Technische Universität München

Institut für Experimentelle Genetik
(Institute of Experimental Genetics)

Identification of Clinical Markers in Colon Cancer by Tissue Based *in situ* Proteomics

Stephan Meding

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. D. Langosch
Prüfer der Dissertation:	1. apl. Prof. Dr. J. Adamski
	2. Univ.-Prof. Dr. B. Küster
	3. apl. Prof. Dr. A. Walch
	(Albert-Ludwigs-Universität Freiburg)

Die Dissertation wurde am 25.10.2011 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 16.02.2012 angenommen.

Table of Contents

1 Abstract

Colon cancer is one of the most frequent cancers. Its carcinogenesis and progression are well studied. However, several clinically relevant aspects are still incompletely understood. For this, human tissue cohorts were analysed by mass spectrometric methods and validated by immunohistochemical tissue staining. Three clinically relevant aspects of colon cancer were addressed: regional lymph node metastasis, relapse, and tumour classification. For identifying proteomic markers associated with regional lymph node metastasis, a cohort of 54 primary colon tumour tissues was analysed by MALDI Imaging and label-free quantitative proteomics. This resulted in 38 discriminating proteins. Three of them were validated by immunohistochemical staining on an independent cohort of 168 primary colon tumour tissues: FXYD3 ($p=0.0110$), S100A11 ($p=0.0071$) and GSTM3 ($p=0.0173$). Addressing the problem of carcinomas of unknown primary, a proof-of-principle study of proteomic tumour classification was performed. At first, a cohort comprising 171 tumour tissues from 6 different adenocarcinoma entities was analysed by MALDI Imaging. The resulting proteomic patterns were used for highly accurate classification (>80% accuracy). Then, a cohort of 55 tumour tissues comprising primary colon cancer, its liver metastasis, and primary hepatocellular carcinoma were also classified with high accuracy (>80%). For identifying proteomic markers for relapse in colon cancer, a cohort of 119 primary tumour tissues was analysed by MALDI Profiling. This resulted in 27 differentially expressed *m/z* species. 13 had a p-value below 0.05, and 14 had a p-value below 0.1. 8 of them could be identified: CRIP1, DEFA1, DEFA2, HBA1, MRPL14, S100A8, S100A9, and S100P. Additionally, methodological improvements of MALDI Imaging were made. Improved protocols for MALDI Imaging on formalin-fixed and alcohol-fixed paraffin-embedded tissues were established. In this thesis, several clinically relevant aspects of colon cancer were addressed. The resultant findings could support new, personalised approaches in diagnostics and therapy.

Abstract (deutsch)

Das Kolonkarzinom gehört zu den häufigsten Krebserkrankungen. Seine Karzinogenese und Progression sind gut untersucht. Leider sind mehrere klinische Aspekte noch unvollständig verstanden. Deshalb wurden humane Gewebeproben massenspektrometrisch analysiert und mit immunhistochemischen Färbungen validiert. Drei klinisch relevante Aspekte des Kolonkarzinoms wurden untersucht: die lymphoregionäre Metastasierung, die Rezidivierung und die Tumorklassifikation. Um mit lymphoregionärer Metastasierung assoziierte proteomische Marker zu identifizieren, wurde ein Kollektiv aus 54 Primärtumoren des Kolonkarzinoms mit MALDI Imaging und markierungsfreier, quantitativer Proteomik untersucht. Dies führte zu 38 diskriminierenden Proteinen. Drei dieser wurden mit immunhistochemischer Färbung an einem unabhängigen Kollektiv aus 168 Primärtumoren des Kolonkarzinoms validiert: FXYD3 (p=0.0110), S100A11 (p=0.0071) und GSTM3 (p=0.0173) Um die klinische Fragestellung des Karzinoms unbekannter Herkunft zu untersuchen, wurde eine “proof-of-principle” Studie zur proteomischen Tumorklassifikation durchgeführt. Zuerst wurde ein Kollektiv bestehend aus 171 Tumorproben sechs verschiedener Adenokarzinome mit MALDI Imaging analysiert. Die erhaltenen proteomischen Muster ermöglichten eine Klassifikation mit einer Genauigkeit von mehr als 80%. Dann wurde ein Kollektiv aus 55 Tumorproben des Primärtumors und der Lebermetastasen des Kolonkarzinoms, sowie des hepatozellulären Karzinoms ebenfalls mit einer Genauigkeit von mehr als 80% klassifiziert. Um proteomische Marker für eine Rezidivierung beim Kolonkarzinom zu identifizieren wurden 119 Primärtumore mit MALDI Profiling untersucht. Dabei konnten 27 differentiell exprimierte *m/z* Spezies gefunden werden, 13 mit einem p-Wert kleiner 0,05 und 14 mit einem p-Wert kleiner 0,1. 8 konnten identifiziert werden: CRIP1, DEFA1, DEFA2, HBA1, MRPL14, S100A8, S100A9 und S100P. Zusätzlich wurde die Methode MALDI Imaging verbessert. Bessere Verfahren zur Analyse Formalin

fixierter und Alkohol fixierter in Paraffin eingebetteter Gewebe wurden entwickelt. In dieser Arbeit wurden mehrere klinisch relevante Aspekte beim Kolonkarzinom bearbeitet. Die Ergebnisse könnten neue, personalisierte Ansätze in Diagnostik und Therapie unterstützen.

2 Introduction

2.1 Impact of cancer

Regarding noncommunicable diseases, cancer is the leading cause of death in economically developed countries and the second leading cause of death in developing countries.[1] About 12.7 million cancer cases and 7.6 million cancer deaths are estimated to have occurred in 2008, worldwide.[2] Only cardiovascular and infectious diseases pose a greater health risk (**Figure 1**).

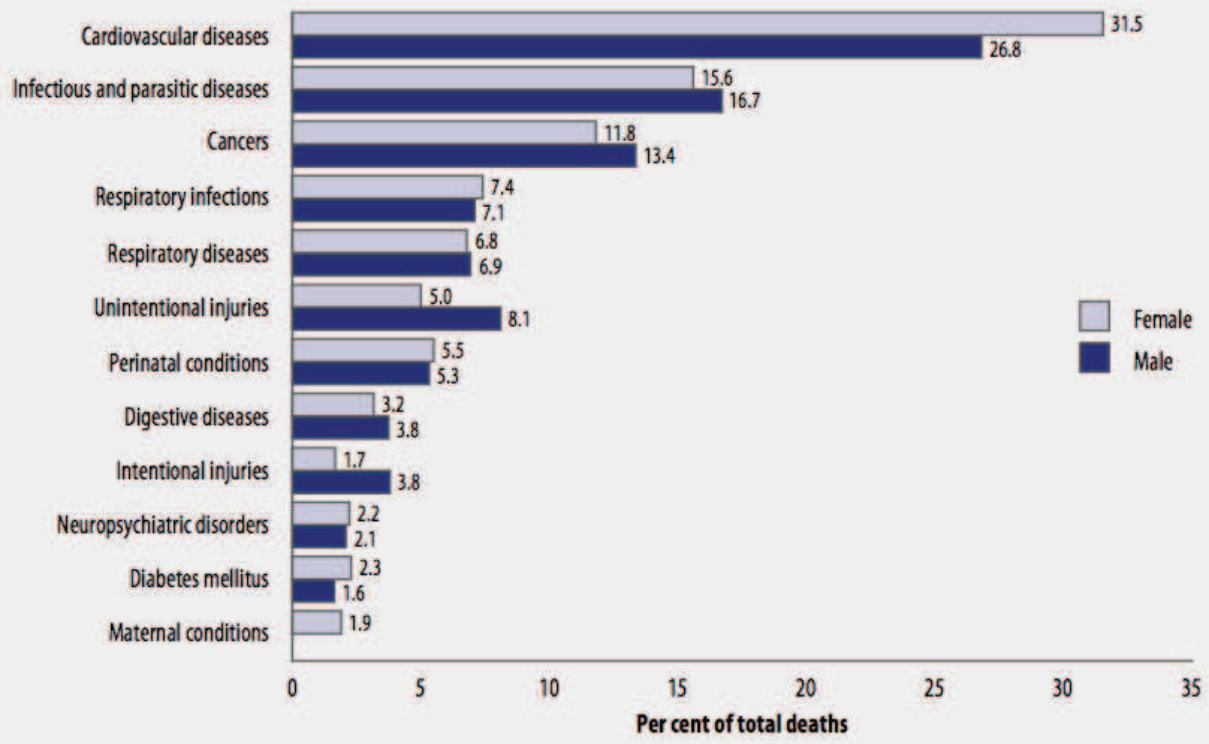


Figure 1 – Leading causes of death, worldwide in 2004.[1]

The burden of cancer is increasing in economically developing and developed countries as a result of population aging and an adoption of cancer-associated lifestyle choices including smoking, physical inactivity and "westernized" diet.[2] Regarding incidence and mortality colorectal cancer is amongst the top three, worldwide and in developed countries (**Figure 2**). In Europe, colorectal cancer is the most common cause of cancer deaths, accounting for

approximately 436,000 incident cases and 212,000 deaths in 2008.[3] Rates of rectal cancer are about 50% higher and colon cancer rates about 20% higher in men than in woman.[4]

Figure 2 – Estimated new cancer cases and deaths worldwide and in developed countries in 2008.[2]

2.2 Definition of cancer

The terms “tumour” and “cancer” are often used as synonyms. However, all cancers are tumours but not vice versa. In general, tumours are neoplastic tissue lesions. This means that their cells underwent genetic alterations changing their fundamental physiological properties. They are able to proliferate independently and indefinitely. Tumours can be either benign which indicates that they remain in the tissue of origin and lack the capability to invade the surrounding tissue or spread to other tissues and organs, or malignant which means that they are invasive and have the potential to spread. Once tumours have become malignant they are termed cancers.[5]

2.3 Carcinogenesis – The development of cancer

Cancers do not occur spontaneously. They usually develop over years and sometimes decades until they become clinically overt. During this time they gradually develop from normal cells into neoplastic lesions. Every cancer and its carcinogenesis are unique. Nevertheless, most cancers share common features and progress by similar steps.[5]

2.3.1 Common features of carcinogenesis

Before genomic alterations occur there is often a hyperplasia of the tissue. It is induced by a dysregulated, increased cell proliferation. Yet, there are no DNA mutations. This hyperplasia is mostly induced by inflammation. It can develop into a neoplasia but it can also regress and become normal tissue again.

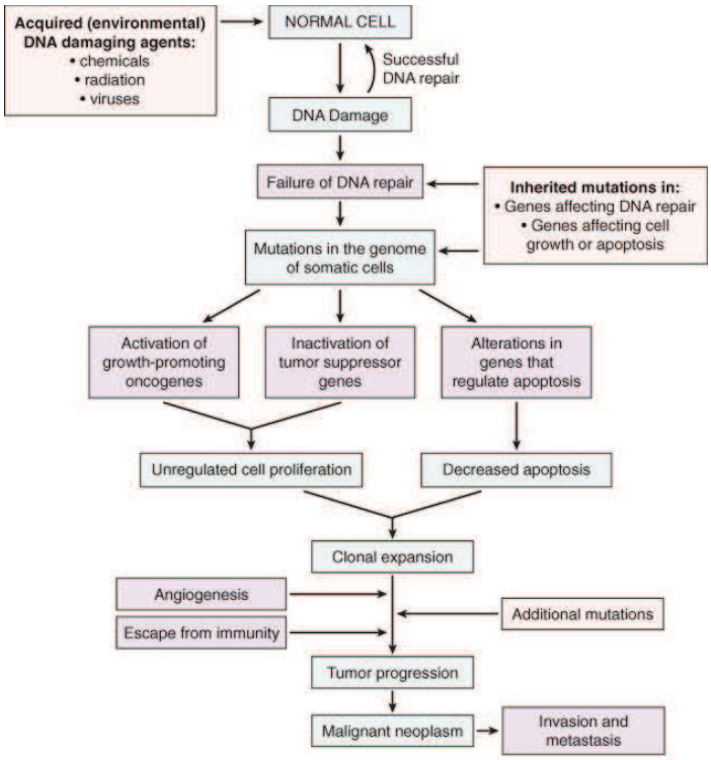


Figure 3 – Molecular basis of cancer.[5]

Once a cell acquires a persistent genomic alteration the carcinogenesis becomes irreversible (**Figure 3**). However, it is indetermined whether the cell will ever reach the stage of a cancer cell. Mutations are rare, random events. An accumulation of mutations within oncogenes, tumour suppressor genes, apoptosis related genes or DNA repair genes is very improbable.

Most cells will die off or remain at an early stage of carcinogenesis. Several factors can promote the carcinogenesis. Hereditary predisposition, such as mutations in DNA repair genes, exposition to carcinogenic substances, ionizing radiation or certain retroviruses can lead to an increased mutation frequency. Once a cell has acquired enough mutations to become a tumour cell it will be driven by selective pressure into increased malignancy.[5] Hanahan and Weinberg suggested six hallmarks of cancer which are alterations in cell physiology that collectively dictate malignant growth: self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (**Figure 4**).[6]

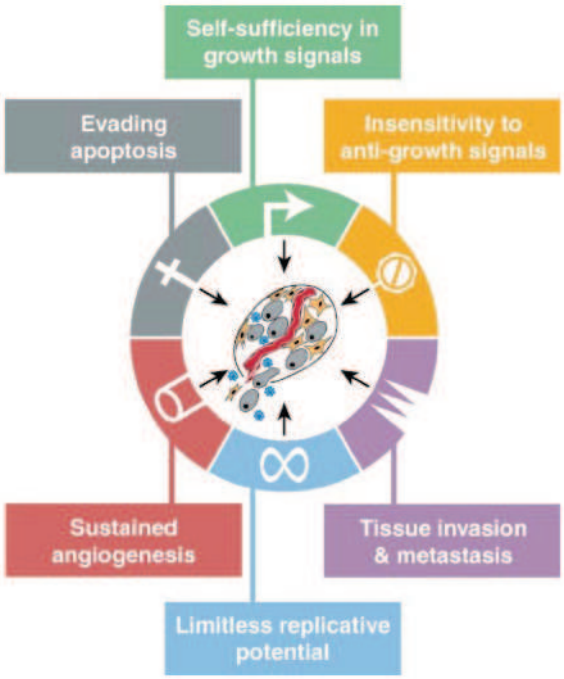


Figure 4 – Hallmarks of cancer.[6]

Normal cells require mitogenic growth signals before they move from a quiescent state into an active proliferative state.[6] Cancer cells have acquired self-sufficiency in growth signals. This can be achieved by several means, for instance by autocrine production of growth factors, by overexpression of or mutations in growth factor receptors, by mutations in signal-transducing proteins or by mutations in transcription factors.[5]

Within a normal tissue, multiple antiproliferative signals operate to maintain cellular quiescence and tissue homeostasis.[6] Antigrowth signals can block proliferation by two distinct mechanisms: by forcing cells into the quiescent (G_0) state from which they may reemerge on some future occasion when extracellular signals permit or by inducing postmitotic states, which are usually associated with the acquisition of specific, differentiation associated traits, that permanently relinquish the proliferative potential of the cells.[6] Insensitivity to antigrowth signals is induced by mutations in tumour suppressor genes and prevents tumour cells from going into an unproliferative state.[5]

The ability of tumour cell populations to expand in number is determined not only by the rate of cell proliferation but also by the rate of cell attrition. Programmed cell death – apoptosis – represents a major source of this attrition.[6] Thus, evading apoptosis is essential for successful expansion of neoplastic cells.

Many, perhaps all types of mammalian cells carry an intrinsic, cell-autonomous program that limits their multiplication.[6] It too must be disrupted in order for a clone of cells to expand to a size which constitutes a macroscopic, life-threatening tumour.[6] Nearly all cancers acquire this limitless replicative potential by upregulation of telomerase expression.[5]

The oxygen and nutrients supplied by the vasculature are crucial for cell function and survival, obligating virtually all cells in a tissue to reside within 100 μm of a capillary blood vessel.[6] Tumours cannot enlarge beyond 1 to 2 mm in diameter unless they are vascularised. Like normal tissues, tumours require delivery of oxygen and nutrients and removal of waste products; presumably the 1 to 2 mm zone represents the maximal distance across which oxygen, nutrients, and waste can diffuse from or to blood vessels.[5] Therefore, inducing angiogenesis is necessary for sustained tumour growth. However, tumour vasculature is abnormal. The vessels are leaky, dilated and have a haphazard pattern of connection.[5]

The capability for invasion and metastasis enables cancer cells to escape the primary tumour mass and colonise new terrain in the body where, at least initially, nutrients and space are not

limiting.[6] By this, the cancer can become a systemic disease. Two aspects are important for tumour spreading: firstly, the invasion of the extra cellular matrix and the successful penetration of several barriers, such as the basement membranes of the host tissue, of the vasculature and of the graft tissue; secondly, the vascular dissemination and the homing of the tumour cells to the target organ. Many tumours metastasise to the organ that represents the first capillary bed they encounter after entering the circulation. However, several tumours, such as lung cancer, spread differently. Thus, it is likely that homing of tumour cells is also regulated by cell adhesion molecules, chemokines and their receptors.[5]

2.3.2 Carcinogenesis of colorectal cancer

The carcinogenesis of colorectal cancer is intensely studied and described. In 1990, Fearon and Vogelstein proposed a model for the carcinogenesis of colorectal carcinoma.[7] They proposed that mutations in at least four to five genes are required for the formation of a malignant tumour.[7] Their model has been refined and several new oncogenes and tumour suppressor genes involved in the carcinogenesis have been identified since then.[8] Nevertheless, the core principle of the proposed model is still valid today and is termed the adenoma – carcinoma sequence (**Figure 5**). This pathway of carcinogenesis accounts for about 80% of sporadic colon tumours.[5]

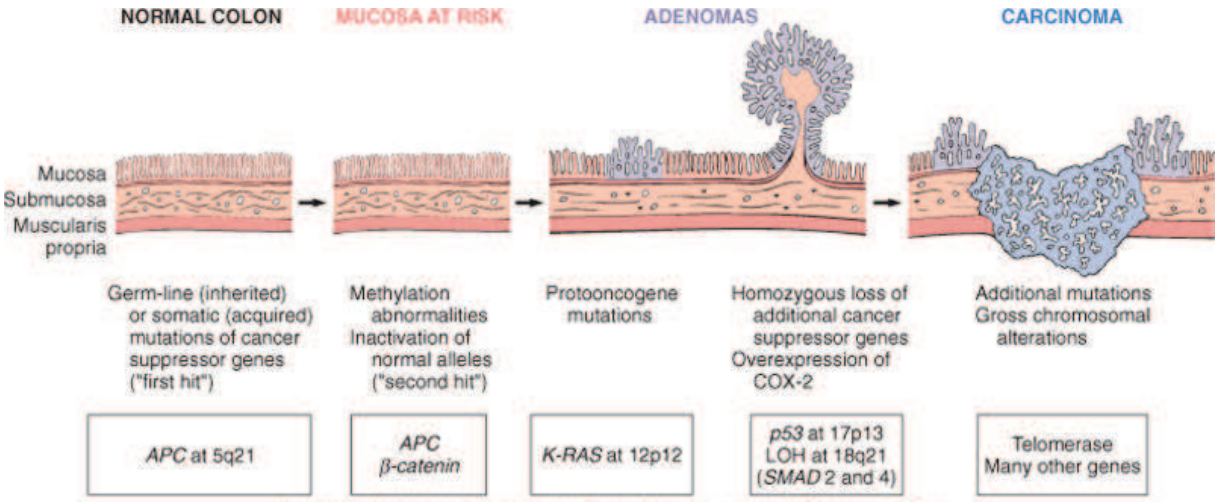


Figure 5 – The adenoma – carcinoma sequence in colorectal cancer carcinogenesis.[5]

A second pathway of carcinogenesis in colorectal cancer has been identified. It is involved in 10% to 15% of sporadic cases.[5] It is termed serrated or mismatch repair pathway since sessile serrated adenomas are an important intermediate step in this carcinogenesis, and since defects in DNA repair are key features of the pathway (**Figure 6**).[9]

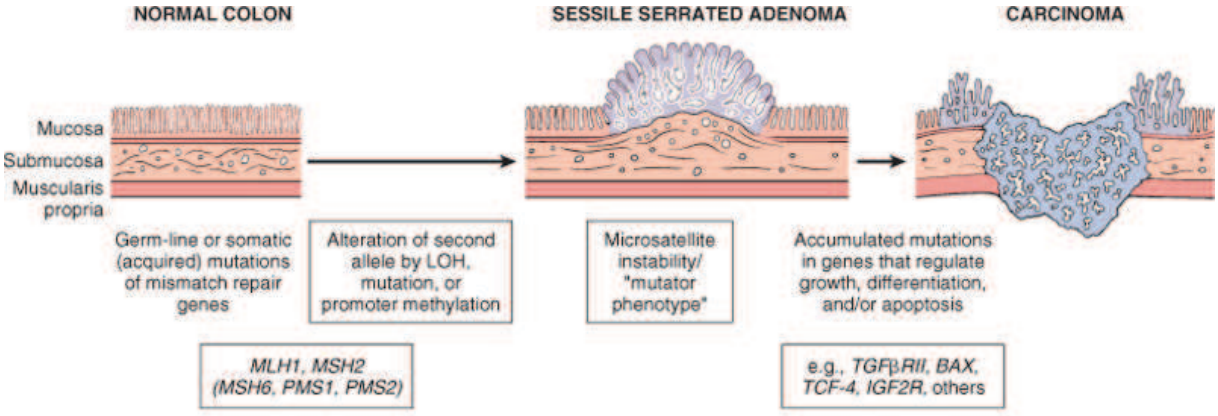


Figure 6 – The serrated pathway in colorectal cancer carcinogenesis.[5]

