDI Imaging Mass Spectrometry on Formalin-fixed Ovarian Cancer Tissues, *J. Proteome Res.* 12 (2012) 308–315.
- [97] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, *Bioinformatics* 26 (2010) 966–968.
- [98] O.J. Gustafsson, J.S. Eddes, S. Meding, S.R. McColl, M.K. Oehler, P. Hoffmann, Matrix-assisted laser desorption/ionization imaging protocol for in situ characterization of tryptic peptide identity and distribution in formalin-fixed tissue, *Rapid Commun. Mass Spectrom.* 27 (2013) 655–670.
- [99] J.O.R. Gustafsson, J.S. Eddes, S. Meding, T. Koudelka, M.K. Oehler, S.R. McColl, P. Hoffmann, Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS, *J. Proteome* 75 (2012) 5093–5105.
- [100] F. Varghese, A.B. Bukhari, R. Malhotra, A. De, IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples, *PLoS One* 9 (2014), e96801.
- [101] R.J. Campello, D. Moulavi, J. Sander, Density-based clustering, in: J. Pei, V.S. Tseng, L. Cao, H. Motoda, G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining*, Springer 2013, pp. 160–172.
- [102] G. Arentz, F. Weiland, M.K. Oehler, P. Hoffmann, State of the art of 2D DIGE, *Proteomics Clin Appl.* (2014).
- [103] H.D. Reyes, K.W. Thiel, M.J. Carlson, X. Meng, S. Yang, J.M. Stephan, K.K. Leslie, Comprehensive profiling of EGFR/HER receptors for personalized treatment of gynecologic cancers, *Mol. Diagn. Ther.* 18 (2014) 137–151.
- [104] P. Mittal, M. Klingler-Hoffmann, G. Arentz, L. Winderbaum, N.A. Lokman, C. Zhang, L. Anderson, J. Scurry, Y. Leung, C.J. Stewart, J. Carter, G. Kaur, M.K. Oehler, P. Hoffmann, Maldi imaging of primary endometrial cancers reveals proteins associated with lymph node metastasis, *Proteomics* (2016).
- [105] E. Solito, C. de Coupade, S. Canaider, N.J. Goulding, M. Perretti, Transfection of annexin 1 in monocytic cells produces a high degree of spontaneous and stimulated apoptosis associated with caspase-3 activation, *Br. J. Pharmacol.* 133 (2001) 217–228.
- [106] J.S. Kang, B.F. Calvo, S.J. Maygarden, L.S. Caskey, J.L. Mohler, D.K. Ornstein, Dysregulation of annexin I protein expression in high-grade prostatic intraepithelial neoplasia and prostate cancer, *Clin. Cancer Res.* 8 (2002) 117–123.
- [107] D. Shen, F. Nooraie, Y. Elshimali, V. Lonsberry, J. He, S. Bose, D. Chia, D. Seligson, H.R. Chang, L. Goodglick, Decreased expression of annexin A1 is correlated with breast cancer development and progression as determined by a tissue microarray analysis, *Hum. Pathol.* 37 (2006) 1583–1591.
- [108] A.F. Rossi, M.C. Duarte, A.B. Poltronieri, M.C. Valsechi, Y.C. Jorge, D.d.-S. Neto, P. Rahal, S.M. Oliani, A.E. Silva, Deregulation of annexin-A1 and galectin-1 expression in precancerous gastric lesions: intestinal metaplasia and gastric ulcer, *Mediators of inflammation* 2014 (2014) 478138.
- [109] Y. Hippo, M. Yashiro, M. Ishii, H. Taniguchi, S. Tsutsumi, K. Hirakawa, T. Kodama, H. Aburatani, Differential gene expression profiles of scirrhous gastric cancer cells with high metastatic potential to peritoneum or lymph nodes, *Cancer Res.* 61 (2001) 889–895.
- [110] N. Hu, M.J. Flaig, H. Su, J.Z. Shou, M.J. Roth, W.J. Li, C. Wang, A.M. Goldstein, G. Li, M.R. Emmert-Buck, P.R. Taylor, Comprehensive characterization of annexin I alterations in esophageal squamous cell carcinoma, *Clin. Cancer Res.* 10 (2004) 6013–6022.
- [111] G. Yu, J. Wang, Y. Chen, X. Wang, J. Pan, Q. Li, K. Xie, Tissue microarray analysis reveals strong clinical evidence for a close association between loss of annexin A1 expression and nodal metastasis in gastric cancer, *Clin. Exp. Metastasis* 25 (2008) 695–702.
- [112] W. Zhang, P. Zhao, X.-L. Xu, L. Cai, Z.-S. Song, D.-Y. Cao, K.-S. Tao, W.-P. Zhou, Z.-N. Chen, K.-F. Dou, Annexin A2 Promotes the Migration and Invasion of Human Hepatocellular Carcinoma Cells *<italic> In Vitro</italic>* by Regulating the Shedding of CD147-Harboring Microvesicles from Tumor Cells, *PLoS One* 8 (2013), e67268.
- [113] J.H. An, S.Y. Lee, J.Y. Jeon, K.G. Cho, S.U. Kim, M.A. Lee, Identification of Gliotropic Factors That Induce Human Stem Cell Migration to Malignant Tumor, *J. Proteome Res.* 8 (2009) 2873–2881.
- [114] M.R. Sharma, L. Koltowski, R.T. Ownbey, G.P. Tuszyński, M.C. Sharma, Angiogenesis-associated protein annexin II in breast cancer: Selective expression in invasive breast cancer and contribution to tumor invasion and progression, *Exp. Mol. Pathol.* 81 (2006) 146–156.
- [115] V.M. Diaz, M. Hurtado, T.M. Thomson, J. Reventós, R. Paciucci, Specific interaction of tissue-type plasminogen activator (t-PA) with annexin II on the membrane of pancreatic cancer cells activates plasminogen and promotes invasion in vitro, *Gut* 53 (2004) 993–1000.
- [116] R.J. Tressler, T.V. Updyke, T. Yeatman, G.L. Nicolson, Extracellular annexin II is associated with divalent cation-dependent tumor cell–endothelial cell adhesion of metastatic RAW117 large-cell lymphoma cells, *J. Cell. Biochem.* 53 (1993) 265–276.
- [117] N.A. Lokman, A.S. Elder, M.P. Ween, C.E. Pyragius, P. Hoffmann, M.K. Oehler, C. Ricciardelli, Annexin A2 is regulated by ovarian cancer–peritoneal cell interaction and promotes metastasis, *Oncotarget* 4 (2013) 1199–1211.
- [118] K. Foley, A.A. Rucki, Q. Xiao, D. Zhou, A. Leubner, G. Mo, J. Kleponis, A.A. Wu, R. Sharma, Q. Jiang, R.A. Anders, C.A. Iacobuzio-Donahue, K.A. Hajjar, A. Maitra, E.M. Jaffee, L. Zheng, Semaphorin 3D autocrine signaling mediates the metastatic role of annexin A2 in pancreatic cancer, *Sci. Signal.* 8 (2015) ra77.
- [119] P. Shetty, A. Bargale, B.R. Patil, R. Mohan, U.S. Dinesh, J.K. Vishwanatha, P.B. Gai, V.S. Patil, T.S. Amsavardani, Cell surface interaction of annexin A2 and galectin-3 modulates epidermal growth factor receptor signaling in Her-2 negative breast cancer cells, *Mol. Cell. Biochem.* (2015).
- [120] J.C. Dunne, D.S. Lamb, B. Delahunt, J. Murray, P. Bethwaite, P. Ferguson, J.N. Nacey, S. Sondhauss, T.W. Jordan, Proteins from formalin-fixed paraffin-embedded prostate cancer sections that predict the risk of metastatic disease, *Clin. Proteomics* 12 (2015) 24.

- [121] S. Rauser, C. Marquardt, B. Balluff, S.O. Deininger, C. Albers, E. Belau, R. Hartmer, D. Suckau, K. Specht, M.P. Ebert, M. Schmitt, M. Aubele, H. Hofler, A. Walch, Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry, *J. Proteome Res.* 9 (2010) 1854–1863.
- [122] S. Muniyan, S.J. Chen, F.F. Lin, Z. Wang, P.P. Mehta, S.K. Batra, M.F. Lin, ErbB-2 signaling plays a critical role in regulating androgen-sensitive and castration-resistant androgen receptor-positive prostate cancer cells, *Cell. Signal.* 27 (2015) 2261–2271.
- [123] J. Zhao, W. Xu, Z. Zhang, R. Song, S. Zeng, Y. Sun, C. Xu, Prognostic role of HER2 expression in bladder cancer: a systematic review and meta-analysis, *Int. Urol. Nephrol.* 47 (2015) 87–94.
- [124] K. Matsuo, T.B. Sheridan, S. Mabuchi, K. Yoshino, K. Hasegawa, K.D. Studeman, D.D. Im, N.B. Rosenshein, L.D. Roman, A.K. Sood, Estrogen receptor expression and increased risk of lymphovascular space invasion in high-grade serous ovarian carcinoma, *Gynecol. Oncol.* 133 (2014) 473–479.
- [125] V. Nagaraja, G.D. Eslick, HER2 expression in gastric and oesophageal cancer: a meta-analytic review, *J. Gastrointest. Oncol.* 6 (2015) 143–154.
- [126] G.F. Fleming, M.W. Sill, K.M. Darcy, D.S. McMeekin, J.T. Thigpen, L.M. Adler, J.S. Berek, J.A. Chapman, P.A. DiSilvestro, I.R. Horowitz, J.V. Fiorica, Phase II trial of trastuzumab in women with advanced or recurrent, HER2-positive endometrial carcinoma: A Gynecologic Oncology Group study, *Gynecol. Oncol.* 116 (2010) 15–20.

5.4 Supplementary File

Supplementary Table 1: List of target proteins with the potential to discriminate EC with or without metastasis

Protein	Accession no	Expression observed as a metastatic protein	Reference
A4D0S9_human	A4D0S9	Up	[148, 149]
ACTN4_human	O43707	Up, Down	[150-152]
ADA1B_human	P35368	Down	TCGA data mining
ALBU_human	P02768	Down	[153, 154]
ALPK2_human	Q86TB3	Down	TCGA data mining
ANXA1_human	P04083	Down	[146, 155-157]
ANXA2_human	P07355	Up	[158-160]
ARSI_human	Q5FYB1	Independently associated	[161]
B2L568_human	B2L568	Up	[98]
C4BPB_human	P20851	Up	TCGA data mining
CCNB1_human	P14635	Down	[162-164]
CCNE1_human	P24864	Down	[165, 166]
CD2A1_human	P42771	Down	[167-169]
CLD3_human	O15551	Up	[170, 171]
CNN1_human	P51911	Down	[98]
CO1A1_human	P02452	Up	[98]
CO1A2_human	P08123	Up	[98]
CP089_human	Q6UX73	Up	TCGA data mining
DESP_human	P15924	Down	[172]
DHYS_human	P49366	Up	[98]
DLX4_human	Q92988	Down	TCGA data mining
DNER_human	Q8NFT8	Down	[161],TCGA data mining
DPB1_human	P04440	Up	[98]
DUS9_human	Q99956	Independently associated	[161]
EF2K_human	O00418	Independently associated	[161]
EGFR_human	P00533	Independently associated	[161, 173-175]
ENPL_human	P14625	Up	[176-178]
ERBB2_human	P04626	Down	[179]
ESR1_human	P03372	Down	[180-182]
EZRI_human	P15311	Up	[183-185]
FBN3_human	Q75N90	Independently associated	[161]
FBX17_human	Q96EF6	Down	TCGA data mining

GBRT_human	Q9UN88	Independently associated	[161]
GRP78_human	P11021	Up	[176, 186-189]
HMGA2_human	P52926	Up	[190, 191]
HNRPM_human	P52272	Up	[192, 193]
HPT_human	P00738	Up	[194-197]
IF4E2_human	O60573	Up	[98]
IMP3_human	Q9NV31	Up	[198-203]
KS6A6_human	Q9UK32	Independently associated	[161]
L1CAM_human	P32004	Up	[204-208]
LMNB1_human	P20700	Up	[98, 209]
MSLN_human	Q13421	Up	[210, 211]
MT3_human	P25713	Down	[98]
MUC6_human	Q6W4X9	Up	[161], TCGA data mining
MYH11_human	P35749	Down	[98]
MYLK_human	Q15746	Down	[98]
NR4A1_human	P22736	Down	[98]
OVGP1_human	Q12889	Up	TCGA data mining
P53_human	P04637	Up	[180, 212-214]
PDPK1_human	O15530	Independently associated	[161]
PTEN_human	P60484	Down	[215, 216]
PTTG1_human	O95997	Up	[98]
RBM5_human	P52756	Down	[98]
RN183_human	Q96D59	Up	[161], TCGA data mining
RUNX1_human	Q01196	Down	[98, 217]
RUXF_human	P62306	Up	[98]
S2535_human	Q3KQZ1	Up	TCGA data mining
TEX19_human	Q8NA77	Independently associated	[161]
UPK2_human	O00526	Down	TCGA data mining
YBOX_human	P67809	Independently associated	[161, 218, 219]

Role of *N*-linked Glycans in Endometrial Cancer

This page intentionally left blank

6.1 Statement of Authorship

Title of Paper	Role of N-linked Glycans in Endometrial Cancer		
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style		
Publication Details	<u>Parul Mittal</u> , Matthew Briggs, Manuela Klingler-Hoffmann, Martin K. Oehler, Peter Hoffmann, Role of N-linked Glycans in Endometrial Cancer, <u>Unpublished work</u>		

Principal Author

Name of Principal Author (Candidate)	Parul Mittal		
Contribution to the Paper	Performed presented MALDI Imaging experiment, interpreted data and wrote the first draft of the manuscript		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	23/7/2017

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Matthew Briggs		
Contribution to the Paper	Helped with the presented MALDI MSI and LC-MS/MS experiment		
Signature		Date	08/08/17

Name of Co-Author	Manuela Klingler-Hoffmann		
Contribution to the Paper	Supervised development of work and manuscript evaluation		
Signature		Date	8/8/17

Name of Co-Author	Martin K. Oehler		
Contribution to the Paper	Supervised development of work and manuscript evaluation		
Signature	M.K. Oehler	Date	8/8/17

Name of Co-Author	Peter Hoffmann		
Contribution to the Paper	Supervised development of work and manuscript evaluation		
Signature		Date	9/8/17

6.2 Role of *N*-linked Glycans in Endometrial Cancer

Parul Mittal^{1,2}, Matthew Briggs^{1,2}, Manuela Klingler-Hoffmann^{1,2}, Martin K. Oehler³, Peter Hoffmann^{1,2,4}

1. Adelaide Proteomics Centre, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, 5005
2. Institute for Photonics & Advanced Sensing (IPAS), The University of Adelaide, Adelaide, South Australia, 5005
3. Department of Gynaecological Oncology, Royal Adelaide Hospital, North Terrace, Adelaide, South Australia, 5000
4. Future Industries Institute, University of South Australia, Adelaide, South Australia, 5001

Keywords: Biomarker, Endometrial cancer, Glycan Imaging

Abbreviations: EC (Endometrial cancer), FFPE (Formalin Fixed Paraffin Embedded), MALDI MSI (Matrix Assisted Laser Desorption/Ionisation Mass Spectrometry Imaging), LC-MS/MS (Liquid Chromatography Tandem Mass Spectrometry), PNGase F (Peptide-N-Glycosidase F)

Abstract

The function of the glycoproteins has been found to be strongly correlated with the structure of their glycan moieties. Altered glycosylation has been observed in many diseases, including cancer. Glycan matrix assisted laser desorption ionisation mass spectrometry imaging (MALDI MSI) is an evolving technique in which mass spectrometry techniques are applied directly on tissues to characterize the spatial distribution of glycans. Here, we applied *N*-glycan MALDI MSI to an endometrial cancer patient's formalin fixed paraffin embedded (FFPE) tissue section to spatially characterize the *N*-glycan expression within the heterogeneous tissue. Identification of the discriminative *m/z* values was achieved using porous graphitized carbon liquid chromatography (PGC-LC) and collision induced electrospray negative mode MS fragmentation analysis (ESI-MS/MS). The high mannose glycan structures were predominately detected in the tumour regions, while complex bi and tri-antennary structures were observed in the normal regions. In summary, tumour and normal regions were clearly distinguished based on their *N*-glycan distributions.

6.2.1 Introduction

Glycosylation is one of the most common protein post-translational modifications (PTM), and its alteration is considered to be a hallmark of cancer [220]. It is estimated that approximately one-fifth of all proteins are glycosylated [221]. Glycoproteins play an important role in cell differentiation, cell-cell interactions, cell growth, adhesion, immune response and other cancer related activities [222]. An increasing number of studies have established that glycosylation varies significantly for cancer cells compared to normal cells [223, 224].

Glycoproteins are mainly classified into two groups: (1) *N*-linked glycans that are attached to asparagine residues by their amine group and (2) *O*-linked glycans that are attached to serine/threonine residues by their hydroxyl group. Changes in the branching of *N*-glycans, truncation of *O*-glycans, changes in the amount, linkage, acetylation and expression of sialic acids have all been suggested to play a vital role in cancer [225]. *N*-linked glycans have been the most comprehensively studied type of protein glycosylation, as they can be easily released from proteins by enzymatic cleavage using Peptide-*N*-glycosidase F (PNGase F) [226]. Alterations in the degree of branching and levels of sialylation and fucosylation in *N*-glycans have been reported as a consequence of diseases [227-229]. An understanding of these alterations could provide a valuable opportunity to exploit cancer-specific protein glycosylation markers for prognosis and diagnosis of endometrial cancer (EC).

All *N*-linked glycans share a common core sugar sequence, $\text{Man}\alpha 1\text{--}6(\text{Man}\alpha 1\text{--}3)\text{Man}\beta 1\text{--}4\text{GlcNAc}\beta 1\text{--}4\text{GlcNAc}\beta 1\text{--Asn-X-Ser/Thr}$, and are classified into three types: (1) High-mannose, in which only mannose residues are attached to the core; (2) complex, in which “antennae” initiated by *N*-acetylglucosaminyltransferases (GlcNAcTs) are attached to the core; and (3) hybrid, in which only mannose residues are attached to the $\text{Man}\alpha 1\text{--}6$ arm of the core and one or two antennae are on the $\text{Man}\alpha 1\text{--}3$ arm [230].

Matrix assisted laser desorption ionisation (MALDI) mass spectrometry imaging (MSI) of glycans is a recent MS-based development which allows the visualization of the tissue specific spatial distribution of glycans without the need for homogenization, fractionation, or application of antibody-based techniques. The proof-of-principle paper describing the measurement of *N*-glycans by MALDI MSI was first published in 2014 [231]. Subsequently, *N*-glycan MALDI MSI has been successfully applied to discriminate specific tissue regions based on their glycan structure distributions. Our group has recently developed and optimised the protocol for glycan MALDI MSI [232]. In this study, we aimed to identify tissue specific

N-linked glycan markers which can discriminate regions of tumour from adjacent normal regions. *N*-glycan MALDI MSI was used to spatially characterize the *N*-glycan expression within the endometrial cancer FFPE tissue section. Identification of the discriminative *m/z* values were achieved using porous graphitized carbon liquid chromatography (PGC-LC) and collision induced electrospray negative mode MS fragmentation analysis (ESI-MS/MS). Our data indicates that the high mannose glycan structures were predominately detected in the tumour regions, while complex bi and tri-antennary structures were observed in the normal regions.

6.2.2 Material and Methods

6.2.2.1 MALDI MSI analysis of *N*-linked glycans

The formalin fixed paraffin embedded (FFPE) EC tissue block was collected from the Royal Adelaide Hospital (RAH), South Australia. The study was approved by the ethics committee of the hospital.

The tissue section was cut at 6 μm thickness, mounted onto an indium tin oxide (ITO) conductive glass slide (Bruker Daltonics), dried overnight in an oven at 37°C and was prepared as described previously [232]. Briefly, upon citric acid antigen retrieval (10 mM citric acid, pH = 6) and printing of 15 nL Peptide-*N*-Glycosidase F (PNGase F, New England Biolabs) at 250 μm spacing using a ChIP-1000 printer (Shimadzu, Japan), the tissue section was incubated overnight at 37°C. 2, 5-Dihydroxybenzoic acid (DHB, Sigma-Aldrich, USA) matrix was prepared at 20 mg/mL in 0.1% TFA and 1 mM NaCl, and sprayed onto tissue sections using a TM-sprayer (HTX Technologies). Mass spectra were acquired with a laser frequency of 2 kHz over *m/z* 800–4500 Da using an ultrafleXtreme MALDI-TOF/TOF (Bruker Daltonics) in positive reflectron mode with flexControl (version 3.4, Bruker Daltonics) and flexImaging software (version 4.0, Bruker Daltonics). Prior to data acquisition, the instrument was externally calibrated using GLY3 standard (Man₃GlcNAc₂, Man₃GlcNAc₅, and Man₃Gal₄GlcNAc₆, at a final concentration of 1 pmol/ μL , Prozyme, USA) [232]. Data was acquired by oversampling at 100 μm lateral resolution overlaid onto the 250 μm PNGase F array. Following MS acquisition, DHB matrix was removed using 70% ethanol, the tissue section was H&E stained and scanned using a Nanozoomer (Hamamatsu, Shimadzu). To align the MS data with the tissue histology, the H&E scanned tissue section was co-registered with the MALDI MSI results.

6.2.2.2 Data analysis

The data was analysed using SCiLS lab software (version 2016b, GmbH, Bremen, Germany). Here, the raw data was processed by baseline removal and total ion count (TIC) normalization and the ion intensity maps were generated. Tumour and adjacent normal regions were grouped based on the annotations provided by a pathologist and discriminatory m/z values were calculated by receiver operating characteristic (ROC) curve.

6.2.2.3 Identification of m/z values by PGC-LC-ESI-Ion Trap MS/MS

To identify the discriminatory m/z values, the FFPE tissue block was sectioned at 6 μm thickness, water bath mounted onto polyethylene naphthalate (PEN) membrane slides (Micro-Dissect, Herborn, Germany), and deparaffinised by heating at 60°C for 5 minutes, stripped of paraffin wax using xylene, and rinsed with 100% Ethanol. The tissue section was manually micro dissected, treated with antigen retrieval (10 mM citric acid, pH = 6), and incubated overnight with PNGase F. The reaction supernatant was reduced, desalted, and purified using in-house packed carbon columns prior to porous graphitized carbon liquid chromatography (PGC-LC) coupled online to a collision induced electrospray negative mode MS fragmentation analysis as described previously [226].

6.2.3 Results

A MALDI MSI experiment was carried out to comprehensively map the tissue specific distribution of the characterized *N*-linked glycans on an EC FFPE tissue section as previously described [232]. This was done using an ultrafleXtreme MALDI-TOF/TOF MS with flexControl and flexImaging software in positive reflectron mode over a detection range of m/z 800-4500 Da. All glycans were detected as sodium adducts and m/z values were selected and visualized in SCiLS lab software (v2016b, GmbH, Bremen, Germany) with edge-preserving image denoising and automatic hotspot removal applied. Statistical analysis by ROC and determination of the area under the ROC curve was carried out to determine the specificity and sensitivity of the discriminative m/z values. Figure 6 shows the H&E stained annotated EC FFPE tissue section. The region annotated in red represents the tumour while the region adjacent to it represents normal tissue.

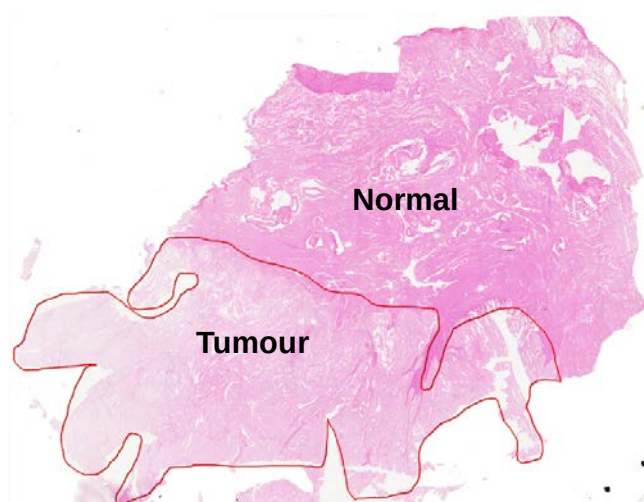
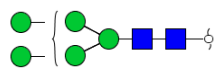
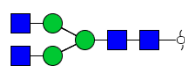
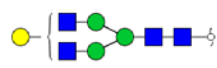
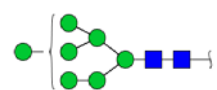
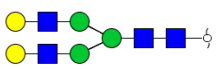
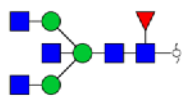
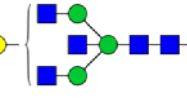
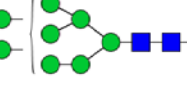
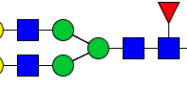
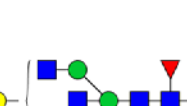
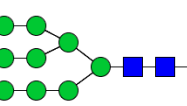
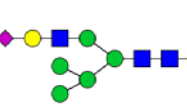
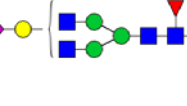
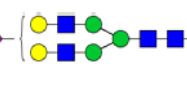


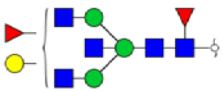
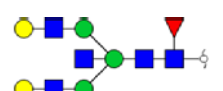
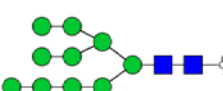
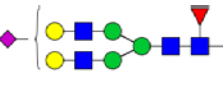
Figure 6: Annotated haematoxylin and eosin stained endometrial cancer formalin fixed paraffin embedded tissue section

N-glycan structures were manually assigned to the corresponding glycan masses from spatially localized MALDI MSI data using the following steps: a) the monosaccharide composition was deduced based on the mass of the released *N*-glycan masses (using GlycoMod, <http://web.expasy.org/glycomod/>) [233], (b) mass-matching of theoretical fragment mass ions produced by tandem MS for each structure, and (c) use of diagnostic fragment ions in negative ion mode that have been previously reported for the identification of specific glycan structural features [232, 234]. Table 1 provides a combined list of the *N*-glycan candidates which were found by MALDI MSI and further identified by PGC-LC-ESI-MS/MS. In total, 32 *N*-glycans were identified by LC-MS/MS. Of the 32 *N*-glycans identified by LC-MS/MS, 23 were detectable in the MALDI MSI dataset. Figure 7 shows a representative MS/MS fragmentation spectrum derived from the doubly charged negative ion of m/z 698.53²⁻ corresponding to the (Hex)3 + (Man)3(GlcNAc)2 (Figure 7A) and m/z 812.55²⁻ corresponding to the (Hex)1 (HexNAc)2 (Deoxyhexose)1 + (Man)3(GlcNAc)2 (Figure 7B). MS/MS spectra were annotated using GlycoWorkbench (version 2.1, Build 146), based on the nomenclature of Domon and Costello [235] viz., fucose (red triangle), mannose (green circle), galactose (yellow circle), *N*-acetylglucosamine (blue square), and *N*-acetylgalactosamine (yellow square).

Table 1: *N*-glycan structures assigned by PGC-LC-ESI-ion trap MS/MS. Proposed *N*-glycan structures detected and characterized by negative ion mode LC-ESI-ion trap MS/MS in tumour and adjacent normal FFPE tissue regions. The *m/z* values from MALDI MSI corresponds to the sodiated mass $[M+Na]^+$. The singly negatively charged monoisotopic masses $[M-H]^{1-}$ were calculated from their observed doubly negatively charged monoisotopic masses $[M-2H]^{2-}$, matching composition and proposed structure. Monosaccharide symbols are as defined by the Consortium for Functional Glycomics [236]. (ND is not detected)

	Glycan composition	Glycan structure	Retention Time	Detected <i>m/z</i> in LC-ESI-ion trap MS/MS		Detected <i>m/z</i> in MALDI MSI
				Normal	Tumour	
1	(Hex) ₂ + (Man) ₃ (GlcNAc) ₂		26.01	617.4	617.4	1257
2	(HexNAc) ₂ + (Man) ₃ (GlcNAc) ₂		21.06	658.8	ND	1339.6
3	(Hex) ₃ + (Man) ₃ (GlcNAc) ₂		21.72	698.3	698.8	1419.6
4	(HexNAc) ₂ (Deoxyhexose) ₁ + (Man) ₃ (GlcNAc) ₂		29.28	731.3	731.8	1485.8
5	(Hex) ₁ (HexNAc) ₂ + (Man) ₃ (GlcNAc) ₂		27.06	739.5	ND	1501.6
6	(Hex) ₄ + (Man) ₃ (GlcNAc) ₂		21.88	779.51	779.8	1581.5
7	(Hex) ₁ (HexNAc) ₂ (Deoxyhexose) ₁ + (Man) ₃ (GlcNAc) ₂		32.15	812.61	812.8	1647.6
8	(Hex) ₂ (HexNAc) ₂ + (Man) ₃ (GlcNAc) ₂		29.99	820.67	820.3	1663.6

9	(HexNAc) ₃ (Deoxyhexose) ₁ + (Man) ₃ (GlcNAc) ₂		21.93	833.09	833.3	1688.6
10	(Hex) ₁ (HexNAc) ₃ + (Man) ₃ (GlcNAc) ₂		17.84	841.22	840.09	1704.6
11	(Hex) ₅ + (Man) ₃ (GlcNAc) ₂		20.90	860.59	860.8	1743.6
12	(Hex) ₂ (HexNAc) ₂ (Deoxyhexose) ₁ + (Man) ₃ (GlcNAc) ₂		36.35	893.6	893.2	1809.6
13	(Hex) ₁ (HexNAc) ₃ (Deoxyhexose) ₁ + (Man) ₃ (GlcNAc) ₂		24.93	914.15	913.8	1850.9
14	(Hex) ₅ + (Man) ₃ (GlcNAc) ₂		26.03	941.22	941.8	1905.6
15	(Hex) ₃ (GlcNAc) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		26.59	ND	945.9	ND
16	(Hex) ₁ (HexNAc) ₂ (Deoxyhexose) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		40.50	958.17	ND	ND
17	(Hex) ₂ (HexNAc) ₂ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		28.75	966.2	966.4	ND

18	(Hex) ₁ (HexNAc) ₃ (Deoxyhexose) ₂ + (Man) ₃ (GlcNAc) ₂		46.86	987.4	987.68	ND
19	(Hex) ₂ (HexNAc) ₃ (Deoxyhexose) ₁ + (Man) ₃ (GlcNAc) ₂		27.88	995.15	995.4	2012.8
20	(Hex) ₇ + (Man) ₃ (GlcNAc) ₂		24.69	1022.58	1022.9	2067.8
21	(Hex) ₂ (HexNAc) ₂ (Deoxyhexose) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		34.79	1039.73	1039.4	2100.8
22	(Hex) ₁ (HexNAc) ₃ (Deoxyhexose) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		33.38	1059.75	1059.69	2163
23	(Hex) ₂ (HexNAc) ₂ (Sulphate) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		39.27	1076.27	1076.4	2175.8
24	(Hex) ₂ (HexNAc) ₂ (NeuAc) ₂ + (Man) ₃ (GlcNAc) ₂		29.69	1111.8	1111.9	2245.8
25	(Hex) ₁ (HexNAc) ₃ (Deoxyhexose) ₂ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		32.02	ND	1132.87	ND

26	(Hex) ₂ (HexNAc) ₃ (Deoxyhexose) ₁ + (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		27.61	1140.78	1141	2304
27	(Hex) ₂ (HexNAc) ₂ (Deoxyhexose) ₁ (NeuAc) ₂ + (Man) ₃ (GlcNAc) ₂		46.46	1184.33	1184.9	2391.8
28	(Hex) ₃ (HexNAc) ₃ (Deoxyhexose) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		35.47	1221.85	1221.9	2465.8
29	(Hex) ₂ (HexNAc) ₃ (Deoxyhexose) ₁ (NeuAc) ₂ + (Man) ₃ (GlcNAc) ₂		27.63	1286.38	1286.4	2594.8
30	(Hex) ₅ (HexNAc) ₅ + (Man) ₃ (GlcNAc) ₂		41	1368.93	1368.3	2759.6
31	(Hex) ₄ (HexNAc) ₄ (Deoxyhexose) ₁ (NeuAc) ₁ + (Man) ₃ (GlcNAc) ₂		41.75	1404.49	1404.7	2831.4
32	(Hex) ₃ (HexNAc) ₃ (NeuAc) ₃ + (Man) ₃ (GlcNAc) ₂		36.55	1439.50	ND	2902.4

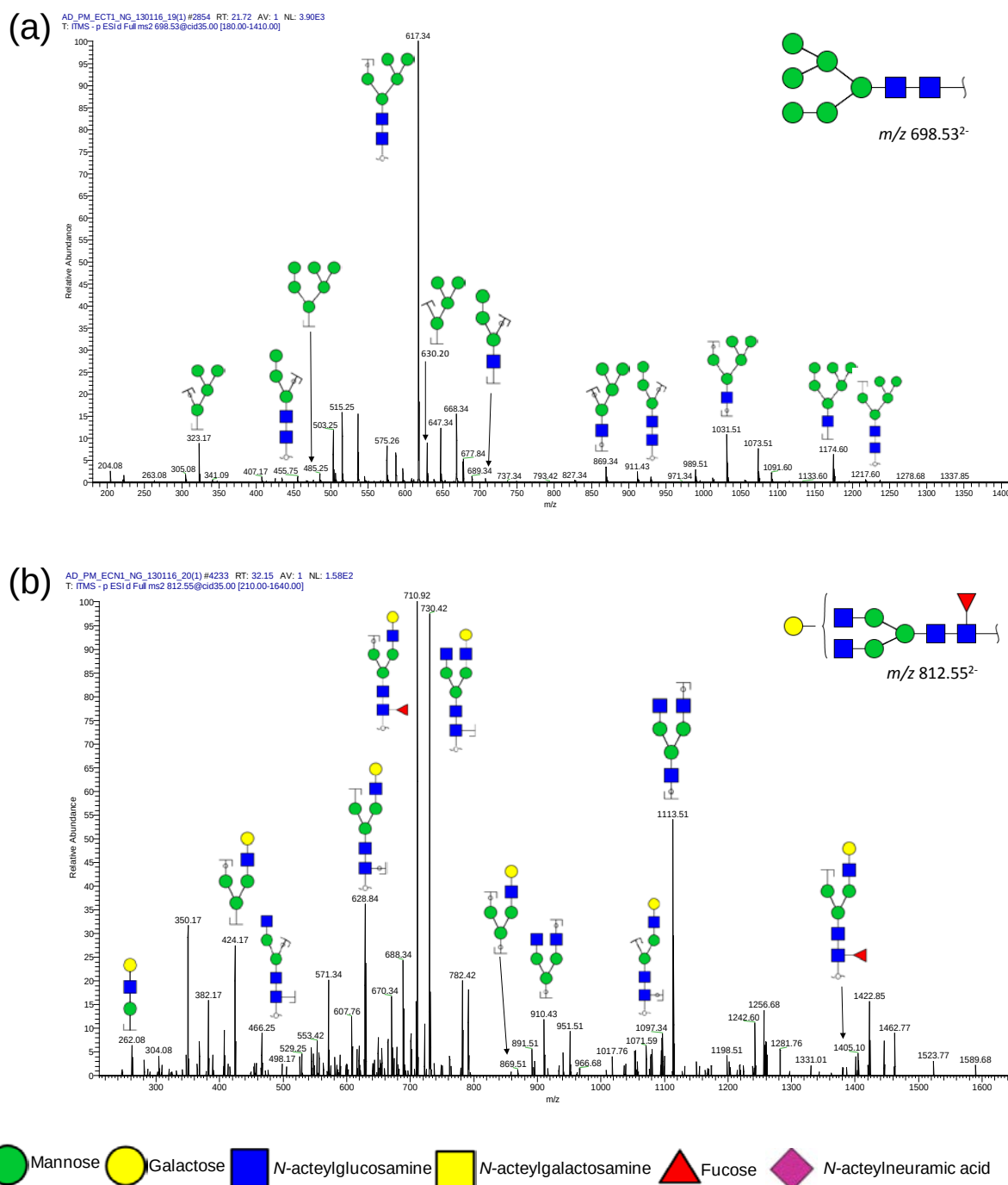


Figure 7: Tandem mass spectrometry fragmentation spectra derived from the doubly charged negative ion isomers of (A) m/z 698.53²⁻ and (B) m/z 812.55²⁻, respectively. The identified structures with their key fragments are shown. Monosaccharide symbols are as defined by the Consortium for Functional Glycomics [236].

The spatial ion intensity maps for four individual $[M+Na]^+$ values which correspond to $(Hex)_3 + (Man)_3(GlcNAc)_2$ (m/z 1419.671 $\pm$ 0.125 Da), $(Hex)_4 + (Man)_3(GlcNAc)_2$ (m/z 1581.726 $\pm$ 0.125 Da), $(Hex)_5 + (Man)_3(GlcNAc)_2$, (m/z 1743.807 $\pm$ 0.125 Da) and $(Hex)_5 + (Man)_3(GlcNAc)_2$ (m/z 1905.871 $\pm$ 0.125 Da), as identified by LC-MS/MS are shown in Figure 8. These *N*-linked glycans belong to the high-mannose glycan family and were observed to be predominantly abundant in the tumour region compared to the adjacent normal region, as shown also in the ROC curve (AUC = 0.768-0.907).

The ion intensity maps for four individual $[M+Na]^+$ values which correspond to $(Hex)_1(HexNAc)_2(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$ (m/z 1647.801 $\pm$ 0.125 Da), $(Hex)_2(HexNAc)_2 + (Man)_3(GlcNAc)_2$ (m/z 1663.781 $\pm$ 0.125 Da), $(Hex)_2(HexNAc)_2(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$ (m/z 1809.853 $\pm$ 0.125 Da), and $(Hex)_1(HexNAc)_3(Deoxyhexose)_1 + (Man)_3(GlcNAc)_2$ (m/z 1850.923 $\pm$ 0.125 Da), are shown in Figure 9. These hybrid/complex *N*-glycans were observed predominately in the adjacent normal region compared to the tumour region. The corresponding ROC curves display a fair accuracy in discriminating the normal region from the tumour region (AUC = 0.041-0.173).

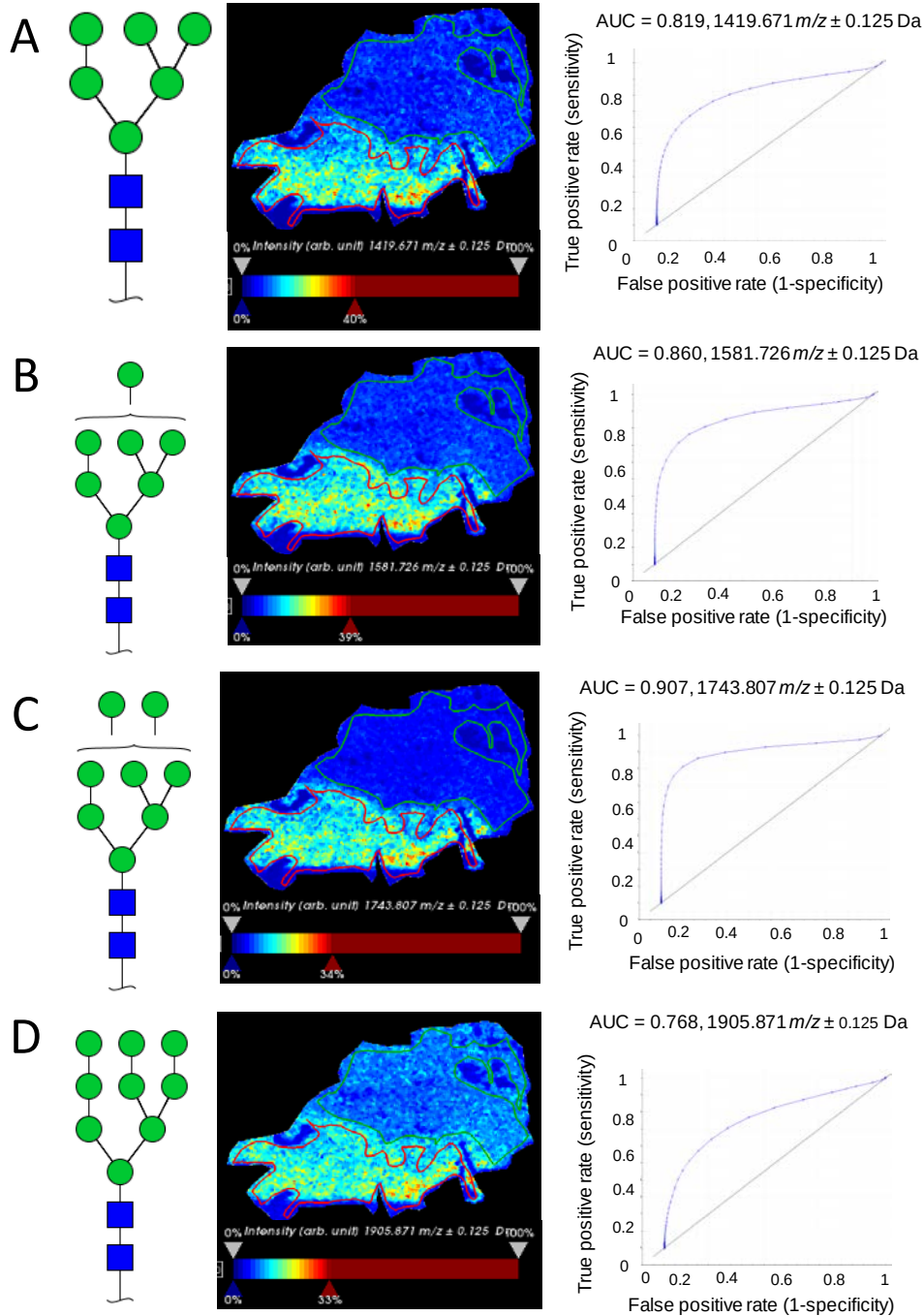


Figure 8: High-mannose glycans are elevated in endometrial cancer. *N*-linked glycans were released *in situ* on FFPE tissue sections using PNGase F and analysed by MALDI-TOF/TOF MS. Monoisotopic glycan masses were measured in the positive ion reflectron mode as $[\text{M}+\text{Na}]^+$ adducts. The m/z values were selected and visualized in SCLS lab software (v2016b, GmbH, Bremen, Germany) with edge-preserving image denoising and automatic hotspot removal applied. Four high-mannose structures 1419.671 (A), 1581.726 (B), 1743.807 (C) and 1905.871 (D), were observed predominantly in the tumour region compared to the adjacent normal region.

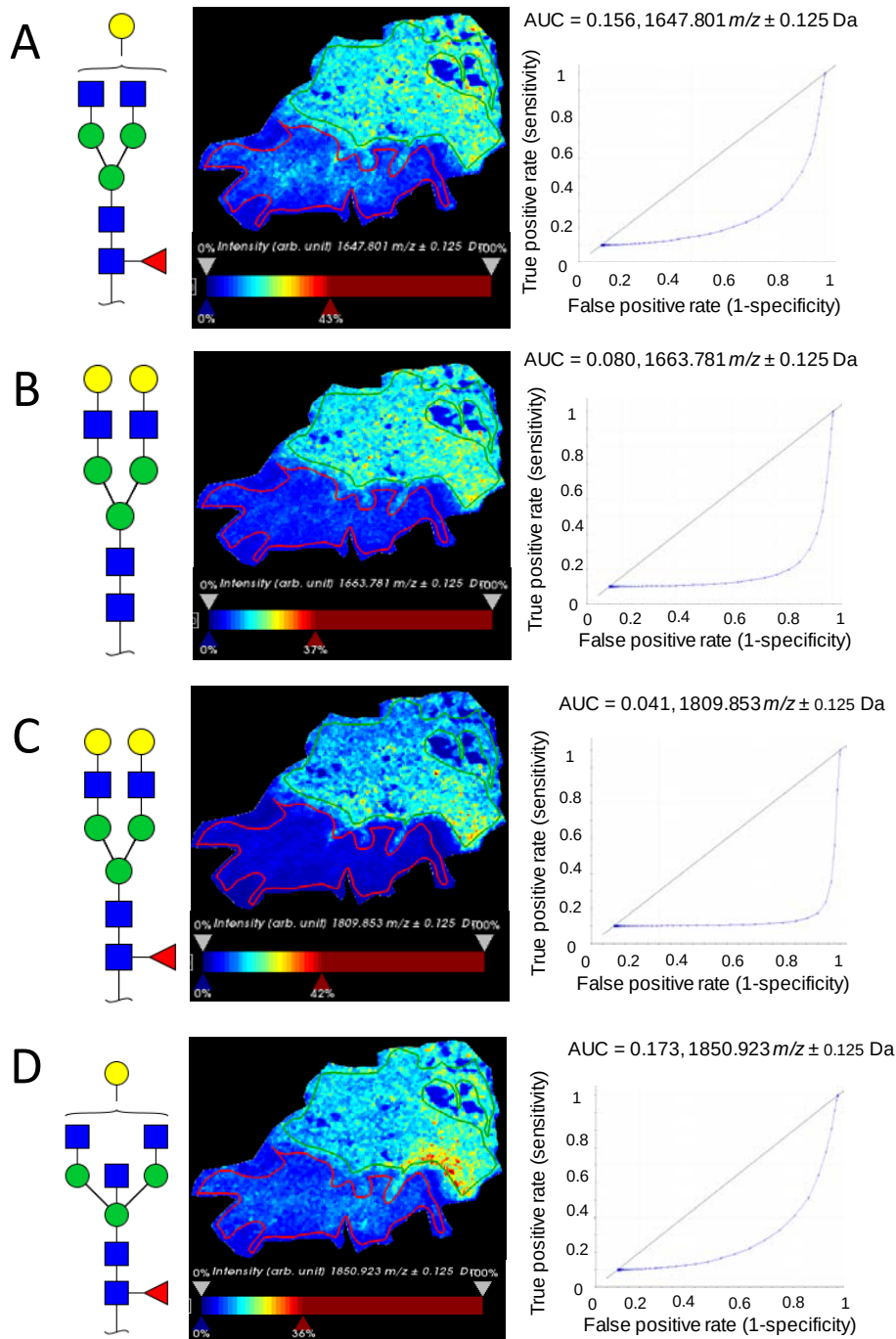


Figure 9: Hybrid/complex *N*-glycan structures were predominantly observed in the adjacent normal region compared to the tumour region. *N*-linked glycans were released *in situ* on FFPE tissue sections using PNGase F and analysed by MALDI-TOF/TOF MS. Monoisotopic glycan masses were measured in the positive ion reflectron mode as $[M+Na]^+$ adducts. The m/z values were selected and visualized in SCoLS lab software (v2016b, GmbH, Bremen, Germany) with edge-preserving image denoising and automatic hotspot removal applied. The hybrid/complex *N*-glycan structures 1647.805 (A), 1663.781 (B), 1809.853 (C) and 1851.923 (D), were observed predominantly in the adjacent normal region compared to the tumour region.

6.2.4 Discussion

It has long been known that glycans are frequently altered in tumorigenesis and therefore the discovery of cancer-specific protein glycosylation markers would be beneficial in the staging and treatment of EC patients. Glycan MALDI MSI is a recent MS-based development and has been successfully applied to discriminate specific tissue regions based on their glycan structure distribution. Our data indicate that the *in situ* PNGase F released *N*-linked glycans were easily able to discriminate the regions of tumour from adjacent normal regions. High-mannose glycan structures (shown in Figure 8) were observed to be significantly more abundant in the tumour region compared to the adjacent normal region. The increase in high-mannose structures has been previously reported in breast cancer, colorectal cancer, hepatocellular carcinoma and ovarian cancer [234, 237-239]. On the other hand, the complex bi-antennary and tri-antennary structures (shown in Figure 9) were significantly more abundant in the adjacent normal region. The glycan masses showed high AUC values, confirming their potential power as discriminators between the tumour and adjacent normal regions.

The results observed in this study offer a unique advantage in the analysis of endometrial cancer. However, future work on larger patient samples will be required to understand the biological relevance of this observation and to discover cancer-specific markers.

6.2.5 Conflict of interest

The authors have declared no conflict of interest.

This page intentionally left blank

DISCUSSION

This page intentionally left blank

7.1 Main Findings and Other Considerations

The clinical relevance of lymphadenectomy is currently one of the most discussed topics in the management of EC. Since FIGO introduced surgical staging of EC, essential questions have remained unanswered including the extent of an optimal lymphadenectomy and which subgroup of patients would benefit. In Australia, clinicians routinely perform systematic pelvic and para-aortic lymphadenectomy up to the level of the renal vein for EC. However, it has been established that patients with a low-risk for recurrence do not benefit from lymphadenectomy [45, 240].

Primary pathological characteristics such as myometrial invasion, cervical stromal invasion, histological type, lympho-vascular space invasion (LVSI), tumour size, and tumour grade have been investigated for their association with LNM [241]. However, only LVSI has been shown to correlate with LNM [40]. Multiple studies including ours have demonstrated the positive association of LVSI and LNM [242]. Unfortunately, the LVSI information can be obtained only after surgery. Clinical use of predictive tissue markers before surgery could prevent unnecessary surgical treatment, thereby directly reducing EC morbidity and complications associated with the removal of lymph nodes.

For over 100 years clinical pathology has relied upon the description of tissue morphology by trained specialists for tumour classification and grading. The information that can be gained from such analyses is descriptive, without providing any molecular information. In the past decade, MALDI MSI has successfully been applied to the discovery of potential biomarkers with the capacity to differentiate tumours of various subtypes, stages, and/or degree of metastasis [243]. Using MALDI MSI, Nipp *et al.* have shown the overexpression of S100A10, thioredoxin, and S100A6 to be highly associated with metastatic potential in papillary thyroid carcinoma (PTC) [244]. Other protein biomarkers have been suggested for the prediction of LNM in several cancers, such as S100A8 and S100A9 for gastric adenocarcinoma [245], and COX7A2, TAGLN2 and S100-A10 for predicting LNM in Barrett's adenocarcinoma [246]. Yanagisawa *et al.* were able to classify primary lung tumours with and without LNM with an accuracy of 85% [247]. Using MALDI MSI, our goal is to improve upon the current system of clinicopathologic risk group stratification that is based on irreproducible variables of stage and grade assignment [248], and is not highly predictive of outcomes [249].

MALDI MSI data are always characterized by a large number of variables i.e. the mass peaks relative to a small sample size. An important issue in analysing such data is to prevent

mistakenly selecting mass peaks whose high discriminatory power is based purely on chance because of artefacts in the data. MALDI MSI on TMAs facilitates the collection of label-free and high-dimensional data, which are difficult to process and interpret. In order to manage and interpret such data, it is necessary to perform some dimension reduction prior to classification. Therefore in the present study, after the MALDI MSI data from TMAs containing a total of 43 patient were acquired, a CCA based classification method proposed by Winderbaum *et.al.* [143] was applied to rank the intensities of the m/z values. This ranking was based on their power to discriminate the primary carcinomas with metastatic potential from those without and to reduce the dimensions of the data by restricting them to only the most highly ranked MSI variables. By doing this, we have identified 20 m/z values which can classify 88% of all tumours correctly. Potential peptide identifications for the top 20 m/z values, were compiled by matching to sequences obtained by data dependent acquisition LC-MS/MS (Chapter 4, Section 4.4, Supplementary Table 2). Of the 20 m/z values, 3 had no matches within the tandem MS data, with the remaining 17 having 2 or more possible sequence matches. In order to confirm peptide identifications, the m/z values were targeted for *in situ* MS/MS directly off the tissue, from which only 2 peptides could be verified; m/z 1198 AVFPSIVGRPR (α -Actin-2), and m/z 976 AGFAGDDAPR (α -Actin-2). Hence, most of the m/z values used in the MALDI MSI classification model remained unidentified. The possible explanation for why not all the m/z values were identified by tandem MS is because a number of exclusion or inclusion criteria were used, such as the number of unique peptides identified, peak intensity thresholds, identifiable modifications, mass tolerance, and/or enzyme specificity. For instance, to obtain the protein identities, it is generally necessary to enzymatically convert the proteins into their constituent peptides and then sequenced by tandem MS experiments, in which parent ions are fragmented to produce daughter ions. The resulting product-ion mass spectra contain the information about the sequence, molecular mass, and abundance of the ions. Based on the degree of correlation between the experimentally observed and the theoretical spectra (*in silico* digests of protein), protein identities are scored and assigned [112]. We believe that in doing so, we could be losing important information in each step. In contrast, the MALDI MSI data has not been subjected to any downstream exclusion or inclusion criteria, thereby avoiding any inherent bias introduction. Future work might consider the clinical evaluation of proteomic patterns without necessarily knowing their identity or sequence, as similar pattern recognition is currently being employed by the automated analysis of microorganisms using MALDI.

The proteins identified as plectin and α -Actin-2, were further validated by immunohistochemistry (IHC). This was chosen as the method for validation given its routine use in clinical pathology laboratories. However, the IHC analysis of plectin failed to identify a difference in staining intensity of tumours with LNM when compared to tumours without LNM. Significantly, this is in contrast to the results produced by the MALDI MSI classification model and data independent acquisition LC-MS/MS. One possible explanation of the discrepancy of results as shown by Bian *et al.* is that S1047 in plectin is a potential phosphorylation site [250]. As a result, it might be difficult to validate the relative abundance of a protein of interest by the clinically approved IHC technique.

In the past two decades, multiple studies have identified potential biomarkers of EC metastasis, however none of them are in clinical use. As a targeted approach in the current study, we aimed to develop a predictive model using previously identified biomarkers. Following a review of the literature and publically available data, we compiled a list of 60 proteins having the potential to display differential abundance between the primary tumours with and without LNM, of which 23 were detected by DDA LC-MS/MS. The possible explanation as to why not all the proposed biomarkers were detectable at the protein level is because a large number of the biomarkers assessed in this study originated from the genomics data and it is well documented that only about two-thirds of protein abundance can be predicted from mRNA measurements [251]. Since 2004, numerous studies have compared correlation between mRNA and protein levels and concluded that the correlation is poor [252-255]. A recent study by Lu, R *et al.* on embryonic stem cells revealed that changes in protein levels are not necessarily accompanied by changes in corresponding mRNAs [256]. Schwanhausser *et al.* performed stable isotope labelling by amino acids in cell culture (SILAC) followed by mass spectrometry to measure protein turnover, and 4-thiouridine to measure mRNA turnover in exponentially growing, non-synchronized mammalian cells and found that proteins were, on average 5 times more stable (median half-life 46 h) than mRNA (9 h) [252, 257]. Authors have found that abundant housekeeping genes coding for ribosomal, glycolytic and TCA cycle proteins tend to have stable mRNAs and stable proteins. In contrast, transcription factors, signalling genes, chromatin modifying genes, and genes with cell-cycle specific functions tend to have unstable mRNA and unstable protein [252]. Thus, the most important and interesting regulators of cellular division and differentiation would be expected to have a poor correlation between mRNA and protein while housekeeping proteins would probably have relatively good correlation [258]. However, Marginantu *et al* showed that even mitochondrial house-keeping

proteins may vary independently of mRNA changes [259]. Stark *et al.* measured the correlation between mRNA and protein levels in breast cancer primary tumours and brain metastasis and observed that the mRNA level of p53 was significantly lower in brain metastasis than primary tumours but protein levels were only slightly lower (not significant) [260]. Sarro *et al.* also observed that the mRNA level did not correlate with protein levels [261].

Of the 23 biomarkers that were detectable at the protein level by DDA LC-MS/MS, the differential abundance of five of those proteins were observed by DIA LC-MS/MS. The proteins identified were further validated by IHC. When this IHC data was used to generate a logistic regression model for LNM in conjunction with LVSI, it resulted in a model with only ACTN4 and LVSI being significant. Likewise, a predictive model using the clinicopathologic features such as age, stage, grade, and LVSI and the top 10 MALDI MSI variables as ranked by the CCA analysis, resulted in a model with 3 MALDI MSI variables and LVSI being significant. Both the IHC data of ACTN4 and MALDI MSI variables, when combined with LVSI, outperformed LVSI alone. However, the MALDI MSI model outperformed the ACTN4 IHC model. Our data analyses have shown that the predictive model using highly ranked *m/z* values identified by MALDI MSI outperforms the current clinicopathologic approach. Therefore, the question arises whether this classification model can be used for new clinical studies and guide treatment decisions.

Clinical mass spectrometric assays generally use internal standards to correct for variable ionization and suppression [262]. As an internal standard reference, it would be desirable to incorporate either known added components or classic and well-established biomarkers such as keratins and vimentins that are found consistently throughout all tissue samples. Spiking this internal standard into calibrators, standards, and samples at an equal concentration provides a means to normalize for variability in sample preparation and MALDI mass spectrometric analysis. Most important, was the observation that randomly selected samples, unknown to the person performing MALDI MSI obtained nearly identical spectral patterns after months or even a year later. Furthermore, it will be critical to determine the analytical sensitivity of this method; especially the reproducibility of MALDI MSI spectra, *i.e.*, mass location and intensity from one patient to other patient in a single lab and between labs, which has to be determined using the control sample. Dekker *et al.* reported the first multicentre clinical study and demonstrated that protein biomarkers of stromal activation in breast cancer could be reproducibly detected in two different biomedical centres, in which different patient collections were analysed by different analysts, using slightly different methods and instruments [263].

Differences were observed in the spectra that were recorded in both centres, as would be expected considering the different sample preparation strategies employed. Nevertheless, all protein ions that were found to be statistically significant in the Munich data set, were also detected and found to be statistically significant in the Leiden data set [263].

7.2 Clinical Applications of MALDI-TOF

The ability of matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) to register biomarker ions in a broad m/z range, that are unique and representative for individual microorganisms, forms the basis of current applications of mass spectrometry in microbiology [264]. Even though the first study for microbial identification based on species-specific spectra of peptides and protein masses by mass spectrometry was first proposed in 1975 by Anhalt & Fenselau [265], MALDI-TOF MS devices designed for use under routine conditions have only recently been commercially introduced. For example the Biotyper, a MALDI-based identification of microorganisms based on standardized MALDI MS analysis of colonies followed by classification, is now an FDA-approved technology and has had a large clinical impact [266], primarily through the reduction in health care costs associated with faster and less expensive methods for microorganism identification [267]. More than 2000 MALDI Biotyper instruments have been installed in hospital laboratories worldwide. It provides a straightforward means to differentiate microorganism species based on their protein profiles (protein mass pattern). The success of this application is due to the performance advantages of the mass spectrometer, which provides data to the clinician faster, more cost effectively, and more accurately than the classical microbial approaches [268]. As the MALDI Biotyper approach is similar to the use of MALDI imaging in tissue-based experiments, we believe that this concept could also be of great interest for classification of EC patient samples. For clinical pathology laboratories currently using MALDI MS for microbial identifications, it is attractive to extend this technology for the identification of cancer biomarkers.

MALDI mass spectrometry imaging (MSI) is a relatively simple technique that has great potential to develop into a valuable diagnostic technique that can complement established histological and histochemical methods. Generation of the mass spectra requires only a small tissue sample that could be obtained by endometrial biopsy or curettage specimens that are routinely obtained during the EC evaluation. If a high degree of accuracy can be achieved through a protein profile, identifying the individual proteins that make up the signature becomes diagnostically unnecessary [269]. By this approach, the protein profile itself, independent of the underlying identities of the proteins or peptides, is the discriminator, and has the ability to direct clinical care. In order to continue to move toward being placed in a clinical setting, further improvements need to be made towards standardizing production of

libraries of tissue signatures for routine matching to patient samples [267]. MALDI MSI is still under development and is a promising tool for discovering clinical biomarkers, strengthening the efforts for optimizing both the prevention and treatment of disease.

7.3 Role of Extracellular Matrix in Endometrial Cancer

The ability of tumour cells to invade into surrounding tissues and spread to distant sites is characteristic of malignant tumour cells. During metastasis, tumour cells cross the basement membrane (BM), the extracellular matrix (ECM) and vessel walls [270]. The ECM is a complex three-dimensional network of cross-linked proteins providing architectural support necessary to maintain the tissue integrity and sustainability [271]. It has been shown previously that the ECM supports proliferation and survival of tumour cells; it contributes to the formation of the cancer stem cell niche and thus sustains primary tumour growth; it participates by its nature and/or architecture in the formation of a pro-invasive environment; and, finally, it contributes to the invasion of distant sites by participating in the formation of a microenvironment that will support tumour cell seeding and growth [272-276]. The ECM thus appears to be of great interest for discovery of biomarkers for disease progression and metastasis.

Our study focused on primary tumours and we didn't analyse the ECM from our patient samples. However, a group from the University of New South Wales, NSW, Australia, have done extensive research on ECM in EC and initiated a collaboration with us. Preliminary results show differences in the ECM of healthy individuals versus cancer patients (manuscript under preparation) and it is worth considering repeating our study with the focus on ECM.

7.4 Conclusion and Future Perspectives

Although MALDI MSI has already had a substantial impact in cancer research, it is not yet widely adopted. In order to establish MALDI MSI as a robust analytical technology for clinical applications, further developments need to be made towards (i) high-throughput methods and instrumentation, (ii) the ability to quantitate MALDI MSI results, and (iii) bioinformatics tools for analysis of large datasets. However to meet the needs of the researchers, work has been done to improve both the instrumentation and the experimental workflow during the past several years. Advances in instrumentation including Fourier transform-based methods and the incorporation of ion mobility, have been shown to be effective for broadening the application of MSI. With the advent of new instruments on the market, resolutions of $< 10\ \mu\text{m}$ can be achieved that have the capacity to analyse subcellular structures [277]. Laser speed has increased from the 200 Hz of 8–10 years ago to 2–10 kHz. Several groups have also introduced methods to obtain absolute quantitative information from MALDI MSI data by incorporating an internal standard of varying concentrations to correct for signal variability. The intensity of the biomarkers is estimated by comparing the ion intensity in the image to the standard curve from the dilution range [278, 279]. Groseclose *et al.* demonstrated that this method is comparable to similar analysis by LC MS/MS [278]. Pirman *et al.* applied an isotopically labelled internal standard uniformly beneath the tissue section, and wide-isolation tandem mass spectrometry was performed. Normalizing the analyte ion signal by the internal standard ion signal resulted in significant improvements in MS images, signal reproducibility, and calibration curve linearity. From the improved MS images, the concentration of protein of interest was determined and plotted producing a concentration-scaled image of the distribution of protein of interest in the piglet brain section [279]. However, data processing and analysis are areas within the MALDI MSI workflow that are currently being rigorously pursued for effective, cross-platform applications [280].