2.4 Classification and staging of colon cancers

Cancer treatment relies upon the accurate description of the respective tumour. For this, a generally accepted and used classification system is needed. The TNM System for the classification of malignant tumours was developed by Pierre Denoix (France) between the years 1943 and 1952.[10] This system was expanded and refined over the next decades by the International Union Against Cancer (UICC). The present seventh edition of the TNM classification contains rules of classification and staging that correspond with those appearing

in the seventh edition of the AJCC (American Joint Committee on Cancer) Cancer Staging Manual (2009)[11], and that have approval of all national TNM committees.[10] The central aim of the TNM system for classification is the unambiguous description of tumours and cancers in respect to their histopathology and anatomical extent. The TNM system is based on three components: (i) the extent of the primary tumour (T), (ii) the absence or presence and extent of regional lymph node metastasis (N) and (iii) the absence or presence of distant metastasis (M) (**Table 1**).[10]

T – Primary tumour		
TX	Primary tumour cannot be assessed	
T0	No evidence of primary tumour	
Tis	Carcinoma in situ: intraepithelial or invasion of lamina propria	
T1	Tumour invades submucosa	
T2	Tumour invades muscularis propria	
T3	Tumour invades subserosa or into non-peritonealised pericolic or perirectal tissue	
T4	Tumour directly invades other organs or structures and/or perforates visceral peritoneum	
	T4a	Tumour perforates visceral peritoneum
	T4b	Tumour directly invades other organs or structures
N – Regional lymph nodes		
NX	Regional lymph nodes cannot be assessed	
N0	No regional lymph node metastasis	
N1	Metastasis in 1 – 3 regional lymph nodes	
	N1a	Metastasis in 1 regional lymph node
	N1b	Metastasis in 2 – 3 regional lymph nodes
	N1c	Tumour deposit(s), i.e., satellites, in the subserosa, or in non-peritonealised pericolic or perirectal soft tissue without regional lymph node metastasis
N2	Metastasis in 4 or more regional lymph nodes	
	N2a	Metastasis in 4 – 6 regional lymph nodes
	N2b	Metastasis in 7 or more regional lymph nodes
M – Distant metastasis		
M0	No distant metastasis	
M1	Distant metastasis	
	M1a	Metastasis confined to one organ (liver, lung, ovary, non-regional lymph node(s))
	M1b	Metastasis in more than one organ or the peritoneum

Table 1 – TNM classification for colon and rectum tumours.[10]

In addition to the TNM classification the primary tumours are also regularly characterised by their histopathological grading (G) which characterises cell differentiation of the cancer cells (**Table 2**).

G – Histopathological grading	
G1	Well differentiated
G2	Moderately differentiated
G3	Poorly differentiated
G4	Undifferentiated

Table 2 – Histopathological grading for colon and rectum tumours.[10]

Upon this TNM classification patients can be grouped into clinical stages, which have similar clinical characteristics and shared prognosis (**Table 3**).

Stage grouping (according to UICC)			
Stage 0	Tis	N0	M0
Stage I	T1, T2	N0	M0
Stage II	T3, T4	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T4a	N0	M0
Stage IIC	T4b	N0	M0
Stage III	Any T	N1, N2	M0
Stage IIIA	T1, T2 T1	N1 N2a	M0
Stage IIIB	T3, T4a T2, T3 T1, T2	N1 N2a N2b	M0
Stage IIIC	T4a T3, T4a T4b	N2a N2b N1, N2	M0
Stage IVA	Any T	Any N	M1a
Stage IVB	Any T	Any N	M1b

Table 3 – Stage grouping for colon and rectum tumours according to UICC.[10]

2.5 Clinical challenges in colon cancer therapy

Colon cancer can be treated efficiently by curative resection of the affected part of the colon if the tumour is diagnosed at an early stage. Unfortunately, most cancers are asymptomatic for many years and thus, are diagnosed at a relatively late stage.[12] For early detection of colon

cancer several tests, such as sigmoidoscopy, colonscopy or stool tests, have been introduced.[12-14] UICC stage I and UICC stage IIA patients (≤pT3N0) have a 5-year survival rate exceeding 80%.[15] In case of local metastasis the 5-year survival rate drops drastically. UICC stage IIIA patients have a 5-year survival rate of approximately 60%, while UICC stage IIIC patients have a 5-year survival rate below 30%.[14,16,17] If local metastasis is diagnosed, adjuvant (post operation) chemotherapy is regarded as beneficial for patients.[18-20] In patients without metastasis adjuvant chemotherapy is not regarded as being generally beneficial since these patients are mostly cured by tumour resection.[18-20] However, a significant fraction of UICC stage I (~5%) and UICC stage II (~15-25%) patients who underwent curative resection have a relapse.[21-24] Mostly, the relapse occurs within 5 years after curative resection in the liver, the local site, the abdomen, or the lung.[25] Patients with a high risk for developing a relapse would probably benefit from adjuvant chemotherapy.[22,26] Therefore, markers that could reliably predict the likelihood for tumour recurrence would be particularly helpful for UICC stage II patients.[21] Unfortunately, no reliable markers for predicting tumour recurrence are sufficiently validated for clinical application.[8,21,27] Once the cancer has spread to distant organ sites, prognosis becomes very poor (5-year survival rate <10%).[14,20] Then, it has to be decided whether a curative or a palliative treatment is adequate.[28] In both cases neoadjuvant chemotherapy and targeted therapy, such as α-EGFR and α-VEGF antibody treatment, are being employed.[19,28,29] Unfortunately, only very few markers, such as KRAS mutations, have been identified which can predict the therapy response.[8,27] So, patient treatment could be improved if colon cancer was better understood in respect to occurrence, progression, relapse and biological behaviour.

2.6 Methods applied in tissue analysis

Clinical problems can be addressed in many ways. The easiest way is looking at the patient's symptoms and by this deducing his ailment. However, for most diseases the symptoms are

either inconclusive or overt symptoms arise very late. Therefore, elaborate non-invasive and invasive methods have been developed. Many invasive methods result in tissue biopsy procurement and consecutive tissue analysis. In general, the invasive methods for tissue analysis can be divided into two groups. *In situ* techniques require intact tissue morphology, while the second group works with tissue lysates. *In situ* techniques use either specific dyes, antibodies for specific labelling with chromogenes and fluorophores, or specific DNA/RNA-probes.[30] By this, the spatial distribution of certain cellular features can be visualised within a tissue section. However, only very few features can be labelled at the same time and the features have to be known in advance. So, these methods lack the capability for extensive multiplexing and screening. Lysate based methods can analyse cellular components, such as DNA, RNA and proteins, in a directed but also in a screening approach. For proteomic analysis several techniques have been established. Especially mass spectrometric analyses have acquired great analytical depth which allows screening for a multitude of proteins and their modifications.[31] Nevertheless, it remains open where these proteins were located within the tissue. So, a combination of *in situ* techniques and proteomic mass spectrometry would be highly beneficial for tissue screening. With laser microdissection this combination is partially achieved. Tissue areas of interest can be excised and then analysed by lysate based techniques. Anyhow, laser microdissection is very time consuming, requires a lot of tissue material and the cellular content of the excised tissue cannot be verified afterwards.[32] These limitations are overcome by MALDI Imaging and with minor restrictions by MALDI Profiling.[33]

2.6.1 MALDI Imaging

Matrix-assisted laser desorption/ionisation (MALDI) Imaging (mass spectrometry) is a powerful tool for the analysis of a variety of different endogeneous and exogeneous molecules directly in tissue sections.[34] It is used for a variety of applications, including biomarker

research or the determination of protein, peptide, endogeneous metabolites and drug distribution within tissues.[35] For clinical research MALDI Imaging is capable of filling the gap between *in situ* techniques which require a defined target for labelling, and lysate based screening techniques which lack the direct relationship between tissue morphology and the expression of a specific protein.[32]

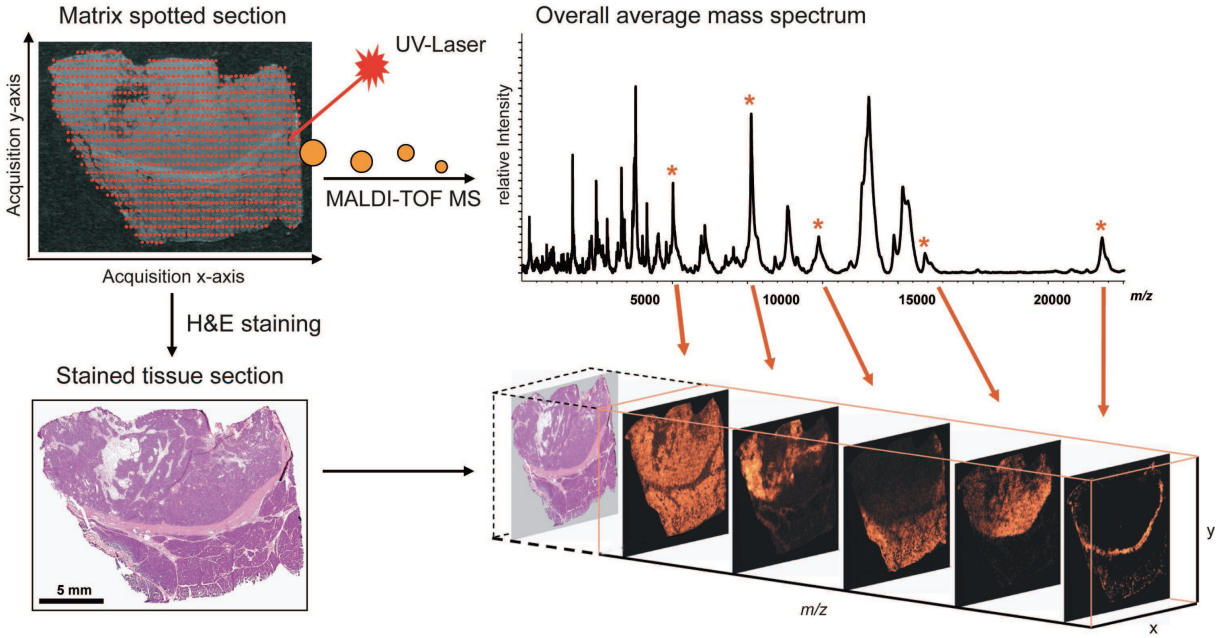


Figure 7 – Concept of MALDI Imaging. A matrix-coated tissue section is spot by spot analysed in a mass spectrometer. After the MALDI Imaging measurement the tissue section is histologically stained. This allows a histology driven analysis of the proteomic pattern acquired by mass spectrometry.[36]

MALDI Imaging of proteins is a straightforward approach. Tissue sections are directly analysed by mass spectrometry without destroying the tissue morphology. For this, native tissue sections are mounted onto a conductive MALDI Imaging slide, briefly washed with ethanol, and then coated with organic MALDI matrix. Afterwards, the tissues are analysed in a mass spectrometer in a spatially resolved, grid-like manner. By this, a mass spectrum is generated for every measurement spot (**Figure 7**). After the measurement the tissue is histologically stained and the mass spectra data can be correlated with the tissue morphology by superimposing the spectra data with the virtual microscopy image of the stained tissue section.[36] The resolution of measurements is usually between 50 and 250 µm with the attainable resolution depending on the matrix coating method.[37] The matrix coating can be done either by an automated spray device or by an automated spotter. Both have their

advantages. Spraying yields a higher measurement resolution. Recently, the resolution could be improved to 20 µm using a spray device and a commercially available mass spectrometer.[38] Spotting yields better mass spectra but resolution is limited to approximately 100 µm due to the matrix spot size.[39] After data analysis, the resulting peak patterns can be used for comparison of patient cohorts in respect to certain clinical endpoints, such as HER2/neu status, therapy response, prognosis or tumour type.[40-48] Only the molecular weight of proteins is measured. So, an additional protein identification step is necessary in order to identify each new marker. For this, a robust, routine method is still lacking but considerable effort is spent on this issue.[35] Taken together, MALDI Imaging is a powerful discovery tool for clinical research which has many strengths but also several limitations (**Table 4**).[32-35,49]

Advantages	Limitations
• Minimal need of sample material • Morphology based analysis • Label-free analysis • Multiplexing capability • Full protein analysis (physiological and posttranslational protein modifications detectable)	• Restricted to low molecular weight (≤25 kDa), soluble proteins • No direct protein identification • Potential ion suppression

Table 4 – Advantages and limitations of MALDI Imaging.

Since most patient tissues are stored as formalin-fixed paraffin-embedded (FFPE) tissue specimens, several groups have attempted to analyse these samples by MALDI Imaging. Using a tryptic digest step before matrix application allowed MALDI Imaging on FFPE tissues.[50-54] However, the protocols will have to be improved, especially since spatial resolution is still inadequate for detailed tissue analysis.* A different approach for tackling the FFPE "problem" is to establish alternative alcohol-based fixation protocols.[55-57†]

* Work on this protocol improvement will be presented in the results section.
† Establishing a protocol will be presented in the results section.

2.6.2 MALDI Profiling

MALDI Profiling is a method very similar to MALDI Imaging. A droplet of matrix is manually applied onto a tissue section, which is prepared and washed in the same way as in MALDI Imaging tissue preparation, and then the spot is analysed by mass spectrometry.[33] This of course reduces the spatial resolution. Nevertheless, the morphology of the analysed tissue remains intact, so the cellular content of the tissue at the site of matrix application can be assessed after measurement and taken into consideration during data analysis. The biggest advantage of MALDI Profiling is the preparation and analysis time. While in MALDI Imaging matrix application and measurement require usually several hours per tissue section, in MALDI Profiling this is reduced to several minutes. Thus, large patient cohorts can be analysed in a few days.[33]

2.6.3 Label-free quantitative proteomics

Quantitative proteomics rely on a lysate based analysis approach. In the beginning, two-dimensional gel electrophoresis was used for protein separation prior to protein identification. Nowadays, liquid chromatography based protein separation methods are state-of-the-art.[58] Quantitative proteomics can be separated into two major approaches; the use of stable isotope labelling – such as isotope-coded affinity tags (ICAT), stable isotope labels with amino acids in cell culture (SILAC) or isobaric tags for relative and absolute quantification (iTRAQ) – and label-free techniques.[59] Labelling strategies were often considered to be more accurate than label-free ones, however they have several limitations, like the complex sample preparation, the requirement for increased sample concentration, the incomplete labelling, the expensive labelling reagents, and the reduced capability for multiplexing.[58-60] With high-performance liquid chromatography, high-end mass spectrometers and novel software analysis tools, label-free strategies are now comparable in quality and accuracy with labelling strategies.[59,60] Label-free quantitative proteomics consists

of three steps: (i) sample preparation including protein extraction, (initial gel based protein separation,) reduction, alkylation, and digestion; (ii) sample separation by liquid chromatography and analysis by tandem mass spectrometry; (iii) data analysis including peptide/protein identification, quantification, and statistical analysis.[58]

Label-free quantitative proteomics based on liquid chromatography coupled mass spectrometry provides a good complementation of the MALDI Imaging approach since it is capable of analysing proteins in the higher mass range.

2.7 Scope of this thesis

Within this thesis several clinically relevant aspects of colon cancer were addressed:

(i) proteomic markers for regional lymph node metastasis were identified,*

(ii) accurate proteomic classification of tumour entities was achieved.†

(iii) proteomic markers for relapse were identified,‡

Additionally, methodological advances in MALDI Imaging were made.[56,61]

2.7.1 Proteomic markers for regional lymph node metastasis in colon cancer primary tumours

The presence of regional lymph node metastasis drastically reduces patients' prognosis in colon cancer. Patients with regional lymph node metastasis have a 5-year survival rate below 50%.[14,16,17] The process of lymphoid metastasis is still not fully understood and proteomic markers indicating the metastatic risk are still scarce.[62] However, such marker proteins could give a better insight into the molecular processes leading to lymphoid metastasis, help finding

---

* The results will be submitted to "The Journal of Pathology".

† The results were submitted to "The Journal of Proteome Research" and are currently in revision.

‡ The results are part of an ongoing study which is intended to be published in a clinical journal.

new therapeutic targets, or be used as diagnostic markers for assessing the metastatic potential of a primary tumour.

For proteomic marker identification, a tissue cohort of 54 primary colon tumours was analysed by MALDI Imaging and label-free quantitative proteomics. Subsequent validation of identified proteins was done by immunohistochemical staining on an independent cohort of colon cancer primary tumour tissues (n = 168).

2.7.2 Discrimination of tumour entities by proteomic classification

Correct, unambiguous tumour diagnosis is the initial step in cancer. However, in a significant number of metastasised cancers (2.3-4.2% of cancer cases, worldwide) the primary tumour cannot be identified and thus, they are diagnosed as cancer of unknown primary (CUP).[63,64] So far, clinical diagnosis relies on histological and often extensive immunohistochemical analyses of tumour biopsies.[65] Classification by gene expression[66-72] and proteomic means[73-77] has been attempted. These methods often need more tissue material than available and consider tissues to be homogeneous. MALDI Imaging could overcome these two limitations and has already been used for discriminating tumours according to their clinical endpoints.[40,45,47,51,78-80]

Therefore, it was tested whether MALDI Imaging can also be used for tumour classification. This was done in a two-step approach. For an initial proof-of-principle, six adenocarcinoma entities were classified according to their proteomic patterns. Then, a setting closer to the clinical problem of cancers of unknown primary was chosen. Colon cancer primary tumours, liver metastases and hepatocellular carcinoma were classified.

2.7.3 Proteomic markers for relapse in colon cancer primary tumours

The majority of colon cancer patients can be treated successfully if the cancer is diagnosed at an early, non-metastasised stage. After curative resection of the primary tumour, patients are considered fully cured if no detectable residual tumour remains. However, a significant

fraction of these "cured" patients develops a relapse within the first 5 years after resection (~5% of UICC stage I and ~15-25% of UICC stage II patients).[21,22] At present, no markers for relapse prediction exist which could be used in clinical diagnostics and aftercare.[8,21,27] Therefore, identifying new markers could lead to individually tailored, personalised therapeutic regimens and better aftercare, and thus could improve the patients' benefit.

Herein, we have analysed a cohort of 119 primary colon cancer tissues samples from UICC stage II patients by MALDI Profiling. The median follow-up for the cohort was 106 months, with a minimum of 3 month and a maximum of 180 month. A subgroup of this cohort (n = 25) developed a relapse while the other patients remained tumour-free (n = 94). This cohort was analysed in order to identify a panel of proteomic markers correlated with the later development of a relapse.

2.7.4 Methodological advances in MALDI Imaging – Opening new fields or research

For MALDI Imaging, cryo preserved tissues are the most common type of sample material. Unfortunately, only a minority of clinical samples is stored as cryo preserved material. Most clinical samples are stored as formalin-fixed paraffin-embedded tissues in the archives of pathology.[35] In order to gain access to these vast tissue archives, several attempts to implement protocols for MALDI Imaging on formalin-fixed paraffin-embedded tissues have been made.[50-54,81-85] Due to the used preparation method, which relied on trypsin and matrix spotting, the lateral resolution could not exceed 200 µm.

To overcome this limitation, a different protocol, which relied on spray application of trypsin and matrix, was tested in this study. After establishing this novel protocol on colon cancer tissue specimens the obtained data sets were used for testing a software algorithm, which could increase the lateral resolution of MALDI Imaging data after the measurement.

In pathology, alcohol-based fixation is a well-known method for tissue preparation. This preparation method could combine the advantages of cryo preservation and formalin fixation.

The intact tissue proteome could be analysed while the samples are easy to handle and store. Herein, a novel alcohol-based fixative, named PAXgene, was tested for its suitability for MALDI Imaging. Afterwards, this fixative was used in a proteomic study of the mammalian retina.