The ongoing technological developments in MALDI MSI have the capability to revolutionize pathology. The possibility of measuring hundreds of molecules at the cellular level with direct correlation to histological features could overcome a multitude of limitations of other approaches in the identification and routine diagnostic measurement of new marker molecules/profiles. However, large scale studies proving the technology to be reliable in a clinical setting and establishing robust sample preparation and measurement strategies that show high agreement between different laboratories are still lacking. In order to take MALDI

MSI to the next level and towards implementation in the clinic, close collaborations between physicians, especially pathologists, molecular biologists and technology developers are needed. We believe that such strategies may allow the replacement in the near future of the traditional methods of identification, which are time-consuming and sometimes not reliable.

7.5 Overall summary

The identification of biomarkers that can predict the status of lymph node metastasis in endometrial cancer assists with the prognosis and management of the disease. A number of potential molecular markers have been identified but their role in clinical decision making still needs to be established. The major limitations in delaying translation of the biomarkers to the clinic is due to (i) The heterogeneity of protein expression among and in tumours creates a challenge since most tissue based assays rely on tissue homogenization (ii) Given the numerous molecular pathways that impact tumorigenesis, it is unlikely that a single marker currently fulfils all criteria [269]. Therefore, combinations of molecular markers or a proteomic signature consisting of multiple differentially expressed peptides are expected to form the most promising tool to help guide treatment decisions.

Using MALDI MSI on formalin-fixed paraffin embedded tissue samples, we have proposed a new system for classification based on the molecular features. This provides independent prognostic information beyond the established risk factors. The proposed molecular classification tool has the potential to be used routinely in guiding treatment for individuals with endometrial carcinoma and in stratifying cases in future clinical trials. If we can demonstrate equivalence or superiority of a molecular classification system in diagnostic endometrial samples and prognostic significance of a classifier, then women and their physicians could have valuable information that would help them guide decision making at the earliest time point in their cancer journey (e.g., at diagnosis). In future it is envisioned that decisions could be made before surgical staging regarding the urgency and extent of surgery, anticipated adjuvant therapy and follow-up plans.

Furthermore, the application of these classification tools to a larger patient's cohort, will ensure the technology's reliability in a clinical setting. We are working towards the confirmation and validation of this molecular classification in an unknown EC patient cohort. It remains to be seen whether this molecular classification tool may be able to classify the metastasis patient from those without with similar accuracy or possibly to exceed current best practice. However, to assure the robustness, the CCA classification algorithm is being verified among a vulvar cancer patient cohort and the classification model was able to distinguish vulvar cancer patients with LNM from those without LNM (for details refer to Supplementary File of Appendix I). We see an opportunity to test this classification model across cancer centres and on preoperative endometrial samples, thus influencing management at time of diagnosis.

In summary, a molecular classification model proposed in this study may have the potential to be used in the clinic. It will likely prove to be more reproducible than histopathological assessment, but this needs to be formally evaluated. We suggest combining the pathologic and molecular classification model, as an option to improve the clinical management of EC patients and this should form the basis for future prospective clinical studies.

Classification of MALDI-MS Imaging Data of
Tissue Microarrays using Canonical Correlation
Analysis-Based Variable Selection

This page intentionally left blank

TECHNICAL BRIEF

Classification of MALDI-MS imaging data of tissue microarrays using canonical correlation analysis-based variable selection

Lyron Winderbaum, Inge Koch, Parul Mittal and Peter Hoffmann

The University of Adelaide, Adelaide, SA, Australia

Applying MALDI-MS imaging to tissue microarrays (TMAs) provides access to proteomics data from large cohorts of patients in a cost- and time-efficient way, and opens the potential for applying this technology in clinical diagnosis. The complexity of these TMA data—high-dimensional low sample size—provides challenges for the statistical analysis, as classical methods typically require a nonsingular covariance matrix that cannot be satisfied if the dimension is greater than the sample size. We use TMAs to collect data from endometrial primary carcinomas from 43 patients. Each patient has a lymph node metastasis (LNM) status of positive or negative, which we predict on the basis of the MALDI-MS imaging TMA data. We propose a variable selection approach based on canonical correlation analysis that explicitly uses the LNM information. We apply LDA to the selected variables only. Our method misclassifies 2.3–20.9% of patients by leave-one-out cross-validation and strongly outperforms LDA after reduction of the original data with principle component analysis.

Received: November 14, 2015

Revised: March 4, 2016

Accepted: March 14, 2016

Keywords:

Bioinformatics / Canonical correlation analysis / Classification / Endometrial cancer / MALDI-MS imaging / Variable ranking



Additional supporting information may be found in the online version of this article at the publisher's web-site

MALDI-MS imaging enables the relative quantification and spatial expression profiling of thousands of peptides within and between tissues. Using tissue microarrays (TMAs) can facilitate the acquisition of data from large cohorts of patients [1], but relative to the number of resolvable ion species, the sample sizes remain small—less than 100 patients represented in a typical study, although there are some notable exceptions [2]. Classification at a patient level of such data is of clinical interest, but classical approaches to classification, such as linear discriminant analysis (LDA), do not

apply without modifications such as prior feature selection or regularization. In this paper we suggest a method for dimension reduction via variable selection prior to classification and demonstrate its effectiveness when applied to the data of [3], which consist of MALDI-MS imaging data collected from TMAs of formalin-fixed, paraffin-embedded (FFPE) primary endometrial carcinoma tissue. For these endometrial cancer data it is of interest to distinguish between patients with lymph node metastasis (LNM) and those without.

Endometrial cancer is the most common gynaecological malignancy in Australia with 2256 diagnosed cases in 2010 and 381 associated deaths in 2011 [4]. The presence or absence of lymph node metastasis (LNM) is the most important prognostic factor in endometrial cancer as patients with localised disease have a 5-year-survival rate of 96%, which drops to just 17% for patients with metastatic disease [5]. Accurately staging endometrial cancer is difficult and a large percentage of patients are misclassified prior to treatment [6]. Although the presence of LNM is confirmed in only around 15% of

Correspondence: Lyron Winderbaum, Adelaide Proteomics Centre, School of Biological Sciences, Faculty of Sciences, North Terrace Campus, Level 1 Molecular Life Sciences, The University of Adelaide, SA 5005, Australia

Fax: +61 (0)88313 4362

E-mail: lyron.winderbaum@student.adelaide.edu.au

Abbreviations: CCA, canonical correlation analysis; FFPE, formalin-fixed, paraffin-embedded; HDLSS, high-dimension, low sample size; LDA, linear discriminant analysis; LNM, lymph node metastasis; LOO, leave-one-out; PCA, principle component analysis; TMA, tissue microarray

Colour Online: See the article online to view Figures 1 and 2 in colour.

cases [7, 8], the majority of endometrial cancer patients undergo radical treatment including the removal of pelvic lymph nodes as a precautionary measure to compensate for our current inability to accurately stage the disease. Lymph node removal is associated with significant complications including lower extremity lymphoedema, which has been described in up to 38% of patients [9]. A classification system based around predictive tissue markers of metastasis would greatly benefit stage I endometrial cancer patients by helping determine optimal treatment strategies that avoid unnecessary, invasive procedures.

The sample preparation of the endometrial TMAs included citric acid antigen retrieval [10], trypsin digestion, addition of internal calibrants [11], and matrix deposition. Further details on the data acquisition for the endometrial data are available in [3]. As part of the data acquisition peak-picking was performed through proprietary software (flexControl v3.0.1 and flexImaging v4.0.1, Bruker Daltonik, <http://www.bruker.com>). All our analyses are carried out solely on the monoisotopic peaks detected by this method. These peak-data are about two orders of magnitude smaller and result in much faster computation times while preserving important information.

Two sections (technical replicates) of two TMAs were analyzed and tumour regions annotated by a pathologist. In this analysis, the only information we use about each spectrum is which patient it belongs to and whether the pathologist's annotations include it as tumor. After consideration of the patient clinical data it was determined that 43 patients suitable for the study are represented across the two TMAs. Of these 43 patients, 16 are LNM positive, 27 are negative. Details on the endometrial cancer project from which these data originate are available in [3].

We construct 0.25 Da wide, nonoverlapping bins that are the variables in our analyses. In addition we repeat the analyses for shifted bins. Our final prediction combines the corresponding results from the individual analyses, using a majority rule. Details regarding the choice of bins, shifted-bin analyses and majority rule are given in the supplementary information.

For each patient, spectra were averaged for each m/z bin. These averages are assembled into data matrices with 43 columns corresponding to the patients represented in the study, and 4582 rows corresponding to m/z bins with at least one peak in them in at least one spectrum (4570 and 4584 in the two shifted-bin analyses, respectively). This resulting data matrix is what will be used in all following analyses. We center the rows of these data matrices, and refer to these centered data as X .

As previously mentioned, classical approaches to classification such as LDA do not apply to high-dimension, low sample size (HDLSS) data without modifications. Different solution paths exist in the literature that include replacing the classical LDA by its Naive Bayes version, see [12, Section 13.2.1]. Other methods that are directly applicable to HDLSS data include Distance Weighted Discrimination [13], kernel

methods – see [14], and regularization approaches that make use of generalized inverses, ridge-like estimates of the covariance matrix, or Lasso and elastic net type classification – see [15, 16] and references therein.

A different solution path consists of reducing the number of variables or features first, and then applying the chosen classifier to the reduced data. In MALDI-MS imaging data principle component analysis (PCA) is a popular feature reduction method, see [17, 18] and references therein. More recently [19] have suggested the use of PCA on MALDI-MS imaging data from TMAs in a classification context. For a generic reference to PCA see [20] and [12, Chapter 2 and Section 4.8].

Other feature reduction approaches used in proteomics include the application of univariate tests such as Wilcoxon rank-sum statistics or t -tests – see [21, 22]. For high-dimensional data [23] ranked the variables by applying t -tests simultaneously to all variables and analyzed the performance of this approach. We follow a related path, based on canonical correlation analysis (CCA), and select the “best” original variables in a supervised manner prior to classification. The selected variables do not require any interpretation, as they are simply the m/z bins. Our approach contains the approach of [23] as a special case, and does not require tuning parameters as regularization-type methods do.

In CCA the data are split into two sets of variables and the linear combinations of variables in each subset with the strongest correlation are determined. Typically, this correlation is much stronger than that of any individual pair of variables from the two subsets. In a regression context the correlation between a response and the predictor variables shows the relative strength of the response on the individual predictors. In a classification framework, we use class labels of observations as one subset of variables, and all other variables as the second subset. For our data, the class labels are the LNM status, and the variables are all m/z bins.

It is beyond the scope of this paper to review the different approaches mentioned above or to show detailed comparisons. However, because of the repeated use of PCA for MALDI-MS data, we include a performance comparison of our proposed method with that based on feature reduction with PCA. We note that our CCA-based variable selection is supervised, while PCA does not use the class labels in its feature reduction. As a consequence PCA may not find features that are optimal for predicting class membership. In addition, our suggested CCA-based approach simply ranks the original variables, here m/z bins, and picks the best, while PCA transforms the original variables and results in features that are linear combinations of all m/z bins. While we expect to relate the “best” CCA masses to biomarkers, no such interpretation is possible for the PCA features. In our results, shown in Figures 1 and 2, we will see the dramatic improvement in prediction that can be achieved with the supervised CCA-based variable selection compared to the unsupervised PCA-based dimension reduction.

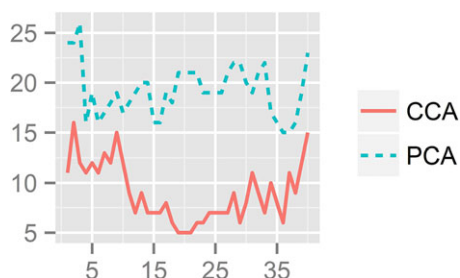


Figure 1. LOO-misclassification for log-intensity data, including zeroes for absent peaks when averaging and restricting to spectra from tumor regions. The x-axis shows the number of variables (m/z bins) 1 – 40 included in the CCA data, or the number of principle components (linear combinations of m/z bins), 1 – 40 included in the PCA data. For each dimension-reduced data, LDA is performed and the LOO-misclassification shown on the y-axis separately for CCA and PCA. The smallest misclassification of five for CCA-LDA occurs when the best 19 variables are selected. The smallest misclassification of 15 for PCA-LDA occurs when the PCA-dimension is 36. A comparison shows that CCA-LDA results in lower misclassification and a more parsimonious model in this case.

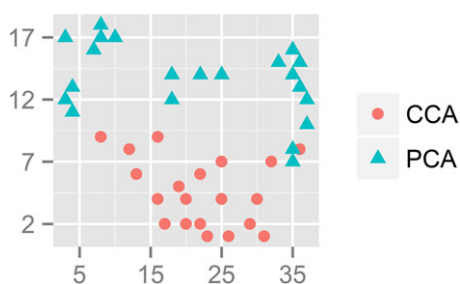


Figure 2. Best LOO-misclassification for each of the 22 combinations of data type, inclusion/exclusion of empty values (for non-binary data types), smoothing (for binary data types), and restriction to spectra from tumor annotated tissue. The x-axis shows the minimum number of variables for the CCA data, or the number of principle components of the PCA data, that results in the smallest LOO-misclassification in each scenario. The y-axis shows the number of misclassified patients for that scenario, with CCA in circles and PCA in triangles. For example the circle at $(x, y) = (23, 1)$ indicates that CCA-LDA results in a LOO misclassification of one when the best 23 CCA-selected variables are used in LDA for this particular scenario (binary data with a 0.25-smooth and restricting to spectra from regions annotated as tumor only). The figure clearly shows that CCA-LDA consistently achieves better LOO misclassification than PCA-LDA on these data.

In the context of MALDI-MS data, CCA has been used in finding correlations between datasets collected from the same tissue using different imaging techniques [24]. These authors use CCA in its standard form: finding correlations between two subsets of variables that belong to a natural partition — m/z values collected using one imaging technique, and m/z values collected using the second imaging technique. CCA has also been used in a classification context [25], although in an exploratory manner—plotting the strongest correlation

between their data and their prediction variable to observe a linear relationship.

In this paper, we also use the first and strongest correlation obtained from CCA, but we deviate from the conventional approach in that we explicitly use the information inherent in the first direction vector for variable selection. CCA-based variable selection as suggested by [26] is further described in [12, Sections 13.3.1 and 13.3.3], and references therein. We modify their approach for our data, essentially by using centered labels. Here, we briefly introduce our variable ranking approach, but further details, motivation, and comments can be found in the supporting information. Let X be a $d \times n$ centered data matrix whose columns represent observations (patients), and whose rows represent variables (m/z bins). Let y be a $1 \times n$ vector of (centered) class labels $\frac{-2n_+}{n}$ and $\frac{2n_-}{n}$ corresponding to the same n observations as the columns of X , where n_+ (n_-) is the number of observations in the positive (negative) class. The key idea is to exploit the relationship between X and y . Details describing how we use CCA are included in the supporting information, but essentially we will be interested in the vector ϕ , which satisfies

$$XX^T \phi = Xy^T. \quad (1)$$

The vector ϕ contains the weights for the linear combination of variables, or rows of X , which gives rise to the strongest absolute correlation with y . Variable reduction to k variables is obtained by selecting variables corresponding to the k highest entries of ϕ by absolute value.

In what follows, LDA refers to Fisher's linear classification as described in [12, Section 4.3]. LDA finds the direction that maximizes the between-class variance and simultaneously minimizes the within-class variance of the data. We will use the term CCA-LDA to refer to the process of first using the CCA variable selection to perform a dimension reduction step, and then apply LDA to the selected variables. Similarly, we will use the term PCA-LDA to refer to the process of first performing PCA dimension reduction, and then apply LDA to the selected variables. We measure the performance of these methods by leave-one-out (LOO) cross-validation. LOO cross-validation is an iterative process where in each iteration one observation is left out, the classification rule is trained on the remaining data, and tested on the left-out observation—assigning it a class label. This process is repeated until all observations have been assigned class labels in the “testing” step. Incorrectly assigned observations are counted as LOO misclassification. We use LOO cross-validation rather than a more traditional training/testing design (where separate datasets are used for the training and testing steps) because of the relatively small number of observations ($n = 43$).

We examined the effect of the parameters such as *regions of tissue* and *data form* as variations of the computational pipeline. In particular we consider the variations listed below and implement all possible combinations of these choices in order to determine which parameters are most important for classification.

- Data form: using raw intensities (I), log-intensities ($\log(I + 1)$), peak areas, signal-to-noise ratios, or binary presence/absence of peaks.
 - For nonbinary data forms, absence of peaks: When averaging spectra in which some m/z bins contains no peaks, the absence of peaks is included as a zero value, or ignored and hence not included in the average.
 - For binary data, spatial smoothing: The spatially smoothed binary data as described in [27] is used with smoothing parameters 0, 0.15, or 0.25 where 0 corresponds to no smoothing.
- Regions of tissue: all spectra, or spectra from annotated tumor regions only. Restricting to only annotated tumor regions represents a reduction in the total number of spectra from 123561 to 45877 (~ 37%) in the endometrial TMA data.

Our analyses showed that *data form* was the only factor that consistently affected classification performance, with log-intensity data performing best, and binary data typically performing second best.

Overall, the CCA-LDA outperformed PCA-LDA. A typical case can be seen in Figure 1 in which we compare our CCA-LDA with PCA-LDA using the same parameter settings—log-intensity data with zeroes included for absent peaks in the averaging step, and restricting to only annotated tumour regions. Figure 1 shows the LOO misclassification (out of 43 patients) of CCA-LDA and PCA-LDA on the y -axis against the number of variables or dimensions on the x -axis (used in the LDA step). In the case shown in Figure 1, CCA-LDA strictly outperforms PCA-LDA and it can be seen that CCA-LDA achieves its most parsimonious minimum misclassification of 5 by selecting 19 variables. Figure 2 shows the most parsimonious minimum misclassifications for each of the 22 combinations listed in the dot points above, when using PCA-LDA and CCA-LDA. The minimum misclassifications of CCA-LDA shown in Figure 2 vary between 1 (2.3%) and 9 (20.9%), while the minimum misclassifications of PCA-LDA vary between 7 (16.3%) and 18 (41.9%).

The improved performance of our CCA-based variable selection over the standard PCA dimension reduction, which uses linear combinations of variables, is likely due to the fact that PCA is a purely variance-based method, which ignores any information about the clinical outcome of interest (LNM). It is possible that the high-variance components of the data, which PCA selects, may not include the crucial classification information necessary in order to predict the clinical outcome accurately. CCA variable selection in contrast explicitly takes into account the clinical outcome when selecting the variables, and this could account for its improved performance in comparison to PCA.

Another advantage of CCA variable selection over that with PCA is the interpretability of the variable-reduced data—the

PCA variable-reduced data is produced by projection and so the new variables are linear combinations of the original variables, making them difficult to interpret. The variables selected via our CCA-based approach are a subset of the original variables and so preserve their interpretation as m/z bins. This interpretation is of use in designing follow-up biomarker studies. Some of the specific m/z bins that are highly ranked in the CCA-based variable selection are discussed in [3].

In conclusion, we have presented a variable selection method for MALDI-MS imaging data on TMAs, and demonstrated that this variable selection method combined with LDA achieves a LOO misclassification of between 2.3 and 20.9% for lymph node metastasis status in the endometrial cancer data of [3].

This research was carried out in collaboration with work supported by the Ovarian Cancer Research Foundation. The authors would also like to thank Professor Martin Oehler for access to the samples, and Associate Professor Gurjeet Kaur for her help with histopathological annotations.

The authors have declared no conflict of interest.

References

- [1] Casadonte, R., Caprioli, R. M., Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry. *Nat Protocols* 2011, 6, 1695–1709.
- [2] Steurer, S., Borkowski, C., Odinga, S., Buchholz, M. et al., MALDI mass spectrometric imaging based identification of clinically relevant signals in prostate cancer using large-scale tissue microarrays. *Int. J. Cancer* 2013, 133, 920–928.
- [3] Mittal, P., Klingler-Hoffmann, M., Arentz, G., Winderbaum, L. et al., MALDI imaging of primary endometrial cancers reveals proteins associated with lymph node metastasis. Accepted for publication in *Proteomics: Imaging Mass Spectrometry Special Issue*.
- [4] Australian Institute of Health and Welfare & Cancer Australia 2012. Gynaecological cancers in Australia: an overview. Cancer series no. 70. Cat. no. CAN 66. Canberra: AIHW.
- [5] Rungruang, B., Olawaiye, A. B., Comprehensive surgical staging for endometrial cancer. *Rev. Obstet. Gynecol.* 2012, 5, 28–34.
- [6] Jacques, S. M., Qureshi, F., Munkarah, A., Lawrence, W. D., Interinstitutional surgical pathology review in gynecologic oncology I. Cancer in endometrial curettages and biopsies. *Int. J. Gynecol. Pathol.* 1998, 17, 36–41.
- [7] Morrow, C. P., Bundy, B. N., Kurman, R. J. Creasman, W. T. et al., Relationship between surgical-pathological risk factors and outcome in clinical stage I and II carcinoma of the endometrium: a gynecologic oncology group study. *Gynecol. Oncol.* 1991, 40, 55–65.
- [8] Creasman, W. T., Odicino, F., Maisonneuve, P., Quinn, M. A. et al., Carcinoma of the corpus uteri. *Int. J. Gynecol. Obstet.* 2006, 95, S105–S143.
- [9] Todo, Y., Yamamoto, R., Minobe, S., Suzuki, Y. et al., Risk factors for postoperative lower-extremity lymphedema in

- endometrial cancer survivors who had treatment including lymphadenectomy. *Gynecol. Oncol.* 2010, *119*, 60–64.
- [10] Gustafsson, J. O. R., Oehler, M. K., McColl, S. R., Peter, H., Citric acid antigen retrieval (CAAR) for tryptic peptide imaging directly on archived formalin-fixed paraffin-embedded tissue. *J. Proteome Res.* 2010, *9*, 4315–4328.
- [11] Gustafsson, J. O. R., Eddes, J. S., Meding, S., Koudelka, T. et al., Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS. *J. Proteomics* 2012, *75*, 5093–5105. Special Issue: Imaging Mass Spectrometry: A Users Guide to a New Technique for Biological and Biomedical Research.
- [12] Koch, I., *Analysis of Multivariate and High-Dimensional Data*, volume 32 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge 2013.
- [13] Marron, J. S., Todd, M. J., Ahn, J., Distance-weighted discrimination. *Am. Stat. Assoc.* 2007, *102*, 1267–1271.
- [14] Mika, S., Rätsch, G., Weston, J., Schölkopf, B., Müller, K.-R., Fisher discriminant analysis with kernels. *Neural Networks Signal Proc. IX* 1999, *1*, 1.
- [15] Xu, P., Brock, G. N., Parrish, R. S., Modified linear discriminant analysis approaches for classification of high-dimensional microarray data. *Comput. Stat. Data Anal.* 2009, *53*, 1674–1687.
- [16] Vincent, M., Hansen, N. R., Sparse group lasso and high dimensional multinomial classification. *Comput. Stat. Data Anal.* 2014, *71*, 771–786.
- [17] McCombie, G., Staab, D., Stoeckli, M., Knochenmuss, R., Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal. Chem.* 2005, *77*, 6118–6124.
- [18] Deininger, S.-O., Ebert, M. P., Marc, G., Arne, F., Rocken, C., MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *J. Proteome Res.* 2008, *7*, 5230–5236. PMID: 19367705.
- [19] Mascini, N. E., Eijkel, G. B., Brugge, P. ter, Jonkers, J. et al., The use of mass spectrometry imaging to predict treatment response of patient-derived xenograft models of triple-negative breast cancer. *J. Proteome Res.* 2015, *14*, 1069–1075.
- [20] Jolliffe, I., *Principal Component Analysis*. John Wiley & Sons, Ltd 2002.
- [21] Rauser, S., Marquardt, C., Balluff, B., Deininger, S.-O. et al., Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. *J. Proteome Res.* 2010, *9*, 1854–1863.
- [22] Groseclose, M. R., Massion, P. P., Chaurand, P., Caprioli, R. M., High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. *Proteomics* 2008, *8*, 3715–3724.
- [23] Fan, J., Fan, Y., High-dimensional classification using features annealed independence rules. *Annal. Stat.* 2008, *36*, 2605–2637.
- [24] Eijkel, G. B., Kükler, K. B., Van Der Wiel, I. M., Kros, J. M., Luider, T. M., Heeren, R. M. A., Correlating MALDI and SIMS imaging mass spectrometric datasets of biological tissue surfaces. *Surf. Interface Anal.* 2009, *41*, 675–685.
- [25] Nicolaou, N., Xu, Y., Goodacre, R., Detection and quantification of bacterial spoilage in milk and pork meat using MALDI-TOF-MS and multivariate analysis. *Anal. Chem.* 2012, *84*, 5951–5958.
- [26] Koch, I., Naito, K., Prediction of multivariate responses with a selected number of principal components. *Comput. Stat. Data Anal.* 2010, *54*, 1791–1807.
- [27] Winderbaum, L. J., Koch, I., Gustafsson, O. J. R., Meding, S. et al., Feature extraction for proteomics imaging mass spectrometry data. *Annal. Appl. Stat.* 2015, *9*, 1973–1996.

PROTEOMICS

Supporting Information for Proteomics

DOI 10.1002/pmic.201500451

Lyron Winderbaum, Inge Koch, Parul Mittal and Peter Hoffmann

**Classification of MALDI-MS imaging data of tissue microarrays using
canonical correlation analysis-based variable selection**

PROTEOMICS

Supporting Information for Proteomics

DOI 10.1002/pmic.201500451

Lyron Winderbaum, Inge Koch, Parul Mittal and Peter Hoffmann

**Classification of MALDI-MS imaging data of tissue microarrays using
canonical correlation analysis-based variable selection**

SUPPORTING INFORMATION FOR

Classification of MALDI-MS imaging data of tissue microarrays using Canonical Correlation Analysis based variable selection

Lyron Winderbaum, Inge Koch, Parul Mittal and Peter Hoffmann

This supporting information includes three sections, brief summaries of which follow.

- **Binning.** In this section we give more details about the binning we use, as well as details describing how and why we use shifted-bin analyses.
- **Canonical Correlation Analysis (CCA).** In this section we provide more detail, background, and motivation for CCA.
- **Replication of Results.** In this section we briefly introduce a different dataset, from a Tissue Microarray (TMA) study of vulvar cancer. We then reproduce the analyses carried out in the main paper on this dataset, thereby demonstrating the robustness of these results.

Binning

We construct 0.25Da wide m/z bins centred at integer multiples of 0.25 and use these bins to identify peaks from different spectra that have similar m/z and so are likely to originate from the same ion species. These bins will represent variables in our data to follow. We choose to place the bins at arbitrary locations rather than centred them around peaks in the data as this allows the binning to remain data-independent. The binning being data-independent means it can be applied to new data simply and this is important in classification as the interest is to apply the rule to new data and the new data needs to have the same variables as the training data. The fact the locations of the bins are arbitrary means that multiple peaks from the same ion species will sometimes occur in different bins and this is a problem. In order to address this problem, we replicate all our analyses but with bin locations shifted by $-\frac{0.25}{3}$ and $+\frac{0.25}{3}$. Having three parallel shifted-bin analyses total means that any given ion species will have all the peaks in the same bin for at least two of these three shifted-bin analyses almost always because the mass error typical in MALDI-MS imaging is much less than 0.25, particular with internal calibration [1]. We apply classification methods to each of the three shifted-bin analyses in parallel and assign the class label reached by majority in each case.

Canonical Correlation Analysis (CCA)

In CCA the data are split into two sets of variables and the linear combinations of variables in each subset with the strongest correlation are determined. CCA can be of particular interest when there is a natural partitioning of the variables by context, as has been demonstrated in the context of genomic data for which multiple sets of measurements from the same samples are available [2]. Typically the strongest correlation between linear combinations of the two subsets of variables is much stronger than that of any individual pair of variables from the two subsets. In a regression context the correlation between a response and the predictor variables shows the relative strength of the response on the individual predictors. In a classification framework, we use class labels of observations as one subset of variables, and all other variables as the second subset. For our data, the class labels are the LNM status, and the variables are all m/z bins.

The following is based on [3, Chapter 3], and details can be found there.

Let X and Y be the two sets of variables based on the same n observations, and for convenience of notation, assume that both X and Y are centred, so each row has mean zero. The crucial

information is contained in the cross-covariance matrix S_{XY} of X and Y and the corresponding correlation matrix C which are given by

$$S_{XY} = \frac{1}{n-1}XY^T \quad \text{and} \quad C = S_X^{-\frac{1}{2}}S_{XY}S_Y^{-\frac{1}{2}}, \quad (1)$$

where S_X and S_Y are the covariance matrices of X and Y respectively. The matrices S_{XY} and C are of size $d_X \times d_Y$, where d_X and d_Y are the number of variables in X and Y respectively, and satisfy

$$C^T \mathbf{q}_i = \lambda_i \mathbf{p}_i \quad \text{for} \quad i \leq \min\{d_X, d_Y\} \quad (2)$$

see [3, (3.5)]. Here $\mathbf{p}_i$ and $\mathbf{q}_i$ are the left and right eigenvectors of C , λ_i are the singular values, normally ordered starting from the largest one, λ_1 . This first value λ_1 equals the highest absolute covariance over linear combinations of the X -variables and the Y -variables. Furthermore, the d_X and d_Y dimensional vectors $\phi_1 = S_X^{-1/2} \mathbf{p}_1$ and $\psi_1 = S_Y^{-1/2} \mathbf{q}_1$, corresponding to λ_1 , contain the weights of the linear combinations applied to the X -variables and Y -variables respectively which result in the highest covariance λ_1 . For details see [3, Section 3.4 and specifically Theorem 3.10]. Of particular interest in the variable ranking and selection will be the vector ϕ_1 , as it tells us which X -variables are most important in determining the strongest correlation with the Y -variables.

In our context, X is the matrix whose rows correspond to m/z bins and columns correspond to patients, as described on page 5 of the manuscript, and Y is the row vector of class labels also described on page 5 of the manuscript. As Y is a row vector in this context, its covariance is a positive number, S_{XY} is a $d \times 1$ vector, and thus C in Equation (1) also reduces to a $d \times 1$ vector. We write $|C|$ for the length of C and the vector ϕ_1 is then given by

$$\phi_1 = S_X^{-\frac{1}{2}} \mathbf{p}_1 = S_X^{-\frac{1}{2}} \left(\frac{1}{|C|} C \right) = \frac{1}{|C|} S_X^{-\frac{1}{2}} \left(S_X^{-\frac{1}{2}} S_{XY} S_Y^{-\frac{1}{2}} \right) \propto S_X^{-1} S_{XY} = (X X^T)^{-1} X Y^T. \quad (3)$$

If S_X is non-invertible, then we replace its inverse by the Moore-Penrose pseudoinverse, see [4]. We will use the entries of ϕ_1 to rank the variables of X , or more precisely, the entries of $\phi = S_X^{-1} S_{XY}$ which is proportional to ϕ_1 , as the ranking is independent of the length of the vector. We sort the entries of ϕ by their decreasing absolute value, and denote this decreasing sequence of d values $\varphi_{(1)} \geq \varphi_{(2)} \geq \dots \varphi_{(d)}$. Using the ordering inherited from ϕ , we rank the variables of X , so the highest ranked variable corresponds to $\varphi_{(1)}$, the second highest to $\varphi_{(2)}$ etc.

The number of variables to select, k , can be chosen to optimise classification performance as described below. For details see [3, Sections 4.3.1 and 4.4.2].

1. Fix the number, k , of variables. Let X_k be the $k \times n$ matrix which consist of the first k ranked variables of X , with the ranking inherited from the $\varphi_{(i)}$ s.
2. Train and test your classification rule on X_k , and record the number of misclassified observations. We do this using LOO-CV.
3. If $k < d$ increase k to $k + 1$ and repeat Steps 1-3.

Figures 1 and S1 plot this misclassification vs the number of variables selected k . We choose the ‘best’ number of variables k as that which achieves the lowest misclassification, and if multiple values of k achieve equal-best misclassification, we choose the smallest of them. Figures 2 and S2 show these ‘best’ choices for each combination of processing parameters such as data form and regions of tissue.

Replication of Results

Here we apply the methods from the paper on a second, independent, dataset relating to vulvar cancer in order to demonstrate the robustness of our claims. Similarly to the endometrial data, the sample preparation of the vulvar TMAs included citric acid antigen retrieval [5], trypsin digestion, addition of internal calibrants [1], and matrix deposition. Also similarly to the endometrial data, peak-picking was performed as part of the data acquisition through proprietary software (flexControl v3.0.1 and flexImaging v4.0.1, Bruker Daltonik, <http://www.bruker.com>).

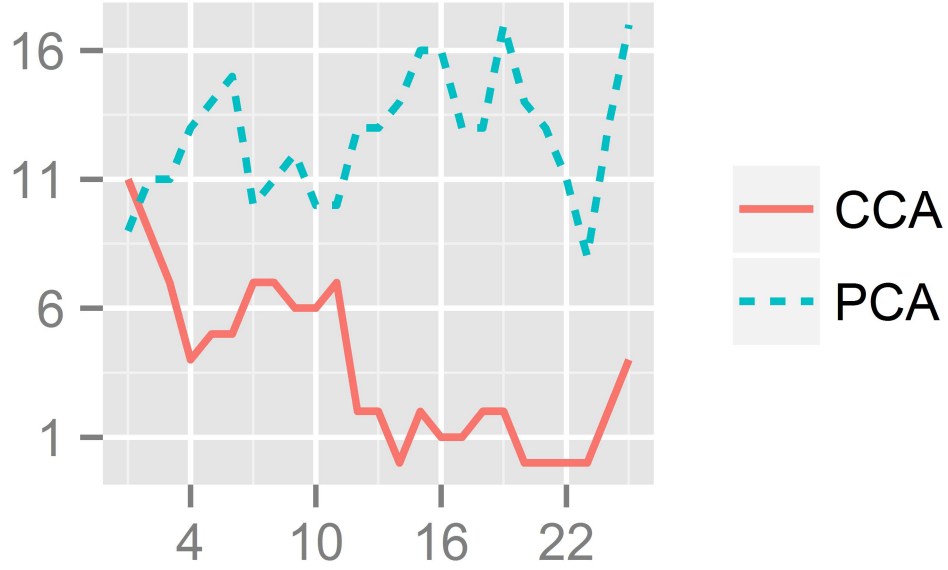


Figure S1: Analogous to Figure 1, but for the vulvar cancer. LOO-misclassification for log-intensity data, including zeroes for absent peaks when averaging and restricting to spectra from tumour regions. The x -axis shows the number of variables (m/z bins) 1 to 25 included in the CCA data, or the number of principle components (linear combinations of m/z bins), 1 to 25 included in the PCA data. For each dimension-reduced data, LDA is performed and the LOO-misclassification shown on the y -axis separately for CCA and PCA. The smallest misclassification of 0 for CCA-LDA occurs when the best 14 variables are selected. The smallest misclassification of 8 for PCA-LDA occurs when the PCA-dimension is 23. A comparison shows that CCA-LDA results in lower misclassification and a more parsimonious model in this case – the same trend as is shown in Figure 1 for the endometrial data.

The vulvar cancer data represent 28 patients, 16 metastasis positive and 12 metastasis negative. Figures S1 and S2 mirror the results of Figures 1 and 2 for the vulvar data. These figures show that the vulvar data reflects the same trends as the endometrial data – CCA variable selection outperforming PCA dimension reduction. Also similarly to the endometrial data, the log-intensity data form seems to achieve the best classification results in the vulvar data, with the binary data also showing some good results. None of the other variations considered seem to show a consistent effect on classification performance in the vulvar data. Overall, the results on the vulvar data agree well with the results on the endometrial data as far as which methods seem to achieve the best classification results.

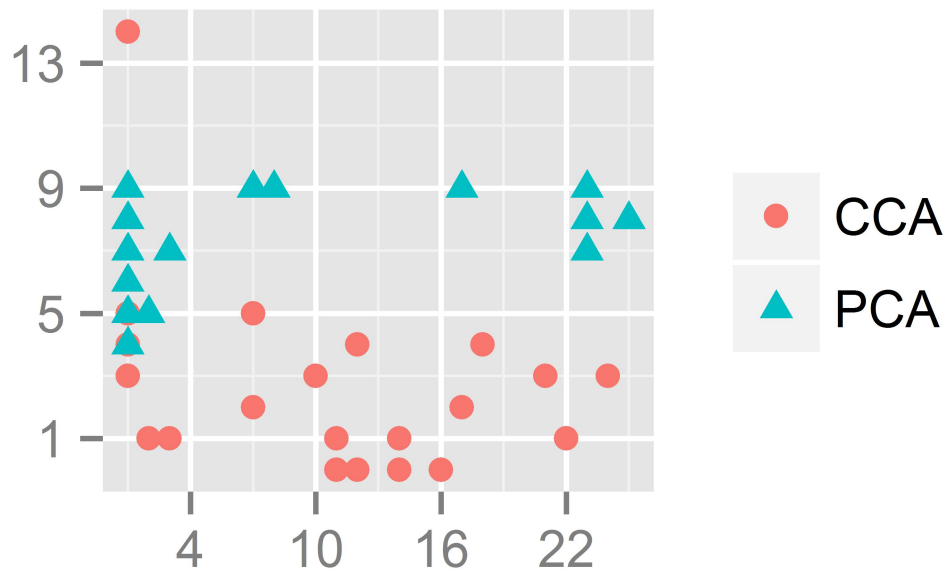


Figure S2: Analogous to Figure 2, but for the vulvar cancer. Best LOO-misclassification for each of the 22 combinations of data type, inclusion/ exclusion of empty values (for non-binary data types), smoothing (for binary data types), and restriction to spectra from tumour annotated tissue. The x -axis shows the minimum number of variables for the CCA data, or the number of principle components of the PCA data, that results in the smallest LOO-misclassification in each scenario. The y -axis shows the number of misclassified patients for that scenario, with CCA in circles and PCA in triangles. The figure clearly shows that CCA-LDA consistently achieves better LOO misclassification than PCA-LDA on these data – the same trend as is shown in Figure 2 for the endometrial data.

Bibliography

- [1] Johan O.R. Gustafsson, James S. Eddes, Stephan Meding, Tomas Koudelka, Martin K. Oehler, Shaun R. McColl, and Peter Hoffmann. Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS. *Journal of Proteomics*, 75(16):5093 – 5105, 2012. Special Issue: Imaging Mass Spectrometry: A Users Guide to a New Technique for Biological and Biomedical Research.
- [2] Daniela M Witten and Robert J Tibshirani. Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical applications in genetics and molecular biology*, 8(1):1–27, 2009.
- [3] Inge Koch. *Analysis of Multivariate and High-Dimensional Data*, volume 32 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2013.
- [4] Adi Ben-Israel and Thomas NE Greville. *Generalized inverses: theory and applications*, volume 15. Springer Science & Business Media, 2003.
- [5] Johan O. R. Gustafsson, Martin K. Oehler, Shaun R. McColl, and Peter Hoffmann. Citric acid antigen retrieval (CAAR) for tryptic peptide imaging directly on archived formalin-fixed paraffin-embedded tissue. *Journal of Proteome Research*, 9(9):4315–4328, July 2010.

MALDI Mass Spectrometry Imaging Reveals
Decreased CK5 Levels in Vulvar Squamous Cell
Carcinomas Compared to the Precursor Lesion
Differentiated Vulvar Intraepithelial Neoplasia

This page intentionally left blank



Article

MALDI Mass Spectrometry Imaging Reveals Decreased CK5 Levels in Vulvar Squamous Cell Carcinomas Compared to the Precursor Lesion Differentiated Vulvar Intraepithelial Neoplasia

Chao Zhang ^{1,2,†}, Georgia Arentz ^{1,2,†}, Lyron Winderbaum ^{1,†}, Noor A. Lokman ^{1,3},
Manuela Klingler-Hoffmann ^{1,2}, Parul Mittal ^{1,2}, Christopher Carter ⁴, Martin K. Oehler ^{3,5}
and Peter Hoffmann ^{1,2,*}

¹ Adelaide Proteomics Centre, School of Biological Sciences, The University of Adelaide, Adelaide 5005, Australia; chao.zhang01@student.adelaide.edu.au (C.Z.); georgia.arentz@adelaide.edu.au (G.A.); lyron.winderbaum@student.adelaide.edu.au (L.W.); noor.lokman@adelaide.edu.au (N.A.L.); manuela.klinglerhoffmann@adelaide.edu.au (M.K.-H.); parul.mittal@adelaide.edu.au (P.M.)

² Institute for Photonics and Advanced Sensing (IPAS), The University of Adelaide, Adelaide 5005, Australia

³ Discipline of Obstetrics and Gynaecology, School of Medicine, Research Centre for Reproductive Health, Robinson Institute, Adelaide 5005, Australia; martin.oehler@adelaide.edu.au

⁴ Department of Cytopathology, SA Pathology, Adelaide 5005, Australia; Christopher.Carter@sa.gov.au

⁵ Department of Gynaecological Oncology, Royal Adelaide Hospital, Adelaide 5005, Australia

* Correspondence: Peter.hoffmann@adelaide.edu.au; Tel.: +61-08-8313-4903; Fax: +61-08-0-8313-4362

† These authors contributed equally to this work.

Academic Editors: Jamal Zweit and Sundaresan Gobalakrishnan

Received: 10 May 2016; Accepted: 30 June 2016; Published: 8 July 2016

Abstract: Vulvar cancer is the fourth most common gynecological cancer worldwide. However, limited studies have been completed on the molecular characterization of vulvar squamous cell carcinoma resulting in a poor understanding of the disease initiation and progression. Analysis and early detection of the precursor lesion of HPV-independent vulvar squamous cell carcinoma (VSCC), differentiated vulvar intraepithelial neoplasia (dVIN), is of great importance given dVIN lesions have a high level of malignant potential. Here we present an examination of adjacent normal vulvar epithelium, dVIN, and VSCC from six patients by peptide Matrix-assisted laser desorption/ionization Mass Spectrometry Imaging (MALDI-MSI). The results reveal the differential expression of multiple peptides from the protein cytokeratin 5 (CK5) across the three vulvar tissue types. The difference observed in the relative abundance of CK5 by MALDI-MSI between the healthy epithelium, dVIN, and VSCC was further analyzed by immunohistochemistry (IHC) in tissue from eight VSCC patients. A decrease in CK5 immunostaining was observed in the VSCC compared to the healthy epithelium and dVIN. These results provide an insight into the molecular fingerprint of the vulvar intraepithelial neoplasia that appears to be more closely related to the healthy epithelium than the VSCC.

Keywords: Matrix-assisted laser desorption/ionization (MALDI); imaging mass spectrometry (MSI); differentiated vulvar intraepithelial neoplasia (dVIN); vulva squamous cell carcinomas (VSCC); Cytokeratin 5 (CK5)

1. Introduction

Vulvar cancer is the fourth most common gynecological cancer worldwide, constituting 5% of all gynecological cancers and 0.6% of all cancer cases in women [1]. In 2015, 5150 new cases were diagnosed in the United States with 1080 recorded deaths [1]. The majority of vulvar cancers are squamous cell carcinomas (VSCC), of which there are two etiological pathways: one linked to

human papilloma virus (HPV) infection, and the other HPV-independent with genetic alterations. The HPV related and non-HPV related forms of vulvar cancer have distinct precursor lesions, known as usual-type vulvar intraepithelial neoplasia and differentiated vulvar intraepithelial neoplasia, respectively. Mutations in the tumor suppressor gene TP16 are associated with HPV related VSCC, while mutations in TP53 have been shown to promote the development of non-HPV related VSCC from the inflammatory condition lichen sclerosis [2].

Recently, the incidence rates of non-HPV related invasive vulvar cancer and its precursor lesion, differentiated vulvar intraepithelial neoplasia (dVIN), have significantly increased amongst young women [3,4]. In Australia, the incidence rates of the disease are extremely high among Aboriginal women from Arnhem Land in the Northern Territory. Studies conducted within this cohort have found no relation between the disease and HPV infection or any underlying genomic irregularities that differentiate this group from other VSCC patients [3,4].

For many years the clinicopathological significance of vulvar intraepithelial neoplasia (VIN) as a precursor to vulvar squamous cell carcinoma (VSCC) was questioned [5] due to the low detection rates of dVIN as a solitary lesion [6]. However, in recent times, dVIN has been confirmed as the direct precursor lesion to HPV-independent vulvar carcinoma, which in the past may have been frequently misdiagnosed as lichen sclerosis or other forms of dermatoses [7,8]. Such misdiagnoses were in part due to numerous changes over the last few decades with regards to the terminology and criteria used to differentiate lichen sclerosis from dVIN and other inflammatory conditions such as squamous cell hyperplasia and lichen simplex chronicus [7,9,10].

The detection of dVIN as a solitary lesion is of great importance given the lesions have a high level of malignant potential [6], with studies showing the median time between the diagnoses of dVIN to VSCC development around 23 to 28 months [6,7]. Mutation of TP53 is a known early event in the development of dVIN [11] and the detection of dVIN adjacent to VSCC is high at around 80%. However, the detection of dVIN as a solitary lesion is scarce [6] and can be difficult to diagnose due to the subtle histological appearance of the lesion [12,13]. Currently the characterization of the dVIN at the molecular level has been poor due to sample rarity, hence the understanding of VSCC progression from healthy tissue to the precursor lesion is low compared to other well studied carcinomas.

MALDI-MSI allows the in situ analysis of tissue sections combining classical mass spectrometry with histological tissue analysis [14]. MALDI-MSI can identify peptides and other molecules from samples whilst retaining the important spatial information of the tissue. Previous studies in the gastrointestinal tract, lung, brain and other gynaecological cancers have shown that MALDI-MSI can acquire a more comprehensive proteomic picture than immunohistochemical testing [14–17]. Here we present an examination of adjacent normal vulvar epithelium, dVIN, and VSCC from six patients by peptide MALDI-MSI. All patients included in this study were HPV negative and lichen sclerosis positive. The results revealed the differential detection of multiple peptides from the protein cytokeratin 5 (CK5) across the three tissue types. The difference observed in the relative abundance of CK5 by MALDI-MSI between the healthy epithelium, dVIN, and VSCC was further investigated by IHC in eight tissues from VSCC patients. This revealed a decrease in CK5 immunostaining in the VSCC as compared to the healthy epithelium and dVIN.

2. Results

2.1. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging (MALDI-MSI)

MALDI-MSI was carried out on tissue sections containing regions of healthy vulvar epithelium, dVIN, and VSCC from six late stage carcinoma patients. Peak groups, representing the imaged peptides across all of the analyzed tissue sections with the same m/z (assuming a mass accuracy of ± 0.02 Da), were detected using density-based clustering of peaks (DBSCAN* [18] with an epsilon of 0.02 Da and a minimum of 100 points) and the stringent criterion of requiring peak groups to contain a minimum of 10,000 peaks. This stringent criterion of a minimum of 10,000 peaks per peak group was initially

put in place to ensure that the peak groups analyzed could be reproducibly detected across a majority of spectra. With this criteria in place, 31 peak groups were detected (Table 1). When the minimum number of peaks per peak group was lowered to 1000, 316 peak groups were detected. For each peak group the abundance weighted mean (AWM) was calculated, representing the apex of the peak group and the m/z of the most intense peptide within the group. The widths of the 31 detected peak groups ranged from ± 0.075 to ± 0.9 Da. The maximum difference, d , in median log intensity between the three tissue regions was calculated for each of the 31 peak groups of interest across the patient cohort (Table 1). Of the 31 peak groups, 19 were found to have a difference in fold change intensity of ≥ 1.4 -fold across the three tissue types.

Table 1. Matrix-assisted laser desorption/ionization mass spectrometry imaging peak groups ranked heuristically by the largest difference in median log intensity between the healthy epithelium, differentiated vulvar intraepithelial neoplasia, and vulvar squamous cell carcinoma.

Peptide AWM [M + H] of the Peak Group ¹	Da Range of the Peak Group	No. Spectra in Peak Group ²	Maximum Difference Median Log Intensity ³
1198.71	± 0.62	23,099	0.75
1028.62	± 0.56	15,354	0.66
1410.72 *	± 0.82	18,784	0.58
1905.95	± 0.52	10,956	0.56
1095.58	± 0.82	21,704	0.55
944.54	± 0.69	19,354	0.53
957.63	± 0.65	10,464	0.52
857.60	± 0.68	14,956	0.50
976.51	± 0.65	15,534	0.49
878.56	± 0.68	11,213	0.46
810.48 *	± 0.76	17,095	0.45
958.59	± 0.69	15,432	0.42
943.62	± 0.82	20,030	0.41
865.42 *	± 0.84	15,617	0.41
1217.66	± 0.55	10,906	0.39
884.52 *	± 0.66	14,337	0.37
856.67	± 0.74	15,483	0.36
1143.65 *	± 0.64	17,685	0.36
1045.62	± 0.65	16,370	0.35
2147.19	± 0.14	20,418	0.33
1105.59	± 0.50	11,479	0.30
1296.68	± 0.10	22,823	0.29
870.64	± 0.85	22,749	0.29
842.60	± 0.90	25,210	0.22
1267.65	± 0.32	11,686	0.20
2932.59	± 0.07	15,500	0.19
816.50	± 0.73	10,983	0.19
971.56 *	± 0.65	15,998	0.18
1111.58	± 0.51	13,473	0.09
1570.70	± 0.17	12,029	0.05
864.63	± 0.70	11,114	0.05

¹ The abundance weighted means (AWM) m/z was calculated for the overall peak group; ² Number of Spectra in each peak group compiled from all acquisition spectra across the patient cohort; ³ Peak groups were heuristically ranked based on the maximum difference, d , in median log intensity between the tissue regions of interest;

* CK5 peptides.

2.2. Cytokeratin 5 (CK5) Identified as a Protein of Interest

In order to gain peptide identifications for the MALDI-MSI peak groups of interest nanoflow liquid chromatography tandem mass spectrometry (nano-LC-MS/MS) was performed on laser microdissected regions of the healthy epithelium, dVIN, and VSCC. Matching between the MALDI-MSI peak groups and nano-LC-MS/MS data was done by aligning the experimental m/z values of the

sequenced peptides that fell between the minimum and maximum m/z of each of the MALDI-MSI peak groups of interest. A table containing all of the matching results for the 31 peak groups of interest is provided in Table S2. Of the 31 MALDI-MSI peak groups of interest, six matched to sequenced peptides from Cytokeratin 5 (CK5) (Table 2), hence CK5 was selected for further analysis. One of the peak groups (AWM [M + H] of 1410.72) matched to two unique CK5 peptides that share the same mass to within 0.009 Da, SFSTASAITPSVSR (1409.7203) and TTAENEFVMLKK (1409.7295). A MALDI-MSI annotated ion intensity map for the peak group m/z 1410.67, which matches to the CK5 peptides SFSTASAITPSVSR and TTAENEFVMLKK (both with a m/z 1410.67), is shown in Figure 1. With regards to the remaining results in Table 1, only three other proteins were detected with more than one unique peptide match as shown in Table S2. Four peptides were detected from the protein Neuroblast differentiation-associated protein AHNK, three peptides were detected from the protein Annexin A4, and two peptides were detected from the protein Nicotinate phosphoribosyltransferase.

Table 2. CK5 matrix-assisted laser desorption/ionization mass spectrometry imaging peak groups of interest with matched nanoflow liquid chromatography tandem mass spectrometry results.

Peptide AWM [M + H] ¹	d ²	LC-MS/MS m/z ³	LC-MS/MS MR ⁴	Peptide Score ⁵	Peptide Significance ⁶	Peptide Sequence ⁷
810.48	0.45	405.7087	809.4028	21.44	0.0072	QSSVSFR
865.42	0.41	433.1998	864.3851	26	0.0025	SGGGGGGGFGR
884.52	0.37	442.7286	883.4427	24.74	0.0034	TSFTSVSR
971.56	0.18	486.2429	970.4712	25.1	0.0031	FVSTSSSR
1143.65	0.36	572.3175	1142.6205	39.59	0.00011	LAELEEALQK
1410.72	0.58	705.8674	1409.7203	115.28	3×10^{-12}	SFSTASAITPSVSR
1410.72	0.58	705.872	1409.7295	150.55	8.8×10^{-16}	TTAENEFVMLKK

¹ The peptide abundance weighted mean (AWM) m/z calculated for each peak group; ² Heuristically ranked maximum difference, d , between the healthy epithelium, dVIN, and VSCC; ³ Experimental m/z of the CK5 matched nano-LC-MS/MS detected peptide; ⁴ Experimental mass of the CK5 matched nano-LC-MS/MS detected peptide; ⁵ Mascot Ions Score; ⁶ Mascot Expected Score; ⁷ CK5 matched peptide sequence.

2.3. Immunohistochemistry (IHC) Analysis of CK5 across the Healthy Epithelium, Differentiated Vulvar Intraepithelial Neoplasia and Vulvar Squamous Cell Carcinoma

The differential detection of CK5 across the three tissue types as determined by MALDI-MSI was further investigated by IHC in tissue from eight VSCC patients. Of these eight tissues, three had been previously analyzed in the MALDI-MSI experiments. IHC analysis of CK5 revealed a significant difference in the level of staining between the dVIN (mean 91.6 ± 6.1) and VSCC (mean 75.4 ± 10.7) tissues ($p = 1.758 \times 10^{-7}$). A similar result was observed for the healthy as compared to VSCC tissue ($p = 0.004$), with increased staining observed in the healthy epithelium (mean 86.7 ± 14.8). Overall, there was no significant difference in the level of staining between the dVIN and healthy epithelium ($p = 0.14$), although the mean staining intensity in the dVIN was slightly higher. Results for the IHC analysis and CK5 staining from two representative vulvar tissues are provided in Figure 2. A direct comparison of the CK5 peptides SFSTASAITPSVSR and TTAENEFVMLKK, both m/z 1410.72, as detected by MALDI-MSI to the detection of CK5 by IHC on consecutive vulvar tissue sections is shown in Figure 3.

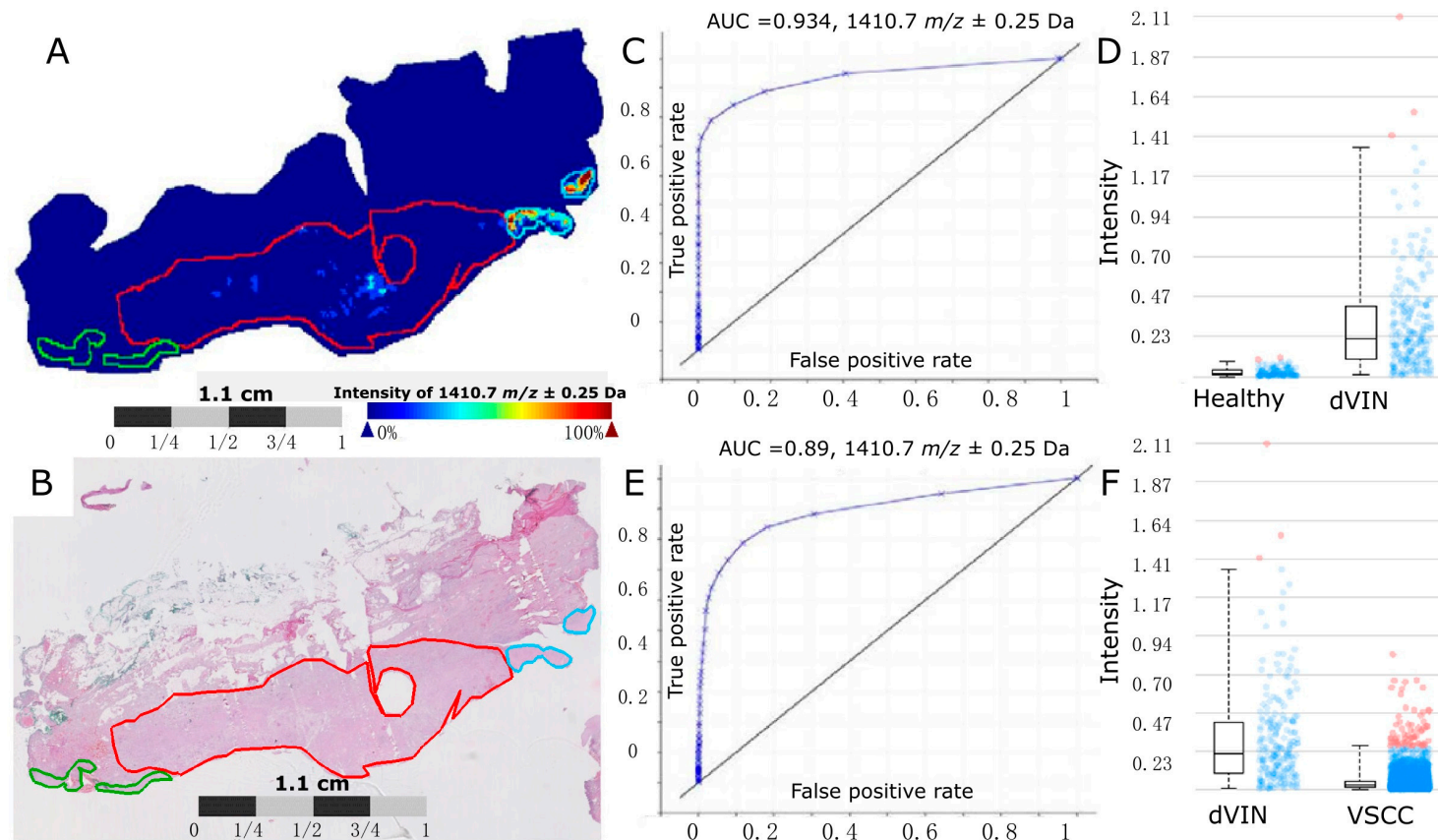


Figure 1. A MALDI-MSI annotated ion intensity map from a representative vulvar tissue section for the peak group m/z 1410.67. (A) An ion intensity map for m/z 1410.67, which matches to the CK5 peptides SFSTASAITPSVSR and TTAENEFVMLKK (both with a $[M + H]^+$ m/z 1410.67). The regions of healthy epithelium are outlined in green, the dVIN is outlined in blue, and the VSCC is outlined in red; (B) H & E stain of Healthy (green), dVIN (blue) and VSCC (red) tissue post MALDI-MSI data acquisition; (C) A ROC curve and (D) boxplot analysis of the MALDI-MSI data for m/z 1410.67 comparing the annotated healthy epithelium to the dVIN region, AUC = 0.934; (E) A ROC curve and (F) boxplot analysis of the MALDI-MSI data for m/z 1410.67 comparing the annotated dVIN to the VSCC region, AUC = 0.89. Blue dots represent the m/z 1410.67 spectra intensity interval between the lower and upper quartiles. Red dots represent the spectra intensities outside of the intervals.

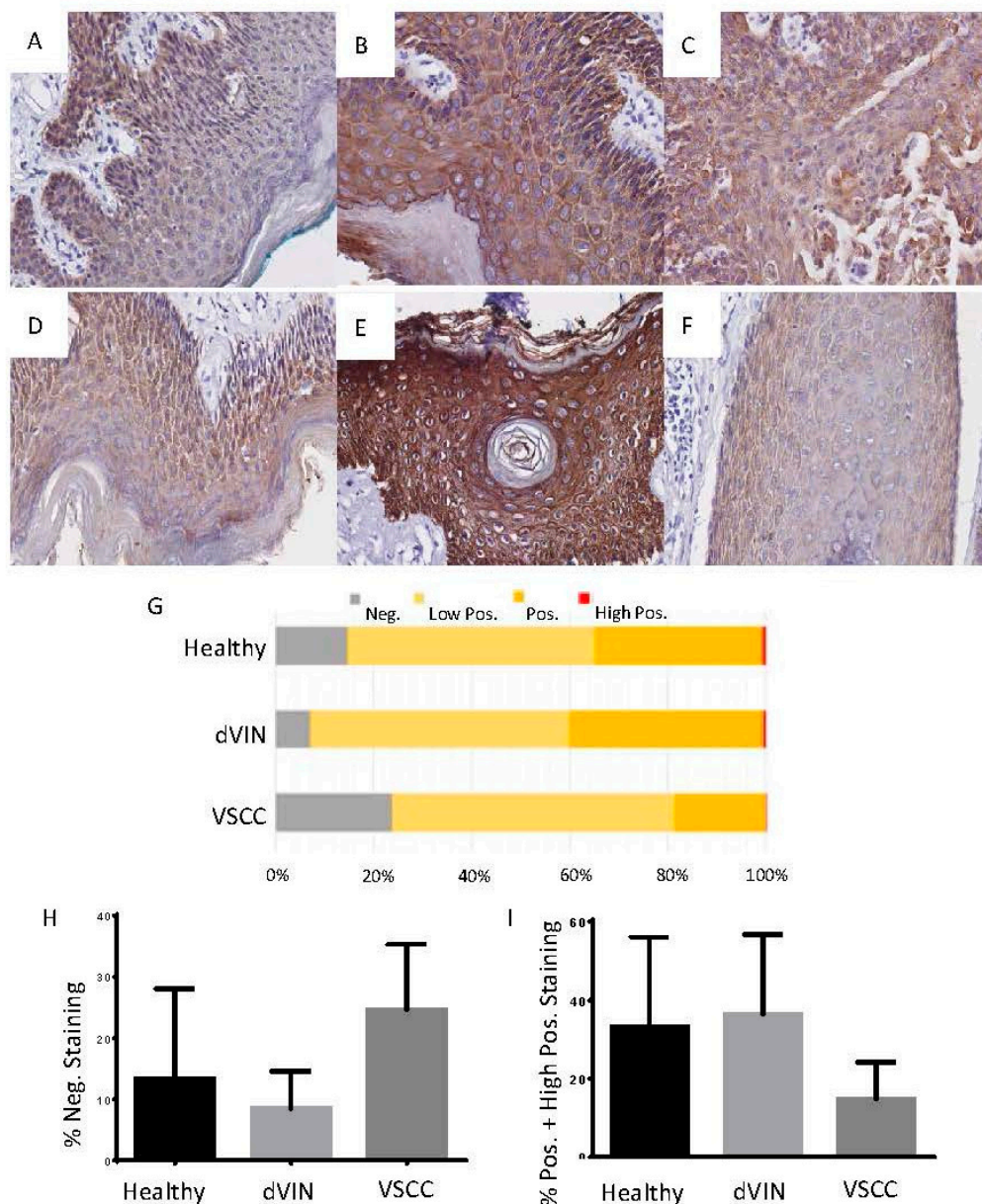


Figure 2. IHC analysis of CK5 on vulvar tissues containing regions of healthy epithelium, dVIN, and VSCC, all at 40× magnification. (A) CK5 staining of healthy epithelium; (B) dVIN; (C) VSCC from a single section; (D) CK5 staining of healthy epithelium; (E) dVIN; (F) VSCC from a single section of a different patient; (G) Quantitative analysis of CK5 staining across the tissue types was performed using IHC profiler-Image J. For each tissue section, three representative photo-micrographic images at 40× magnification were used and each image was assigned a staining score of negative, low positive, positive, and high positive; (H) The percent of negative staining was significantly different between the dVIN and the VSCC ($p < 0.0001$), but not between the dVIN and healthy epithelium ($p = 0.11$); (I) The percent of positive and high positive staining was significantly different between the dVIN and the VSCC ($p < 0.0001$), but not between the dVIN and healthy epithelium ($p = 0.59$).