3 Materials

3.1 Chemicals, reagents and ready-to-use solutions

Chemicals	Manufacturer
Acetic acid, p.a.	Merck, Darmstadt, Germany
Acetonitrile, HPLC grade	Sigma-Aldrich, Steinheim, Germany
Acrylamide/bis solution, 37.5:1	Serva, Heidelberg, Germany
Ammonium hydrogen carbonate, ≥99.5%	Sigma-Aldrich, Steinheim, Germany
Ammonium hydrogen phosphate, ≥99.5%	Sigma-Aldrich, Steinheim, Germany
Ammonium persulfate, ≥98%	National Diagnostics, Atlanta, GA, USA
Bromophenol blue, for electrophoresis	Sigma-Aldrich, Steinheim, Germany
Citric acid, ≥99.0%	Sigma-Aldrich, Steinheim, Germany
Coomassie brilliant blue R250	Serva, Heidelberg, Germany
α-Cyano-4-hydroxycinnamic acid, ≥99.0%	Sigma-Aldrich, Steinheim, Germany
2,5-Dihydroxybenzoic acid, ≥99.0%	Sigma-Aldrich, Steinheim, Germany
2,4-Dithiothreitol	Merck, Darmstadt, Germany
EDTA, ≥99.0%	Carl Roth, Karlsruhe, Germany
Eosin Y disodium, for microscopy	Sigma-Aldrich, Steinheim, Germany
Ethanol, p.a.	Merck, Darmstadt, Germany
Formalin solution, neutral buffered, 10%	Sigma-Aldrich, Steinheim, Germany
Formic acid, p.a.	Sigma-Aldrich, Steinheim, Germany
Glycine, ≥99.0%	Sigma-Aldrich, Steinheim, Germany
Glycerol, ≥99.5%	Merck, Darmstadt, Germany
2-Iodoacetamide	Merck, Darmstadt, Germany
Isopropanol, p.a.	Merck, Darmstadt, Germany
Mayer's haematoxylin solution	Carl Roth, Karlsruhe, Germany
β-Mercaptoethanol, for electrophoresis	Sigma-Aldrich, Steinheim, Germany
Methanol, p.a.	Merck, Darmstadt, Germany
NP40, ≥99.5%	Sigma-Aldrich, Steinheim, Germany
Peptide calibration standard II	Bruker Daltonik GmbH, Bremen, Germany
Pertex mounting medium	Medite, Burgdorf, Germany
Poly-L-lysine, 0.1% (w/v) in water	Sigma-Aldrich, Steinheim, Germany

Protein calibration standard I	Bruker Daltonik GmbH, Bremen, Germany
Sodium dodecyle sulfate (SDS), ≥99.0%	Sigma-Aldrich, Steinheim, Germany
Sinapinic acid, ≥99.0%	Sigma-Aldrich, Steinheim, Germany
TBS pH 8.0, powder	Sigma-Aldrich, Steinheim, Germany
TEMED	Bio-Rad, München, Germany
Triethylammonium bicarbonate buffer (TEAB), 1M, for HPLC	Sigma-Aldrich, Steinheim, Germany
Trifluoroacetic acid	Applied Biosystems, Carlsbad, CA, USA
TRIS, ≥99.0%	Sigma-Aldrich, Steinheim, Germany
Trypsin Gold, mass spectrometry grade	Promega, Madison, WI, USA
Xylene, p.a.	Merck, Darmstadt, Germany

3.2 Commercial solutions and kits

Reagent	**Manufacturer**
BCA Protein Assay	Pierce Protein Research, Thermo Fisher Scientific, Rockford, IL, USA
DAB Map detection kit	Ventana Medical Systems, Tuscon, AZ, USA
NuPAGE Novex 4-12% Bis-Tris Mini Gel	Invitrogen, Life Technologies, Carlsbad, CA, USA
NuPAGE LDS Sample Buffer (4X)	Invitrogen, Life Technologies, Carlsbad, CA, USA
PAXgene Tissue Container	Qiagen, Hilden, Germany
Tissue homogenizing CKMix	Bertin Technologies, Montigny-le-Bretonneux, France

3.3 Solutions and buffers

Solution	Constituents	Concentration
ABC solution		in water
	ammonium bicarbonate	50 mM
Antigen retrieval buffer	pH 9.0	in water
	TRIS	10 mM
	EDTA	1 mM
DDT solution		in water
	2,4-dithiothreitol	5 mM
Destaining solution II		in water
	ethanol	50%
	TEAB	2.5 mM
Eluent A		in water
	acetonitrile	2%
	formic acid	0.1%
Eluent B		in acetonitrile
	water	2%
	formic acid	0.1%
Eosin solution		in water
	eosin Y disodium	1 g/L
Extraction solution I		in water
	acetonitrile	50%
	trifluoroacetic acid	0.5%
Extraction solution II	acetonitrile	99.5%
	trifluoroacetic acid	0.5%
Extraction solution III		in water
	acetonitrile	7.5%
	trifluoroacetic acid	0.1%
Extraction solution IV		in water
	acetonitrile	60%
	trifluoroacetic acid	0.1%

Fixation solution / Destaining solution I	methanol acetic acid	in water 50% 12%
IAA solution I	2-iodoacetamide	in water 25 mM
IAA solution II	2-iodoacetamide	in water 100 mM
Laemmli buffer 5x	pH 6.8 TRIS SDS glycerol β-mercaptoethanol bromophenol blue	in water 250 mM 5% 50% 500 mM 0.04%
LC solution	acetonitrile trifluoroacetic acid	in water 2% 0.5%
Lysis buffer	TBS NP40	in water 0.5 mM 1%
Peptide calibration standard	(125 µL per vial) acetonitrile trifluoroacetic acid	in water 33% 0.7%
Poly-lysine solution	poly-L-lysine NP40	in water 50% 0.1%
Protein calibration standard	(125 µL per vial) acetonitrile trifluoroacetic acid	in water 33% 0.07%
Reconstitution buffer	NuPAGE LBS buffer 2,4-dithiothreitol	in water 2x 50 mM

Running buffer for PAGE	pH 8.6	in water
	TRIS	25 mM
	glycine	192 mM
	SDS	0.1%
SDS-PAGE (15% resolving gel)	pH 8.8	in water
	TRIS	375 mM
	acrylamide/bis	15%
	SDS	0.1%
	TEMED	0.1%
	ammonium persulfate	0.05%
SDS-PAGE (4% stacking gel)	pH 6.8	in water
	TRIS	125 mM
	acrylamide/bis	4%
	SDS	0.1%
	TEMED	0.2%
	ammonium persulfate	0.1%
Staining solution		in water
	methanol	50%
	acetic acid	12%
	Coomassie brilliant blue R250	0.04%
TBS buffer	pH 7.4	in water
	TRIS	50 mM
	NaCl	150 mM
Trypsin solution (for MALDI Imaging)		in water
	ammonium hydrogen carbonate	80 mM
	trypsin (from stock solution)	1 µg/gL
Trypsin solution I (for peptide mass spectrometry)		in water
	ammonium hydrogen carbonate	50 mM
	trypsin (from stock solution)	0.01 µg/gL
Trypsin solution II (for peptide mass spectrometry)		in water
	ammonium hydrogen carbonate	5 mM
	trypsin (from stock solution)	0.01 µg/gL
Trypsin stock solution		in water
	acetic acid	50 mM

3.4 Antibodies

Antibody	Dilution	Buffer	Manufacturer
FXYD3	1:25	Tris-EDTA	Polyclonal, R01849, Sigma Life Science, St. Louis, MO, USA
GSTM3	1:100	Tris-EDTA	Monoclonal, 730.A17.1, NCI-CPTC, Developmental Studies Hybridoma Bank, Iowa City, IA, USA
PSAT1	1:100	Tris-EDTA	Monoclonal, 707.A45.1, NCI-CPTC, Developmental Studies Hybridoma Bank, Iowa City, IA, USA
S100A11	1:50	Tris-EDTA	Polyclonal, rabbit, 10237-1-A8, ProteinTech, Chicago, IL, USA
S100A4	1:500	Tris-EDTA	Polyclonal, rabbit, ab41532, Abcam, Cambridge, UK
TRX	1:500	Tris-EDTA	Monoclonal, rabbit, C63C6, #2429, Cell Signaling Technology, Danvers, MA, USA
UGDH	1:750	Tris-EDTA	Polyclonal, rabbit, HPA036656, R34779, Sigma Life Science, St. Louis, MO, USA

3.5 Matrices

Matrix	Concentration (w/v)	Solvent
α-Cyano-4-hydroxy-cinnamic acid (CHCA)	7 g/L	50% acetonitrile in water, 0.2% trifluoroacetic acid
2,5-Dihydroxybenzoic acid (DHB)	30 g/L	50% methanol in water, 0.2% trifluoroacetic acid
Sinapinic acid (SA)	10 g/L	60% acetonitrile in water, 0.2% trifluoroacetic acid 60% methanol in water, 0.2% trifluoroacetic acid

3.6 Instrumentation

Instrument	Manufacturer
Automated Stainer	Ventana Discovery XT; Ventana Medical Systems, Tuscon, AZ, USA
Cryostat Microtome	CM 1950; Leica Microsystems GmbH, Wetzlar, Germany
Mass Spectrometer for LC-MS/MS	LTQ OrbiTrapXL with nano spray ion source, Thermo Fisher Scientific, Rockford, IL, USA
Mass Spectrometer for MALDI Imaging	Ultraflex III TOF/TOF; Bruker Daltonik GmbH, Bremen, Germany
Matrix Spray Roboter	ImagePrep; Bruker Daltonik GmbH, Bremen, Germany
Microtome	HM340E; Microm International, Walldorf, Germany
Nano-HPLC (for tissue proteomics)	Ultimate3000 nano HPLC system with PepMap column (15 cm x 75 µm ID, 3 µm/100Å pore size); Dionex, Sunnyvale, CA, USA
Nano-HPLC (for matrix proteomics)	nanoLC-1D plus; Eksigent, Dublin, CA, USA with a in-house packed Reprosil Pur (15 cm x 75 µm, 3µm/60Å pore size) column; Dr. Maisch, Ammerbuch-Entringen, Germany
Slide Scanner	Mirax Desk System; Carl Zeiss MicroImaging GmbH, Göttingen, Germany
Tabletop Scanner	CanoScan 8800F, Canon, Tokyo, Japan
Tissue Arrayer	MTA-1; Beecher Instruments, Sun Prairie, WI, USA
Tissue homogeniser	Precellys24; Bertin Technologies, Montigny-le-Bretonneux, France
Vacuum concentrator	Univapo 150 ECH, Uniequip GmbH, Planegg, Germany

3.7 Software

Program	Manufacturer
ClinProTools (2.2)	Bruker Daltonics, Bremen, Germany
FlexAnalysis (3.3)	Bruker Daltonics, Bremen, Germany
FlexControl (3.0)	Bruker Daltonics, Bremen, Germany
FlexImaging (2.1 and 3.0)	Bruker Daltonics, Bremen, Germany
Mascot (2.2 and 2.3)	Matrix Science, Boston, MA, USA
Mascot Distiller (2.2)	Matrix Science, Boston, MA, USA
Progenesis LC-MS (3.0)	Nonlinear Dynamics, Newcastle upon Tyne, UK
R (2.13)	R Foundation for Statistical Computing
Scaffold Viewer (3.0 and 3.1)	Proteome Software, Portland, OR, USA

4 Methods

4.1 Tissue fixation and storage

4.1.1 Cryo fixation – Native tissues

Native tissues were snap-frozen in liquid nitrogen. Tissues were stored in liquid nitrogen for long-term storage or at -80 °C for short to medium term storage.

4.1.2 Formalin fixation – Formalin-fixed paraffin-embedded (FFPE) tissues

Formalin fixation and paraffin embedding was done in a standardised manner according to the protocols for diagnostic tissue material preparation of the Institute of Pathology, Technische Universität München. Tissue preparation was done at the histology facility of the Institute of Pathology, Helmholtz Zentrum München. FFPE tissues were stored at room temperature.

4.1.3 PAXgene fixation – PAXgene-fixed paraffin-embedded (PFPE) tissues

PAXgene fixation was done in a standardised manner according to manufacturer's protocol. Tissues were fixed for 3 h and stabilised overnight. Tissue preparation was done at the histology facility of the Institute of Pathology, Helmholtz Zentrum München, using formalin free reagents. PFPE tissues were stored at 4 °C.

4.2 Construction of tissue microarrays

Tissue microarrays were made using a manual tissue arrayer according to manufacturer’s protocol. The punch size was 1 mm in diameter. One representative core per patient sample was used.

4.3 Tissue sectioning

For histological staining, immunohistochemical staining, or MALDI Imaging the tissues were sectioned in the same way. For histological and immunohistochemical staining adhesive microscopy slides were used, for MALDI Imaging poly-lysine coated, conductive slides were used.

4.3.1 Sectioning of native tissues

Native tissues were sectioned in a cryostat microtome. Depending on the tissue type and its consistency the temperature while sectioning was set between -16 and -25 °C. The samples were mounted onto precooled specimen discs using a droplet of ice-cold, deionised water and the quick-freeze shelf of the cryostat. Sections were cut with a thickness of 12 µm.

4.3.2 Sectioning of FFPE and PFPE tissues

FFPE or PFPE tissues were sectioned in a microtome. The tissue blocks were precooled for 30 min at 4 °C. Sections were cut with a thickness of 3.5 µm and transferred to a water-bath (45 °C) before mounting.

4.4 Histological staining of tissues with haematoxylin and eosin

4.4.1 Staining of native tissues

Native tissue sections were fixed for 15 min in formalin solution and rinsed afterwards with deionised water. Then, they were stained for 1 min in haematoxylin, washed 2 min with tap water and stained for 1 min in eosin. The sections were rinsed in a graded alcohol series (70%, 90% and 100% ethanol, isopropanol, xylene twice; 30 sec each). Each tissue section was covered with mounting medium and a coverslip, which was pressed gently onto the section to remove excess mounting medium. Then, the tissue sections were dried at 60 °C.

4.4.2 Staining of native, FFPE and PFPE tissues after MALDI Imaging

For MALDI Imaging slides, the fixation step with formalin solution was omitted. Instead, the matrix was rinsed off with 70% ethanol. The further steps were done as described in section 4.4.1.

4.4.3 Staining of FFPE and PFPE tissues

FFPE or PFPE tissues were deparaffinised with xylene (10 min, twice). Then, the sections were rehydrated in a graded alcohol series (isopropanol, 100%, 90%, 70% and 50% ethanol; 1 min each). The further steps were done as described in section 4.4.1.

4.5 MALDI Imaging

MALDI Imaging consists of four steps: (i) tissue sectioning and preparation, (ii) matrix application, (iii) mass spectrometric measurement and (iv) data processing.[36]

4.5.1 Tissue preparation

For all tissue types MALDI Imaging slides were used because they have a conductive indium-tin oxide covering which allows MALDI mass spectrometry. The slides were coated with poly-lysine for better tissue adherence. For this, 20 µL of poly-lysine solution were streaked out on a MALDI Imaging slide with a Drigalski spatula. The slide was then dried on a hot plate (60 °C).

4.5.1.1 Preparation of native tissues

Tissue samples were sectioned as described in section 4.3.1. The sections were mounted onto MALDI Imaging slides, which have been kept cold in the cryostat. The respective slide was hand-warmed at the site of application before mounting the tissue. The mounted tissue was dried by hand warming the backside of the slide. Then, the slide was kept cold in the cryostat. After removing the slide from the cryostat the condensed water was evaporated using a cold airflow. Then, the slide was rinsed for 1 min each in 70% and 100% ethanol and air-dried.[36]

4.5.1.2 Preparation of FFPE tissues

Two methods were applied for FFPE sample preparation. Tissue samples were sectioned as described in section 4.3.2.

4.5.1.2.1 Preparation by antigen retrieval

FFPE tissue sections were placed for 15 min on a hot plate (60 °C). Then, they were deparaffinised with xylene (5 min, twice), rehydrated in a graded alcohol series (isopropanol, 100%, 90%, 70% and 50% ethanol; 3 min each), rinsed for 5 min in TBS buffer and boiled for 40 min in antigen retrieval buffer. Afterwards, the sections were washed for 5 min with deionised water and dried for 10 min at 40 °C.

4.5.1.2.2 Preparation by antigen retrieval and tryptic digest

The *in situ* tryptic digest was done with the ImagePrep spraying device according to manufacturer's protocol. First, the FFPE tissues were prepared as described in section 4.5.1.2.1. Then, 200 µL trypsin solution were applied onto the tissue in 12 cycles which consisted each of an initial spraying followed by 8 min of incubation.

4.5.1.3 Preparation of PFPE tissues

Tissue samples were sectioned as described in section 4.3.2. PFPE tissue sections were deparaffinised with xylene (10 min, twice), rehydrated in a graded alcohol series (isopropanol, 100%, 90%, 70% and 50% ethanol; 5 min each), and dried for 10 min at 40 °C.[36,56]

4.5.2 Matrix application

For measurement, the tissue section had to be coregistered with a light scan of it. For this, teach markings were painted onto the glass slide using Tipp-Ex. Then, the slide was scanned with a tabletop scanner with 2400 dpi resolution. Afterwards, the MALDI matrix was applied according to manufacturer's protocol using the ImagePrep spraying device. For each matrix the recommended spray program was used. After matrix application the glass slide was

immersed for 5 sec in ice-cold ammonium hydrogen phosphate solution and dried at room temperature. Matrix coated tissues were stored until the measurement in a vacuum chamber.

4.5.3 MALDI mass spectrometry measurement of tissue samples

A protein standard for linear mode measurement, a peptide standard for reflectron mode measurement was mixed with matrix (0.5 µL each) and spotted adjacent to the tissue section. Using the teach markings, the each tissue section was coregistered with its light scan image in the FlexControl software. Then, measurement regions were selected and the spotted standard was used for calibration.

4.5.3.1 Measurement in linear positive mode

For linear positive mode measurement, a mass range of 2,500 – 25,000 *m/z*, a sampling rate of 0.1 GS/s, and a lateral resolution between 50 and 200 µm (with adequate laser focus; medium or large) was selected. 200 laser shots were accumulated per measurement spot. The lateral resolution and the laser focus were kept constant within each study. The laser intensity and the detector sensitivity were adjusted for each measurement.

4.5.3.2 Measurement in reflectron mode

For reflectron mode measurement, a mass range of 600 – 4,000 *m/z*, a sampling rate of 1.0 GS/s, and a lateral resolution between 50 and 200 µm (with adequate laser focus; medium or large) was selected. 200 laser shots were accumulated per measurement spot. The lateral resolution and the laser focus were kept constant within each study. The laser intensity and the reflector detector sensitivity were adjusted for each measurement.

4.5.4 Data processing

After the mass spectrometric measurement, the tissue sections were stained as described in section 4.4.2 and a virtual microscopy scan was made with the light microscopy scanner. The MALDI Imaging file was opened with the FlexImaging software and coregistered with the

virtual microscopy scan. A first quality assessment was made by looking at the intensity of the dominant peak within the single spectra. If the unstandardised intensity was above 5 (arbitrary units) the spectra quality was considered as adequate. The data set was standardised to the total ion count (TIC). Then, regions of interest were selected. By default, the whole tissue section, the tumour region, the carcinoma cells only, and, if present, the mucosa region were selected and exported. If any selection consisted of more than one region of interest it was pooled into one region of interest by a self-written Perl programme after exporting. Further data analysis is described in the section 4.10.

4.6 MALDI Profiling

MALDI Profiling consists of four steps: (i) tissue sectioning and preparation, (ii) matrix application, (iii) mass spectrometric measurement and (iv) data processing.

4.6.1 Tissue preparation

Native tissue preparation was done as described in section 4.5.1.

4.6.2 Matrix application

For measurement, the tissue section had to be coregistered with a light scan image of it. For this, teach markings were painted onto the glass slide using Tipp-Ex. Then, the MALDI matrix was applied manually. Sinapinic acid (0.5 µL) was spotted onto the tissue at the site of interest. The matrix application was repeated three times with drying in between. For each section, 1 to 3 matrix spots were applied to sites of interest. Afterwards, the slide was scanned with a tabletop scanner with 2400 dpi resolution.

4.6.3 MALDI mass spectrometry measurement of tissue samples

A protein standard was mixed with matrix (0.5 µL each) and spotted adjacent to the tissue section. Using the teach markings, the section was coregistered with its light scan image in

the FlexControl software. Then, the measurement regions were selected and the spotted standard was used for calibration.

The measurement was done in linear positive mode with a mass range of 2,500 – 25,000 *m/z*, a sampling rate of 0.1 GS/s, and a lateral resolution of 70 µm (medium laser focus). 200 laser shots were accumulated per measurement spot. The laser intensity and the detector sensitivity were adjusted for each measurement.

4.6.4 Data processing

After measurement the tissue sections were stained as described in section 4.4.2 and a virtual microscopy scan was made with the light microscopy scanner. The MALDI Profiling file was opened with the FlexImaging software and coregistered with the virtual microscopy scan. The data set was standardised to the total ion count (TIC). Then, the cellular composition of the measurement region was checked. The spectra of the region were only exported for later analysis if the region contained an adequate amount of appropriate cellular components. If more than one region was exported from a single sample, the regions were pooled into a single region by a self-written Perl programme. Further data analysis is described in the section 4.10.