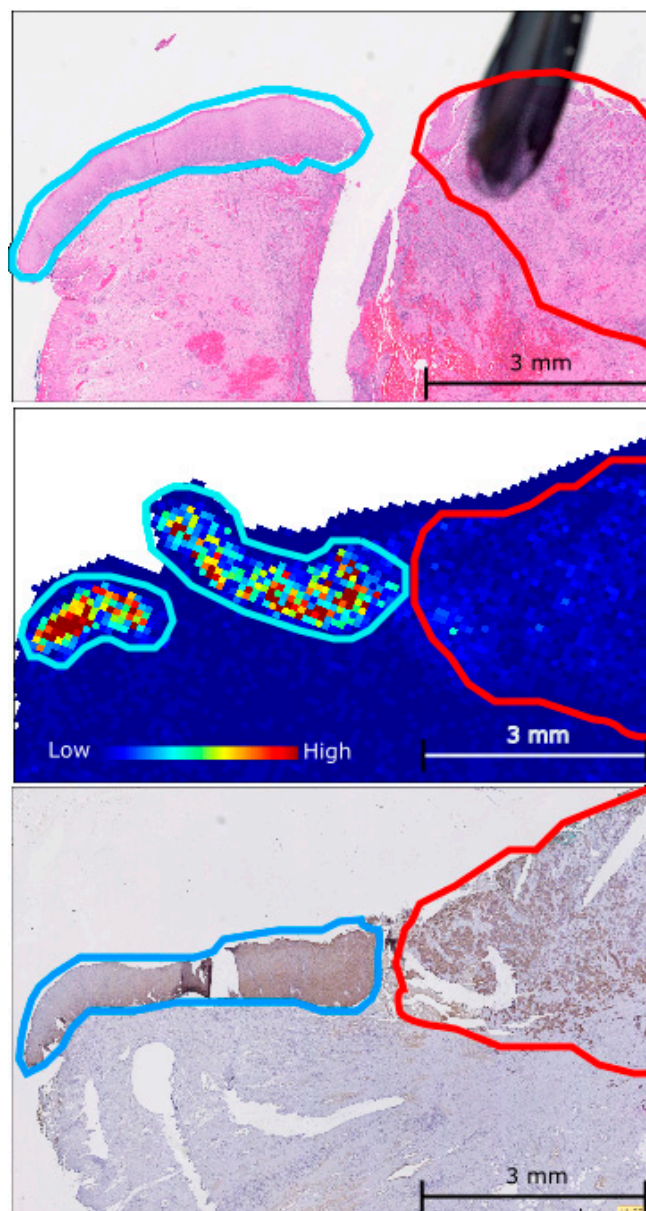


Figure 3. A comparison of the expression pattern observed for the CK5 peptides SFSTASAITPSVSR and TTAENEFVMLKK, both m/z 1410.72, in the dVIN (blue) and VSCC (red) using: MALDI-MSI ion intensity map (**middle**); H & E staining (**top**); and CK5 IHC staining (**bottom**) on consecutively cut vulvar tissue sections.

3. Discussion

Due to the relative rarity of vulvar cancer, characterization of the disease at the molecular level has been poor. A greater understanding of vulvar cancer progression is required, however, especially with regards to dVIN as the precursor lesion is known to have a high level of malignant potential [6] and is difficult to diagnose [12,13]. Generally, dVIN is characterized by thickened parakeratotic epithelium with elongation and anastomosing rete ridges, little to no atypia above the basal or parabasal layers of the epidermis, with the basal cell layer containing abnormal squamous cells undergoing atypical mitosis and high levels of epithelial cell differentiation in the upper epithelium [13,19,20]. The aim of this study was to in situ characterize the dVIN lesions at the peptide level using MALDI-MSI for

comparison to healthy vulvar epithelium and VSCC. To the best of our knowledge, this is the first report performing such an analysis from FFPE vulvar cancer tissues.

MALDI-MSI was carried out on tissue sections containing regions of healthy vulvar epithelium, dVIN, and VSCC from six late stage carcinoma patients. Thirty-one peak groups, representing the in situ imaged peptides, could be reproducibly detected across the majority of acquired spectra (Table 1). The median log intensity for each of the 31 peak groups was calculated for the regions of healthy epithelium, dVIN, and VSCC, and the maximum difference in intensity between the tissue types was calculated. This revealed 19 of the peak groups to have a change in intensity of ≥ 1.4 -fold across the three tissue types, of which six matched to unique peptides from the protein CK5 (Table 2).

CK5 is a 58 kDa intermediate filament protein that dimerizes with cytokeratin 14 to form the cytoskeleton of basal epithelial cells [21]. CK5 is known to have high levels of expression in squamous cell carcinomas and can be used as a marker of differentiation, with weaker to no expression in adenocarcinomas [22]. Given squamous cell carcinomas account for over 90% of all vulvar cancers and that CK5 is a structural protein of basal epithelial cells [23], the detection of CK5 by both MALDI-MSI was to be expected. A study characterizing the expression of CK5 in serous gynecological carcinomas of the ovaries, endometrium, fallopian tube, primary peritoneum, and cervix observed the highest level of CK5 expression in the cervical and ovarian serous carcinomas, with little to no reactivity in the endometrial, fallopian tube, or primary peritoneum serous carcinomas [24].

In order to determine if there was a genuine difference in CK5 levels across the three tissue types, IHC staining was performed on vulvar tissue from eight VSCC patients. Of these eight tissues, three had been previously analyzed in the MALDI-MSI experiments, six were moderately differentiated, one was well differentiated, and one was poorly differentiated. A degree of keratinization was reported for all of the VSCC with the exception of the poorly differentiated tumor and one of the moderately differentiated tumors. Generally, the expression of CK5 appeared to be diffuse and cytosolic across the three tissue types, with no distinct tissue specific staining patterns observed. IHC analysis revealed a significant increase in the level of CK5 staining in the dVIN as compared to the VSCC lesions, and in the healthy epithelium as compared to the VSCC lesions (Figure 2). There was a slight increase in the level of CK5 staining detected in the dVIN as compared to the healthy epithelium, but the difference was not significant.

To verify the spatial expression pattern of the peptides detected by MALDI-MSI, consecutive tissue sections that had been H&E stained and analysed by CK5 IHC were compared to ion intensity maps generated for the CK5 peptides from the MSI experiments. A representative example of H&E staining, CK5 IHC, and an ion intensity map for the CK5 peptides SFSTASAITPSVSR and TTAENEFVMLKK, both m/z 1410.72, from consecutive vulvar tissue sections is shown in Figure 3. The intensity and spatial distribution of the CK5 peptides as detected by MALDI-MSI corroborated well with the CK5 immunostaining observed. These results demonstrate the power of MALDI-MSI in the ability to combine classical mass spectrometry with histological tissue analysis.

This study provides an insight into the molecular fingerprint of the VSCC precursor lesion dVIN, which appears to more closely resemble healthy epithelium than it does cancerous tissue, even in later stage patients. To our knowledge, this is the first MALDI imaging mass spectrometry analysis of vulva cancer.

4. Materials and Methods

4.1. Sample Cohort

Formalin-fixed paraffin-embedded (FFPE) samples diagnosed with advanced VSCC from 2001 to 2013 were retrieved from the archive of SA Pathology at the Royal Adelaide Hospital (RAH). Ethics approval for the study was granted by the RAH. The age of patients ranged from 37 to 83 years with median age of 66.5 years, all patients were HPV negative, and each tissue analyzed contained regions of healthy vulvar epithelium, dVIN, and invasive vulvar squamous carcinoma (VSCC). Nine of the

vulvar carcinomas analyzed were stage III and three were stage IB. Clinical information for the patients is provided in Table S1. All tissues analyzed in this study were annotated by a pathologist.

4.2. MALDI-MSI Preparation and Acquisition

MALDI-MSI of FFPE tissue sections was performed on tissue from 6 VSCC patients as previously described [25]. Briefly, tissues were sectioned 8 μm thick and mounted on the Indium-Tin-Oxide (ITO) coated glass slides (Bruker Daltonics, Bremen, Germany) by heating at 60 $^{\circ}\text{C}$ for 1 h. Tissues were deparaffinised in xylene for 5 min, following by two 2 min incubations in 100% ethanol, and two 5 min incubations in 100 mM NH_4HCO_3 . The tissues were then subjected to heat induced citric acid antigen retrieval (CAAR) (10 mM citric acid, pH = 6) [26] followed by digestion with trypsin gold (Promega, Madison, WI, USA) using an ImagePrep station (Bruker Daltonics, Bremen, Germany) at 37 $^{\circ}\text{C}$ for 2 h. Peptide Internal calibrants (Angiotensin I, [Glu1]-Fibrinopeptide B, Dynorphin A and ACTH fragment [1–24]) and α -cyano-4-hydroxycinnamic acid (CHCA) matrix was overlaid onto the tissue sections using an ImagePrep station (Bruker Daltonics, Bremen, Germany) [27]. An ultrafleXtreme MALDI-TOF/TOF MS system (Bruker Daltonics, Bremen, Germany) was used for data acquisition in positive reflectron mode as monitored by flexControl (V3.0.1 Bruker Daltonics, Bremen, Germany) using following settings: 2 kHz, m/z 800–4000, with 100 μm spatial resolution. Following data acquisition slides were Hematoxylin and Eosin (H & E) stained and scanned using a NanoZommer (Hamamatsu, Japan). The H & E stained images were then co-registered with the collected MALDI MSI spectra using the flexImaging software (V4.0.1 Bruker Daltonics, Bremen, Germany).

4.3. MALDI-MSI Data Analysis

Acquired spectra were recalibrated using the internal calibrants and pre-processed using flexAnalysis and flexImaging software (V4.0.1 Bruker Daltonics, Bremen, Germany) in order to produce peak lists. During the pre-processing the Snap Algorithm with a signal to noise ratio of 2 was used for peak detection, a baseline subtraction was performed using the TopHat method, and baseline smoothing was performed using the Gauss algorithm. Tissue specific regions of interest (ROI) (i.e., healthy epithelium, dVIN and VSCC) were selected and exported into flexImaging (Bruker, Germany). Peaklists from all ROI were combined and density based clustering (DBSCAN* [18], with an epsilon of 0.02 and minimum density of 100 peaks) was used to cluster peaks into peak groups based on their m/z values. Initially only peak groups containing at least 10,000 peaks were considered for further analysis. This strict criterion was put in place to ensure that all peak groups analyzed could be reproducibly detected in many spectra. The median log intensity in each of the 3 tissue types was calculated for each peak group and for each patient. Peak groups were heuristically ranked by the largest difference in median log intensity between tissue types. The raw data was also analyzed using the SCiLS lab software (SCiLS, GmbH, Bremen, Germany, 2015b) where the processing steps of baseline removal and normalization were carried out [28]. Ion intensity maps for the peptides of interest were generated using the SCiLS lab software.

4.4. Peptide Identification by Nanoflow Liquid Chromatography Tandem Mass Spectrometry (nano-LC-MS/MS)

Tissue areas of healthy, dVIN, and VSCC were collected using laser capture microdissection (LCM) for protein digestion and identification by nano-LC-MS/MS. The FFPE tissue was sectioned at 8 μm thickness, water bath mounted onto PEN membrane slides (Micro-Dissect, Herborn, Germany), and deparaffinised as described above. Areas of healthy epithelium, dVIN, and VSCC were dissected using a Leica AS LCM microscope (Leica Microsystems, Wetzlar, Germany) into 20 μL of 10 mM citric acid buffer (pH = 6) and subjected to heat induced antigen retrieval as described above. Samples were buffered with 10 mM of NH_4HCO_3 and digested with 100 ng of trypsin gold (Promega, Madison, WI, USA) overnight at 37 $^{\circ}\text{C}$.

Nano-LC-MS/MS was performed using an Ultimate 3000 RSLC system (Thermo-Fisher Scientific, MA, USA) coupled to an Impact IITM QTOF mass spectrometer (Bruker Daltonics, Billerica, MA, USA)

via an Advance CaptiveSpray source (Bruker Daltonics). Peptide samples were pre-concentrated onto a C18 trapping column (Acclaim PepMap100 C18 75 μm $\times$ 20 mm, Thermo-Fisher Scientific) at a flow rate of 5 $\mu\text{L}/\text{min}$ in 2% ACN 0.1% TFA for 10 min. Peptide separation was performed using a 75 μm ID C18 column (Acclaim PepMap100 C18 75 μm $\times$ 50 cm, Thermo-Fisher Scientific) at a flow rate of 0.2 $\mu\text{L}/\text{min}$ using a linear gradient from 5% to 45% B (A: 5% ACN 0.1% FA, B: 80% ACN 0.1% FA) over 130 min, followed by a 20 min wash with 90% B, and a 20 min equilibration with 5% A. MS scans were acquired in the mass range of 300 to 2200 m/z in a data-dependent fashion using Bruker's Shotgun Instant Expertise™ method. Singly charged precursor ions were excluded from acquisition. Collision energy ranged from 23% to 65% as determined by the m/z of the precursor ion.

Acquired spectra were subjected to peak detection, de-convolution, and re-calibration according to a lock mass using DataAnalysis (Version 4.2, Bruker Daltonics). Processed spectra were then exported to Mascot generic format and submitted to Mascot (Version 2.3.02) for identification. Search parameters were as follows; SwissProt Homo sapiens database was searched, the digestion enzyme was specified as trypsin with up to 2 missed cleavages, variable modification of oxidation of methionine, MS mass tolerance of 40 ppm, and MS/MS mass tolerance of 0.2 Da. In Mascot, the peptide false discovery rate was set to <1% using Percolator, the peptide ion score cut off was set to 20, and the peptide significance score was set to <0.05.

4.5. Matching the MALDI-MSI Peak Groups to the Nanoflow Liquid Chromatography Tandem Mass Spectrometry (Nano-LC-MS/MS) Results

Matching between the two data sets was completed manually by comparing the experimental m/z values of the nano-LC-MS/MS sequenced peptides that fell between the minimum and maximum m/z of each of the MALDI-MSI peak groups.

4.6. Immunohistochemistry

For the analysis of Cytokeratin 5 by IHC, tissue from 8 VSCC patients was analyzed as previously described [29]. Briefly, 8 μm tissue sections were placed on plain glass slides, de-waxed, rehydrated and subjected to microwave antigen retrieval in 10 mM citric acid buffer (pH = 6) for 10 min at 100 °C in a steam microwave (Sixth Sense, Whirlpool, VIC, Australia). Tissue sections were then incubated overnight at 4 °C with the Cytokeratin 5 antibody (rabbit monoclonal antibody at a 1/200 dilution, Abcam, Cambridge, MA, USA) in 5% goat serum blocking buffer, followed by incubation with biotinylated anti-rabbit immunoglobulin (1/400, Dako, North Sydney, NSW, Australia) and streptavidin-HRP (1/500, Dako, Australia). Diaminobenzidine (DAB)/H₂O₂ (Sigma Aldrich) substrate was added before counterstaining with haematoxylin (Sigma Aldrich), dehydrating and mounting in Pertex (Meditate Medizintechnik, Germany). The CK5 stained sections were scanned using a NanoZommer (Hamamatsu, Japan) and viewed in NDP View (Hamamatsu, Japan). For each tissue section, three representative photo-micrographic images from areas of healthy epithelium, dVIN, and VSCC were captured at 40 $\times$ magnification. Analysis was carried out using the IHC Profiler in ImageJ which determines areas of high positive, positive, low positive, and negative staining [30]. The levels of CK5 positive staining for each tissue type were summed and compared across patients using GraphPad Prism 6 v008 (GraphPad Software, La Jolla, CA, USA), where the mean, standard error of the mean, and significance (paired t -test) were calculated.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/17/7/1088/s1>.

Acknowledgments: The authors gratefully acknowledge the financial support of Bioplatforms Australia and the Government of South Australia. The authors thank Chris Cursaro for outstanding technical support. We would like to acknowledge Johan O.R. Gustafsson for supervising the MALDI-MSI and experiments and providing the experiment protocol and Inge Koch (School of Mathematical Science, University of Adelaide) for mass spectra statistics analysis and classification.

Author Contributions: Martin K. Oehler, Christopher Carter, and Peter Hoffmann contributed to study design and sample collection; Chao Zhang, Noor A. Lokman, Parul Mittal, and Georgia Arentz contributed to the sample preparation and execution of experiments; Lyron Winderbaum performed the bioinformatics analysis; and Georgia Arentz, Martin K. Oehler, Peter Hoffmann and Manuela Klingler-Hoffmann wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2015. *CA Cancer* **2015**, *65*, 5–29. [[CrossRef](#)] [[PubMed](#)]
2. Brooks, L.A.; Tidy, J.A.; Gusterson, B.; Hiller, L.; O’Nions, J.; Gasco, M.; Marin, M.C.; Farrell, P.J.; Kaelin, W.G.; Crook, T. Preferential retention of codon 72 arginine p53 in squamous cell carcinomas of the vulva occurs in cancers positive and negative for human papillomavirus. *Cancer Res.* **2000**, *60*, 6875–6877. [[PubMed](#)]
3. Condon, J.; Rumbold, A.; Thorn, J.; O’Brien, M.; Davy, M.; Zardawi, I. A cluster of vulvar cancer and vulvar intraepithelial neoplasia in young australian indigenous women. *Cancer Causes Control* **2009**, *20*, 67–74. [[CrossRef](#)] [[PubMed](#)]
4. Joura, E.A. Epidemiology, diagnosis and treatment of vulvar intraepithelial neoplasia. *Curr. Opin. Obstet. Gynecol.* **2002**, *14*, 39–43. [[CrossRef](#)] [[PubMed](#)]
5. Liegl, B.; Regauer, S. P53 immunostaining in lichen sclerosus is related to ischaemic stress and is not a marker of differentiated vulvar intraepithelial neoplasia (d-VIN). *Histopathology* **2006**, *48*, 268–274. [[CrossRef](#)] [[PubMed](#)]
6. Van de Nieuwenhof, H.P.; Massuger, L.F.; van der Avoort, I.A.; Bekkers, R.L.; Casparie, M.; Abma, W.; van Kempen, L.C.; de Hullu, J.A. Vulvar squamous cell carcinoma development after diagnosis of vin increases with age. *Eur. J. Cancer* **2009**, *45*, 851–856. [[CrossRef](#)] [[PubMed](#)]
7. Van de Nieuwenhof, H.P.; Bulten, J.; Hollema, H.; Dommerholt, R.G.; Massuger, L.F.; van der Zee, A.G.; de Hullu, J.A.; van Kempen, L.C. Differentiated vulvar intraepithelial neoplasia is often found in lesions, previously diagnosed as lichen sclerosus, which have progressed to vulvar squamous cell carcinoma. *Mod. Pathol.* **2011**, *24*, 297–305. [[CrossRef](#)] [[PubMed](#)]
8. Kokka, F.; Singh, N.; Faruqi, A.; Gibbon, K.; Rosenthal, A.N. Is differentiated vulval intraepithelial neoplasia the precursor lesion of human papillomavirus-negative vulval squamous cell carcinoma? *Int. J. Gynecol. Cancer* **2011**, *21*, 1297–1305. [[CrossRef](#)] [[PubMed](#)]
9. New Nomenclature for Vulvar Disease. Report of the committee on terminology of the international society for the study of vulvar disease. *J. Reprod. Med.* **1990**, *35*, 483–484.
10. Ridley, C.M.; Frankman, O.; Jones, I.S.; Pincus, S.H.; Wilkinson, E.J.; Fox, H.; Friedrich, E.G.J.R.; Kaufman, R.H.; Lynch, P.J. New nomenclature for vulvar disease: International society for the study of vulvar disease. *Hum. Pathol.* **1989**, *20*, 495–496. [[CrossRef](#)]
11. Pinto, A.P.; Miron, A.; Yassin, Y.; Monte, N.; Woo, T.Y.; Mehra, K.K.; Medeiros, F.; Crum, C.P. Differentiated vulvar intraepithelial neoplasia contains tp53 mutations and is genetically linked to vulvar squamous cell carcinoma. *Mod. Pathol.* **2010**, *23*, 404–412. [[CrossRef](#)] [[PubMed](#)]
12. Aide, S.; Lattario, F.R.; Almeida, G.; do Val, I.C.; Carvalho Mda, G. Promoter hypermethylation of death-associated protein kinase and p16 genes in vulvar lichen sclerosus. *J. Low. Genit. Tract Dis.* **2012**, *16*, 133–139. [[CrossRef](#)] [[PubMed](#)]
13. Fox, H.; Wells, M. Recent advances in the pathology of the vulva. *Histopathology* **2003**, *42*, 209–216. [[CrossRef](#)] [[PubMed](#)]
14. Gustafsson, J.O.R.; Oehler, M.K.; Ruszkiewicz, A.; McColl, S.R.; Hoffmann, P. Maldi imaging mass spectrometry (MALDI-IMS)—Application of spatial proteomics for ovarian cancer classification and diagnosis. *Int. J. Mol. Sci.* **2011**, *12*, 773–794. [[CrossRef](#)] [[PubMed](#)]
15. Balluff, B.; Rauser, S.; Meding, S.; Elsner, M.; Schöne, C.; Feuchtinger, A.; Schuhmacher, C.; Novotny, A.; Jütting, U.; Maccarrone, G. Maldi imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer. *Am. J. Pathol.* **2011**, *179*, 2720–2729. [[CrossRef](#)] [[PubMed](#)]
16. Marko-Varga, G.; Fehninger, T.E.; Rezeli, M.; Döme, B.; Laurell, T.; Végvári, Á. Drug localization in different lung cancer phenotypes by maldi mass spectrometry imaging. *J. Proteom.* **2011**, *74*, 982–992. [[CrossRef](#)] [[PubMed](#)]

17. Crecelius, A.C.; Cornett, D.S.; Caprioli, R.M.; Williams, B.; Dawant, B.M.; Bodenheimer, B. Three-dimensional visualization of protein expression in mouse brain structures using imaging mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1093–1099. [[CrossRef](#)] [[PubMed](#)]
18. Campello, R.J.; Moulavi, D.; Sander, J. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining*; Springer: Berlin & Heidelberg, Germany, 2013; pp. 160–172.
19. Reyes, M.C.; Cooper, K. An update on vulvar intraepithelial neoplasia: Terminology and a practical approach to diagnosis. *J. Clin. Pathol.* **2014**, *67*, 290–294. [[CrossRef](#)] [[PubMed](#)]
20. Trietsch, M.D.; Nooij, L.S.; Gaarenstroom, K.N.; van Poelgeest, M.I. Genetic and epigenetic changes in vulvar squamous cell carcinoma and its precursor lesions: A review of the current literature. *Gynecol. Oncol.* **2015**, *136*, 143–157. [[CrossRef](#)] [[PubMed](#)]
21. Vasca, V.; Vasca, E.; Freiman, P.; Marian, D.; Luce, A.; Mesolella, M.; Caraglia, M.; Ricciardiello, F.; Duminica, T. Keratin 5 expression in squamocellular carcinoma of the head and neck. *Oncol. Lett.* **2014**, *8*, 2501–2504. [[CrossRef](#)] [[PubMed](#)]
22. Chu, P.G.; Weiss, L.M. Expression of cytokeratin 5/6 in epithelial neoplasms: An immunohistochemical study of 509 cases. *Mod. Pathol.* **2002**, *15*, 6–10. [[CrossRef](#)] [[PubMed](#)]
23. Khanna, N.; Rauh, L.A.; Lachiewicz, M.P.; Horowitz, I.R. Margins for cervical and vulvar cancer. *J. Surg. Oncol.* **2016**, *113*, 304–309. [[CrossRef](#)] [[PubMed](#)]
24. Nofech-Mozes, S.; Khalifa, M.A.; Ismiil, N.; Saad, R.S.; Hanna, W.M.; Covens, A.; Ghorab, Z. Immunophenotyping of serous carcinoma of the female genital tract. *Mod. Pathol.* **2008**, *21*, 1147–1155. [[CrossRef](#)] [[PubMed](#)]
25. Gustafsson, O.J.; Eddes, J.S.; Meding, S.; McColl, S.R.; Oehler, M.K.; Hoffmann, P. Matrix-assisted laser desorption/ionization imaging protocol for *in situ* characterization of tryptic peptide identity and distribution in formalin-fixed tissue. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 655–670. [[CrossRef](#)] [[PubMed](#)]
26. Gustafsson, J.O.; Oehler, M.K.; McColl, S.R.; Hoffmann, P. Citric acid antigen retrieval (CAAR) for tryptic peptide imaging directly on archived formalin-fixed paraffin-embedded tissue. *J. Proteom. Res.* **2010**, *9*, 4315–4328. [[CrossRef](#)] [[PubMed](#)]
27. Meding, S.; Martin, K.; Gustafsson, O.J.; Eddes, J.S.; Hack, S.; Oehler, M.K.; Hoffmann, P. Tryptic peptide reference data sets for maldi imaging mass spectrometry on formalin-fixed ovarian cancer tissues. *J. Proteom. Res.* **2012**, *12*, 308–315. [[CrossRef](#)] [[PubMed](#)]
28. Thiele, H.; Heldmann, S.; Trede, D.; Strehlow, J.; Wirtz, S.; Dreher, W.; Berger, J.; Oetjen, J.; Kobarg, J.H.; Fischer, B. 2D and 3D MALDI-imaging: Conceptual strategies for visualization and data mining. *Biochim. Biophys. Acta* **2014**, *1844*, 117–137. [[CrossRef](#)] [[PubMed](#)]
29. Lokman, N.A.; Elder, A.S.; Ween, M.P.; Pyragius, C.E.; Hoffmann, P.; Oehler, M.K.; Ricciardelli, C. Annexin a2 is regulated by ovarian cancer-peritoneal cell interactions and promotes metastasis. *Oncotarget* **2013**, *4*, 1199–1211. [[CrossRef](#)] [[PubMed](#)]
30. Varghese, F.; Bukhari, A.B.; Malhotra, R.; De, A. Ihc profiler: An open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS ONE* **2014**, *9*, e96801. [[CrossRef](#)] [[PubMed](#)]



Supplementary File Appendix II

Supplementary Table 1. Clinical information for the patient cohort

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11	Patient 12
Age	52	74	76	48	67	61	74	87	75	66	57	64
HPV status	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative
Lymph Node Status	Positive	Positive	Positive	Negative	Negative	Positive	Positive	Positive	Positive	Positive	Negative	Positive
Stage	III	III	III	IB	IB	III	III	III	III	III	IB	III
Analysed by MALDI-MSI	Yes	Yes	Yes	No	Yes	No	No	No	Yes	Yes	No	No
Analysed by ICH	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes

Peptide AWM [M+H] ⁺ of the peak group ¹	± Da Range of the peak group	No. spectra in peak group ²	Maximum difference median log intensity ³	LC-MS/MS m/z	LC-MS/MS MR ⁴	Peptide Score ⁵	Peptide Significance ⁶	Peptide Sequence ⁸	Protein Accession ⁹	Protein Description ¹⁰
1198.71	0.62	23099	0.75	599.8097	1197.6049	53.49	4.50E-06	IVLANDPDAHR	PGM2_HUMAN	Phosphoglucuronase-2 OS-Homo sapiens GN=PGM2 PE=1 SV=4
				400.2284	1197.6634	26.48	0.0022	IHNFLQJQEK	ACADV_HUMAN	Very long chain specific acyl-CoA dehydrogenase, mitochondrial OS-Homo sapiens GN=ACADVL PE=1 SV=1
				599.8406	1197.6667	69.65	1.10E-07	ASGPVSELITK	H12_HUMAN	Histone H1.2 OS-Homo sapiens GN=HIST1H1C PE=1 SV=2
				599.8406	1197.6667	69.65	1.10E-07	ASGPVSELITK	H14_HUMAN	Histone H1.4 OS-Homo sapiens GN=HIST1H1E PE=1 SV=2
				599.8597	1197.7048	94.81	3.30E-10	AVFPPVGRPR	ACTC_HUMAN	Actin, alpha cardiac muscle 1 OS-Homo sapiens GN=ACTC1 PE=1 SV=1
				599.8597	1197.7048	94.81	3.30E-10	AVFPPVGRPR	ACTB_HUMAN	Actin, cytoplasmic 1 OS-Homo sapiens GN=ACTB PE=1 SV=1
1028.62	0.56	15354	0.66	514.7932	1027.5719	52.15	6.10E-06	LADALQELR	LMNA_HUMAN	Prelamin-A/C OS-Homo sapiens GN=LMNA PE=1 SV=1
				514.7932	1027.5718	47.29	1.90E-05	AIENELLAR	GSHB_HUMAN	Glutathione synthetase OS-Homo sapiens GN=GSS PE=1 SV=1
				514.8309	1027.6473	37.68	0.00017	VVINVPFIK	GSH1_HUMAN	Glutamate-cysteine ligase catalytic subunit OS-Homo sapiens GN=GCLC PE=1 SV=2
1410.72*	0.82	18784	0.58	705.8674	1409.7203	115.28	3.00E-12	SFSTASATPVSRR	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
				705.8725	1409.7304	154.1	3.90E-16	TTAEENFVLMKK	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
				470.925	1409.7531	31.08	0.00078	VGLIFHSTFGR	NOP56_HUMAN	Nucleolar protein 56 OS-Homo sapiens GN=NOP56 PE=1 SV=4
				705.8915	1409.7684	25.84	0.0026	QLFHPEQLITGK	TBA1B_HUMAN	Tubulin alpha-1B chain OS-Homo sapiens GN=TUBA1B PE=1 SV=1
				705.8915	1409.7684	25.84	0.0026	QLFHPEQLITGK	TBA1A_HUMAN	Tubulin alpha-1A chain OS-Homo sapiens GN=TUBA1A PE=1 SV=1
				705.8915	1409.7684	25.84	0.0026	QLFHPEQLITGK	TBA4A_HUMAN	Tubulin alpha-4A chain OS-Homo sapiens GN=TUBA4A PE=1 SV=1
				705.924	1409.8334	59.93	1.00E-06	LNIHIAEGADIR	K6PL_HUMAN	6-phosphofructokinase, liver type OS-Homo sapiens GN=PFKL PE=1 SV=6
1905.95	0.52	10956	0.56	953.9953	1905.976	154.1	3.90E-16	ETDILLDDSLVSFGNR	PRDX5_HUMAN	Peroxiorexin-5, mitochondrial OS-Homo sapiens GN=PRDX5 PE=1 SV=4
				636.0074	1905.0004	30.44	0.0009	LATQSNITPVTFESR	HSPB1_HUMAN	Heat shock protein beta-1 OS-Homo sapiens GN=HSPB1 PE=1 SV=2
				636.3176	1905.9309	39.08	0.00012	VDEFVTNLSFDEINK	ADHX_HUMAN	Alcohol dehydrogenase class-3 OS-Homo sapiens GN=ADH5 PE=1 SV=4
1095.58	0.82	21704	0.55	477.0164	1904.0366	75.9	2.60E-08	AAIYQETGKPLDETLLK	ANXA1_HUMAN	Annexin A1 OS-Homo sapiens GN=ANXA1 PE=1 SV=2
				635.6879	1904.0419	26.34	0.0023	VQLPTETLQELDLHLR	GRP1_HUMAN	Interferon-induced guanylate-binding protein 1 OS-Homo sapiens GN=GRP1 PE=1 SV=2
				953.5051	1904.9956	154.1	3.90E-16	LATQSNITPVTFESR	HSPB1_HUMAN	Heat shock protein beta-1 OS-Homo sapiens GN=HSPB1 PE=1 SV=2
944.54	0.69	19354	0.53	472.7213	943.5286	27.71	0.0017	AGLQFPVGR	H2A2A_HUMAN	Histone H2A type 2-A OS-Homo sapiens GN=HIST2H2AA3 PE=1 SV=3
				472.7716	943.5286	27.71	0.0017	AGLQFPVGR	H2A1D_HUMAN	Histone H2A type 1-D OS-Homo sapiens GN=HIST1H2AD PE=1 SV=2
				472.7716	943.5286	27.71	0.0017	AGLQFPVGR	H2A1B_HUMAN	Histone H2A type 1-BE OS-Homo sapiens GN=HIST1H2AB PE=1 SV=2
				472.7716	943.5286	27.71	0.0017	AGLQFPVGR	H2AX_HUMAN	Histone H2AX OS-Homo sapiens GN=H2AFX PE=1 SV=2
				472.7716	943.5286	27.71	0.0017	AGLQFPVGR	H2AV_HUMAN	Histone H2A.V OS-Homo sapiens GN=H2AV PE=1 SV=3
957.63	0.65	10464	0.52	479.2679	956.5213	37.42	0.00018	LEGPDVSLK	AHNK_HUMAN	Neuroblast differentiation-associated protein AHNAK OS-Homo sapiens GN=AHNAK PE=1 SV=2
				479.286	956.5574	26.09	0.0025	LLDQIVEK	FUBP1_HUMAN	Far upstream element-binding protein 1 OS-Homo sapiens GN=FUBP1 PE=1 SV=3
				479.2892	956.5639	26.21	0.0024	LDLTLNLR	PRP8_HUMAN	Pre-mRNA-processing-splicing factor 8 OS-Homo sapiens GN=PRP8 PE=1 SV=2
				479.292	956.5694	60.22	9.50E-07	ILAVTIDGR	CO6A1_HUMAN	Collagen alpha-1(VI) chain OS-Homo sapiens GN=COL6A1 PE=1 SV=3
857.6	0.68	14956	0.5	429.2439	856.4733	21.9	0.0065	SLSYFK	ANXA4_HUMAN	Annexin A4 OS-Homo sapiens GN=ANXA4 PE=1 SV=4
				429.279	856.5434	34.25	0.00038	LTELILLR	ADRO_HUMAN	NADPH:adenodoxin oxidoreductase, mitochondrial OS-Homo sapiens GN=FDXR PE=1 SV=3
976.51	0.65	15534	0.49	NA	NA	NA	NA	NA	NA	NA
878.56	0.68	11213	0.46	NA	NA	NA	NA	NA	NA	NA
810.48*	0.76	17095	0.45	405.7087	809.4028	21.44	0.0072	QSSVSFR	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
958.59	0.69	15432	0.42	479.7771	957.5396	45.28	3.00E-05	GVYITIK	ROAA_HUMAN	Heterogeneous nuclear ribonucleoprotein A/B OS-Homo sapiens GN=HNRNPAB PE=1 SV=2
				479.7899	957.5652	37.91	0.00016	ALAQLSLR	PNCB_HUMAN	Nicotinate phosphoribosyltransferase OS-Homo sapiens GN=NAPRT1 PE=1 SV=2
				479.7897	957.5649	23.96	0.004	VIVSLSAGGR	ANXA4_HUMAN	Annexin A4 OS-Homo sapiens GN=ANXA4 PE=1 SV=4
				479.8008	957.587	33.68	0.00043	SGVGLALLIK	TRR1_HUMAN	Transferin receptor protein 1 OS-Homo sapiens GN=TRRC PE=1 SV=2
943.62	0.82	20030	0.41	472.7213	942.5681	35.37	0.00029	VPPPPPIAR	HNRPC_HUMAN	Heterogeneous nuclear ribonucleoproteins CLC2 OS-Homo sapiens GN=HNRNPC PE=1 SV=4
865.42*	0.84	15617	0.41	433.1998	864.3851	26	0.0025	SGGGGGQZGR	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
1217.66	0.55	10906	0.39	609.3218	1216.629	22.98	0.005	DAGTIAGLNVMR	GRP78_HUMAN	78 kDa glucose-regulated protein OS-Homo sapiens GN=HSPA5 PE=1 SV=2
				609.3444	1216.6742	42.57	5.50E-05	IIQLLDVYPK	RLA0_HUMAN	60S acidic ribosomal protein P0 OS-Homo sapiens GN=RLP0 PE=1 SV=1
				609.3636	1216.7125	32.6	0.00055	FYSVSDLLVPK	GDIB_HUMAN	Rab GDP dissociation inhibitor beta OS-Homo sapiens GN=GD12 PE=1 SV=2
				406.578	1216.7122	48.29	1.50E-05	KLIYFVQHR	TIF1B_HUMAN	Transcription intermediary factor 1-beta OS-Homo sapiens GN=TRIM28 PE=1 SV=5
884.52*	0.66	14337	0.37	442.7286	883.4427	24.74	0.0034	TFTSFSR	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
				442.7835	883.5525	23.85	0.0041	IULLAEGR	SAHH_HUMAN	Adenosylhomocysteinase OS-Homo sapiens GN=AHCY PE=1 SV=4
856.67	0.74	15483	0.36	NA	NA	NA	NA	NA	NA	NA
1143.65*	0.64	17685	0.36	572.3089	1142.6033	23.56	0.0044	IGAEVYHNLK	ENO4_HUMAN	Alpha-enolase OS-Homo sapiens GN=ENO1 PE=1 SV=2
				572.3183	1142.622	43.56	4.40E-05	LAELEALQK	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
				572.3231	1142.6317	45.39	2.90E-05	LAVNMVFPFR	TBB5_HUMAN	Tubulin beta chain OS-Homo sapiens GN=TUBB PE=1 SV=2
				572.3231	1142.6317	45.39	2.90E-05	LAVNMVFPFR	TBB4B_HUMAN	Tubulin beta-4B chain OS-Homo sapiens GN=TUBB4B PE=1 SV=1
				572.3231	1142.6317	45.39	2.90E-05	LAVNMVFPFR	TBB3_HUMAN	Tubulin beta-3 chain OS-Homo sapiens GN=TUBB3 PE=1 SV=2
				572.3231	1142.6317	45.39	2.90E-05	LAVNMVFPFR	TBB2A_HUMAN	Tubulin beta-2A chain OS-Homo sapiens GN=TUBB2A PE=1 SV=1
				572.3231	1142.6317	45.39	2.90E-05	LAVNMVFPFR	TBB4A_HUMAN	Tubulin beta-4A chain OS-Homo sapiens GN=TUBB4A PE=1 SV=2
				572.3231	1142.6317	45.39	2.90E-05	LAVNMVFPFR	TBB6_HUMAN	Tubulin beta-6 chain OS-Homo sapiens GN=TUBB6 PE=1 SV=1
				572.3816	1142.7486	45.43	2.90E-05	LNLILGLIFK	CND1_HUMAN	Condensin complex subunit 1 OS-Homo sapiens GN=NCAPD2 PE=1 SV=3
1045.62	0.65	16370	0.35	523.2619	1044.5092	23.25	0.0047	GRGFVSFR	PABP1_HUMAN	Polyadenylate-binding protein 1 OS-Homo sapiens GN=PABPC1 PE=1 SV=2
				523.2882	1044.5618	57.7	1.70E-06	GVDSTQQR	DESP_HUMAN	Desmoplakin OS-Homo sapiens GN=DSP PE=1 SV=3
2147.19	0.14	20418	0.33	716.3459	2146.0159	75.77	2.60E-08	SADESGQALLAAGHYASDEVRR	SPTN1_HUMAN	Spectrin alpha chain, non-erythrocytic 1 OS-Homo sapiens GN=SPTAN1 PE=1 SV=3
1105.59	0.5	11479	0.3	553.799	1105.5835	55.58	2.80E-06	SEIDLFPNR	ANXA5_HUMAN	Annexin A5 OS-Homo sapiens GN=ANXA5 PE=1 SV=2
				553.8077	1105.6009	22.67	0.0054	DFLQLFAR	TGFI1_HUMAN	Transforming growth factor beta-1-induced transcript 1 protein OS-Homo sapiens GN=TFGB11 PE=1 SV=2
				553.8115	1105.6083	46.18	2.40E-05	ADLFTDAIK	F10A1_HUMAN	Hsc70-interacting protein OS-Homo sapiens GN=HSC713 PE=1 SV=2
				553.8536	1105.6926	76.29	2.30E-08	VILPFIAR	COX2_HUMAN	Cytochrome c oxidase subunit 2 OS-Homo sapiens GN=MT-CO2 PE=1 SV=1
1296.68	0.1	22823	0.29	648.8565	1295.6986	81.56	7.00E-09	AAFYVAIAFPR	PNCB_HUMAN	Nicotinate phosphoribosyltransferase OS-Homo sapiens GN=NAPRT1 PE=1 SV=2
870.64	0.85	22749	0.29	435.769	869.5234	30.25	0.00094	LLELTGPK	A1BG_HUMAN	Alpha-1B-glycoprotein OS-Homo sapiens GN=A1BG PE=1 SV=4
				435.776	869.5375	25.71	0.0027	LAVEAVLR	TCPB_HUMAN	T-complex protein 1 subunit beta OS-Homo sapiens GN=CTC2 PE=1 SV=4
				435.775	869.5354	24.24	0.0038	LAVEALVR	HYAS1_HUMAN	Hyaluronan synthase 1 OS-Homo sapiens GN=HAS1 PE=1 SV=2
				435.776	869.5375	25.71	0.0027	LAVEALVR	HYAS1_HUMAN	Hyaluronan synthase 1 OS-Homo sapiens GN=HAS1 PE=1 SV=2
				435.7761	869.5377	25.68	0.0027	ALLEVLGR	GLO2_HUMAN	Hydroxycarboxylate hydrolase, mitochondrial OS-Homo sapiens GN=HAGH PE=1 SV=2
842.6	0.9	25210	0.22	421.7795	841.5427	20.78	0.0084	VLAYAVIR	PUR2_HUMAN	Trifunctional purine biosynthetic protein adenosine-3 OS-Homo sapiens GN=GART PE=1 SV=1
1267.65	0.32	11686	0.2	634.3339	1266.6333	111.75	6.70E-12	AEQFEDVNLPK	AHNK_HUMAN	Neuroblast differentiation-associated protein AHNAK OS-Homo sapiens GN=AHNAK PE=1 SV=2
				634.338	1266.6614	45.24	3.00E-05	MVVEASVEYIK	LDHB_HUMAN	L-lactate dehydrogenase B chain OS-Homo sapiens GN=LDHB PE=1 SV=2
				423.2312	1266.6719	35.87	0.00026	ISMPDVLHLK	AHNK_HUMAN	Neuroblast differentiation-associated protein AHNAK OS-Homo sapiens GN=AHNAK PE=1 SV=2
				634.3753	1266.7361	94.8	3.30E-10	IAAAILNTPDLR	AATM_HUMAN	Aspartate aminotransferase, mitochondrial OS-Homo sapiens GN=GOAT2 PE=1 SV=3
				423.2598	1266.7575	48.53	1.40E-05	ALPAPIEKTIK	IGHG1_HUMAN	Ig gamma-1 chain C region OS-Homo sapiens GN=IGHG1 PE=1 SV=1
2932.59	0.07	15500	0.19	NA	NA	NA	NA	NA	NA	NA
816.5	0.73	10983	0.19	408.7213	815.4281	25	0.0032	FIDDFK	SYAC_HUMAN	Alanine-tRNA ligase, cytoplasmic OS-Homo sapiens GN=AARS PE=1 SV=2
				408.7416	815.4687	22.38	0.0058	GFPINIR	FRN1_HUMAN	Fibrillin-1 OS-Homo sapiens GN=FRN1 PE=1 SV=3
971.56*	0.65	15998	0.18	486.244	970.4735	29.12	0.0012	FVSTTSRR	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
				486.2931	970.5716	26.86	0.0021	VLAEEVALK	EGLN2_HUMAN	Egl nine homolog 2 OS-Homo sapiens GN=EGLN2 PE=1 SV=1
				486.2995	970.5844	58.97	1.30E-06	VIVVITDGR	COEAL_HUMAN	Collagen alpha-1(XIV) chain OS-Homo sapiens GN=COL14A1 PE=1 SV=3
1111.58	0.51	13473	0.09	556.2794	1110.5442	28.87	0.0013	IJEDAHNSRR	ATZ11_HUMAN	Ethanolamine-phosphate phospho-lyase OS-Homo sapiens GN=ETNPPL PE=1 SV=1
				556.2841	1110.5536	45.4	2.90E-05	QDIAFYQRR	ANXA2_HUMAN	Annexin A2 OS-Homo sapiens GN=ANXA2 PE=1 SV=2
				556.293	1110.5714	36.07	0.00025	ISSTSGSRR	K2C5_HUMAN	Keratin, type II cytoskeletal 5 OS-Homo sapiens GN=KRT5 PE=1 SV=3
				556.2955	1110.5764	35.35	0.00029	DLLYFDGLR	PKP3_HUMAN	Plakophilin-3 OS-Homo sapiens GN=PKP3 PE=1 SV=1
				556.2967	1110.5788	47.39	1.80E-05	DLLYFDGLR	PKP3_HUMAN	Plakophilin-3 OS-Homo sapiens GN=PKP3 PE=1 SV=1
				556.3096	1110.6047	22.46	0.0057	DGLGLPDIVR	UBP5_HUMAN	Ubiquitin carboxyl-terminal hydrolase 5 OS-Homo sapiens GN=USP5 PE=1 SV=2
				371.2152	1110.6239	54.45	3.60E-06	FLLSPEHR	PSMD6_HUMAN	26S proteasome non-ATPase regulatory subunit 6 OS-Homo sapiens GN=PSMD6 PE=1 SV=1
				556.3198	1110.6251	21.79	0.0066	TPTAVVAPVEK	CLIP1_HUMAN	CAP-GP domain-containing linker protein 1 OS-Homo sapiens GN=CLIP1 PE=1 SV=2
				556.3287	1110.6428	66.76	2.10E-07	TGAAPIDIVR	RL27A_HUMAN	60S ribosomal protein L27a OS-Homo sapiens GN=RLP27A PE=1 SV=2
1570.7	0.17	12029	0.05	524.2693	1569.7861	26.02	0.0025	GSVEGFGAAPDISVK	AHNK_HUMAN	Neuroblast differentiation-associated protein AHNAK OS-Homo sapiens GN=AHNAK PE=1 SV=2
				785.9101	1569.8057	136.03	2.50E-14	GAGTGGGLAVEGSEAK	FLNA_HUMAN	Filamin-A OS-Homo sapiens GN=FLNA PE=1 SV=4
				393.4601	1569.8115	46.4	2.30E-05	NHLLHVFDEYKR	ANXA4_HUMAN	Annexin A4 OS-Homo sapiens GN=ANXA4 PE=1 SV=4
				524.283	1569.8271	27.48	0.0018	TDLHAFENLEIR	EGRF_HUMAN	Epidermal growth factor receptor OS-Homo sapiens GN=EGRF PE=1 SV=2
				785.9281	1569.8416	52.29	5.90E-06	ALEEANADLEVKIR	KIC13_HUMAN	Keratin, type I cytoskeletal 13 OS-Homo sapiens GN=KRT13 PE=1 SV=4
				785.9281	1569.8416	52.29</				

NA No LC-MS/MS match

* Peak groups matched to Cytokeratin 5 peptides

Non-unique peptides with matches to multiple proteins

¹ The abundance weighted means (AWM) m/z was calculated for the overall peak group

² Number of Spectra in each peak group compiled from all acquisition spectra across the patient cohort

³ Heuristically ranked maximum difference, d, between the healthy epithelium, dVIN, and VSCC

⁴ Experimental m/z of the CK5 matched nanoLC-ESI-MS/MS detected peptide

⁵ Experimental mass of the CK5 matched nanoLC-ESI-MS/MS detected peptide

⁶ Mascot Ions Score

⁷ Mascot Expected Score

⁸ Matched LC-MSMS peptide sequence

⁹ Uniprot Protein Accession

This page intentionally left blank

REFERENCES

This page intentionally left blank

References

- [1] S. Bendifallah, G. Canlorbe, P. Collinet, E. Arsène, F. Huguet, C. Coutant, D. Hudry, O. Graesslin, E. Raimond, C. Touboul, E. Daraï, M. Ballester, Just how accurate are the major risk stratification systems for early-stage endometrial cancer?, *British Journal of Cancer*, 112 (2015) 793-801.
- [2] S. Bendifallah, E. Daraï, M. Ballester, Predictive Modeling: A New Paradigm for Managing Endometrial Cancer, *Annals of surgical oncology*, 23 (2016) 975-988.
- [3] R.L. Siegel, K.D. Miller, A. Jemal, *Cancer statistics, 2015*, CA: A Cancer Journal for Clinicians, 65 (2015) 5-29.
- [4] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012, *International journal of cancer*, 136 (2015) E359-E386.
- [5] F. Amant, P. Moerman, P. Neven, D. Timmerman, E. Van Limbergen, I. Vergote, Endometrial cancer, *Lancet (London, England)*, 366 (2005) 491-505.
- [6] C. Australian Institute of Health and Welfare, *Cancer in Australia An overview 2012*, in, the Australian Institute of Health and Welfare Canberra ACT 2601, 2012, pp. <http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129542353>.
- [7] H.M. Pelikan, J.W. Trum, F.C. Bakers, R.G. Beets-Tan, L.J. Smits, R.F. Kruitwagen, Diagnostic accuracy of preoperative tests for lymph node status in endometrial cancer: a systematic review, *Cancer imaging : the official publication of the International Cancer Imaging Society*, 13 (2013) 314-322.
- [8] J.-H. Yoon, S.-C. Yoo, W.Y. Kim, S.-J. Chang, K.-H. Chang, H.-S. Ryu, Para-aortic lymphadenectomy in the management of preoperative grade 1 endometrial cancer confined to the uterine corpus, *Annals of surgical oncology*, 17 (2010) 3234-3240.
- [9] S.L. Antonsen, L.N. Jensen, A. Loft, A.K. Berthelsen, J. Costa, A. Tabor, I. Qvist, M.R. Hansen, R. Fisker, E.S. Andersen, MRI, PET/CT and ultrasound in the preoperative staging of endometrial cancer—a multicenter prospective comparative study, *Gynecologic oncology*, 128 (2013) 300-308.
- [10] J.V. Bokhman, Two pathogenetic types of endometrial carcinoma, *Gynecologic oncology*, 15 (1983) 10-17.
- [11] D. Denschlag, U. Ulrich, G. Emons, The Diagnosis and Treatment of Endometrial Cancer: Progress and Controversies, *Deutsches Ärzteblatt International*, 108 (2011) 571-577.
- [12] Endometrial (Uterine cancer), in, American Cancer Society, 2015.
- [13] W. Creasman, Revised FIGO staging for carcinoma of the endometrium, *International Journal of Gynecology & Obstetrics*, 105 (2009) 109-109.
- [14] A.J. O'Hara, D.W. Bell, The genomics and genetics of endometrial cancer, *Advances in genomics and genetics*, 2012 (2012) 33-47.
- [15] D.W. Bell, Our changing view of the genomic landscape of cancer, *The Journal of pathology*, 220 (2010) 231-243.
- [16] S.F. Lax, R.J. Kurman, A dualistic model for endometrial carcinogenesis based on immunohistochemical and molecular genetic analyses, *Verhandlungen der Deutschen Gesellschaft für Pathologie*, 81 (1997) 228-232.
- [17] X. Matias-Guiu, L. Catasus, E. Bussaglia, H. Lagarda, A. Garcia, C. Pons, J. Munoz, R. Arguelles, P. Machin, J. Prat, Molecular pathology of endometrial hyperplasia and carcinoma, *Hum Pathol*, 32 (2001) 569-577.
- [18] A. Yeramian, G. Moreno-Bueno, X. Dolcet, L. Catasus, M. Abal, E. Colas, J. Reventos, J. Palacios, J. Prat, X. Matias-Guiu, Endometrial carcinoma: molecular alterations involved in tumor development and progression, *Oncogene*, 32 (2013) 403-413.
- [19] X. Matias-Guiu, J. Prat, Molecular pathology of endometrial carcinoma, *Histopathology*, 62 (2013) 111-123.
- [20] C. Kandoth, N. Schultz, A.D. Cherniack, R. Akbani, Y. Liu, H. Shen, A.G. Robertson, I. Pashtan, R. Shen, C.C. Benz, C. Yau, P.W. Laird, L. Ding, W. Zhang, G.B. Mills, R. Kucherlapati, E.R. Mardis, D.A. Levine, Integrated genomic characterization of endometrial carcinoma, *Nature*, 497 (2013) 67-73.

- [21] M. Le Gallo, D.W. Bell, The emerging genomic landscape of endometrial cancer, *Clinical chemistry*, 60 (2014) 98-110.
- [22] R.A. Soslow, High-grade endometrial carcinomas - strategies for typing, *Histopathology*, 62 (2013) 89-110.
- [23] P.A. Sanderson, H.O. Critchley, A.R. Williams, M.J. Arends, P.T. Saunders, New concepts for an old problem: the diagnosis of endometrial hyperplasia, *Human reproduction update*, (2016).
- [24] K. Tomczak, P. Czerwinska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, *Contemp Oncol (Pozn)*, 19 (2015) A68-A77.
- [25] C.G.A.R. Network, Integrated genomic characterization of endometrial carcinoma, *Nature*, 497 (2013) 67-73.
- [26] Biomarkers and surrogate endpoints: preferred definitions and conceptual framework, *Clinical pharmacology and therapeutics*, 69 (2001) 89-95.
- [27] T. Clark, P. Barton, A. Coomarasamy, J. Gupta, K. Khan, Gynaecological oncology: Investigating postmenopausal bleeding for endometrial cancer: cost-effectiveness of initial diagnostic strategies, *BJOG: An International Journal of Obstetrics & Gynaecology*, 113 (2006) 502-510.
- [28] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical chemistry*, 39 (1993) 561-577.
- [29] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L.M. Irwig, D. Moher, D. Rennie, H.C. de Vet, J.G. Lijmer, The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration, *Annals of internal medicine*, 138 (2003) W1-12.
- [30] K. Kitajima, K. Murakami, E. Yamasaki, S. Hagiwara, I. Fukasawa, N. Inaba, Y. Kaji, K. Sugimura, Performance of FDG-PET/CT in the diagnosis of recurrent endometrial cancer, *Annals of nuclear medicine*, 22 (2008) 103-109.
- [31] S.H. Kim, H.D. Kim, Y.S. Song, S.B. Kang, H.P. Lee, Detection of deep myometrial invasion in endometrial carcinoma: comparison of transvaginal ultrasound, CT, and MRI, *Journal of computer assisted tomography*, 19 (1995) 766-772.
- [32] P.T. Ramirez, M. Frumovitz, M.R. Milam, M. Deavers, R. dos Reis, R.B. Iyer, P. Bhosale, K.M. Schmeler, The limited utility of magnetic resonance imaging in determining the primary site of disease in patients with inconclusive endometrial biopsy, *International journal of gynecological cancer: official journal of the International Gynecological Cancer Society*, 20 (2010) 1344.
- [33] L. Savelli, M. Ceccarini, M. Ludovisi, E. Fruscella, P. De Iaco, E. Salizzoni, M. Mabrouk, R. Manfredi, A.C. Testa, G. Ferrandina, Preoperative local staging of endometrial cancer: transvaginal sonography vs. magnetic resonance imaging, *Ultrasound in Obstetrics & Gynecology*, 31 (2008) 560-566.
- [34] L.J. Havrilesky, G.L. Maxwell, E.R. Myers, Cost-effectiveness analysis of annual screening strategies for endometrial cancer, *American journal of obstetrics and gynecology*, 200 (2009) 640. e641-640. e648.
- [35] P. Uharcek, Prognostic factors in endometrial carcinoma, *The journal of obstetrics and gynaecology research*, 34 (2008) 776-783.
- [36] C.L. Creutzberg, R.A. Nout, The role of radiotherapy in endometrial cancer: current evidence and trends, *Current oncology reports*, 13 (2011) 472-478.
- [37] J.V. Bokhman, Two pathogenetic types of endometrial carcinoma, *Gynecologic oncology*, 15 (1983) 10-17.
- [38] K.J. Dedes, D. Wetterskog, A. Ashworth, S.B. Kaye, J.S. Reis-Filho, Emerging therapeutic targets in endometrial cancer, *Nature reviews Clinical oncology*, 8 (2011) 261-271.
- [39] M. Frumovitz, D.K. Singh, L. Meyer, D.H. Smith, I. Wertheim, E. Resnik, D.C. Bodurka, Predictors of final histology in patients with endometrial cancer, *Gynecologic oncology*, 95 (2004) 463-468.
- [40] R.W. Holloway, S. Gupta, N.M. Stavitzski, X. Zhu, E.L. Takimoto, A. Gubbi, G.E. Bigsby, L.A. Brudie, J.E. Kendrick, S. Ahmad, Sentinel lymph node mapping with staging lymphadenectomy for patients with endometrial cancer increases the detection of metastasis, *Gynecologic oncology*, 141 (2016) 206-210.
- [41] S. Pecorelli, Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium, *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, 105 (2009) 103-104.

- [42] C.P. Morrow, B.N. Bundy, R.J. Kurman, W.T. Creasman, P. Heller, H.D. Homesley, J.E. Graham, Relationship between surgical-pathological risk factors and outcome in clinical stage I and II carcinoma of the endometrium: a Gynecologic Oncology Group study, *Gynecologic oncology*, 40 (1991) 55-65.
- [43] N.L. Neubauer, L.J. Havrilesky, B. Calingaert, A. Bulusu, M.Q. Bernardini, N.D. Fleming, A.E. Bland, A.A. Secord, The role of lymphadenectomy in the management of preoperative grade 1 endometrial carcinoma, *Gynecologic oncology*, 112 (2009) 511-516.
- [44] M. Frumovitz, B.M. Slomovitz, D.K. Singh, R.R. Broaddus, J. Abrams, C.C. Sun, M. Bevers, D.C. Bodurka, Frozen section analyses as predictors of lymphatic spread in patients with early-stage uterine cancer¹, *Journal of the American College of Surgeons*, 199 (2004) 388-393.
- [45] P. Benedetti Panici, S. Basile, F. Maneschi, A. Alberto Lissoni, M. Signorelli, G. Scambia, R. Angioli, S. Tateo, G. Mangili, D. Katsaros, G. Garozzo, E. Campagnutta, N. Donadello, S. Greggi, M. Melpignano, F. Raspagliesi, N. Ragni, G. Cormio, R. Grassi, M. Franchi, D. Giannarelli, R. Fossati, V. Torri, M. Amoroso, C. Croce, C. Mangioni, Systematic pelvic lymphadenectomy vs. no lymphadenectomy in early-stage endometrial carcinoma: randomized clinical trial, *Journal of the National Cancer Institute*, 100 (2008) 1707-1716.
- [46] H. Kitchener, A.M. Swart, Q. Qian, C. Amos, M.K. Parmar, Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study, *Lancet (London, England)*, 373 (2009) 125-136.
- [47] E.L. Pollom, C.M. Conklin, R. von Eyben, A.K. Folkins, E.A. Kidd, Nomogram to Predict Risk of Lymph Node Metastases in Patients With Endometrioid Endometrial Cancer, *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists*, 35 (2016) 395-401.
- [48] M. Koskas, M. Fournier, A. Vanderstraeten, F. Walker, D. Timmerman, I. Vergote, F. Amant, Evaluation of models to predict lymph node metastasis in endometrial cancer: A multicentre study, *European journal of cancer*, 61 (2016) 52-60.
- [49] S. Kang, W.D. Kang, H.H. Chung, D.H. Jeong, S.-S. Seo, J.-M. Lee, J.-K. Lee, J.W. Kim, S.-M. Kim, S.-Y. Park, K.T. Kim, Preoperative Identification of a Low-Risk Group for Lymph Node Metastasis in Endometrial Cancer: A Korean Gynecologic Oncology Group Study, *Journal of Clinical Oncology*, 30 (2012) 1329-1334.
- [50] S. Bendifallah, A.S. Genin, I. Naoura, N. Chabbert Buffet, F. Clavel Chapelon, B. Haddad, D. Luton, E. Darai, R. Rouzier, M. Koskas, A nomogram for predicting lymph node metastasis of presumed stage I and II endometrial cancer, *American journal of obstetrics and gynecology*, 207 (2012) 197.e191-197.e198.
- [51] W.T. Creasman, C.P. Morrow, B. Bundy, H.D. Homesley, J. Graham, P. Heller, Surgical pathologic spread patterns of endometrial cancer, A gynecologic oncology group study. *Cancer*, 60 (1987) 2035-2041.
- [52] M.M. AlHilli, K.C. Podratz, S.C. Dowdy, J.N. Bakkum-Gamez, A.L. Weaver, M.E. McGree, G.L. Keeney, W.A. Cliby, A. Mariani, Risk-scoring system for the individualized prediction of lymphatic dissemination in patients with endometrioid endometrial cancer, *Gynecologic oncology*, 131 (2013) 103-108.
- [53] T. Jacks, R.A. Weinberg, Cell-cycle control and its watchman, *Nature*, 381 (1996) 643-644.
- [54] C.A. Hudis, Trastuzumab--mechanism of action and use in clinical practice, *The New England journal of medicine*, 357 (2007) 39-51.
- [55] K.R. Schultz, W.P. Bowman, A. Aledo, W.B. Slayton, H. Sather, M. Devidas, C. Wang, S.M. Davies, P.S. Gaynon, M. Trigg, R. Rutledge, L. Burden, D. Jorstad, A. Carroll, N.A. Heerema, N. Winick, M.J. Borowitz, S.P. Hunger, W.L. Carroll, B. Camitta, Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a children's oncology group study, *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27 (2009) 5175-5181.
- [56] I. Vandenput, J. Trovik, K. Leunen, E. Wik, I. Stefansson, L. Akslen, P. Moerman, I. Vergote, H. Salvesen, F. Amant, Evolution in endometrial cancer: evidence from an immunohistochemical study, *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, 21 (2011) 316-322.