4.7 Tissue lysate based label-free quantitative proteomics and protein identification

Label-free quantitative proteomics and protein identification were done in a combined approach in close collaboration with Alexander Schäfer and Dr. Stefanie Hauck, Research Unit Protein Science, Helmholtz Zentrum München. For identification, only a single sample is needed. For the quantitative proteomics, at least two samples are needed. In this case six were used, three in each of the two groups. (In this case: UICC stage II and UICC stage III primary colon cancer tissues.)

4.7.1 Protein extraction, sample preparation and peptide mass spectrometry

The tissue samples were cryo-sectioned (12 µm; 60 mg; ≥80% cancer cells within the sections) and lysed in 1 mL TBS buffer (10x) with 1% NP40 using a Precellys24 tissue homogeniser with the appropriate tissue homogenising CKMix. The lysates were spun-down twice (30 min, 16,000 g) and the supernatant was collected. The protein concentrations were determined with a BCA assay. For protein separation, equal amounts of protein (60 µg) were mixed with Laemmli buffer, boiled for 10 min and loaded onto a SDS-PAGE gel (4% stacking and 15% resolving gel). The electrophoresis was performed with 80 V for the stacking gel and 160 V for the resolving gel. Running distance of the dye was 5 cm. Then the gel was incubated for 30 min in fixation solution, overnight in staining solution, for 30 min in destaining solution I, and twice for 10 min in deionised water. The gel was horizontally cut into 6 equidistant fractions before the lanes were separated. The fractions were washed in 60% acetonitrile and deionised water (100 µL, 10 min each). The supernatant was removed and the fractions were incubated in 100 µL DDT solution for 15 min at 60 °C. The supernatant was removed, and the fractions were incubated in 100 µL IAA solution I for 15 min in the dark. The fractions were washed with 100 µL deionised water (5 min), 100% acetonitrile (10 min), ammonium bicarbonate solution (10 min), 60% acetonitrile, and 100% acetonitrile. The gel fractions were covered with trypsin solution I and digested overnight at 37 °C. The resultant peptides were eluted from the gel by two consecutive extraction steps. Each gel fraction was first incubated in ~100 µL extraction solution I for 15 min and the liquid was collected. Then, the gel fraction was incubated in ~100 µL of extraction solution II for 15 min and the liquid was collected. Both fractions were combined and dried in a vacuum concentrator. The samples were dissolved directly prior to LC-MS/MS analyses in LC solution. Then the peptide mixture within the samples was separated by reversed phase chromatography operated on a nano-HPLC system with a nonlinear 170 min gradient using eluent A and eluent B with a flow rate of 250 nL/min. The gradient settings were: 0-140 min:

2-5-31% B, 140-154 min: 31-95% B, 145-150 min: constant at 95% B, 150-155 min: 95-5% B. The nano-HPLC was connected to a linear quadrupole ion trap-Orbitrap mass spectrometer equipped with a nano-ESI source. The mass spectrometer was operated in the data-dependent mode to automatically switch between Orbitrap-MS and LTQ-MS/MS acquisition. Survey full scan mass spectra (from *m/z* 300 to 1500) were acquired in the Orbitrap with a resolution of R = 60,000 at *m/z* 400 (after accumulation to a target of 1,000,000 charges in the LTQ). The used method allowed sequential isolation of the most intense ions, up to ten, depending on signal intensity, for fragmentation on the linear ion trap using collision induced dissociation at a target value of 100,000 ions. High resolution mass spectrometry scans in the Orbitrap, and MS/MS scans in the linear ion trap were performed in parallel. Target peptides already selected for MS/MS were dynamically excluded for 30 seconds. General mass spectrometry conditions were: electrospray voltage, 1.25-1.4 kV; no sheath and auxiliary gas flow. Ion selection threshold was 500 counts for MS/MS, and an activation Q-value of 0.25 and activation time of 30 ms were also applied for MS/MS.

4.7.2 Label-free peptide quantification

The acquired spectra were loaded (Thermo raw files) into the Progenesis software and label-free quantification was performed as described previously.[86] For each fraction from the SDS-PAGE, the profile data of the MS scans as well as MS/MS spectra were transformed to peak lists with Progenesis LC-MS using a proprietary algorithm and then stored in peak lists comprising *m/z* and abundance. One sample was set as a reference, and the retention times of all other samples within the experiment were aligned (3 to 5 manual landmarks, followed by automatic alignment) to create maximal overlay of the two-dimensional feature maps. Features with only one charge or more than 7 charges were masked at this point and excluded from further analyses and all remaining features were used to calculate a normalisation factor for each sample which corrects for experimental variation. Samples were then allocated to

their experimental group. (In this case: UICC stage II and UICC stage III primary colon cancer tissues.) For quantification, all peptides (with Mascot score ≥ 30 and $p < 0.01$, see below) of an identified protein were included and the total cumulative abundance was calculated by summing the abundances of all peptides allocated to the respective protein. No minimal thresholds were set for the method of peak picking or selection of data to use for quantification. Statistical analysis was performed using the "between subject design" and p-values were calculated by a repeated measures ANOVA using the sum of the normalised abundances across all runs. After processing of all samples from the fractionation, the quantification files were merged into a complete data set.

4.7.3 Database search and protein identification

MS/MS spectra were exported from the Progenesis software as Mascot Generic file (mgf) and used for peptide identification with Mascot in the Ensembl database for human (Homo sapiens; GRCh37, download from ftp://ftp.ensembl.org/pub/current_fasta/homo_sapiens/pep/) containing a total of 77,460 protein sequences. The used search parameters were: 10 ppm peptide mass tolerance and 0.6 Da fragment mass tolerance, one missed cleavage allowed, carbamidomethylation was set as fixed modification and methionine oxidation, as well as deamidation of asparagine and glutamine were allowed as variable modifications. For the UICC stage II vs. UICC stage III data set, the Mascot integrated decoy database search calculated a false discovery of ≤1.11% when searching was performed on the concatenated mgf files with an ion score cut-off of 30 and a significance threshold of $p \leq 0.01$. Only peptides with ion scores of 30 and above and only proteins with at least one unique peptide ranked as top candidate (bold red in Mascot) were considered and re-imported into Progenesis software. For accumulation of relative protein abundances, only peptides unique for a given protein were used.

4.7.4 Protein identification for MALDI Imaging

MALDI Imaging allows no direct identification of proteins. Therefore, the protein identity of the *m/z* species has to be identified afterwards. For this, the MS/MS data of the quantitative proteomics approach was used. The identified molecular weights of the proteins were compared with the *m/z* species of the MALDI Imaging and with protein identification results published in other MALDI Imaging articles. If the molecular weights of the proteins identified by MS/MS were in high accordance (≤5 Da) with the *m/z* species derived by MALDI Imaging, the protein was assumed to be identified if the subsequent immunohistochemical validation supported the previously made findings. Thus, the protein identification was highly stringent which resulted in few identified *m/z* species but implied also a very low false-positive rate.

4.8 Matrix proteome based protein identification

Matrix proteome based protein identification was done in close collaboration with Stefan Maier and Prof. Dr. Bernhard Küster, Chair of Proteomics and Bioanalytics, Technische Universität München. For reduction of the proteomic complexity only the proteins which are embedded in the matrix of a MALDI Imaging sample are analysed. For a robust result more than one sample is required. (In the case of the relapse cohort, 3 samples were used for identifying the matrix proteome.)

4.8.1 Tissue sample preparation

Native tissues were sectioned as described in section 4.3.1 and mounted onto a plain microscope slide. Only tissue which contained more than 80% tumour cells were considered adequate. Approximately, 5 to 6 tissue sections of medium size (1 cm^2) per tissue sample were mounted onto a single microscopy slide and used of proteomic analysis. Then, the sinapinic acid matrix was applied onto the tissue sections according to manufacturer's protocol using the ImagePrep spraying device.

4.8.2 Protein extraction, sample preparation and peptide mass spectrometry

For the recovery of peptides and proteins co-crystallised with the matrix layer a two step extraction procedure was employed. For the first extraction step the matrix directly above the tissue section was covered with a thin layer of extraction solution III. After some seconds the liquid was removed and collected. This procedure was repeated with a total amount of 300 µL for one microscopic slide. For the second extraction step the same procedure was repeated with extraction solution IV. The remaining tissue was removed using a scalpel and collected in a tube containing 150 µL extraction solution IV. All three kinds of sample extracts were dried and reconstituted afterwards in reconstitution buffer. The sample extracts were mixed thoroughly and shortly spun-down. Reduction of disulfide bonds was carried out at 90 °C for 45 min. For the alkylation of free sulfhydryl groups 20 µL of IAA solution II was added and incubated for 30 min at room temperature in the dark. Subsequently the sample extracts were centrifuged at 13,000 rpm for 10 min and heated for 10 min at 90 °C. 25 µL supernatant of each sample extract was loaded onto a NuPAGE gel. For the matrix extracts, separation was carried out at 200V for 10 minutes and for the tissue extracts for 45 minutes. Then, the gels were incubated for 30 min in fixation solution, overnight in staining solution, for 30 min in destaining solution I, and twice for 10 min in deionised water. For the matrix extract gel separation, the protein containing region was excised. For the tissue extract gel separation, the samples lanes were cut into 12 slices. The proteins within the samples were in-gel digested with trypsin according to the manufacturer's protocol.

Each gel slice was destained twice with destaining solution II for 90 min each. The destaining solution II was discarded and the gel slides were rehydrated with 100% ethanol. The ethanol was removed and the proteins within the gel slices digested. At first, 12 µL of trypsin solution II were added to each slice and the slices were incubated for 15 min. Then, 20 µL of 5 mM TEAB were added to each sample and the samples were digested overnight at 37 °C. The trypsin solution II was removed and the gel slices were acidified with 5 µL of 5% formic

acid. The peptides were extracted in three steps. The gel slices were twice incubated in 30 µL of 1% formic acid for 30 min and the liquid was collected. Then, the gel slices were incubated in 20 µL of extraction solution IV for 30 min and the liquid was collected. The three liquid fractions of each gel slice were united and dried down. For LC-MS analysis the samples were reconstituted in 0.1% formic acid.

LC-MS/MS measurements were performed on an LTQ-Orbitrap XL mass spectrometer coupled to a nano-HPLC. For the peptide samples derived from the matrix extracts a 200 minute linear gradient (0-35 % acetonitrile in 0.1 % formic acid) with a flow rate of 300 nL/min was used. For the peptide samples derived from tissue extracts a 100 min linear gradient (0-35 % acetonitrile in 0.1 % formic acid) with a flow rate of 300 nL/min was used. Intact masses of eluting peptides were determined in the orbitrap in a mass range from 300 to 1300 *m/z* at a resolution of 60,000. The 15 most intensive species were selected for further fragmentation by collision-induced dissociation (CID) in the linear iontrap with a normalised collision energy of 35%. Singly charged ions as well as ions with unknown charge state were rejected. Dynamic exclusion was set to 30 seconds.

4.8.3 Database search and protein identification

Peak lists were created using Mascot Distiller and database searches were performed using the Mascot search engine with a parent ion tolerance of 10 ppm and a fragment ion tolerance of 0.6 Da against the IPI Database (http://www.ebi.ac.uk/IPI). Carbamidomethylation of cysteine residues was considered as a fixed modification and the oxidation of methionine as a variable modification. Up to 2 missed cleavages were accepted. Search result files were imported into Scaffold and filters for peptide and protein probability were adjusted to get a protein false discovery rate of smaller or equal to 1%.

4.8.4 Protein identification for MALDI Imaging

MALDI Imaging allows no direct identification of proteins. Therefore, the protein identity of the *m/z* species has to be identified afterwards. For this, the MS/MS data of the matrix proteome was used. The identified molecular weights of the proteins were compared with the *m/z* species of the MALDI Imaging and with protein identification results published in other MALDI Imaging studies. If the molecular weight of a protein identified by MS/MS was in high accordance (≤5 Da) with the *m/z* species derived by MALDI Imaging, the protein was assumed to be identified.

4.9 Immunohistochemical staining

Immunohistochemical staining of tissue sections was carried out using an automated stainer (see section 3.6) with its respective reagent kit (see section 3.2) according to the manufacturer's protocol. The used antibodies, dilutions and the buffers are listed in section 3.4.

4.10 Statistical analysis

4.10.1 Statistical analysis of MALDI Imaging and Profiling sample cohorts

The following settings were applied for data analysis with the ClinProTools software. A resolution of 800, a Top Hat baseline subtraction with 10% minimal baseline width, a Savitsky Golay smoothing with 2.0 *m/z* width and 5 cycles, a recalibration with 1000 ppm maximal peak shift and 20% match to calibrant peaks was employed. Null spectra exclusion was enabled. Spectra grouping was supported if more than one case per group was analysed. Peak pickling was applied on the total average spectrum with a signal to noise threshold of 5.00 and 0.000% relative threshold base peak. For peak calculation, intensities were used.

4.10.1.1 Statistical analysis of single samples

Single samples were analysed for assessing the measurement quality. This was done by hierarchic clustering of the acquired spectra using Euclidean distance measurement and ward linkage mode using the ClinProTools software. The clustering results were visualised in the FlexImaging software. If the clustering reflected the histology the spectra quality was assumed as being adequate for further analysis.

4.10.1.2 Statistical analysis of sample cohorts

MALDI Imaging and MALDI Profiling sample cohorts were analysed in several ways: by (i) statistical comparison using a Wilcoxon rank-sum-test, (ii) hierarchic clustering, (iii) classification, and (iv) principal component analysis.

4.10.1.2.1 Statistical comparison of sample cohorts using a Wilcoxon rank-sum test

Identical numbers of tumour spectra were randomly selected from each single sample data set. These data sets were grouped according to the examined clinical endpoint into two groups. This sample cohort data set was statistically compared using the ClinProTools software. Statistical comparison relied on a Wilcoxin rank-sum test with subsequent p-value correction according to Benjamini-Hochberg.[87] This resulted in a panel of differentially expressed *m/z* species. Additionally, the AUC values for the differentially expressed *m/z* species were computed using the R software and its ROCR package.

4.10.1.2.2 Hierarchic clustering of sample cohorts

Identical numbers of tumour spectra were randomly selected from each single sample data set. These data sets were grouped according to the examined clinical endpoint into two groups. Then, a peak list was created using the ClinProTools software and exported as CART file. This peak list was used for hierarchic clustering in the R software using stats and ROCR

package. Several settings were iteratively applied and the clustering with the highest specificity and sensitivity in regard to the sample grouping was identified.

4.10.1.2.3 Classification of sample cohorts

Identical numbers of tumour spectra (40) were randomly selected from each single sample data set. These data sets were grouped according to the examined clinical endpoint into groups. Then, a peak list was created using the ClinProTools software and exported as CART file. This peak list was used for classification in the R software using the boot, caret, e1071, randomForest and ROCR packages. First, the sample cohort was split into a training set for establishing the classifier and a test set for validating it. The training set comprised two thirds of each group of samples, the test set the other third. The sample selection was done randomly. Then, the discriminating peaks were selected by pairwise comparison of the groups within the training set using a Wilcoxon rank-sum test and consecutive Benjamini-Hochberg p-value correction.[87] Only peaks were selected which had a p-value smaller than 0.05 and an AUC higher than 0.8 in at least one of the pairwise comparisons. Support vector machine and Random Forest were employed as classifiers. The classifiers were generated using the selected peaks and the training set. Then, it was validated on the test set. For the Random Forest classifier the calculation was repeated 100 times and the average of the individual results was calculated. To reduce the sampling error, sampling for the training and test set was repeated 50 times, and a classification was performed on each set. The final classification result is the average of the 50 individual classification results.

For correlation of the misclassification rate with the respective tumour depth (T) or grading (G) a Spearman's rank correlation was done using the R software and its stats package. The frequency of misclassifications in all samplings and in case of the Random Forest algorithm for all repeats was counted for each sample within the test set. The resulting numbers were correlated with the tumour depth (T) or grading (G). For the combined analysis a weighted

average of the frequency of both classifications was used. Since the Random Forest classification had 100 additional repeats for each sampling its frequency was divided by 100 before uniting it with the frequency of misclassifications of the Support Vector Machine classification.

4.10.1.2.4 Principal Component Analysis of sample cohorts

Identical numbers of tumour spectra (40) were randomly selected from each single sample data set. These data sets were grouped according to the examined clinical endpoint into groups. Then, a peak list was created using the ClinProTools software and exported as CART file. This peak list was used for the Principal Component Analysis in the R software using the stats and scatterplot3d packages. The discriminatory power of the first three principal components was graphically visualised in a three dimensional plot of the samples. Additionally, the score (% sum explained variance) was exported into a table.

4.10.2 Statistical analysis of immunohistochemically stained sample cohorts

Immunohistochemically stained sections of the tissue sample cohorts were evaluated using a visual scoring system with scores ranging from 0 to 3. The scoring algorithm was set up before analysis and then applied to the respective tissue cohort. The staining pattern of the whole tumour cells, of the cytoplasm, of the nucleus, or of the membranes was determined. A cut-off for the scores was determined and the samples grouped accordingly in high and low intensity staining.

4.10.2.1 Univariate analysis by Fisher's exact test or by Pearson's χ^2 test

The staining data and the data for one clinical endpoint of the samples were entered into a contingency table. Then, the statistical significance (p-value) was determined by Fisher's exact test or by Pearson's χ^2 test using the R software and its stats package.

4.10.2.2 Multivariate analysis with generalised linear model

The staining data and the data of multiple clinical endpoints of the samples were entered into a data table. Then, multivariate analysis was performed with the generalised linear model. For this, the R software and its stats package were used. By this, the statistical significance of the multiple clinical endpoints in respect to the staining data and in respect to each other could be determined.

5 Results

5.1 Proteomic markers for regional lymph node metastasis in colon cancer primary tumours

The presence of regional lymph node metastasis drastically reduces patients' prognosis in colon cancer.[14,16,17] The molecular processes leading to lymph node metastasis are still not fully understood and no diagnostic markers for risk assessment have been established.[62] Therefore, proteomic markers could give a better insight into the molecular processes leading to lymphoid metastasis, help finding new therapeutic targets, or be used as diagnostic markers for assessing the metastatic potential of a primary tumour.

For finding such proteomic markers an approach comprising two complementary mass spectrometric methods, MALDI Imaging and label-free quantitative proteomics, was applied on a first cohort of primary colon cancer tissues. For validation of the identified proteomic markers, immunohistochemical staining was performed on a second, independent cohort of primary colon cancer tissues (**Figure 8**).*

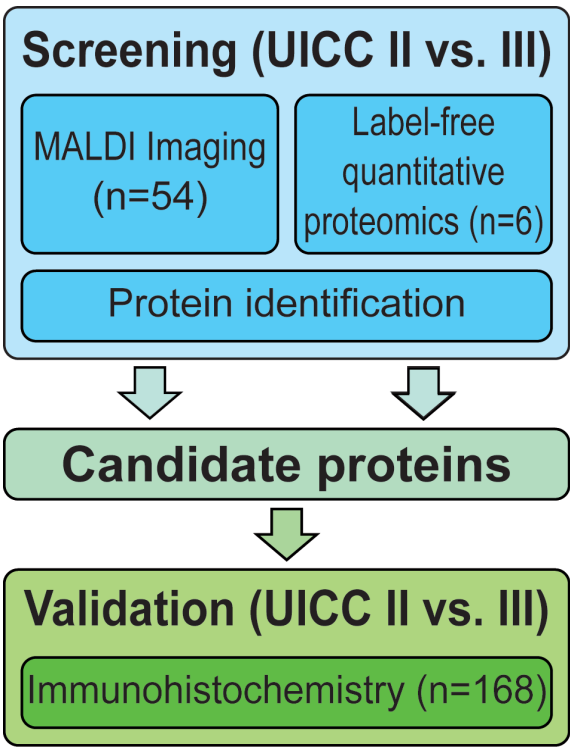


Figure 8 – Work-flow for the identification of proteomic markers for regional lymph node metastasis. First, candidate proteins are identified by MALDI Imaging and label-free quantitative proteomics on a cohort of primary colon tumour tissues. In a consecutive step, the candidate proteins are validated by immunohistochemical staining on a second, independent cohort of primary colon tumour tissues.

* The results will be submitted to "The Journal of Pathology".

5.1.1 Identifying markers for regional lymph node metastasis by MALDI Imaging

In order to define proteomic markers indicating regional lymph node metastasis a tissue cohort of primary colon tumours (UICC stage II, pN0, n = 21; UICC stage III, pN2, n = 33) was analysed by MALDI Imaging (**Table 5**). The tissue samples were taken from the tumour bank of the Department of Medicine, Technische Universität München. The clinical and pathological data were provided by the Department of Surgery, Klinikum rechts der Isar, Technische Universität München, and the Institute of Pathology, Technische Universität München.

Characteristics	Patient cohort for MALDI Imaging	Patient cohort for label-free quantitative proteomics	Patient cohort for immunohistochemical validation
Patients	54	6	168
Mean age [years] (range)	67.3 (41.4 – 89.3)	66.6 (56.1 – 80.4)	70.9 (28.9 – 95.0)
Gender			
male	33	5	101
female	21	1	67
Tumour depth (pT)			
pT1	0	0	1
pT2	1	0	1
pT3	23	1	129
pT4	30	5	37
Nodal status (pN)			
pN0	21	3	87
pN2	33	3	81
Tumour grading (G)			
G1	1	0	2
G2	23	3	80
G3	30	3	86

Table 5 – Histopathological characteristics of the patient cohorts.