- [57] A.K. Sood, R.E. Buller, R.A. Burger, J.D. Dawson, J.I. Sorosky, M. Berman, Value of preoperative CA 125 level in the management of uterine cancer and prediction of clinical outcome, *Obstetrics & Gynecology*, 90 (1997) 441-447.
- [58] E. Ambeba, F. Linkov, Advancements in the use of blood tests for cancer screening in women at high risk for endometrial and breast cancer, *Future oncology*, 7 (2011) 1399-1414.
- [59] S.L. Antonsen, E. Høgdall, I.J. Christensen, M. Lydolph, A. Tabor, A. Loft Jakobsen, C.L. Fagö-Olsen, E.S. Andersen, K. Jochumsen, C. Høgdall, HE4 and CA125 levels in the preoperative assessment of endometrial cancer patients: a prospective multicenter study (ENDOMET), *Acta obstetrica et gynecologica Scandinavica*, 92 (2013) 1313-1322.
- [60] A.C. Staff, J. Trovik, E.A.G. Zahl, E. Wik, K.C. Wollert, T. Kempf, H.B. Salvesen, Elevated plasma growth differentiation factor-15 correlates with lymph node metastases and poor survival in endometrial cancer, *Clinical cancer research*, 17 (2011) 4825-4833.
- [61] S. Lo, D. Cheng, T. Ng, L. Wong, H. Ngan, Prognostic significance of tumour markers in endometrial cancer, *Tumor biology*, 18 (1997) 241-249.
- [62] D.J. Dotters, Preoperative CA 125 in endometrial cancer: is it useful?, *American journal of obstetrics and gynecology*, 182 (2000) 1328-1334.
- [63] N. Koper, L. Massuger, C. Thomas, L. Kiemeney, A. Verbeek, Serum CA 125 measurements to identify patients with endometrial cancer who require lymphadenectomy, *Anticancer research*, 18 (1997) 1897-1902.
- [64] J.M. Niloff, T.L. Klug, E. Schaetzel, V.R. Zurawski, Jr., R.C. Knapp, R.C. Bast, Jr., Elevation of serum CA125 in carcinomas of the fallopian tube, endometrium, and endocervix, *American journal of obstetrics and gynecology*, 148 (1984) 1057-1058.
- [65] T. Jiang, L. Huang, S. Zhang, Preoperative serum CA125: a useful marker for surgical management of endometrial cancer, *BMC Cancer*, 15 (2015) 396.
- [66] G. Scambia, A. Gadducci, P.B. Panici, E. Foti, M. Ferdeghini, G. Ferrandina, Combined use of CA 125 and CA 15-3 in patients with endometrial carcinoma, *Gynecologic oncology*, 54 (1994).
- [67] C.H. Hsieh, C.C. ChangChien, H. Lin, E.Y. Huang, C.C. Huang, K.C. Lan, Can a preoperative CA 125 level be a criterion for full pelvic lymphadenectomy in surgical staging of endometrial cancer?, *Gynecologic oncology*, 86 (2002).
- [68] H.S. Kim, C.Y. Park, J.M. Lee, J.K. Lee, C.H. Cho, S.M. Kim, Evaluation of serum CA-125 levels for preoperative counseling in endometrioid endometrial cancer: a multi-center study, *Gynecologic oncology*, 118 (2010).
- [69] M. Kanat-Pektas, O. Yenicesu, T. Gungor, U. Bilge, Predictive power of sexual hormones and tumor markers in endometrial cancer, *Archives of gynecology and obstetrics*, 281 (2010) 709-715.
- [70] K.D. Hatch, J.S. Berek, Intraepithelial disease of the cervix, vagina and vulva, *Novak's gynecology*, 12 (2002) 447-486.
- [71] N. Takeshima, Y. Shimizu, S. Umezawa, Y. Hirai, J.-T. Chen, I. Fujimoto, K. Yamauchi, K. Hasumi, Combined assay of serum levels of CA125 and CA19-9 in endometrial carcinoma, *Gynecologic oncology*, 54 (1994) 321-326.
- [72] G. Scambia, A. Gadducci, P.B. Panici, E. Foti, M. Ferdeghini, G. Ferrandina, M. Amoroso, C. Castellani, V. Facchini, S. Mancuso, Combined use of CA 125 and CA 15-3 in patients with endometrial carcinoma, *Gynecologic oncology*, 54 (1994) 292-297.
- [73] P. Benedetti Panici, G. Scambia, G. Baiocchi, L. Perrone, S. Greggi, F. Battaglia, S. Mancuso, Multiple serum markers in patients with endometrial cancer, *Gynecologic and obstetric investigation*, 27 (1989) 208-212.
- [74] A. Hakala, B.M. Kacinski, E.R. Stanley, E.I. Kohorn, U. Puistola, J. Risteli, L. Risteli, C. Tomás, A. Kauppila, Macrophage colony-stimulating factor 1, a clinically useful tumor marker in endometrial adenocarcinoma: comparison with CA 125 and the aminoterminal propeptide of type III procollagen, *American journal of obstetrics and gynecology*, 173 (1995) 112-119.
- [75] A. Chao, Y.H. Tang, C.H. Lai, C.J. Chang, S.C. Chang, T.I. Wu, S. Hsueh, C.J. Wang, H.H. Chou, T.C. Chang, Potential of an age-stratified CA125 cut-off value to improve the prognostic classification of patients with endometrial cancer, *Gynecologic oncology*, 129 (2013) 500-504.
- [76] J.M. Niloff, T.L. Klug, E. Schaetzel, V.R. Zurawski, R.C. Knapp, R.C. Bast, Elevation of serum CA125 in carcinomas of the fallopian tube, endometrium, and endocervix, *American journal of obstetrics and gynecology*, 148 (1984) 1057-1058.

- [77] J.M. Duk, J.G. Aalders, G.J. Fleuren, H.W. de Bruijn, CA 125: a useful marker in endometrial carcinoma, *American journal of obstetrics and gynecology*, 155 (1986) 1097-1102.
- [78] R.G. Moore, A.K. Brown, M.C. Miller, D. Badgwell, Z. Lu, W.J. Allard, C. Granai, R.C. Bast, K. Lu, Utility of a novel serum tumor biomarker HE4 in patients with endometrioid adenocarcinoma of the uterus, *Gynecologic oncology*, 110 (2008) 196-201.
- [79] E. Bignotti, M. Ragnoli, L. Zanotti, S. Calza, M. Falchetti, S. Lonardi, S. Bergamelli, E. Bandiera, R. Tassi, C. Romani, Diagnostic and prognostic impact of serum HE4 detection in endometrial carcinoma patients, *British journal of cancer*, 104 (2011) 1418-1425.
- [80] P. Cherchi, S. Dessole, G. Ruiu, G. Ambrosini, M. Farina, G. Capobianco, A. Ambrosini, The value of serum CA 125 and association CA 125/CA 19-9 in endometrial carcinoma, *Eur J Gynnc Oncol*~ MN, 392 (1999) 2936.
- [81] H. Hareyama, N. Sakuragi, S. Makinoda, S. Fujimoto, Serum and tissue measurements of CA72-4 in patients with endometrial carcinoma, *Journal of clinical pathology*, 49 (1996) 967-970.
- [82] E.P. Beck, M. Wagner, L. Anselmino, F.-j. Xu, R.C. Bast, W. Jaeger, Is OVX1 a suitable marker for endometrial cancer?, *Gynecologic oncology*, 65 (1997) 291-296.
- [83] S. Lo, U. Khoo, D. Cheng, T. Ng, L. Wong, H. Ngan, Role of serial tumor markers in the surveillance for recurrence in endometrial cancer, *Cancer detection and prevention*, 23 (1998) 397-400.
- [84] S.S. Lo, D.K. Cheng, T.Y. Ng, L.C. Wong, H.Y. Ngan, Prognostic significance of tumour markers in endometrial cancer, *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 18 (1997) 241-249.
- [85] T. Nagai, H. Niikura, S. Okamoto, K. Nakabayashi, M. Matoda, H. Utsunomiya, S. Nagase, M. Watanabe, N. Takeshima, N. Yaegashi, A new diagnostic method for rapid detection of lymph node metastases using a one-step nucleic acid amplification (OSNA) assay in endometrial cancer, *Annals of surgical oncology*, 22 (2015) 980-986.
- [86] P. Mhawech-Fauceglia, F.R. Herrmann, H. Rai, N. Tchabo, S. Lele, I. Izevbaye, K. Odunsi, R.T. Cheney, IMP3 distinguishes uterine serous carcinoma from endometrial endometrioid adenocarcinoma, *American journal of clinical pathology*, 133 (2010) 899-908.
- [87] M. Voss, N. Gordon, S. Maloney, R. Ganesan, L. Ludeman, K. McCarthy, R. Gornall, G. Schaller, W. Wei, F. Berditchevski, Tetraspanin CD151 is a novel prognostic marker in poor outcome endometrial cancer, *British journal of cancer*, 104 (2011) 1611-1618.
- [88] E. Cocco, S. Bellone, K. El-Sahwi, M. Cargnelutti, F. Casagrande, N. Buza, F. Tavassoli, E. Siegel, I. Visintin, E. Ratner, Serum amyloid A (SAA): a novel biomarker for uterine serous papillary cancer, *British journal of cancer*, 101 (2009) 335-341.
- [89] K. Tanabe, M. Matsumoto, S. Ikematsu, S. Nagase, A. Hatakeyama, T. Takano, H. Niikura, K. Ito, K. Kadomatsu, S.i. Hayashi, Midkine and its clinical significance in endometrial carcinoma, *Cancer science*, 99 (2008) 1125-1130.
- [90] G. Farias-Eisner, F. Su, T. Robbins, J. Kotlerman, S. Reddy, R. Farias-Eisner, Validation of serum biomarkers for detection of early-and late-stage endometrial cancer, *American journal of obstetrics and gynecology*, 202 (2010) 73. e71-73. e75.
- [91] E. Wik, M.B. Ræder, C. Krakstad, J. Trovik, E. Birkeland, E.A. Hoivik, S. Mjos, H.M. Werner, M. Mannelqvist, I.M. Stefansson, Lack of estrogen receptor- α is associated with epithelial–mesenchymal transition and PI3K alterations in endometrial carcinoma, *Clinical Cancer Research*, 19 (2013) 1094-1105.
- [92] E. Wik, J. Trovik, O.E. Iversen, I.B. Engelsen, I.M. Stefansson, L.C. Vestrheim, H.K. Haugland, L.A. Akslen, H.B. Salvesen, Deoxyribonucleic acid ploidy in endometrial carcinoma: a reproducible and valid prognostic marker in a routine diagnostic setting, *American journal of obstetrics and gynecology*, 201 (2009) 603. e601-603. e607.
- [93] K. Banno, Y. Nogami, I. Kisu, M. Yanokura, K. Umene, K. Masuda, Y. Kobayashi, W. Yamagami, N. Susumu, D. Aoki, Candidate Biomarkers for Genetic and Clinicopathological Diagnosis of Endometrial Cancer, *International Journal of Molecular Sciences*, 14 (2013) 12123-12137.
- [94] Y. Todo, R. Yamamoto, S. Minobe, Y. Suzuki, U. Takeshi, M. Nakatani, Y. Aoyagi, Y. Ohba, K. Okamoto, H. Kato, Risk factors for postoperative lower-extremity lymphedema in endometrial cancer survivors who had treatment including lymphadenectomy, *Gynecologic oncology*, 119 (2010) 60-64.

- [95] L.G. Seamon, J.M. Fowler, D.E. Cohn, Lymphadenectomy for endometrial cancer: the controversy, *Gynecologic oncology*, 117 (2010) 6-8.
- [96] C.S. Walsh, B.Y. Karlan, Lymphadenectomy's role in early endometrial cancer: prognostic or therapeutic?, *Journal of the National Cancer Institute*, 100 (2008) 1660-1661.
- [97] B. Patsner, G.W. Yim, Predictive value of preoperative serum CA-125 levels in patients with uterine cancer: The Asian experience 2000 to 2012, *Obstetrics & Gynecology Science*, 56 (2013) 281-288.
- [98] S. Ramaswamy, K.N. Ross, E.S. Lander, T.R. Golub, A molecular signature of metastasis in primary solid tumors, *Nature genetics*, 33 (2003) 49-54.
- [99] M.Q. McCreery, K.D. Halliwill, D. Chin, R. Delrosario, G. Hirst, P. Vuong, K.Y. Jen, J. Hewinson, D.J. Adams, A. Balmain, Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers, *Nature medicine*, 21 (2015) 1514-1520.
- [100] N.Q. Liu, C. Stingl, M.P. Look, M. Smid, R.B. Braakman, T. De Marchi, A.M. Sieuwerts, P.N. Span, F.C. Sweep, B.K. Linderholm, A. Mangia, A. Paradiso, L.Y. Dirix, S.J. Van Laere, T.M. Luider, J.W. Martens, J.A. Foekens, A. Umar, Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer, *Journal of the National Cancer Institute*, 106 (2014) djt376.
- [101] S.N. Zanoaruddin, A. Saleh, Y.H. Yang, S. Hamid, W.M. Mustafa, A.A. Khairul Bariah, R.B. Zain, S.H. Lau, S.C. Cheong, Four-protein signature accurately predicts lymph node metastasis and survival in oral squamous cell carcinoma, *Human pathology*, 44 (2013) 417-426.
- [102] M.P. Pavlou, A. Dimitromanolakis, E. Martinez-Morillo, M. Smid, J.A. Foekens, E.P. Diamandis, Integrating meta-analysis of microarray data and targeted proteomics for biomarker identification: application in breast cancer, *Journal of proteome research*, 13 (2014) 2897-2909.
- [103] H.H. Cao, S.Y. Zhang, J.H. Shen, Z.Y. Wu, J.Y. Wu, S.H. Wang, E.M. Li, L.Y. Xu, A three-protein signature and clinical outcome in esophageal squamous cell carcinoma, *Oncotarget*, (2014).
- [104] M. Shipitsin, C. Small, E. Giladi, S. Siddiqui, S. Choudhury, S. Hussain, Y.E. Huang, H. Chang, D.L. Rimm, D.M. Berman, T.P. Nifong, P. Blume-Jensen, Automated quantitative multiplex immunofluorescence in situ imaging identifies phospho-S6 and phospho-PRAS40 as predictive protein biomarkers for prostate cancer lethality, *Proteome science*, 12 (2014) 40.
- [105] N.E. Mascini, G.B. Eijkel, P. Ter Brugge, J. Jonkers, J. Wesseling, R.M. Heeren, The use of mass spectrometry imaging to predict treatment response of patient-derived xenograft models of triple-negative breast cancer, *Journal of proteome research*, 14 (2015) 1069-1075.
- [106] S. Hanash, Disease proteomics, *Nature*, 422 (2003) 226-232.
- [107] M. Uhlen, L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A. Szilagarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Ponten, Proteomics. Tissue-based map of the human proteome, *Science*, 347 (2015) 1260419.
- [108] D. Baltimore, Our genome unveiled, *Nature*, 409 (2001) 814-816.
- [109] J. Godovac-Zimmermann, L.R. Brown, Perspectives for mass spectrometry and functional proteomics, *Mass spectrometry reviews*, 20 (2001) 1-57.
- [110] M. Tyers, M. Mann, From genomics to proteomics, *Nature*, 422 (2003) 193-197.
- [111] E.F. Petricoin, L.A. Liotta, Proteomic approaches in cancer risk and response assessment, *Trends in molecular medicine*, 10 (2004) 59-64.
- [112] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, *Nature*, 422 (2003) 198-207.
- [113] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, Use of proteomic patterns in serum to identify ovarian cancer, *The lancet*, 359 (2002) 572-577.
- [114] E.F. Petricoin, K.C. Zoon, E.C. Kohn, J.C. Barrett, L.A. Liotta, Clinical proteomics: translating benchside promise into bedside reality, *Nature reviews Drug discovery*, 1 (2002) 683-695.
- [115] W. Wang, H. Zhou, H. Lin, S. Roy, T.A. Shaler, L.R. Hill, S. Norton, P. Kumar, M. Anderle, C.H. Becker, Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards, *Analytical chemistry*, 75 (2003) 4818-4826.

- [116] H. Zhang, Y. Fan, L. Xia, C. Gao, X. Tong, H. Wang, L. Sun, T. Ji, M. Jin, B. Gu, B. Fan, The impact of advanced proteomics in the search for markers and therapeutic targets of bladder cancer, *Tumor Biology*, 39 (2017) 1010428317691183.
- [117] R.M. Caprioli, T.B. Farmer, J. Gile, Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS, *Analytical chemistry*, 69 (1997) 4751-4760.
- [118] B. Balluff, S. Rauser, S. Meding, M. Elsner, C. Schone, A. Feuchtinger, C. Schuhmacher, A. Novotny, U. Jutting, G. Maccarrone, H. Sarioglu, M. Ueffing, H. Braselmann, H. Zitzelsberger, R.M. Schmid, H. Hofler, M.P. Ebert, A. Walch, MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer, *The American journal of pathology*, 179 (2011) 2720-2729.
- [119] G. Marko-Varga, T.E. Fehniger, M. Rezeli, B. Dome, T. Laurell, A. Vegvari, Drug localization in different lung cancer phenotypes by MALDI mass spectrometry imaging, *J Proteomics*, 74 (2011) 982-992.
- [120] T. Gemoll, U.J. Roblick, J.K. Habermann, MALDI mass spectrometry imaging in oncology (Review), *Molecular medicine reports*, 4 (2011) 1045-1051.
- [121] S. Meding, U. Nitsche, B. Balluff, M. Elsner, S. Rauser, C. Schone, M. Nipp, M. Maak, M. Feith, M.P. Ebert, H. Friess, R. Langer, H. Hofler, H. Zitzelsberger, R. Rosenberg, A. Walch, Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging, *J Proteome Res*, 11 (2012) 1996-2003.
- [122] N.E. Mascini, R.M.A. Heeren, Protein identification in mass-spectrometry imaging, *TrAC Trends in Analytical Chemistry*, 40 (2012) 28-37.
- [123] L. Theron, D. Centeno, C. Coudy-Gandilhon, E. Pujos-Guillot, T. Astruc, D. Remond, J.C. Barthelemy, F. Roche, L. Feasson, M. Hebraud, D. Bechet, C. Chambon, A Proof of Concept to Bridge the Gap between Mass Spectrometry Imaging, Protein Identification and Relative Quantitation: MSI~LC-MS/MS-LF, *Proteomes*, 4 (2016).
- [124] A. Hu, W.S. Noble, A. Wolf-Yadlin, Technical advances in proteomics: new developments in data-independent acquisition, *F1000Research*, 5 (2016) F1000 Faculty Rev-1419.
- [125] B. Domon, R. Aebersold, Mass Spectrometry and Protein Analysis, *Science*, 312 (2006) 212-217.
- [126] A.I. Nesvizhskii, Protein identification by tandem mass spectrometry and sequence database searching, *Mass Spectrometry Data Analysis in Proteomics*, (2007) 87-119.
- [127] H. Liu, R.G. Sadygov, J.R. Yates, 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics, *Analytical chemistry*, 76 (2004) 4193-4201.
- [128] A. Hu, W. Noble, A. Wolf-Yadlin, Technical advances in proteomics: new developments in data-independent acquisition [version 1; referees: 3 approved], 2016.
- [129] J.D. Venable, M.-Q. Dong, J. Wohlschlegel, A. Dillin, J.R. Yates, Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra, *Nat Meth*, 1 (2004) 39-45.
- [130] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, *Bioinformatics*, 26 (2010) 966-968.
- [131] A.H. Coons, H.J. Creech, R.N. Jones, Immunological Properties of an Antibody Containing a Fluorescent Group.*, *Proceedings of the Society for Experimental Biology and Medicine*, 47 (1941) 200-202.
- [132] F. Varghese, A.B. Bukhari, R. Malhotra, A. De, IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples, *PloS one*, 9 (2014) e96801.
- [133] S. Steurer, J.M. Singer, M. Rink, F. Chun, R. Dahlem, R. Simon, E. Burandt, P. Stahl, L. Terracciano, T. Schlomm, W. Wagner, W. Höppner, M. Omid, O. Kraus, M. Kwiatkowski, O. Doh, M. Fisch, A. Soave, G. Sauter, M. Wurlitzer, H. Schlüter, S. Minner, MALDI imaging-based identification of prognostically relevant signals in bladder cancer using large-scale tissue microarrays, *Urologic Oncology: Seminars and Original Investigations*, 32 (2014) 1225-1233.
- [134] S. Steurer, A.S. Seddiqi, J.M. Singer, A.S. Bahar, C. Eichelberg, M. Rink, R. Dahlem, H. Huland, G. Sauter, R. Simon, S. Minner, E. Burandt, P.R. Stahl, T. Schlomm, M. Wurlitzer, H.

- Schluter, MALDI imaging on tissue microarrays identifies molecular features associated with renal cell cancer phenotype, *Anticancer research*, 34 (2014) 2255-2261.
- [135] S. Steurer, C. Borkowski, S. Odinga, M. Buchholz, C. Koop, H. Huland, M. Becker, M. Witt, D. Trede, M. Omid, MALDI mass spectrometric imaging based identification of clinically relevant signals in prostate cancer using large-scale tissue microarrays, *International Journal of Cancer*, 133 (2013) 920-928.
- [136] R. Casadonte, M. Kriegsmann, F. Zweynert, K. Friedrich, G. Bretton, M. Otto, S.-O. Deininger, R. Paape, E. Belau, D. Suckau, D. Aust, C. Pilarsky, J. Kriegsmann, Imaging mass spectrometry to discriminate breast from pancreatic cancer metastasis in formalin-fixed paraffin-embedded tissues, *PROTEOMICS*, 14 (2014) 956-964.
- [137] Y. Morita, K. Ikegami, N. Goto-Inoue, T. Hayasaka, N. Zaima, H. Tanaka, T. Uehara, T. Setoguchi, T. Sakaguchi, H. Igarashi, H. Sugimura, M. Setou, H. Konno, Imaging mass spectrometry of gastric carcinoma in formalin-fixed paraffin-embedded tissue microarray, *Cancer Science*, 101 (2010) 267-273.
- [138] M.R. Groseclose, P.P. Massion, P. Chaurand, R.M. Caprioli, High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry, *Proteomics*, 8 (2008) 3715-3724.
- [139] S. Meding, K. Martin, O.J. Gustafsson, J.S. Eddes, S. Hack, M.K. Oehler, P. Hoffmann, Tryptic peptide reference data sets for MALDI imaging mass spectrometry on formalin-fixed ovarian cancer tissues, *Journal of proteome research*, 12 (2013) 308-315.
- [140] A. Quaas, A.S. Bahar, K. von Loga, A.S. Seddiqi, J.M. Singer, M. Omid, O. Kraus, M. Kwiatkowski, M. Trusch, S. Minner, E. Burandt, P. Stahl, W. Wilczak, M. Wurlitzer, R. Simon, G. Sauter, A. Marx, H. Schluter, MALDI imaging on large-scale tissue microarrays identifies molecular features associated with tumour phenotype in oesophageal cancer, *Histopathology*, 63 (2013) 455-462.
- [141] J. Kononen, L. Bubendorf, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M.J. Mihatsch, G. Sauter, O.P. Kallioniemi, Tissue microarrays for high-throughput molecular profiling of tumor specimens, *Nature medicine*, 4 (1998) 844-847.
- [142] H. Battifora, The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing, *Laboratory investigation; a journal of technical methods and pathology*, 55 (1986) 244-248.
- [143] L. Winderbaum, I. Koch, P. Mittal, P. Hoffmann, Classification of MALDI-MS imaging data of tissue microarrays using canonical correlation analysis-based variable selection, *Proteomics*, 16 (2016) 1731-1735.
- [144] P. Mittal, M. Klingler-Hoffmann, G. Arentz, C. Zhang, G. Kaur, M.K. Oehler, P. Hoffmann, Proteomics of endometrial cancer diagnosis, treatment, and prognosis, *PROTEOMICS – Clinical Applications*, 10 (2016) 217-229.
- [145] P. Mittal, M. Klingler-Hoffmann, G. Arentz, L. Winderbaum, N.A. Lokman, C. Zhang, L. Anderson, J. Scurry, Y. Leung, C.J. Stewart, J. Carter, G. Kaur, M.K. Oehler, P. Hoffmann, Lymph node metastasis of primary endometrial cancers: Associated proteins revealed by MALDI imaging, *Proteomics*, 16 (2016) 1793-1801.
- [146] R. Luthra, R.R. Singh, M.G. Luthra, Y.X. Li, C. Hannah, A.M. Romans, B.A. Barkoh, S.S. Chen, J. Ensor, D.M. Maru, R.R. Broaddus, A. Rashid, C.T. Albarracin, MicroRNA-196a targets annexin A1: a microRNA-mediated mechanism of annexin A1 downregulation in cancers, *Oncogene*, 27 (2008) 6667-6678.
- [147] H.D. Reyes, K.W. Thiel, M.J. Carlson, X. Meng, S. Yang, J.M. Stephan, K.K. Leslie, Comprehensive profiling of EGFR/HER receptors for personalized treatment of gynecologic cancers, *Molecular diagnosis & therapy*, 18 (2014) 137-151.
- [148] Y. Li, G. Francia, J.Y. Zhang, p62/IMP2 stimulates cell migration and reduces cell adhesion in breast cancer, *Oncotarget*, 6 (2015) 32656-32668.
- [149] A. Barghash, V. Helms, S.M. Kessler, Overexpression of IGF2 mRNA-Binding Protein 2 (IMP2/p62) as a Feature of Basal-like Breast Cancer Correlates with Short Survival, *Scandinavian journal of immunology*, 82 (2015) 142-143.

- [150] S. Yamamoto, H. Tsuda, K. Honda, T. Kita, M. Takano, S. Tamai, J. Inazawa, T. Yamada, O. Matsubara, Actinin-4 expression in ovarian cancer: a novel prognostic indicator independent of clinical stage and histological type, *Mod Pathol*, 20 (2007) 1278-1285.
- [151] K. Honda, T. Yamada, Y. Hayashida, M. Idogawa, S. Sato, F. Hasegawa, Y. Ino, M. Ono, S. Hirohashi, Actinin-4 increases cell motility and promotes lymph node metastasis of colorectal cancer, *Gastroenterology*, 128 (2005) 51-62.
- [152] K. Honda, T. Yamada, R. Endo, Y. Ino, M. Gotoh, H. Tsuda, Y. Yamada, H. Chiba, S. Hirohashi, Actinin-4, a novel actin-bundling protein associated with cell motility and cancer invasion, *J Cell Biol*, 140 (1998) 1383-1393.
- [153] Z. Jiang, Y. Li, G. Han, J. Zhang, Z. Li, D. Wang, Y. Liu, [Association of serum albumin level with clinicopathologic features and prognosis in colon cancer], *Zhonghua wei chang wai ke za zhi = Chinese journal of gastrointestinal surgery*, 19 (2016) 80-83.
- [154] D.H. Suh, H.S. Kim, H.H. Chung, J.W. Kim, N.H. Park, Y.S. Song, S.B. Kang, Pre-operative systemic inflammatory response markers in predicting lymph node metastasis in endometrioid endometrial adenocarcinoma, *European journal of obstetrics, gynecology, and reproductive biology*, 162 (2012) 206-210.
- [155] L.Y. Xue, L.H. Teng, S.M. Zou, L.Q. Ren, S. Zheng, W. Luo, R. Bi, N. Lu, [Expression of annexin I in different histological types of carcinomas], *Zhonghua zhong liu za zhi [Chinese journal of oncology]*, 29 (2007) 444-448.
- [156] J. Da, X. Meng, P. Wang, Z. Yang, Y. Zhu, [Significance on expressions of Annexin-I and its correlative gene proteins in endometrial hyperplasia, atypical hyperplasia and endometrial carcinoma], *Zhonghua bing li xue za zhi Chinese journal of pathology*, 30 (2001) 256-259.
- [157] S.N. Voisin, O. Krakovska, A. Matta, L.V. DeSouza, A.D. Romaschin, T.J. Colgan, K.W. Siu, Identification of novel molecular targets for endometrial cancer using a drill-down LC-MS/MS approach with iTRAQ, *PloS one*, 6 (2011) e16352.
- [158] L. Deng, Y. Gao, X. Li, M. Cai, H. Wang, H. Zhuang, M. Tan, S. Liu, Y. Hao, B. Lin, Expression and clinical significance of annexin A2 and human epididymis protein 4 in endometrial carcinoma, *Journal of experimental & clinical cancer research : CR*, 34 (2015) 96.
- [159] L. Alonso-Alconada, M. Santacana, P. Garcia-Sanz, L. Muinelo-Romay, E. Colas, C. Mirantes, M. Monge, J. Cueva, E. Oliva, R.A. Soslow, M.A. Lopez, J. Palacios, J. Prat, J. Valls, C. Krakstad, H. Salvesen, A. Gil-Moreno, R. Lopez-Lopez, X. Dolcet, G. Moreno-Bueno, J. Reventos, X. Matias-Guiu, M. Abal, Annexin-A2 as predictor biomarker of recurrent disease in endometrial cancer, *International journal of cancer. Journal international du cancer*, 136 (2015) 1863-1873.
- [160] S. Zhou, T. Yi, R. Liu, C. Bian, X. Qi, X. He, K. Wang, J. Li, X. Zhao, C. Huang, Y. Wei, Proteomics identification of annexin A2 as a key mediator in the metastasis and proangiogenesis of endometrial cells in human adenomyosis, *Molecular & cellular proteomics : MCP*, 11 (2012) M112 017988.
- [161] N. Bou Zgheib, D.C. Marchion, S.H. Bush, P.L. Judson, R.M. Wenham, S.M. Apte, J.M. Lancaster, J. Gonzalez-Bosquet, Molecular determinants for lymph node metastasis in clinically early-stage endometrial cancer, *Oncology letters*, 11 (2016) 323-329.
- [162] Y. Fang, X. Liang, W. Jiang, J. Li, J. Xu, X. Cai, Cyclin b1 suppresses colorectal cancer invasion and metastasis by regulating e-cadherin, *PloS one*, 10 (2015) e0126875.
- [163] C.Y. Lei, W. Wang, Y.T. Zhu, W.Y. Fang, W.L. Tan, The decrease of cyclin B2 expression inhibits invasion and metastasis of bladder cancer, *Urologic oncology*, (2015).
- [164] D. Korenaga, F. Takesue, M. Yasuda, M. Honda, T. Nozoe, S. Inutsuka, The relationship between cyclin B1 overexpression and lymph node metastasis in human colorectal cancer, *Surgery*, 131 (2002) S114-120.
- [165] K. Nakayama, M.T. Rahman, M. Rahman, K. Nakamura, M. Ishikawa, H. Katagiri, E. Sato, T. Ishibashi, K. Iida, N. Ishikawa, S. Kyo, CCNE1 amplification is associated with aggressive potential in endometrioid endometrial carcinomas, *International journal of oncology*, 48 (2016) 506-516.
- [166] S.F. Shariat, R. Ashfaq, A.I. Sagalowsky, Y. Lotan, Correlation of cyclin D1 and E1 expression with bladder cancer presence, invasion, progression, and metastasis, *Human pathology*, 37 (2006) 1568-1576.