Subsequent histological staining of the analysed tissue sections with haematoxylin and eosin allowed selecting mass spectra specific for the tumour regions by designating regions of interest. The mass spectra specific for the tumour regions were extracted and used for statistical comparisons. Spectra alignment and calibration of the mass spectra of the sample cohort resulted in 136 *m/z* species present in all samples. Several *m/z* species showed different intensity levels between the UICC stage II and UICC stage III patients. Statistical analysis

using non-parametric Wilcoxon rank-sum test with subsequent p-value correction according to Benjamini-Hochberg[87] yielded a panel of 10 significantly discriminating *m/z* species (**Figure 9** and **Table 6**).

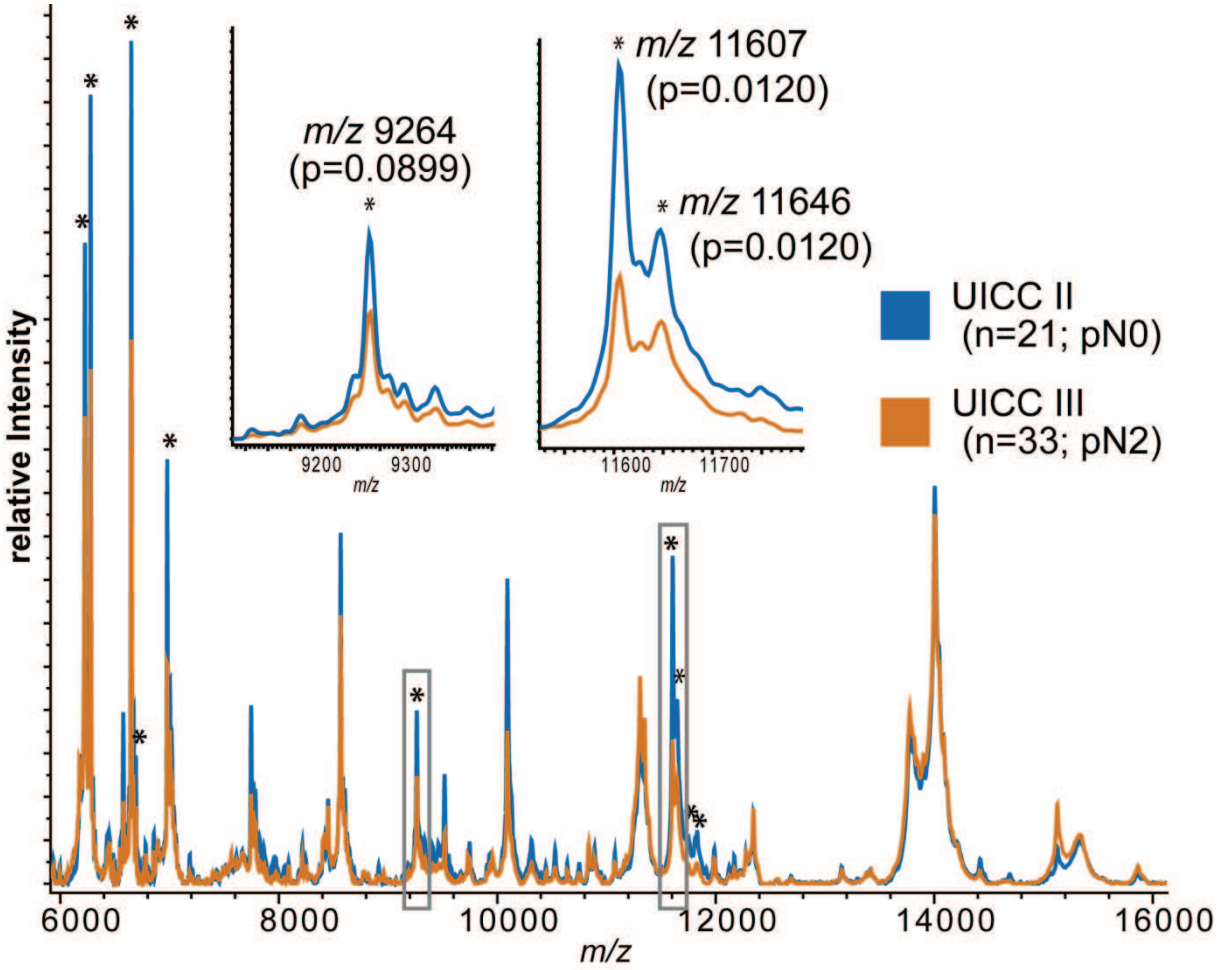


Figure 9 - MALDI Imaging of primary colon cancer tissue specimens. The average spectra of UICC stage II (blue) and UICC stage III (orange) patient tissue samples are displayed. Asterisks indicate significantly differentially expressed *m/z* species. Three *m/z* species (*m/z* 9264, *m/z* 11607 and *m/z* 11646) are enlarged. All three are stronger expressed in UICC stage II patient tissues than in UICC stage III patient ones.

m/z	p-value[1]	AUC value	Protein name (Uniprot ID)	Associated with cancer	Associated with colon cancer
11607	0.0120	0.79	-	-	-
11646	0.0120	0.79	S100A11 (P31949)	Associated with differentiation in breast cancer[88] Upregulated in gastric cancer[89] Association with carcinogenesis in hepatocellular carcinoma[90]	Upregulated in colorectal cancer[91,92] Associated with tumour progression in colorectal cancer[93]
11747	0.0120	0.79	-	-	-
11827	0.0206	0.77	-	-	-
6277	0.0436	0.75	-	-	-
6670	0.0572	0.74	-	-	-
6649	0.0670	0.73	-	-	-
6224	0.0892	0.72	-	-	-
6979	0.0899	0.72	-	-	-
9264	0.0899	0.72	FXYD3 (Q14802)	Downregulation in lung cancer[94] Upregulation in pancreatic cancer [95] Prognostic factor in rectal cancer[96]	Associated with tumour progression in colorectal cancer[97]

Table 6 – MALDI Imaging derived candidate *m/z* species discriminating the nodal status. ([1]Wilcoxon rank-sum test with Benjamini-Hochberg p-value correction)

M/z 11607, *m/z* 11646, *m/z* 11747, *m/z* 11827, *m/z* 6277 had a p-value below 0.05 ($p < 0.05$), and *m/z* 6670, *m/z* 6649, *m/z* 6224, *m/z* 6979, *m/z* 9264 had a p-value between 0.05 and 0.1 ($0.05 \leq p < 0.1$). The AUC value, which is a second indicator for the discriminating power of the identified *m/z* species, was calculated by receiver operator characteristics (ROC) analysis. For all listed *m/z* species the AUC value was greater than or equal to 0.72, for the *m/z* species with a p-value lower than 0.05 it was even greater than or equal to 0.75 (**Table 6**). For assigning the differentially expressed *m/z* species to the tissue regions in which they are expressed they were visualised in the analysed tissue sections. *M/z* 9264, *m/z* 11607 and *m/z* 11646 were predominantly expressed in the tumour regions. Additionally, they were higher expressed in the tumour regions of the tissue sections of the UICC stage II patients than in the tumour sections of the UICC stage III patients. Representative ion images for UICC stage II and UICC stage III patients are given in **Figure 10**.

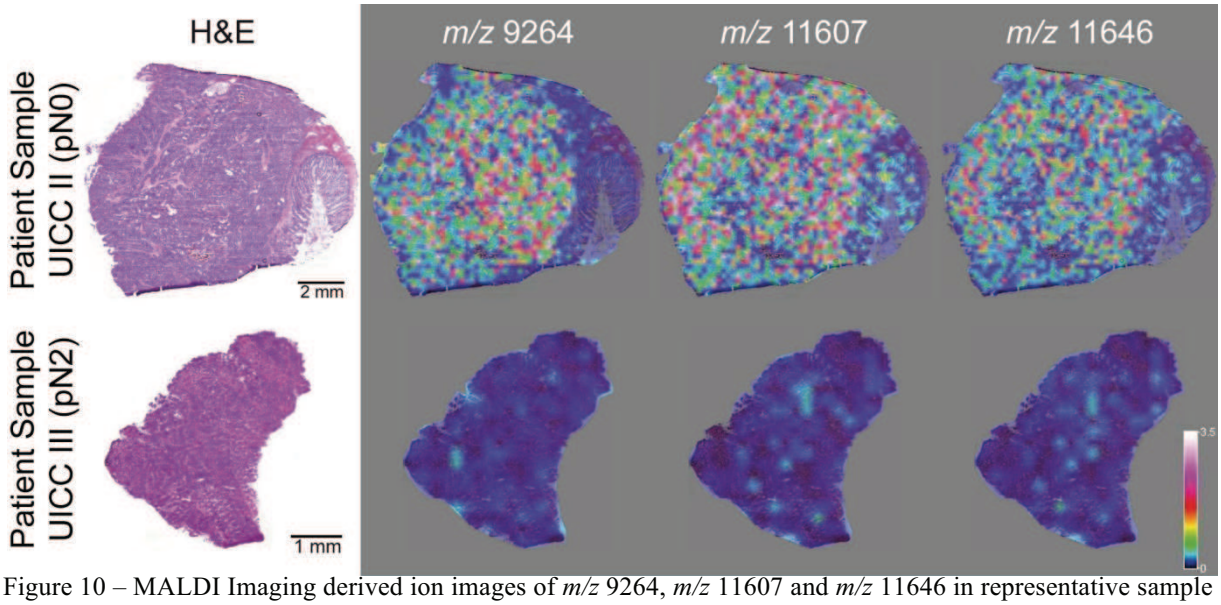


Figure 10 – MALDI Imaging derived ion images of *m/z* 9264, *m/z* 11607 and *m/z* 11646 in representative sample sections. On the left, the haematoxylin and eosin staining, on the right, ion images indicating the expression levels of the *m/z* species within the tissue sections are displayed. The expression levels of the three *m/z* species are higher in the tumour area of the UICC stage II sample tissue than in the tumour area of the UICC stage III sample tissue.

5.1.2 Protein identification of MALDI Imaging derived proteomic markers for regional lymph node metastasis

MALDI Imaging allows no direct identification of proteins. Therefore, the protein identity of the differentially expressed *m/z* species had to be identified by bottom-up tandem mass spectrometric methods and consecutive immunohistochemical staining for validation. Bottom-up protein identification was done in collaboration with Alexander Schäfer and Dr. Stefanie Hauck, Research Unit Protein Science, Helmholtz Zentrum München. For this, a set of 6 primary colon tumours was used (UICC stage II, pN0, n = 3; UICC stage III, pN2, n = 3). The specimens are representative samples of the cohort used for MALDI Imaging and will also be used for label-free quantitative proteomics (**Table 5**). By this, two proteins could be identified (**Table 6** and **Figure 11**).

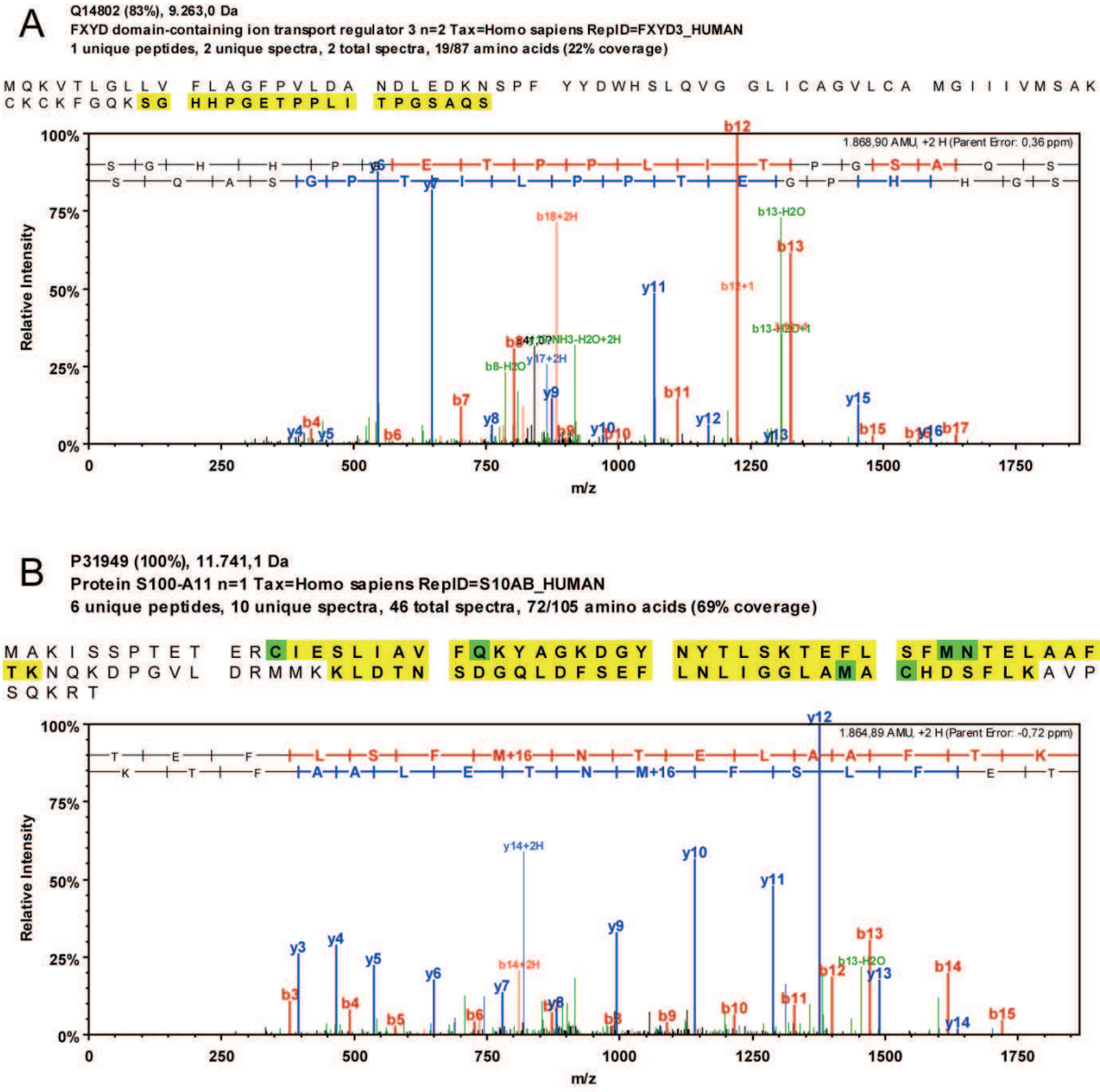


Figure 11 – Sequence coverage and fragment spectra for the identification of MALDI Imaging derived proteins. (A) Identification of FXYD3. FXYD3 (Q14802) was identified with a Mascot score of 91 ($p < 0.01$). FXYD3 could only be identified on a single peptide because only one tryptic peptide is in the detection range of the analysis. The other tryptic peptides are either smaller than 500 Da, or larger than 2460 Da. FXYD3 could be assigned to *m/z* 9264 in the MALDI Imaging experiment. (B) Identification of S100A11. S100A11 (P31949) was identified with a Mascot score of 1318 ($p < 0.01$). By theoretical considerations and according to literature[98] it could be assigned to *m/z* 11646 in the MALDI Imaging experiment.

FXYD3 (*m/z* 9264) has not been reported in any other MALDI Imaging study. S100A11 (*m/z* 11646) was identified in a previous MALDI Imaging study investigating a different scientific context.[98]

5.1.3 Identifying markers for regional lymph node metastasis by label-free quantitative proteomics

As a complementary approach for detecting protein markers discriminating the nodal status in colon cancer, label-free quantitative proteomics was performed on a set of 6 primary colon tumours (UICC stage II, pN0, n = 3; UICC stage III, pN2, n = 3; **Table 5**). The specimens are representative samples of the cohort used for MALDI Imaging and were also used for protein identification. Whole protein tissue extracts were first processed for mass spectrometry, then LC-MS/MS was done, and finally, quantitative analysis was performed using the Progenesis software package. A total of 1682 proteins could be identified, 1217 of them with 2 or more peptides. Only proteins with 2 or more identified peptides were considered for quantitative comparison of their expression levels. 28 proteins were differentially expressed ($p < 0.05$). 12 of them were more than 2-fold up- or downregulated. 12 proteins were upregulated in UICC stage II samples, 16 proteins were upregulated in UICC stage III samples (**Table 7**).

Gene name (*Protein name*; UniProt ID)	**p-value**	**Expression difference [fold]**	**Overexpressed in**	**Identified peptides (peptides for quant. comparison)**	**Associated with cancer**	**Associated with colon cancer**
GSTM3 (*Glutathione S-transferase Mu 3*; **P21266**)	0.003	3.5	N2	4(4)	Associated with prognosis in bladder cancer[99] Downregulated in ovarian cancer[100]	Gene polymorphism associated with progression and prognosis in colorectal cancer[101]
CSRP1 (*Cysteine and glycine-rich protein 1*; **P21291**)	0.006	2.1	N0	2(2)	Downregulated in hepatocellular carcinoma[102]	Downregulated in colorectal cancer[103]
PSAT1 (*Phosphoserine aminotransferase*; **Q9Y617**)	0.008	2.6	N2	5(4)	Therapy response in breast cancer[104]	Upregulated in colorectal cancer[105]
PTPLAD1 (*3-hydroxyacyl-CoA dehydratase 3*; **Q9P035**)	0.009	2.2	N2	2(2)	-	-
UGDH (*UDP-glucose 6-dehydrogenase*; **O60701**)	0.014	2.8	N2	8(8)	Upregulated in prostate cancer[106]	-
OGFR (*Opioid growth factor receptor*; **Q9NZT2**)	0.019	1.6	N2	2(2)	Downregulated in head and neck squamous cell carcinoma[107]	-
MTTP (*Microsomal triglyceride transfer protein large subunit*; **P55157**)	0.019	7.3	N2	10(9)	-	-
MOGS (*Mannosyl-oligosaccharide glucosidase*; **Q13724**)	0.021	1.8	N0	2(2)	-	-

PRDX3 (*Thioredoxin-dependent peroxide reductase, mitochondrial*; **P30048**)	0.026	1.7	N0	6(6)	Upregulated in hepatocellular carcinoma[108] Upregulated in cervical cancer[109] Upregulated in prostate cancer[110]	-
AHSA1 (*Activator of 90 kDa heat shock protein ATPase homolog 1*; **O95433**)	0.033	1.5	N2	2(2)	Associated with drug sensitivity in cancer cell lines[111]	-
HSPA1B (*Heat shock 70 kDa protein 1A/1B*; **P08107**)	0.034	1.4	N2	23(21)	Associated with progression and grade in cervical cancer[112]	Downregulated in colorectal cancer[113] Association with survival in colorectal cancer[114]
NUDC (*Nuclear migration protein nudC*; **Q9Y266**)	0.034	2.7	N2	2(2)	Associated with carcinogenesis and metastasis in esophageal cancer[115]	-
HSP90AB1 (*Heat shock protein HSP 90-beta*; **P08238**)	0.034	1.7	N2	41(22)	Upregulated in gastric cancer[116]	-
DCD (*Dermcidin*; **P81605**)	0.040	2.4	N2	3(3)	Expressed in cancer cell lines and tissues[117]	-
PACAP (*Plasma cell-induced resident endoplasmic reticulum protein*; **Q8WU39**)	0.041	2.8	N0	4(4)	-	-
COPG (*Coatomer subunit gamma*; **Q9Y678**)	0.042	1.9	N0	3(3)	Upregulated in lung cancer[118]	-
TKT (*Transketolase*; **P29401**)	0.042	1.9	N0	12(12)	Downregulated in HER2 negative breast cancer[119] Upregulated in thyroid cancer[120]	Associated with progression in colorectal cancer[121]
YWHAQ (*14-3-3 protein theta*; **P27348**)	0.043	1.4	N2	11(8)	Upregulated in lung cancer[122] Serum marker for lung cancer[123]	-
KARS (*Lysyl-tRNA synthetase*; **Q15046**)	0.043	1.6	N2	9(9)	-	-
PDCD6IP (*Programmed cell death 6-interacting protein*; **Q8WUM4**)	0.043	1.7	N0	8(8)	-	-
NDUFAF3 (*NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 3*; **Q9BU61**)	0.044	4.2	N0	2(2)	-	-
NDUFB5 (*NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial*; **O43674**)	0.044	2.1	N0	2(2)	-	-
GDI2 (*Rab GDP dissociation inhibitor beta*; **P50395**)	0.045	1.6	N2	21(20)	Upregulated in gastric cancer[124]	-
ACOT7 (*Cytosolic acyl coenzyme A thioester hydrolase*; **O00154**)	0.046	2.7	N2	5(5)	-	-
TNPO1 (*Transportin-1*; **Q92973**)	0.046	1.4	N2	2(2)	-	-
FSCN1 (*Fascin*; **Q16658**)	0.046	1.6	N0	2(2)	Associated with lymph node metastasis and prognosis in oesophageal squamous cell carcinoma[125] Associated with prognosis in breast cancer[126] Upregulated in prostate cancer[127]	Associated with lymph node metastasis in colorectal cancer[128]
FH (*Fumarate hydratase, mitochondrial*; **P07954)**	0.047	1.3	N0	3(3)	Downregulated in clear cell renal cancer[129]	-
SSR1(*Translocon-associated protein subunit alpha*; **P43307)**	0.049	1.5	N0	2(2)	-	-

Table 7 – Label-free quantitative proteomics derived candidate proteins discriminating the nodal status

5.1.4 Immunohistochemical validation of FXYD3, S100A11 and GSTM3

Three proteins, which discriminated the nodal status of colon cancer samples in the initial screening by MALDI Imaging or label-free quantitative proteomics, were validated by immunohistochemical staining of an independent patient cohort (n = 168). For this, a tissue microarray comprising non-metastasised (UICC stage II, pN0, n = 87) and regionally metastasised (UICC stage III, pN2, n = 81) primary tumour tissue samples was established (**Table 5**). The tissue samples were taken from the tumour bank of the Department of Medicine, Technische Universität München. Pathological data were provided by the Institute of Pathology, Technische Universität München. Two proteins identified by MALDI Imaging, FXYD3 and S100A11, and GSTM3, which had the highest discriminating power (p = 0.003) in label-free quantitative proteomics, were validated by immunohistochemical staining of this tissue microarray (**Figure 12**).

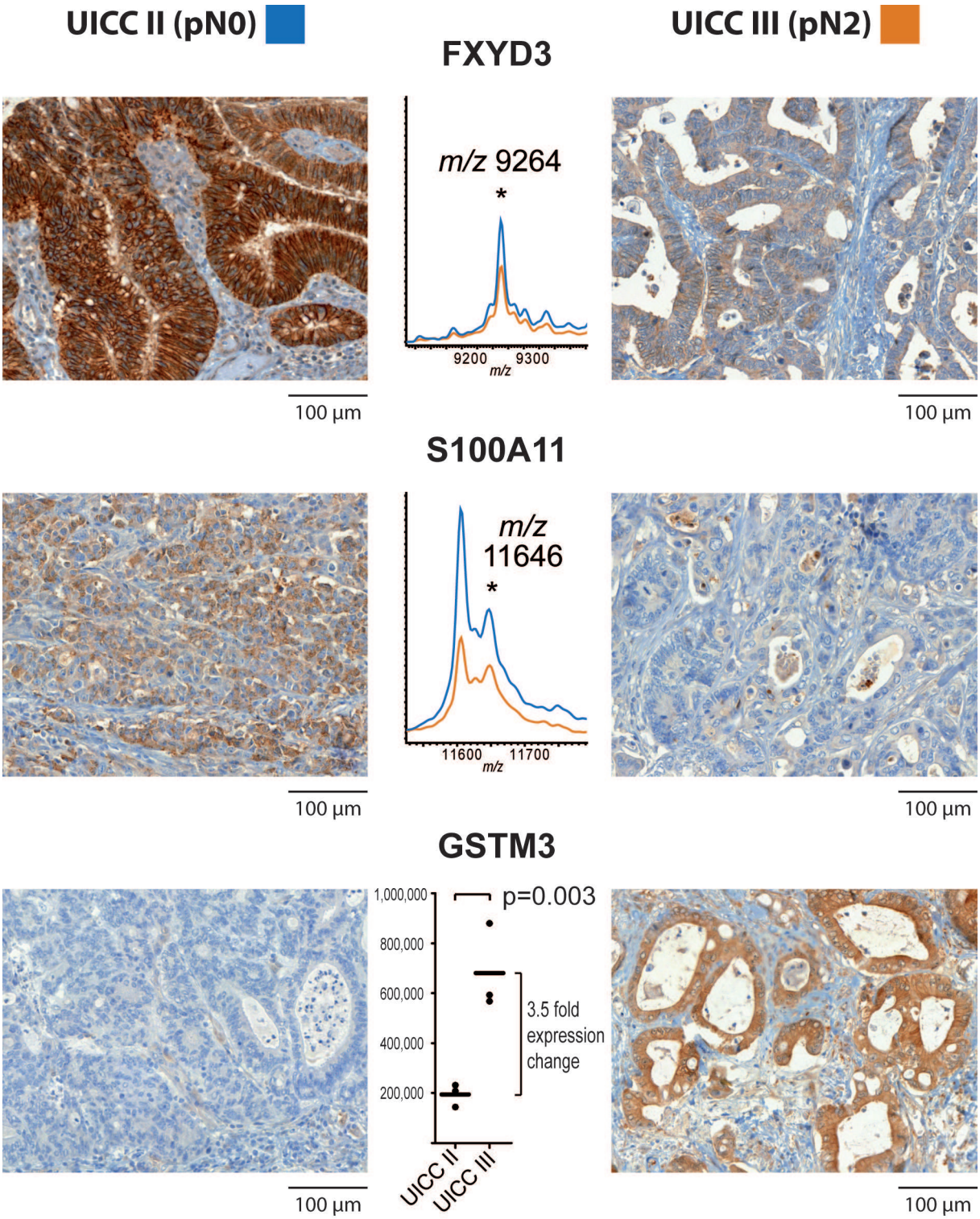


Figure 12 – Immunohistochemical validation of FXYD3, S100A11 and GSTM3. Representative tissues of UICC stage II (left) and UICC stage III (right) patient samples are shown. For FXYD3 (top) and S100A11 (middle) the respective *m/z* species are displayed. For GSTM3 (bottom) the expression levels determined in label-free quantitative proteomics are shown. The immunohistochemical staining displays clear differences in expression in UICC stage II and UICC stage III patients. FXYD3 and S100A11 are stronger expressed in UICC stage II patients. GSTM3 is stronger expressed in UICC stage III patients.

For analysis, the staining intensities of the tumours were visually assessed, a cut-off between low and high staining intensity was determined, and the samples were grouped accordingly. For FXYD3 the cut-off was set between low and intermediate expression, for S100A11

between no detectable and weak expression, and for GSTM3 between no detectable and weak expression. Then, a correlation of the immunohistochemical staining intensity with the nodal status (pN), the tumour depth (pT), the tumour grading (G) and the patients' gender was tested first in a univariate setting with Fisher's exact test and then in a multivariate setting with a generalised linear model (**Table 8**).

	FXYD3				**S100A11**				**GSTM3**			
	low	high	p-value (uni.)[1]	p-value (mult.)[2]	low	high	p-value (uni.)[1]	p-value (mult.)[2]	low	high	p-value (uni.)[1]	p-value (mult.)[2]
Total number of patients (ratio)	96 (0.64)	54 (0.36)			96 (0.6)	64 (0.4)			23 (0.14)	139 (0.86)		
Gender			0.8627	0.8166			0.1711	0.1597			0.8696	0.8019
male	56	33			17	80			56	39		
female	40	21			6	59			40	25		
Tumour depth			**0.0210**	0.1256			0.2486	0.7538			0.8893	0.7859
pT1	1	0			0	1			1	0		
pT2	1	0			1	0			1	0		
pT3	67	48	**0.0141**[3]		17	109	0.7861[3]		73	52	0.6913[3]	
pT4	27	6			5	29			21	12		
Nodal status			**0.0110**	**0.0683**			**0.0071**	**0.0232**			**0.0063**	**0.0173**
pN0	41	35			6	79			59	25		
pN2	55	19			17	60			37	39		
Tumour grading			**0.0293**	0.1669			0.3995	0.5247			0.3208	0.5404
pG1	0	2			0	2			2	0		
pG2	41	30	**0.0881**[4]		8	69	0.1836[4]		49	27	0.2574[4]	
pG3	55	22			15	68			45	37		

Table 8 – Immunohistochemical validation – Univariate and multivariate analysis. ([1]Fisher's exact test [uni. = univariate]; [2]Generalised Linear Model [mult. = multivariate]; [3]pT3 vs. pT4; [4]G2 vs. G3)

With Fisher's exact test, FXYD3 displayed the highest correlation with the nodal status ($p = 0.0110$) and lower correlations with the tumour depth ($p = 0.0210$) and the tumour grading ($p = 0.0293$). S100A11 displayed a significant correlation with the nodal status ($p = 0.0071$) and no significant correlation with the tumour depth or the tumour grading. GSTM3 displayed a significant correlation with the nodal status ($p = 0.0063$) and no significant correlation with tumour depth or the tumour grading. No correlation between the staining intensities of the three proteins and the patients' gender was observed.

With the generalised linear model, a significant correlation with the nodal status was observed for S100A11 ($p = 0.0232$) and GSTM3 ($p = 0.0173$), and a correlation by trend for FXYD3 ($p = 0.0683$). For all three proteins, no significant correlation of the staining intensity with the tumour depth, the tumour grading, or the patients' sex was determined in multivariate analysis (**Table 8**).

5.2 Discrimination of tumour entities by proteomic classification

Correct, unambiguous tumour diagnosis is the initial step in cancer therapy since the patient's regimen is based on the correct tumour classification. So far, clinical diagnosis relies on histological and often extensive immunohistochemical analyses of tumour biopsies.[65] Proteomic classification might support this standard diagnostic approach in the future.

The suitability of MALDI Imaging for tumour classification was assessed in a two-step approach.* At first, it was tested whether it is possible to discriminate tumour entities located in different organ sites by their proteomic profiles. For this, primary tumour tissue samples of six adenocarcinoma entities were analysed by MALDI Imaging (**Table 9**) and a proteomic classifier that could discriminate these tumour entities with high accuracy was generated. After this proof-of-principle, this approach was tested in the more clinically relevant context of identifying the tumour origin of a metastasis with an unknown primary tumour. Therefore, it was tested whether it is possible to discriminate tumour entities which are either located within the same organ site (liver) or which have the same origin (colon).

* The results of this classification approach have been submitted to the Journal of Proteome Research and are currently in revision.

Organ site	Tumour origin	Tumour type	Subtype	Grading	Number of samples
Distal oesophagus	Primary tumour	Adenocarcinoma	- (Barrett's)	G1 G2 G3	3 11 19
Breast	Primary tumour	Adenocarcinoma	Invasive Ductal	G2 G3	2 28
Colon	Primary tumour	Adenocarcinoma		G2 G3	13 8
Liver	Primary tumour	Adenocarcinoma	Hepato-cellular	G1 G2 G3	1 9 5
	Metastasis of colon carcinoma	Adenocarcinoma	-	G2 G3	16 3
Stomach	Primary tumour	Adenocarcinoma	Intestinal	G2 G3	9 14
			Diffuse	G3	9
			Mixed	G3	3
			NA	G2 G3	3 5
Thyroid gland	Primary tumour	Adenocarcinoma	Papillary	-	29

Table 9 – Characteristics of the tumour samples for proteomic classification.

5.2.1 Classification of six tumour entities located in different organ sites

Tumour samples (n = 171) of six tumour entities located in different organ sites (Barrett's cancer, n = 33; breast cancer, n = 30; colon cancer, n = 21; hepatocellular carcinoma, n = 15; gastric cancer, n = 43; thyroid cancer, n = 29) were analysed by MALDI Imaging (**Table 9** and **Figure 13**). The tissue samples were taken from the tumour bank of the Department of Medicine, Technische Universität München. The pathological data were provided by the Institute of Pathology, Technische Universität München.

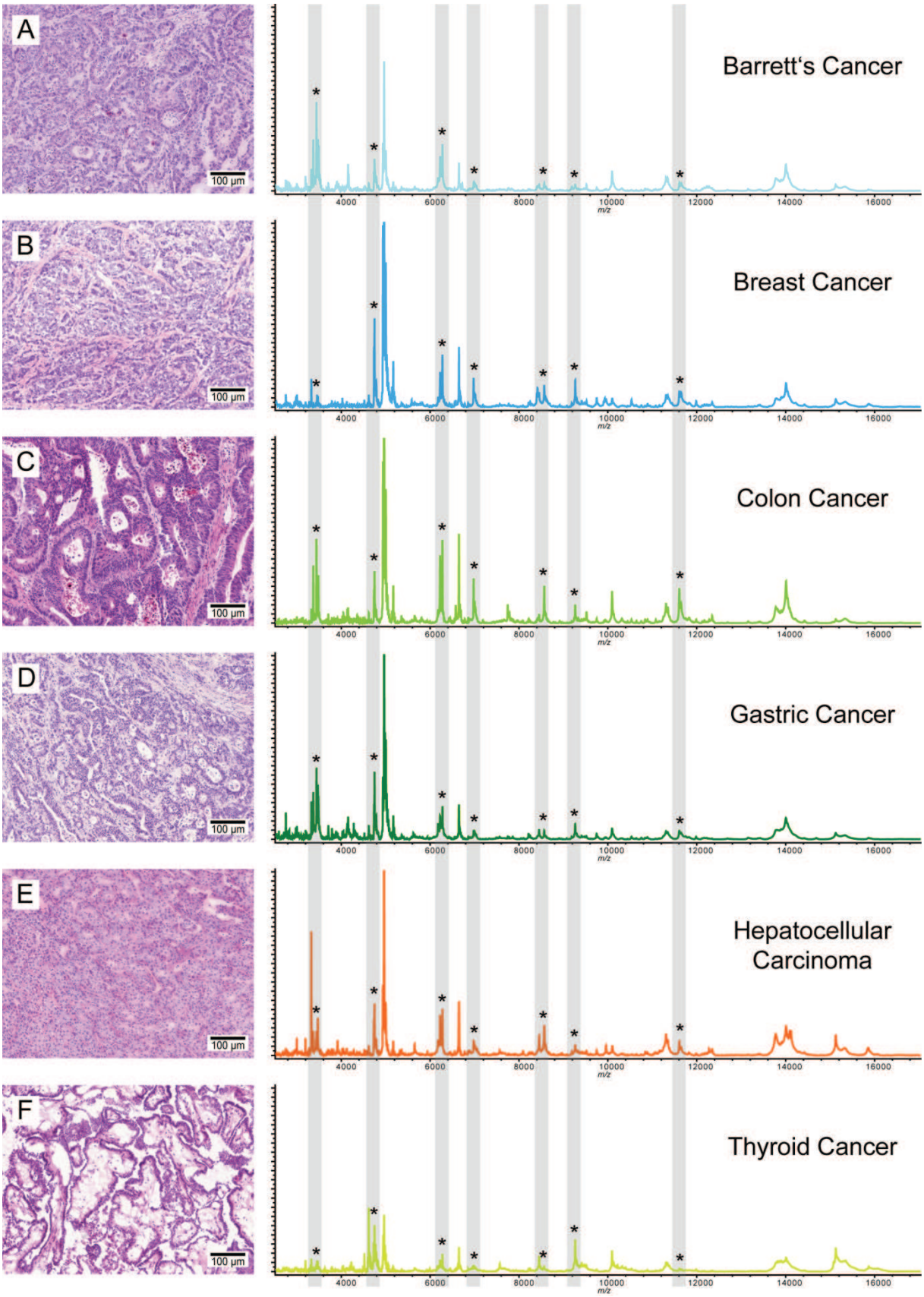

A
100 µm
Barrett's Cancer
B
100 µm
Breast Cancer
C
100 µm
Colon Cancer
D
100 µm
Gastric Cancer
E
100 µm
Hepatocellular Carcinoma
F
100 µm
Thyroid Cancer
4000
6000
8000
10000
12000
14000
16000
m/z

Figure 13 – Histological staining (haematoxylin and eosin) of representative samples of each tumour entity (A: Barrett's Cancer, B: Breast Cancer, C: Colon Cancer, D: Gastric Cancer, E: Hepatocellular Carcinoma, F: Thyroid Cancer) and their average proteomic spectra. The classification was based on 112 to 123 *m/z* species with an average over the 50 repeats of 117 *m/z* species. Exemplarily, 7 of them are highlighted in grey and marked with asterisks.

This resulted in differential mass spectra, which could already be discriminated by sight. The cancer cell specific spectra were extracted and classified (**Figure 14** and **Table 10**).

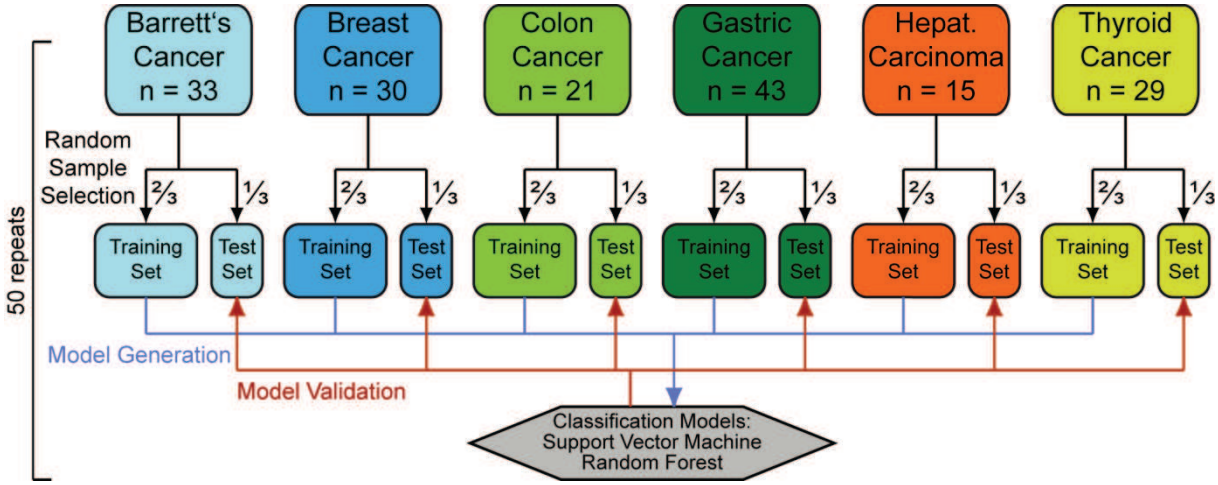


Figure 14 – Schematic display of the classification of the six adenocarcinoma entities. First, a model is generated using a training set. Then, this model is validated on a test set. For reducing the sampling error the classification results are the average of 50 independent samplings. (hepat. carcinoma = hepatocellular carcinoma)

	Barrett's cancer	**Breast cancer**	**Colon cancer**	**Gastric cancer**	**Hepat. carcinoma**	**Thyroid cancer**	**Total number of samples**
Total number of samples	33	30	21	43	15	29	171
Samples in training set	22	20	14	29	10	19	114
Samples in test set	11	10	7	14	5	10	57

Table 10 – Classification of six tumour entities located in different organ sites – Training set and test set make-up. (hepat. = hepatocellular)

The classification was based on 112 to 123 *m/z* species with an average over the 50 repeats of 117 *m/z* species. For the training set the overall accuracy was 99.33% for the Support Vector Machine and 100% for the Random Forest algorithm. The sensitivities, specificities, and accuracies for the individual tumour entity subsets were higher than 98% for the Support Vector Machine (SVM) and 100% for the Random Forest (RF) algorithm (**Table 11**).

	Barrett's cancer		Breast cancer		Colon cancer		Gastric cancer		Hepat. carcin.		Thyroid cancer		Overall result	
	SVM	RF	SVM	RF	SVM	RF	SVM	RF	SVM	RF	SVM	RF	SVM	RF
Training set														
Sensitivity [%]	100	100	99.70	100	98.43	100	98.97	100	98.20	100	100	100		
Specificity [%]	100	100	99.81	100	100	100	99.98	100	99.79	100	99.64	100		
Accuracy [%]	100	100	99.79	100	99.81	100	99.72	100	99.65	100	99.70	100	99.33	100
Test set														
Sensitivity [%]	85.82	84.96	80.40	85.27	82.86	88.69	81.00	76.25	73.20	54.36	88.80	87.96		
Specificity [%]	98.26	94.51	96.85	97.80	98.56	97.13	90.00	92.36	98.31	99.08	96.47	95.81		
Accuracy [%]	95.86	92.66	93.96	95.60	96.63	96.09	87.79	88.40	96.11	95.16	95.12	94.44	82.74	81.18

Table 11 – Classification of six tumour entities located in different organ sites – Classification results. (hepat. carcin. = hepatocellular carcinoma)

Applying the classifiers to the test set yielded an overall accuracy of 82.74% for the Support Vector Machine and 81.18% for the Random Forest algorithm. The individual sensitivities were mostly above 80%. The sensitivity for the hepatocellular carcinoma sample subset was lower for both classification algorithms. For the gastric cancer sample subset it was slightly below 80% for the Random Forest algorithm (**Table 11**). The individual specificities for all tumour entity subsets and both classifiers were higher than 90%, mostly even higher than 95% (**Table 11**). The individual accuracies were higher than 85%, in most of the cases even higher than 95% (**Table 11**).