- [167] R. Smigiel, M. Sasiadek, T. Krecicki, D. Ramsey, J. Jagielski, N. Blin, Inactivation of the cyclin-dependent kinase inhibitor 2A (CDKN2A) gene in squamous cell carcinoma of the larynx, *Molecular carcinogenesis*, 39 (2004) 147-154.
- [168] P. Schulz, A. Scholz, A. Rexin, P. Hauff, M. Schirner, B. Wiedenmann, K. Detjen, Inducible re-expression of p16 in an orthotopic mouse model of pancreatic cancer inhibits lymphangiogenesis and lymphatic metastasis, *British journal of cancer*, 99 (2008) 110-117.
- [169] F. Wang, Y.Z. Yang, C.Z. Shi, P. Zhang, M.P. Moyer, H.Z. Zhang, Y. Zou, H.L. Qin, UHRF1 promotes cell growth and metastasis through repression of p16(ink(4)a) in colorectal cancer, *Annals of surgical oncology*, 19 (2012) 2753-2762.
- [170] M.C. Todd, H.M. Petty, J.M. King, B.N. Piana Marshall, R.A. Sheller, M.E. Cuevas, Overexpression and delocalization of claudin-3 protein in MCF-7 and MDA-MB-415 breast cancer cell lines, *Oncology letters*, 10 (2015) 156-162.
- [171] X.Y. Pan, B. Wang, Y.C. Che, Z.P. Weng, H.Y. Dai, W. Peng, Expression of claudin-3 and claudin-4 in normal, hyperplastic, and malignant endometrial tissue, *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, 17 (2007) 233-241.
- [172] H. Nei, T. Saito, H. Tobioka, E. Itoh, M. Mori, R. Kudo, Expression of component desmosomal proteins in uterine endometrial carcinoma and their relation to cellular differentiation, *Cancer*, 78 (1996) 461-470.
- [173] H. Niikura, H. Sasano, G. Matsunaga, K. Watanabe, K. Ito, S. Sato, A. Yajima, Prognostic value of epidermal growth factor receptor expression in endometrioid endometrial carcinoma, *Human pathology*, 26 (1995) 892-896.
- [174] G. Scambia, P. Benedetti Panici, G. Ferrandina, F. Battaglia, M. Distefano, G. D'Andrea, R. De Vincenzo, F. Maneschi, F.O. Ranelletti, S. Mancuso, Significance of epidermal growth factor receptor expression in primary human endometrial cancer, *International journal of cancer. Journal international du cancer*, 56 (1994) 26-30.
- [175] L. Albitar, G. Pickett, M. Morgan, J.A. Wilken, N.J. Mairhle, K.K. Leslie, EGFR isoforms and gene regulation in human endometrial cancer cells, *Molecular cancer*, 9 (2010) 166.
- [176] G.N. Chen, Y. Ma, Z.L. Yang, [Expression of GRP78 and GRP94 in the liver tissues and their clinicopathological significance in children with hepatoblastoma], *Zhongguo dang dai er ke za zhi = Chinese journal of contemporary pediatrics*, 12 (2010) 634-636.
- [177] T.R. Cawthorn, J.C. Moreno, M. Dharsee, D. Tran-Thanh, S. Ackloo, P.H. Zhu, G. Sardana, J. Chen, P. Kupchak, L.M. Jacks, N.A. Miller, B.J. Youngson, V. Iakovlev, C.J. Guidos, K.A. Vallis, K.R. Evans, D. McCready, W.L. Leong, S.J. Done, Proteomic analyses reveal high expression of decorin and endoplasmin (HSP90B1) are associated with breast cancer metastasis and decreased survival, *PloS one*, 7 (2012) e30992.
- [178] X. Xiong, Y. Wang, C. Liu, Q. Lu, T. Liu, G. Chen, H. Rao, S. Luo, Heat shock protein 90beta stabilizes focal adhesion kinase and enhances cell migration and invasion in breast cancer cells, *Experimental cell research*, 326 (2014) 78-89.
- [179] H. Sakamoto, K. Kinoshita, Y. Nakayama, T. Takami, K. Ohtani, S. Izuka, K. Satoh, Gene expression in primary, metastatic and recurrent lesions of endometrial cancer, *Gynecologic and obstetric investigation*, 48 (1999) 215-220.
- [180] C. Bartosch, S. Monteiro-Reis, R. Vieira, A. Pereira, M. Rodrigues, C. Jeronimo, J.M. Lopes, Endometrial Endometrioid Carcinoma Metastases Show Decreased ER-Alpha and PR-A Expression Compared to Matched Primary Tumors, *PloS one*, 10 (2015) e0134969.
- [181] M.T. Rahman, K. Nakayama, M. Rahman, M. Ishikawa, H. Katagiri, A. Katagiri, T. Ishibashi, E. Sato, K. Iida, N. Ishikawa, N. Nakayama, K. Miyazaki, ESR1 gene amplification in endometrial carcinomas: a clinicopathological analysis, *Anticancer research*, 33 (2013) 3775-3781.
- [182] E. Curtit, V. Nerich, L. Mansi, L. Chaigneau, L. Cals, C. Villanueva, F. Bazan, P. Montcuquet, N. Meneveau, S. Perrin, M.-P. Algros, X. Pivot, Discordances in Estrogen Receptor Status, Progesterone Receptor Status, and HER2 Status Between Primary Breast Cancer and Metastasis, *The Oncologist*, 18 (2013) 667-674.
- [183] K. Ohtani, H. Sakamoto, T. Rutherford, Z. Chen, K. Satoh, F. Naftolin, Ezrin, a membrane-cytoskeletal linking protein, is involved in the process of invasion of endometrial cancer cells, *Cancer letters*, 147 (1999) 31-38.

- [184] M. Slater, M. Cooper, C.R. Murphy, The cytoskeletal proteins alpha-actinin, Ezrin, and talin are De-expressed in endometriosis and endometrioid carcinoma compared with normal uterine epithelium, *Applied immunohistochemistry & molecular morphology : AIMM / official publication of the Society for Applied Immunohistochemistry*, 15 (2007) 170-174.
- [185] K. Ohtani, H. Sakamoto, T. Rutherford, Z. Chen, A. Kikuchi, T. Yamamoto, K. Satoh, F. Naftolin, Ezrin, a membrane-cytoskeletal linking protein, is highly expressed in atypical endometrial hyperplasia and uterine endometrioid adenocarcinoma, *Cancer letters*, 179 (2002) 79-86.
- [186] H.H. Lee, C.A. Lim, Y.T. Cheong, M. Singh, L.H. Gam, Comparison of protein expression profiles of different stages of lymph nodes metastasis in breast cancer, *International journal of biological sciences*, 8 (2012) 353-362.
- [187] W. Zhu, L. Owusu, S. Zang, Y. Zhang, Y. Xin, C. Yan, GRP78 and GAL3, differentially regulated by lymph node homogenates, as potential biomarkers for lymph node metastasis in mouse hepatocellular carcinoma cells, *Oncology letters*, 4 (2012) 1374-1378.
- [188] C. Koumenis, ER stress, hypoxia tolerance and tumor progression, *Current molecular medicine*, 6 (2006) 55-69.
- [189] B. Luvsandagva, K. Nakamura, Y. Kitahara, H. Aoki, T. Murata, S. Ikeda, T. Minegishi, GRP78 induced by estrogen plays a role in the chemosensitivity of endometrial cancer, *Gynecologic oncology*, 126 (2012) 132-139.
- [190] N. Takaha, L.M. Resar, D. Vindivich, D.S. Coffey, High mobility group protein HMGI(Y) enhances tumor cell growth, invasion, and matrix metalloproteinase-2 expression in prostate cancer cells, *The Prostate*, 60 (2004) 160-167.
- [191] E.S. Nam, D.H. Kim, S.J. Cho, S.W. Chae, H.Y. Kim, S.M. Kim, J.J. Han, H.S. Shin, Y.E. Park, Expression of HMGI(Y) associated with malignant phenotype of human gastric tissue, *Histopathology*, 42 (2003) 466-471.
- [192] Y. Xu, X.D. Gao, J.H. Lee, H. Huang, H. Tan, J. Ahn, L.M. Reinke, M.E. Peter, Y. Feng, D. Gius, K.P. Siziopikou, J. Peng, X. Xiao, C. Cheng, Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing, *Genes & development*, 28 (2014) 1191-1203.
- [193] Y.Q. Ouyang, A. zur Hausen, M. Orłowska-Volk, M. Jager, H. Bettendorf, M. Hirschfeld, X.W. Tong, E. Stickeler, Expression levels of hnRNP G and hTra2-beta1 correlate with opposite outcomes in endometrial cancer biology, *International journal of cancer. Journal international du cancer*, 128 (2011) 2010-2019.
- [194] Y.D. Chai, L. Zhang, Y. Yang, T. Su, P. Charugundla, J. Ai, D. Messadi, D.T. Wong, S. Hu, Discovery of potential serum protein biomarkers for lymph node metastasis in oral cancer, *Head Neck*, 38 (2016) 118-125.
- [195] U. Tabassum, O. Reddy, G. Mukherjee, Elevated serum haptoglobin is associated with clinical outcome in triple-negative breast cancer patients, *Asian Pacific journal of cancer prevention : APJCP*, 13 (2012) 4541-4544.
- [196] L. Sun, S. Hu, L. Yu, C. Guo, L. Sun, Z. Yang, J. Qi, Y. Ran, Serum Haptoglobin as a novel molecular biomarker predicting colorectal cancer hepatic metastasis, *International journal of cancer. Journal international du cancer*, (2016).
- [197] H. Nabli, E. Tuller, K.L. Sharpe-Timms, Haptoglobin expression in endometrioid adenocarcinoma of the uterus, *Reproductive sciences (Thousand Oaks, Calif.)*, 17 (2010) 47-55.
- [198] J. Yan, Q. Wei, W. Jian, B. Qiu, J. Wen, J. Liu, B. Fu, X. Zhou, T. Zhao, IMP3 Predicts Invasion and Prognosis in Human Lung Adenocarcinoma, *Lung*, 194 (2016) 137-146.
- [199] Q. Wei, X. Huang, B. Fu, J. Liu, L. Zhong, Q. Yang, T. Zhao, IMP3 expression in biopsy specimens of colorectal cancer predicts lymph node metastasis and TNM stage, *International journal of clinical and experimental pathology*, 8 (2015) 11024-11032.
- [200] C.C. Pasiliao, C.W. Chang, B.W. Sutherland, S.M. Valdez, D. Schaeffer, D.T. Yapp, S.S. Ng, The involvement of insulin-like growth factor 2 binding protein 3 (IMP3) in pancreatic cancer cell migration, invasion, and adhesion, *BMC Cancer*, 15 (2015) 266.
- [201] Y.S. Sheen, Y.H. Liao, M.H. Lin, C.Y. Chu, B.Y. Ho, M.C. Hsieh, P.C. Chen, S.T. Cha, Y.M. Jeng, C.C. Chang, H.C. Chiu, S.H. Jee, M.L. Kuo, C.Y. Chu, IMP-3 promotes migration and invasion of melanoma cells by modulating the expression of HMGA2 and predicts poor prognosis in melanoma, *The Journal of investigative dermatology*, 135 (2015) 1065-1073.

- [202] P. Su, J. Hu, H. Zhang, W. Li, M. Jia, X. Zhang, X. Wu, H. Cheng, L. Xiang, G. Zhou, IMP3 expression is associated with epithelial-mesenchymal transition in breast cancer, *International journal of clinical and experimental pathology*, 7 (2014) 3008-3017.
- [203] O. Fadare, W. Zheng, Insights into endometrial serous carcinogenesis and progression, *International journal of clinical and experimental pathology*, 2 (2009) 411-432.
- [204] J. Zhang, F. Yang, Y. Ding, L. Zhen, X. Han, F. Jiao, J. Tang, Overexpression of L1 cell adhesion molecule correlates with aggressive tumor progression of patients with breast cancer and promotes motility of breast cancer cells, *International journal of clinical and experimental pathology*, 8 (2015) 9240-9247.
- [205] X. Yu, F. Yang, D.L. Fu, C. Jin, L1 cell adhesion molecule as a therapeutic target in cancer, *Expert review of anticancer therapy*, 16 (2016) 359-371.
- [206] D.H. Suh, M.A. Kim, H.S. Kim, H.H. Chung, N.H. Park, Y.S. Song, S.B. Kang, L1 cell adhesion molecule expression is associated with pelvic lymph node metastasis and advanced stage in diabetic patients with endometrial cancer: a matched case control study, *Journal of cancer prevention*, 19 (2014) 231-239.
- [207] M. Fogel, M. Huszar, P. Altevogt, A. Ben-Arie, L1 (CD171) as a novel biomarker for ovarian and endometrial carcinomas, *Expert review of molecular diagnostics*, 4 (2004) 455-462.
- [208] T.H. Dellinger, D.D. Smith, C. Ouyang, C.D. Warden, J.C. Williams, E.S. Han, L1CAM is an independent predictor of poor survival in endometrial cancer — An analysis of The Cancer Genome Atlas (TCGA), *Gynecologic oncology*.
- [209] H. Kashima, R.C. Wu, Y. Wang, A.K. Sinno, T. Miyamoto, T. Shiozawa, T.L. Wang, A.N. Fader, M. Shih Ie, Laminin C1 expression by uterine carcinoma cells is associated with tumor progression, *Gynecologic oncology*, 139 (2015) 338-344.
- [210] S. Yamasaki, Y. Miura, J. Davydova, S.M. Vickers, M. Yamamoto, Intravenous genetic mesothelin vaccine based on human adenovirus 40 inhibits growth and metastasis of pancreatic cancer, *International journal of cancer. Journal international du cancer*, 133 (2013) 88-97.
- [211] Y. Wang, L. Wang, D. Li, H.B. Wang, Q.F. Chen, Mesothelin promotes invasion and metastasis in breast cancer cells, *The Journal of international medical research*, 40 (2012) 2109-2116.
- [212] R.A. Soslow, S.L. Wethington, M. Cesari, D. Chiappetta, N. Olvera, J. Shia, D.A. Levine, Clinicopathologic analysis of matched primary and recurrent endometrial carcinoma, *Am J Surg Pathol*, 36 (2012) 1771-1781.
- [213] J.W. Huh, H.C. Kim, S.H. Kim, Y.A. Park, Y.B. Cho, S.H. Yun, W.Y. Lee, H.K. Chun, Mismatch repair system and p53 expression in patients with T1 and T2 colorectal cancer: predictive role of lymph node metastasis and survival, *Journal of surgical oncology*, 109 (2014) 848-852.
- [214] W. Cheng, X. Li, C. Quan, X. Guo, [Role of P53, nm23 proteins and vascular endothelial growth factor in angiogenesis and metastasis of laryngeal cancer], *Zhonghua er bi yan hou ke za zhi*, 35 (2000) 147-149.
- [215] H.Y. Lou, Q. Lin k, D.F. Ye, X. Xie, X. Yu, [Possibility of PTEN expression on predicting pathologic risk factors of endometrioid adenocarcinoma before operation], *Ai zheng = Aizheng = Chinese journal of cancer*, 24 (2005) 748-750.
- [216] Y. Xi, Y. Chen, Oncogenic and Therapeutic Targeting of PTEN Loss in Bone Malignancies, *Journal of cellular biochemistry*, 116 (2015) 1837-1847.
- [217] A. Doll, M. Gonzalez, M. Abal, M. Llauro, M. Rigau, E. Colas, M. Monge, J. Xercavins, G. Capella, B. Diaz, A. Gil-Moreno, F. Alameda, J. Reventos, An orthotopic endometrial cancer mouse model demonstrates a role for RUNX1 in distant metastasis, *International journal of cancer. Journal international du cancer*, 125 (2009) 257-263.
- [218] T. Ohga, K. Koike, M. Ono, Y. Makino, Y. Itagaki, M. Tanimoto, M. Kuwano, K. Kohno, Role of the human Y box-binding protein YB-1 in cellular sensitivity to the DNA-damaging agents cisplatin, mitomycin C, and ultraviolet light, *Cancer research*, 56 (1996) 4224-4228.
- [219] T. Ise, G. Nagatani, T. Imamura, K. Kato, H. Takano, M. Nomoto, H. Izumi, H. Ohmori, T. Okamoto, T. Ohga, T. Uchiumi, M. Kuwano, K. Kohno, Transcription factor Y-box binding protein 1 binds preferentially to cisplatin-modified DNA and interacts with proliferating cell nuclear antigen, *Cancer research*, 59 (1999) 342-346.
- [220] J.C.M. de Freitas Junior, J.A. Morgado-Díaz, The role of N-glycans in colorectal cancer progression: potential biomarkers and therapeutic applications, *Oncotarget*, 7 (2016) 19395-19413.

- [221] G.A. Khoury, R.C. Baliban, C.A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database, *Scientific reports*, 1 (2011).
- [222] R. Peracaula, S. Barrabes, A. Sarrats, P.M. Rudd, R. de Llorens, Altered glycosylation in tumours focused to cancer diagnosis, *Disease markers*, 25 (2008) 207-218.
- [223] D.H. Dube, C.R. Bertozzi, Glycans in cancer and inflammation--potential for therapeutics and diagnostics, *Nature reviews. Drug discovery*, 4 (2005) 477-488.
- [224] M.M. Fuster, J.D. Esko, The sweet and sour of cancer: glycans as novel therapeutic targets, *Nat Rev Cancer*, 5 (2005) 526-542.
- [225] Varki, ACR.; Esko, JD.; Freeze, HH.; Stanley, P.; Bertozzi, CR.; Hart, GW.; Etzler, ME. Cold Spring Harbor Laboratory Press 2009. Vol. 2. 2009. *Essentials of Glycobiology*. Comprehensive book of principles and applications of glycobiology. It discusses the different classes of glycans and the glycosylation changes observed in diseases.
- [226] P.H. Jensen, N.G. Karlsson, D. Kolarich, N.H. Packer, Structural analysis of N- and O-glycans released from glycoproteins, *Nat Protoc*, 7 (2012) 1299-1310.
- [227] M.L.A. de Leoz, H.J. An, S. Kronewitter, J. Kim, S. Beecroft, R. Vinall, S. Miyamoto, R. de Vere White, K.S. Lam, C. Lebrilla, Glycomic Approach for Potential Biomarkers on Prostate Cancer: Profiling of N-Linked Glycans in Human Sera and pRNS Cell Lines, *Disease markers*, 25 (2008) 243-258.
- [228] C. Kirmiz, B. Li, H.J. An, B.H. Clowers, H.K. Chew, K.S. Lam, A. Ferrige, R. Alecio, A.D. Borowsky, S. Sulaimon, C.B. Lebrilla, S. Miyamoto, A serum glycomics approach to breast cancer biomarkers, *Molecular & cellular proteomics : MCP*, 6 (2007) 43-55.
- [229] K. Kim, L.R. Ruhaak, U.T. Nguyen, S.L. Taylor, L. Dimapasoc, C. Williams, C. Stroble, S. Ozcan, S. Miyamoto, C.B. Lebrilla, G.S. Leiserowitz, Evaluation of glycomic profiling as a diagnostic biomarker for epithelial ovarian cancer, *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 23 (2014) 611-621.
- [230] Stanley P, Schachter H, Taniguchi N. N-Glycans. In: Varki A, Cummings RD, Esko JD, et al., editors. *Essentials of Glycobiology*. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. Chapter 8. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK1917/>.
- [231] T.W. Powers, B.A. Neely, Y. Shao, H. Tang, D.A. Troyer, A.S. Mehta, B.B. Haab, R.R. Drake, MALDI Imaging Mass Spectrometry Profiling of N-Glycans in Formalin-Fixed Paraffin Embedded Clinical Tissue Blocks and Tissue Microarrays, *PloS one*, 9 (2014) e106255.
- [232] O.J. Gustafsson, M.T. Briggs, M.R. Condina, L.J. Winderbaum, M. Pelzing, S.R. McColl, A.V. Everest-Dass, N.H. Packer, P. Hoffmann, MALDI imaging mass spectrometry of N-linked glycans on formalin-fixed paraffin-embedded murine kidney, *Analytical and bioanalytical chemistry*, 407 (2015) 2127-2139.
- [233] C.A. Cooper, E. Gasteiger, N.H. Packer, GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data, *Proteomics*, 1 (2001) 340-349.
- [234] A.V. Everest-Dass, M.T. Briggs, G. Kaur, M.K. Oehler, P. Hoffmann, N.H. Packer, N-glycan MALDI Imaging Mass Spectrometry on Formalin-Fixed Paraffin-Embedded Tissue Enables the Delineation of Ovarian Cancer Tissues, *Molecular & cellular proteomics : MCP*, 15 (2016) 3003-3016.
- [235] B. Domon, C.E. Costello, A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates, *Glycoconjugate Journal*, 5 (1988) 397-409.
- [236] D.J. Harvey, A.H. Merry, L. Royle, M.P. Campbell, R.A. Dwek, P.M. Rudd, Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds, *Proteomics*, 9 (2009) 3796-3801.
- [237] M.L.A. de Leoz, L.J.T. Young, H.J. An, S.R. Kronewitter, J. Kim, S. Miyamoto, A.D. Borowsky, H.K. Chew, C.B. Lebrilla, High-Mannose Glycans are Elevated during Breast Cancer Progression, *Molecular & cellular proteomics : MCP*, 10 (2011) M110.002717.
- [238] M.K. Sethi, M. Thaysen-Andersen, J.T. Smith, M.S. Baker, N.H. Packer, W.S. Hancock, S. Fanayan, Comparative N-Glycan Profiling of Colorectal Cancer Cell Lines Reveals Unique Bisecting GlcNAc and α -2,3-Linked Sialic Acid Determinants Are Associated with Membrane Proteins of the More Metastatic/Aggressive Cell Lines, *Journal of proteome research*, 13 (2014) 277-288.

- [239] A. Amin, A. Bashir, N. Zaki, D. McCarthy, S. Ahmed, M. Lotfy, Insights into glycan biosynthesis in chemically-induced hepatocellular carcinoma in rats: A glycomic analysis, *World journal of gastroenterology : WJG*, 21 (2015) 6167-6179.
- [240] M. Ballester, G. Canlorbe, A. Cortez, J. Gonin, E. Laas, S. Bendifallah, O. Graesslin, E. Daraï, Histological and immunohistochemical profiles predict lymph node status in women with low-intermediate risk endometrial cancer, *Gynecologic oncology*, 130 (2013) 457-462.
- [241] J.J. Chee, T.H. Ho, E.H. Tay, J.J. Low, K.L. Yam, Endometrioid adenocarcinoma of the uterus: surgico-pathological correlations and role of pelvic lymphadenectomy, *Annals of the Academy of Medicine, Singapore*, 32 (2003) 670-675.
- [242] P. Mittal, M. Klingler-Hoffmann, G. Arentz, L. Winderbaum, G. Kaur, L. Anderson, J. Scurry, Y. Leung, C.J. Stewart, J. Carter, P. Hoffmann, M.K. Oehler, Annexin A2 and alpha actinin 4 expression correlates with metastatic potential of primary endometrial cancer, *Biochimica et biophysica acta*, (2016).
- [243] G. Arentz, P. Mittal, C. Zhang, Y.Y. Ho, M. Briggs, L. Winderbaum, M.K. Hoffmann, P. Hoffmann, Chapter Two - Applications of Mass Spectrometry Imaging to Cancer, in: R.D. Richard, A.M. Liam (Eds.) *Advances in Cancer Research*, vol. Volume 134, Academic Press, 2017, pp. 27-66.
- [244] M. Nipp, M. Elsner, B. Balluff, S. Meding, H. Sarioglu, M. Ueffing, S. Rauser, K. Unger, H. Höfler, A. Walch, H. Zitzelsberger, S100-A10, thioredoxin, and S100-A6 as biomarkers of papillary thyroid carcinoma with lymph node metastasis identified by MALDI Imaging, *Journal of Molecular Medicine*, 90 (2011) 163-174.
- [245] J.H. Choi, N.R. Shin, H.J. Moon, C.H. Kwon, G.H. Kim, G.A. Song, T.Y. Jeon, D.H. Kim, D.H. Kim, Y. Park do, Identification of S100A8 and S100A9 as negative regulators for lymph node metastasis of gastric adenocarcinoma, *Histology and histopathology*, 27 (2012) 1439-1448.
- [246] M. Elsner, S. Rauser, S. Maier, C. Schöne, B. Balluff, S. Meding, G. Jung, M. Nipp, H. Sarioglu, G. Maccarrone, M. Aichler, A. Feuchtinger, R. Langer, U. Jütting, M. Feith, B. Küster, M. Ueffing, H. Zitzelsberger, H. Höfler, A. Walch, MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma, *Journal of Proteomics*, 75 (2012) 4693-4704.
- [247] K. Yanagisawa, Y. Shyr, B.J. Xu, P.P. Massion, P.H. Larsen, B.C. White, J.R. Roberts, M. Edgerton, A. Gonzalez, S. Nadaf, J.H. Moore, R.M. Caprioli, D.P. Carbone, Proteomic patterns of tumour subsets in non-small-cell lung cancer, *The Lancet*, 362 (2003) 433-439.
- [248] H. Guan, A. Semaan, S. Bandyopadhyay, H. Arabi, J. Feng, L. Fathallah, V. Pansare, A. Qazi, F. Abdul-Karim, R.T. Morris, Prognosis and reproducibility of new and existing binary grading systems for endometrial carcinoma compared to FIGO grading in hysterectomy specimens, *International Journal of Gynecological Cancer*, 21 (2011) 654-660.
- [249] S. Bendifallah, G. Canlorbe, P. Collinet, E. Arsene, F. Huguet, C. Coutant, D. Hudry, O. Graesslin, E. Raimond, C. Touboul, E. Daraï, M. Ballester, Just how accurate are the major risk stratification systems for early-stage endometrial cancer[quest], *Br J Cancer*, 112 (2015) 793-801.
- [250] Y. Bian, C. Song, K. Cheng, M. Dong, F. Wang, J. Huang, D. Sun, L. Wang, M. Ye, H. Zou, An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome, *J Proteomics*, 96 (2014) 253-262.
- [251] Y. Liu, A. Beyer, R. Aebersold, On the Dependency of Cellular Protein Levels on mRNA Abundance, *Cell*, 165 (2016) 535-550.
- [252] B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control, *Nature*, 473 (2011) 337-342.
- [253] E. Lundberg, L. Fagerberg, D. Klevebring, I. Matic, T. Geiger, J. Cox, C. Algenas, J. Lundberg, M. Mann, M. Uhlen, Defining the transcriptome and proteome in three functionally different human cell lines, *Molecular systems biology*, 6 (2010) 450.
- [254] Q. Tian, S.B. Stepaniants, M. Mao, L. Weng, M.C. Feetham, M.J. Doyle, E.C. Yi, H. Dai, V. Thorsson, J. Eng, D. Goodlett, J.P. Berger, B. Gunter, P.S. Linseley, R.B. Stoughton, R. Aebersold, S.J. Collins, W.A. Hanlon, L.E. Hood, Integrated genomic and proteomic analyses of gene expression in Mammalian cells, *Molecular & cellular proteomics : MCP*, 3 (2004) 960-969.
- [255] C. Vogel, S. Abreu Rde, D. Ko, S.Y. Le, B.A. Shapiro, S.C. Burns, D. Sandhu, D.R. Boutz, E.M. Marcotte, L.O. Penalva, Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line, *Molecular systems biology*, 6 (2010) 400.

- [256] R. Lu, F. Markowetz, R.D. Unwin, J.T. Leek, E.M. Airolidi, B.D. MacArthur, A. Lachmann, R. Rozov, A. Ma'ayan, L.A. Boyer, O.G. Troyanskaya, A.D. Whetton, I.R. Lemischka, Systems-level dynamic analyses of fate change in murine embryonic stem cells, *Nature*, 462 (2009) 358-362.
- [257] B. Schwanhaussner, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Corrigendum: Global quantification of mammalian gene expression control, *Nature*, 495 (2013) 126-127.
- [258] N. Kendrick, A gene's mRNA level does not usually predict its protein level, Madison, WI: Kendrick Laboratories, (2014).
- [259] D.H. Margineantu, C.B. Emerson, D. Diaz, D.M. Hockenbery, Hsp90 inhibition decreases mitochondrial protein turnover, *PloS one*, 2 (2007) e1066.
- [260] A.M. Stark, S. Pfannenschmidt, H. Tscheslog, N. Maass, F. Rosel, H.M. Mehdorn, J. Held-Feindt, Reduced mRNA and protein expression of BCL-2 versus decreased mRNA and increased protein expression of BAX in breast cancer brain metastases: a real-time PCR and immunohistochemical evaluation, *Neurological research*, 28 (2006) 787-793.
- [261] S.M. Sarro, T.L. Unruh, J. Zuccolo, R. Sanyal, J.M. Luidier, I.A. Auer-Grzesiak, A. Mansoor, J.P. Deans, Quantification of CD20 mRNA and protein levels in chronic lymphocytic leukemia suggests a post-transcriptional defect, *Leukemia research*, 34 (2010) 1670-1673.
- [262] G.L. Hortin, Can Mass Spectrometric Protein Profiling Meet Desired Standards of Clinical Laboratory Practice?, *Clinical chemistry*, 51 (2005) 3-5.
- [263] T.J.A. Dekker, B.D. Balluff, E.A. Jones, C.D. Schöne, M. Schmitt, M. Aubele, J.R. Kroep, V.T.H.B.M. Smit, R.A.E.M. Tollenaar, W.E. Mesker, A. Walch, L.A. McDonnell, Multicenter Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging (MALDI MSI) Identifies Proteomic Differences in Breast-Cancer-Associated Stroma, *Journal of proteome research*, 13 (2014) 4730-4738.
- [264] C. Fenselau, P.A. Demirev, Characterization of intact microorganisms by MALDI mass spectrometry, *Mass spectrometry reviews*, 20 (2001) 157-171.
- [265] J.P. Anhalt, C. Fenselau, Identification of bacteria using mass spectrometry, *Analytical chemistry*, 47 (1975) 219-225.
- [266] M.L. Faron, B.W. Buchan, J. Hyke, N. Madisen, J.L. Lillie, P.A. Granato, D.A. Wilson, G.W. Procop, S. Novak-Weekley, E. Marlowe, J. Cumpio, C. Griego-Fullbright, S. Kindig, K. Timm, S. Young, N.A. Ledebore, Multicenter Evaluation of the Bruker MALDI Biotyper CA System for the Identification of Clinical Aerobic Gram-Negative Bacterial Isolates, *PloS one*, 10 (2015) e0141350.
- [267] L.A. McDonnell, P.M. Angel, S. Lou, R.R. Drake, Chapter Eleven - Mass Spectrometry Imaging in Cancer Research: Future Perspectives, in: R.D. Richard, A.M. Liam (Eds.) *Advances in Cancer Research*, vol. Volume 134, Academic Press, 2017, pp. 283-290.
- [268] M. Aichler, A. Walch, MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice, *Laboratory investigation; a journal of technical methods and pathology*, 95 (2015) 422-431.
- [269] T.M. Morgan, E.H. Seeley, O. Fadare, R.M. Caprioli, P.E. Clark, Imaging the clear cell renal cell carcinoma proteome, *The Journal of urology*, 189 (2013) 1097-1103.
- [270] E.I. Deryugina, J.P. Quigley, Matrix metalloproteinases and tumor metastasis, *Cancer metastasis reviews*, 25 (2006) 9-34.
- [271] E. Lambert, E. Dasse, B. Haye, E. Petitfrere, TIMPs as multifacial proteins, *Critical reviews in oncology/hematology*, 49 (2004) 187-198.
- [272] P. Lu, V.M. Weaver, Z. Werb, The extracellular matrix: a dynamic niche in cancer progression, *J Cell Biol*, 196 (2012) 395-406.
- [273] M.W. Pickup, J.K. Mouw, V.M. Weaver, The extracellular matrix modulates the hallmarks of cancer, *EMBO reports*, 15 (2014) 1243-1253.
- [274] A. Neve, F.P. Cantatore, N. Maruotti, A. Corrado, D. Ribatti, Extracellular Matrix Modulates Angiogenesis in Physiological and Pathological Conditions, *BioMed Research International*, 2014 (2014) 10.
- [275] D. Belotti, C. Foglieni, A. Resovi, R. Giavazzi, G. Taraboletti, Targeting angiogenesis with compounds from the extracellular matrix, *The international journal of biochemistry & cell biology*, 43 (2011) 1674-1685.

- [276] A. Naba, K.R. Clauser, C.A. Whittaker, S.A. Carr, K.K. Tanabe, R.O. Hynes, Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver, *BMC Cancer*, 14 (2014) 518.
- [277] N. Ogrinc Potocnik, T. Porta, M. Becker, R.M. Heeren, S.R. Ellis, Use of advantageous, volatile matrices enabled by next-generation high-speed matrix-assisted laser desorption/ionization time-of-flight imaging employing a scanning laser beam, *Rapid communications in mass spectrometry : RCM*, 29 (2015) 2195-2203.
- [278] M.R. Groseclose, S. Castellino, A Mimetic Tissue Model for the Quantification of Drug Distributions by MALDI Imaging Mass Spectrometry, *Analytical chemistry*, 85 (2013) 10099-10106.
- [279] D.A. Pirman, R.A. Yost, Quantitative Tandem Mass Spectrometric Imaging of Endogenous Acetyl-l-carnitine from Piglet Brain Tissue Using an Internal Standard, *Analytical chemistry*, 83 (2011) 8575-8581.
- [280] M.M. Gessel, J.L. Norris, R.M. Caprioli, MALDI imaging mass spectrometry: Spatial molecular analysis to enable a new age of discovery, *Journal of Proteomics*, 107 (2014) 71-82.