5.2.2 Classification of three tumour entities which are either located within the same organ site (liver) or which have the same origin (colon)

After it has become clear that the different tumour entities located in different organ sites could be discriminated with high confidence, this proteomic classification approach was applied to a second cohort which emulated the clinical context of identifying the tumour

origin of a metastasis with an unknown primary tumour more closely. We tested whether it is possible to discriminate different tumour entities which are located within the same organ site or have the same origin. The cohort (n = 55) consisted of colon cancer primary tumour (n = 21), colon cancer liver metastasis (n = 19) and hepatocellular carcinoma (n = 15) samples. The patient samples were analysed by MALDI Imaging, the spectra specific for cancer cells were extracted and the cohort was classified as previously (**Table 12**).

	Colon cancer primary tumour	**Colon cancer liver metastasis**	**Hepat. carcinoma**	**Total number of samples**
Total number of samples	21	19	15	55
Samples in training set	14	13	10	37
Samples in test set	7	6	5	18

Table 12 – Classification of three tumour entities either located within the same organ site (liver) or having the same origin (colon) – Training set and test set make-up. (hepat. = hepatocellular)

The classification was based on 36 to 63 *m/z* species with an average over the 50 repeats of 50 *m/z* species. For the training set the overall accuracy was 94.92% for the Support Vector Machine (SVM) and 100% for the Random Forest (RF) algorithm. The individual sensitivities, specificities, and accuracies for the three subsets were higher than 95% for the Support Vector Machine. Only the sensitivity for the hepatocellular carcinoma subset (87.20%) and the specificity for the colon cancer liver metastasis subset (92.42%) were slightly lower. The individual sensitivities, specificities and accuracies for the three subsets were 100% for the Random Forest algorithm (**Table 13**).

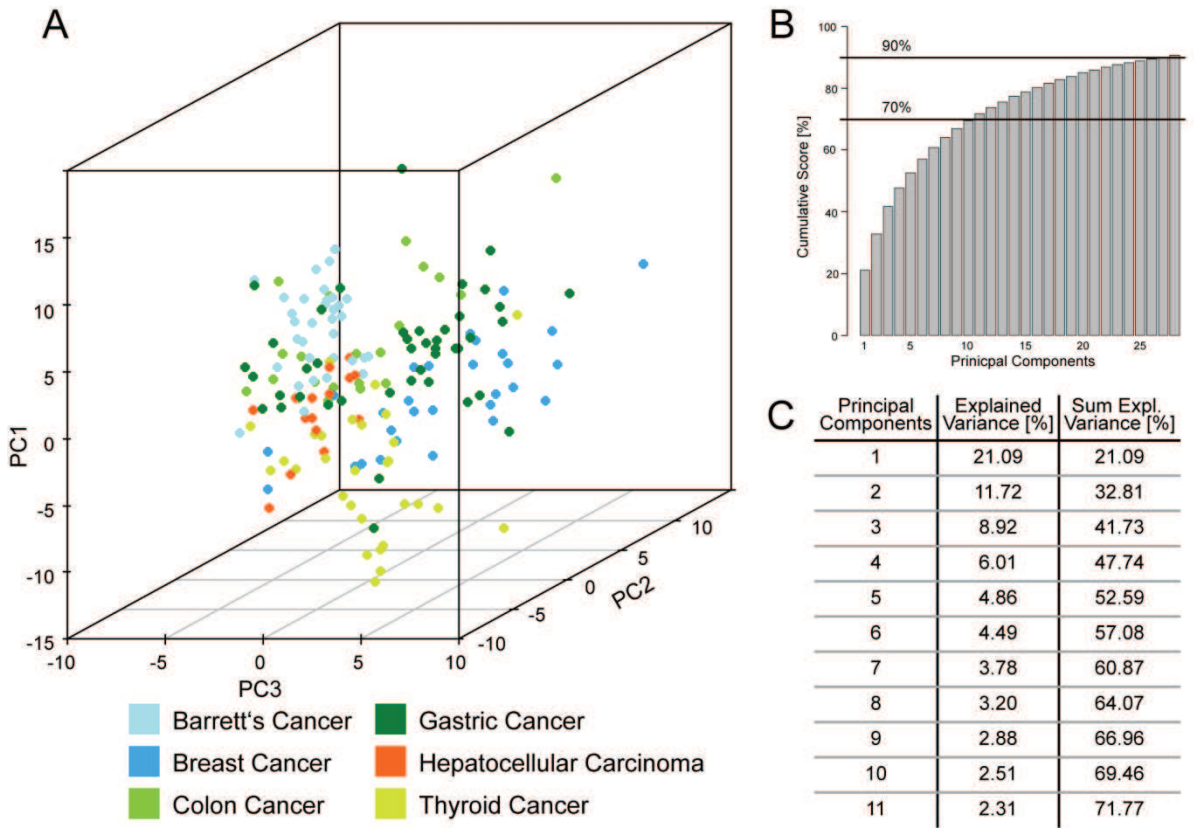
	Colon cancer primary tumour		Colon cancer liver metastasis		Hepatocellular carcinoma		Overall result	
	SVM	**RF**	**SVM**	**RF**	**SVM**	**RF**	**SVM**	**RF**
Training set								
Sensitivity [%]	95.71	100	100	100	87.20	100		
Specificity [%]	100	100	92.42	100	99.78	100		
Accuracy [%]	98.38	100	95.08	100	96.38	100	94.92	100
Test set								
Sensitivity [%]	85.71	86.93	91.33	81.42	73.20	76.94		
Specificity [%]	93.09	89.62	85.50	86.99	97.23	96.30		
Accuracy [%]	90.02	88.57	87.44	85.14	90.56	90.92	84.11	82.32

Table 13 – Classification of three tumour entities either located within the same organ site (liver) or having the same origin (colon) – Classification results.

Applying the classifiers to the test set yielded an overall accuracy of 84.11% for the Support Vector Machine and 82.32% for the Random Forest algorithm. The individual sensitivities for the colon cancer primary tumour and for the colon cancer liver metastasis were higher than 80% for both classifiers. The sensitivities for hepatocellular carcinoma were higher than 70%. The individual specificities and accuracies were higher than 85% for the three subsets and for both classifiers (**Table 13**).

5.2.3 Principal component analysis for discrimination of different tumour entities

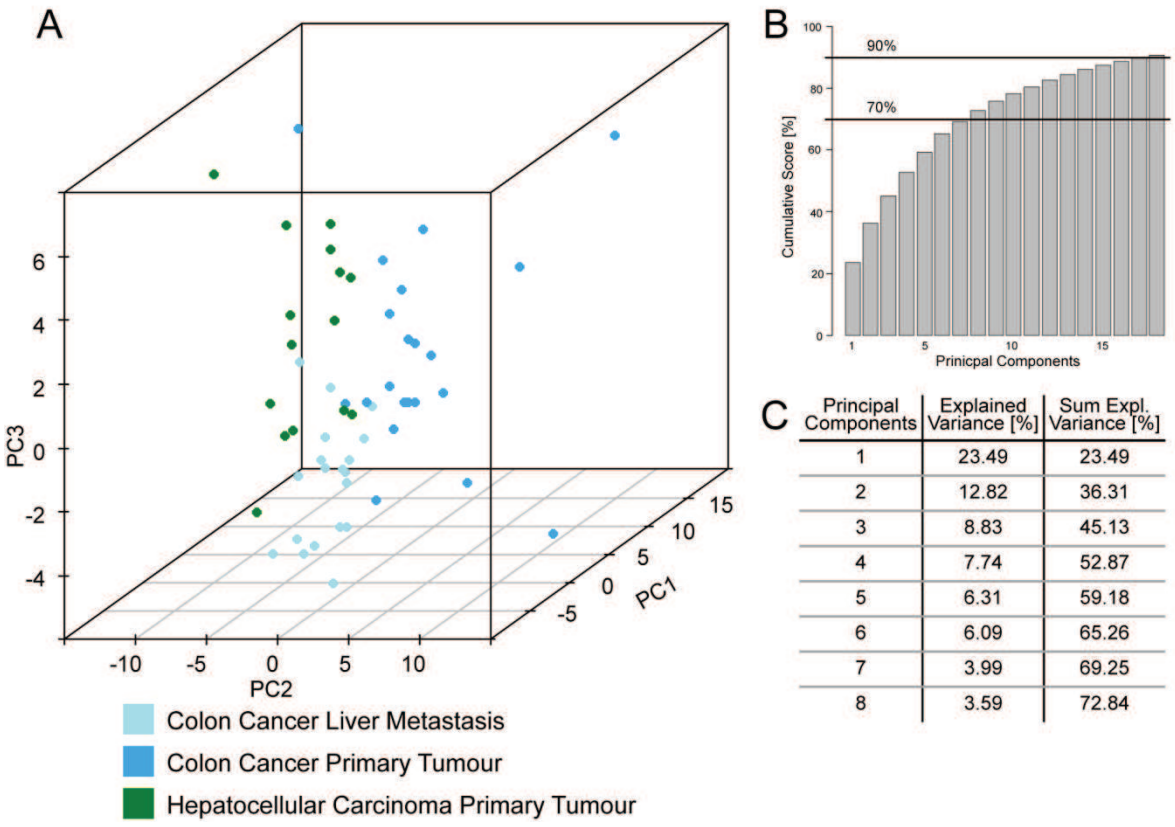
Additionally, a principal component analysis of each of the two cohorts was performed. First, a principal component analysis was performed on the six adenocarcinoma entities located in different organ sites (**Figure 15**).



Principal Components	Explained Variance [%]	Sum Expl. Variance [%]
1	21.09	21.09
2	11.72	32.81
3	8.92	41.73
4	6.01	47.74
5	4.86	52.59
6	4.49	57.08
7	3.78	60.87
8	3.20	64.07
9	2.88	66.96
10	2.51	69.46
11	2.31	71.77

Figure 15 – Principal Component Analysis (PCA) of the six adenocarcinoma entities. (A) Visualisation of the separation of the six tumour entities of the sample cohort by the first three principal components. (B) Histogram displaying the cumulative score of the first 28 principal components, which constitute 90% of the cumulative score. (C) Table of the single and cumulative score of the first 11 principal components, which constitute 70% of the cumulative score. The entities could only be partially separated. This can be explained by the fact that the first three principal components accounted only for 41.74% of the cumulative score. Principal component analysis seems to be less suitable for discriminating tumour entities.

The first three principal components accounted for 41.74% of the variance (score) and thus for the contained information (**Figure 15C**). For explaining 90% of the variance (score), the first 28 principal components were needed (**Figure 15B**). This explains why the six tumour entities were only partially separated in a three dimensional, graphic display (**Figure 15A**). Breast cancer, colon cancer and hepatocellular carcinoma samples were grouped together, while Barrett's cancer, gastric cancer and thyroid cancer samples could be separated well. Then, a principal component analysis was performed on the three tumour entities, which are located within the same organ site or which have the same origin (**Figure 16**).



Principal Components	Explained Variance [%]	Sum Expl. Variance [%]
1	23.49	23.49
2	12.82	36.31
3	8.83	45.13
4	7.74	52.87
5	6.31	59.18
6	6.09	65.26
7	3.99	69.25
8	3.59	72.84

Figure 16 – Principal Component Analysis (PCA) of colon cancer primary tumours, liver metastases and hepatocellular carcinoma primary tumours. (A) Visualisation of the separation of the three tumour entities of the sample cohort by the first three principal components. (B) Histogram displaying the cumulative score of the first 18 principal components, which constitute 90% of the cumulative score. (C) Table of the single and cumulative score of the first 8 principal components, which constitute 70% of the cumulative score. The three entities could be clearly, yet not fully separated. This can be explained by the fact that the first three principal components accounted only for 45.13% of the cumulative score.

The first three principal components accounted for 45.13% of the variance (score) (**Figure 16C**). For explaining 90% of the variance (score), the first 18 principal components were needed (**Figure 16B**). Upon visualisation, the three tumour entities were clearly, yet not fully separated (**Figure 16A**). So, classification by Support Vector Machine or Random Forest has a higher discriminating power than principal component analysis.

5.3 Proteomic markers for relapse in colon cancer primary tumours

Most patients with UICC stage I or stage II colon cancer can be fully cured by resection of the primary tumour. However, a significant fraction of initially "cured" patients develops a relapse with the first 5 years after resection (~5% of UICC stage I and ~15-25% of UICC stage II patients).[21,22] At present, no markers for relapse prediction exist which could be used in clinical diagnostics and aftercare.[8,21,27]

For identifying proteomic markers predicting the risk of tumour recurrence, a tissue based proteomic approach was undertaken using primary tumour tissues of a cohort of UICC stage II patients. Screening for candidate markers was done with MALDI Profiling. For protein identification, a novel bottom-up proteomics approach, which relied on analysing the matrix proteome, was employed. The resulting marker proteins will be validated in the near future.

5.3.1 Identifying markers predicting relapse by MALDI Profiling

For initial identification of proteomic markers that correlate with tumour recurrence in primary tumour tissues of colon cancer, a cohort comprising 119 tumour tissue specimens of UICC stage II patients was analysed by MALDI Profiling. The tissue samples were taken from the tumour bank of the Department of Medicine, Technische Universität München. The clinical and pathological data were provided by the Department of Surgery, Klinikum rechts der Isar, Technische Universität München, and the Institute of Pathology, Technische Universität München. For all primary tumour tissues, detailed histopathological and clinical data of the respective patient were available (**Table 14**).

Characteristics	Relapse developing patients	Relapse free patients
Patients	25	94
Mean age [years] (range)	63 (46 – 87)	62 (23 – 87)
Gender		
male	15	54
female	10	40
Tumour depth (pT)		
pT2	1	0
pT3	17	79
pT4	7	15
Median overall survival [months] (range)	55 (3 – 122)	111 (83 – 180)
Median disease free survival [months] (range)	22 (1 – 97)	-

Table 14 – Relevant histopathological and clinical data of the cohort for relapse marker identification.

The patient cohort contained primary tumour tissues of 25 patients who developed a relapse, and 96 patients who remained tumour free (no relapse). The median for disease free survival (until relapse occurrence) was 22 months and overall survival 55 month for patients who developed a relapse. The median overall survival was 111 months for patients who remained tumour free. Subsequent histological staining of the analysed tissue sections with haematoxylin and eosin allowed a visual assessment of the tissue areas that had been analysed by MALDI Profiling. One to three profiling spots had been applied onto each tissue section depending on its size and tissue morphology. Only measurement areas (profiling spots) that contained tumour cells were used for later analyses. The mass spectra of these measurement areas were extracted. For each patient sample, 50 mass spectra were randomly selected and used for statistical comparison. Spectra alignment and calibration of the mass spectra of the sample cohort resulted in 142 *m/z* species present in all samples. The average spectra of relapse free and relapse developing patients exhibited differential proteomic patterns (**Figure 17**).

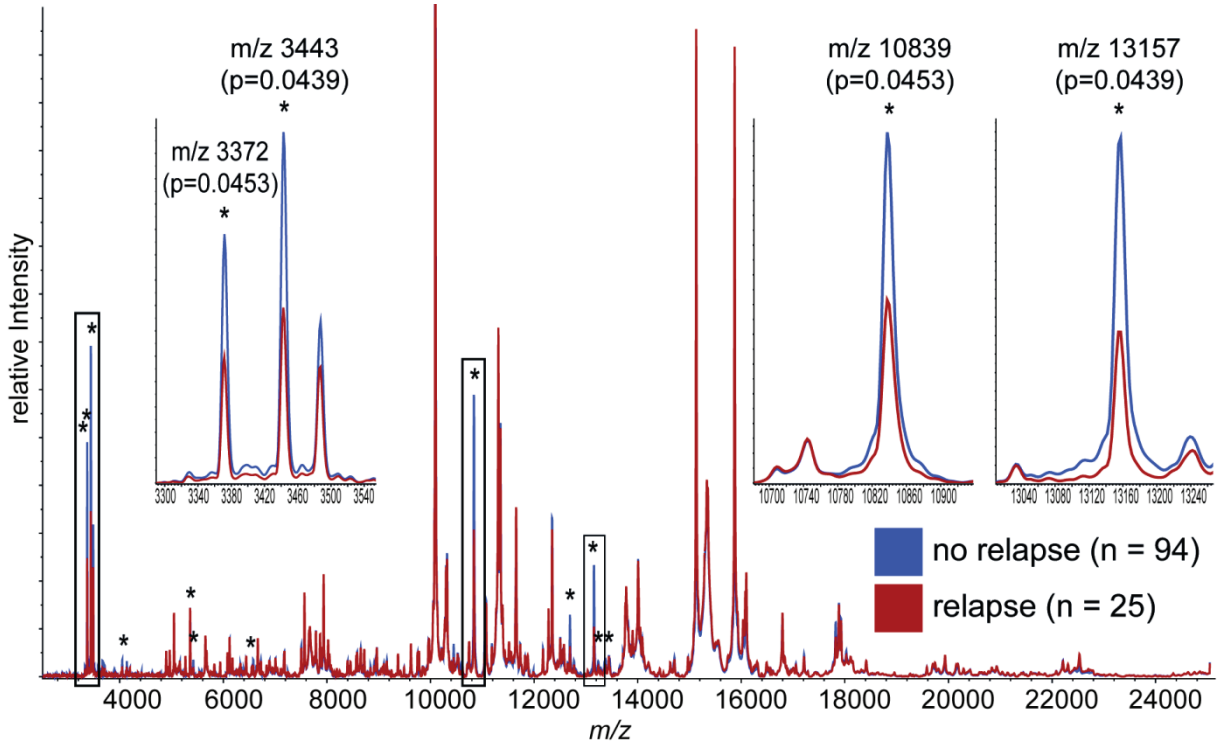


Figure 17 – MALDI Profiling of the relapse cohort. The average spectra of relapse free (blue, n = 94) and relapse developing (red, n = 25) patients are displayed. Differentially expressed *m/z* species ($p < 0.05$) are highlighted with asterisks. Four differentially expressed *m/z* species (*m/z* 3372, *m/z* 3443, *m/z* 10839, and *m/z* 13157) are enlarged and their p-values are given.

Statistical analysis using non-parametric Wilcoxon rank-sum test with subsequent p-value correction according to Benjamini-Hochberg[87] yielded a panel of 27 significantly discriminating *m/z* species (**Table 15**).

Observed *m/z*	**p-value**	**AUC value**	**Putative protein name *(Uniprot ID)***	**Theoretical molecular weight (±Δ) [Da]**	**Associated with cancer**
13157	0.0439	0.72	S100A9 *(P06702)*	13153 (-4)	Unfavourable prognosis in lung cancer[130] Poor tumour differentiation in lung cancer[131] Poor tumour differentiation in breast cancer[132,133]
13240	0.0439	0.72	-	-	-
6577	0.0439	0.71	-	-	-
3329	0.0439	0.70	-	-	-
3443	0.0439	0.70	DEFA1 *(P59665)*	3448 (+5)	Tumour progression in colorectal cancer[134,135] Upregulated in gastric cancer[136]
4046	0.0439	0.70	-	-	-

5268	0.0439	0.70	-	-	-
5420	0.0439	0.70	-	-	-
12693	0.0439	0.70	MRPL14 *(Q6P1L8)*	12688 (-5)	-
3372	0.0453	0.69	DEFA2 *(P59665)*	3377 (+5)	Tumour progression in colorectal cancer[134,135] Upregulated in gastric cancer[136]
10839	0.0453	0.69	S100A8 *(P05109)*	10835 (-4)	Poor tumour differentiation in breast cancer[132] Overpressed in breast, gastric and colorectal cancer[137-139]
3398	0.0461	0.69	-	-	-
13375	0.0491	0.69	-	-	-
3357	0.053	0.68	-	-	-
7567	0.053	0.68	-	-	-
11047	0.053	0.68	-	-	-
4137	0.0562	0.68	-	-	-
8045	0.0562	0.68	-	-	-
3650	0.0576	0.67	-	-	-
10447	0.0688	0.67	S100P *(P25815)*	10442 (-5)	Unfavourable prognosis and metastasis in breast cancer[140] Tumour progression and metastasis in prostate cancer[141] Predictor of distant metastasis and prognosis in lung cancer[142]
3668	0.0843	0.66	-	-	-
16036	0.0843	0.66	-	-	-
7937	0.0855	0.66	-	-	-
8400	0.0855	0.66	CRIP1 *(P50238)*	8402 (+2)	Upregulated in pancreatic cancer,[143] cervical cancer,[144] breast cancer[145] Discriminating HER2 status in breast cancer[45]
15130	0.0855	0.66	HBA1 *(P69905)*	15126 (-4)	-
16306	0.0855	0.66	-	-	-
3465	0.0983	0.65	-	-	-

Table 15 – *M/z* species differentially expressed between primary tumours of patients who developed a relapse and who stayed tumour free. The p-values are calculated by Wilcoxon rank-sum test with subsequent p-value correction according to Benjamini-Hochberg. Additionally, the putative protein names, the theoretical molecular weights, and the reported association with cancer are given.

M/z 13157, *m/z* 13240, *m/z* 6577, *m/z* 3329, *m/z* 3443, *m/z* 4046, *m/z* 5268, *m/z* 5420, *m/z* 12693, *m/z* 3372, *m/z* 10839, *m/z* 3398, and *m/z* 13375 had a p-value below 0.05 ($p < 0.05$), and *m/z* 3357, *m/z* 7567, *m/z* 11047, *m/z* 4137, *m/z* 8045, *m/z* 3650, *m/z* 10447,

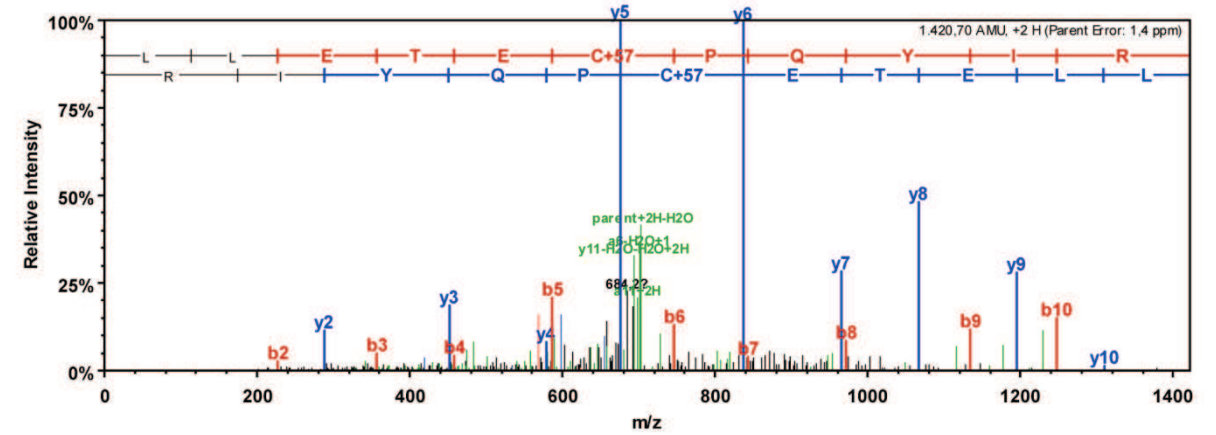
m/z 3668, *m/z* 16036, *m/z* 7937, *m/z* 8400, *m/z* 15130, *m/z* 16306, and *m/z* 3465 had a p-value between 0.05 and 0.1 ($0.05 \leq p < 0.1$) (**Table 15**). The AUC value, which is a second indicator for the discriminating power of the identified *m/z* species, was calculated by receiver operator characteristics (ROC) analysis. For all significant *m/z* species the AUC value was greater than or equal to 0.65, for the *m/z* species with a p-value lower than 0.05 it was even greater than or equal to 0.69 (**Table 15**).

5.3.2 Protein identification of MALDI Profiling derived proteomic markers for relapse prediction

MALDI Profiling – like MALDI Imaging – allows no direct identification of proteins. Therefore, the protein identity of the differentially expressed *m/z* species had to be identified by bottom-up tandem mass spectrometric methods. The employed identification approach varied from the previous one. For identifying proteins correlating with regional lymph node metastasis, tissue lysates were used. This time, the matrix proteome was analysed. For this, tissue sections were prepared in the same way as tissue sections for MALDI Imaging. Then, the proteins were extracted from the matrix and identified by bottom-up proteomics. This was done in collaboration with Stefan Maier and Prof. Dr. Bernhard Küster, Chair of Proteomics and Bioanalytics, Technische Universität München. Sections from three representative patient samples of the relapse cohort were analysed. Proteins were only considered identified if 2 or more peptides were identified. 200 proteins could be identified from the matrix extracts, and 916 proteins could be identified from the tissue extracts. Uniting the matrix and tissue extract results, 938 proteins could be identified. The molecular weights of the identified proteins were calculated using the protein sequences, and the known protein processing (such as methionine removal or amino-terminal acetylation) listed in the Uniprot database (www.uniprot.org). Protein names could be assigned to 8 of the 27 differential *m/z* species: S100A9, DEFA1, MRPL14, DEFA2, S100A8, S100P, CRIP1, and HBA1 (**Table 15**).

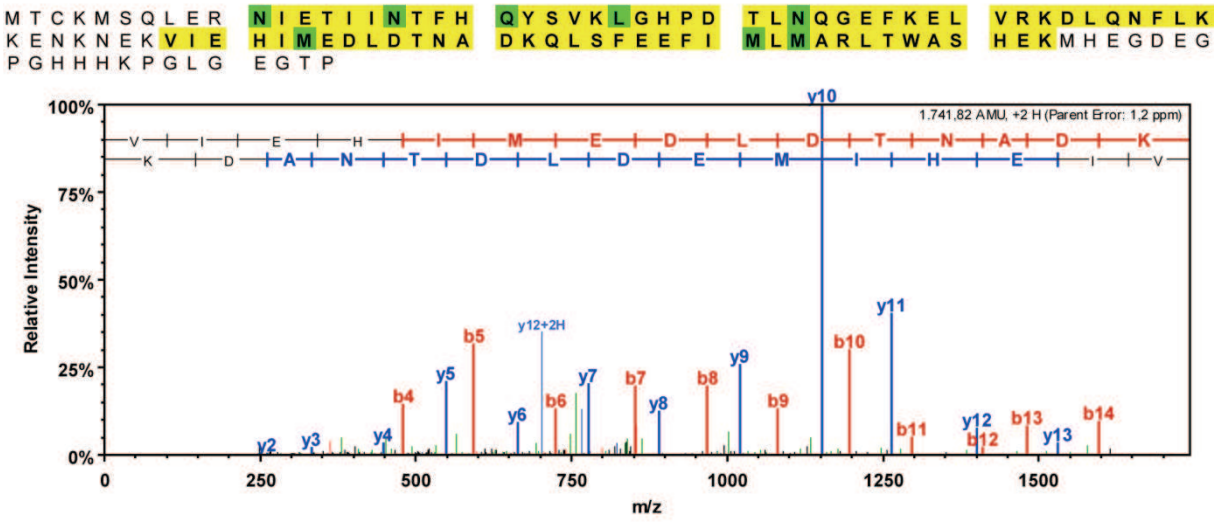
A

IPI00007047 (100%), 10.835,0 Da
Protein S100-A8
11 unique peptides, 12 unique spectra, 15 total spectra, 56/93 amino acids (60% coverage)



B

IPI00027462 (100%), 13.242,3 Da
Protein S100-A9
8 unique peptides, 13 unique spectra, 32 total spectra, 76/114 amino acids (67% coverage)



C

IPI00005721 (100%), 10.201,0 Da
Neutrophil defensin 1
5 unique peptides, 6 unique spectra, 12 total spectra, 25/94 amino acids (27% coverage)

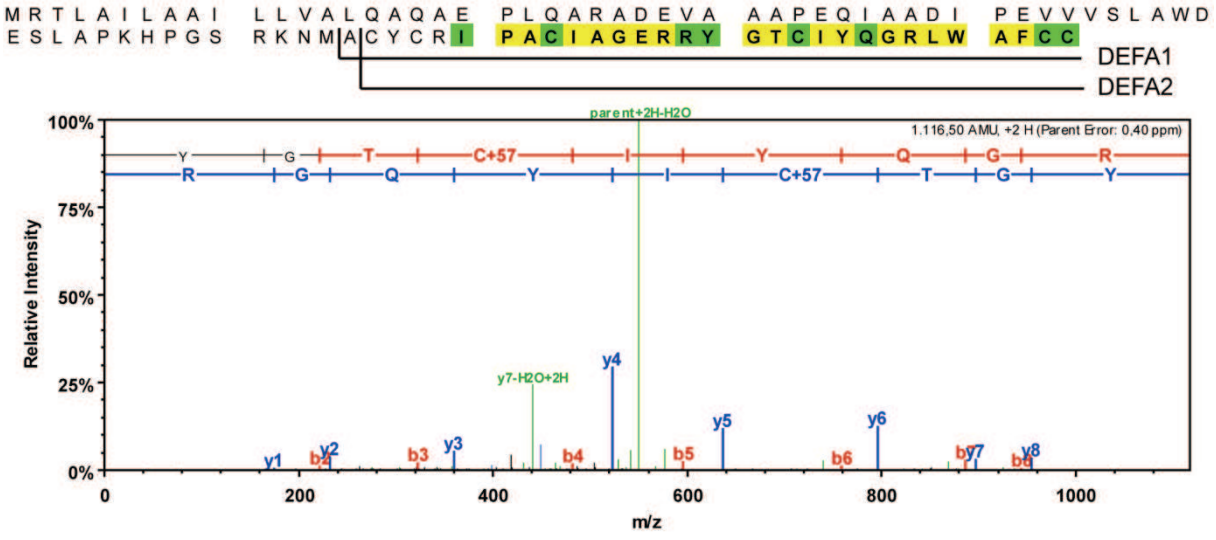


Figure 18 – Sequence coverage and fragment spectra for the identification of MALDI Profiling derived proteins. (A) Identification of S100A8. S100A8 (P05109) was identified with 11 unique peptides, 12 unique spectra, a p-value below 0.01 ($p < 0.01$) and a sequence coverage of 60% (56/93 amino acids). S100A8 could be assigned to *m/z* 10839 in the MALDI Profiling experiment. (B) Identification of S100A9. S100A9 (P06702) was identified with 8 unique peptides, 13 unique spectra, a p-value below 0.01 ($p < 0.01$) and a sequence coverage of 67% (76/114 amino acids). S100A9 could be assigned to *m/z* 13157 in the MALDI Profiling experiment. (C) Identification of DEFA1 and DEFA2. DEFA1 and DEFA2 (P59665) were identified with 5 unique peptides, 6 unique spectra, a p-value < 0.01 and a sequence coverage of 27% (25/94 amino acids) for the whole, unprocessed protein. DEAF1 and DEFA2 comprise the last 30 or 29 amino acids of the unprocessed sequence. So, for this part of the sequence the coverage was actually 83% (25/30 amino acids) or 86% (25/29 amino acids), respectively. DEFA1 and DEFA2 could be assigned to *m/z* 3443 and *m/z* 3377 in the MALDI Profiling experiment.

5.4 Methodological advances in MALDI Imaging – Opening new fields of research

The vast majority of clinical tissue samples, especially those with associated clinical data, are stored as formalin-fixed and paraffin-embedded (FFPE) tissue specimens in the archives of pathology departments.[35] Thus, preparation protocols for MALDI Imaging on this kind of tissue material would open new fields research. So far, several groups have worked on establishing protocols for formalin-fixed tissues.[50-54,81-85] Nevertheless, further improvement of theses protocols is needed, especially in regard to the achievable spatial resolution of the MALDI Imaging measurement and the reproducibility of the tissue preparation. Herein, MALDI Imaging on formalin-fixed paraffin-embedded tissues is attempted using a spray device for trypsin and matrix application onto the tissue sections in order to guarantee a high spatial resolution of the measurement.

5.4.1 MALDI Imaging on formalin-fixed paraffin-embedded tissues – Gaining access to the tissue archives of pathology

At first, MALDI Imaging on formalin-fixed paraffin-embedded tissues was done relying only on antigen retrieval for tissue preparation. For this, an endoscopic biopsy of the distal oesophagus was analysed. Tissue preparation consisted of antigen retrieval by heat and consecutive spray coating of the tissue section with cyanohydroxycinnamic acid matrix. MALDI Imaging was performed in linear mode and with a lateral resolution of 50 µm. The resulting overall spectrum and three differentially expressed *m/z* species are depicted in **Figure 19**.

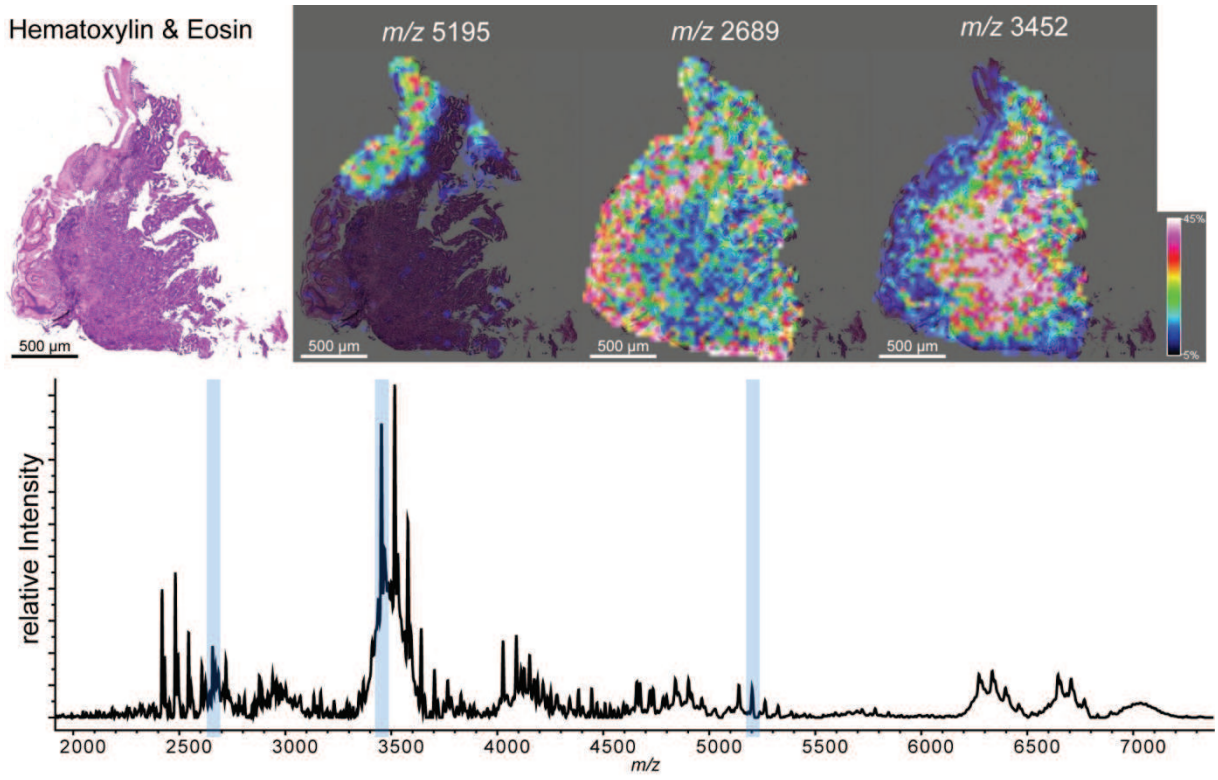


Figure 19 – MALDI Imaging of an endoscopic biopsy of the distal oesophagus. In the upper part, haematoxylin and eosin staining of the tissue section after measurement (left) and visualisation of three *m/z* species located predominantly in the squamous epithelium (*m/z* 5295), the Barrett's metaplasia (*m/z* 2689) and the Barrett's carcinoma (*m/z* 3452). In the lower part, the average spectrum of the endoscopic biopsy, in which the three above visualised *m/z* species are highlighted in blue, is displayed.

In **Figure 19**, three *m/z* species, which were predominantly expressed in the squamous epithelium (*m/z* 5295), the Barrett's metaplasia (*m/z* 2689) and the Barrett's carcinoma (*m/z* 3452), are displayed. Several other *m/z* species had a high correlation with the morphological features of the tissue, too. Many *m/z* species were predominately expressed in the squamous epithelium and the Barrett's carcinoma. Only few *m/z* species had a specific expression in the Barrett's metaplasia, and these *m/z* species had less pronounced differences in their expression levels comparing the expression intensity in Barrett's metaplasia and in Barrett's carcinoma or squamous epithelium. A total of 276 *m/z* species were detected in the average spectrum. After first visual assessment of the data set, a hierarchic clustering was performed on the data set (**Figure 20**).

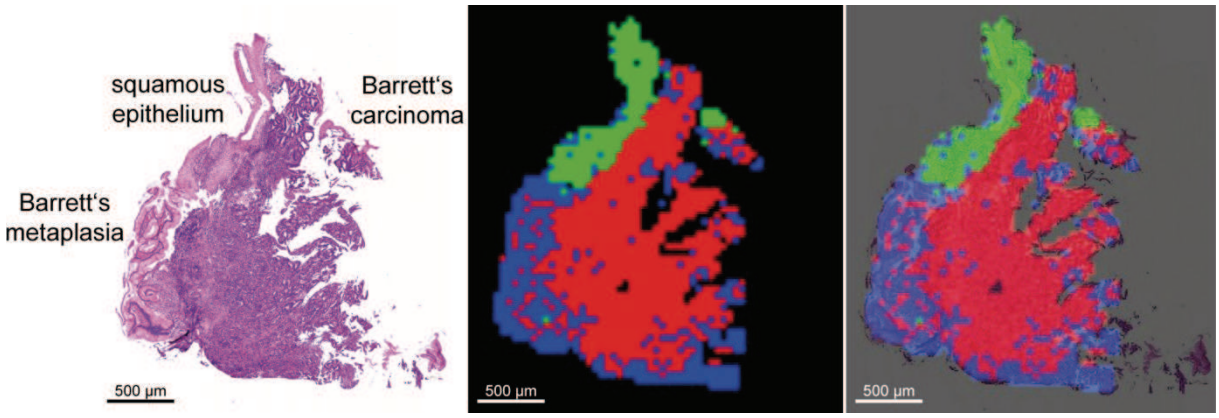


Figure 20 – Hierarchic clustering of the MALDI Imaging data set of the endoscopic biopsy of the distal oesophagus. Haematoxylin and eosin staining of the endoscopic biopsy (left), hierarchic clustering (middle), and superimposition of both (right) are displayed. Squamous epithelium (green), Barrett's metaplasia (blue) and Barrett's carcinoma (red) could be clearly separated by the clustering algorithm.

The hierarchic clustering resulted in a highly accurate separation of the three morphological features present in the endoscopic biopsy. Squamous epithelium, Barrett's metaplasia and Barrett's carcinoma specific spectra were grouped into different branches of the clustering tree (**Figure 20**). So, the tissue preparation protocol relying only on antigen retrieval resulted in specific and measurable expression patterns. However, the quality of the spectrum was suboptimal. It displayed broad peaks, which are a clear indication for protein degradation (**Figure 19**). Additionally, the number of differentially expressed *m/z* species was lower than in comparable, cryo preserved tissues. Therefore, tissue preparation with tryptic digest was attempted in the next step.

For assessing the quality of tryptic tissue preparation, colon tumour samples were used. The carcinogenesis of colon cancer has distinct, well-described steps starting from healthy mucosa, progressing to adenoma, and ending in carcinoma. These steps in carcinogenesis can be easily identified and discriminated by histological tissue staining. This allowed a rapid and reliable analysis of the preparation quality. If the detected *m/z* species were differentially expressed in the different tissue types within a single tissue section the preparation could be assumed as successful. If the same *m/z* species were detectable in the same tissue types in different tissue sections, a statistical comparison of the different tissue types would result in statistically significant, discriminating *m/z* species. If discriminating *m/z* species were

detectable, the preparation could be assumed as robust and reliable. For testing the preparation quality, a cohort of 14 samples was analysed (**Table 16**). Suitable colon cancer tissue samples were taken from the tumour bank of the Department of Medicine, Technische Universität München.

Case	**Morphology of the sample**	**Number of peaks (s/n > 10)**	**Number of peaks (s/n > 5)**	**Number of peaks (s/n > 3)**	**Evaluation**	
					Diff. exp. *m/z* species	Overlap of morph. and clust.
1	Mucosa / Adenoma	30	84	151	+++	+++
2	Mucosa / Adenoma	17	70	160	+++	
3	Mucosa / Adenoma	36	90	176	+++	+++
4	Adenoma	23	73	164	+	
5	Mucosa / Adenoma	31	90	171	+	
6	Mucosa / Carcinoma	37	83	183	++	
7	Carcinoma	33	82	153	+++	+++
8	Adenoma	16	77	152	++	
9	Mucosa / Carcinoma	29	92	188	+++	
10	Mucosa / Adenoma / Carcinoma	29	81	163	+++	++
11	Mucosa / Adenoma / Carcinoma	19	75	176	++	
12	Mucosa / Adenoma / Carcinoma	38	85	174	+	
13	Mucosa / Carcinoma	35	89	159	++	++
14	Mucosa / Adenoma	20	85	175	++	
Av. Ran.	-	28.1 [16-38]	82.6 [70-92]	167.5 [151-188]		

Table 16 – The cohort of formalin-fixed paraffin-embedded colon cancer samples. The morphology of the tissue samples, the number of the detected *m/z* species (depending on the s/n ratio), and a general assessment of the measurement quality in respect to the correlation of *m/z* species expression and tissue morphology are given. (s/n = signal/noise, diff. exp. = differentially expressed, morph. = morphology, clust. = clustering, av. = average, ran. = range)

Tissue preparation was done by antigen retrieval, *in situ* tryptic digest and spray coating with cyanohydroxycinnamic acid matrix. MALDI Imaging was performed in reflectron mode and with a lateral resolution of 150 µm. The average spectra of two representative samples are depicted in **Figure 21**.

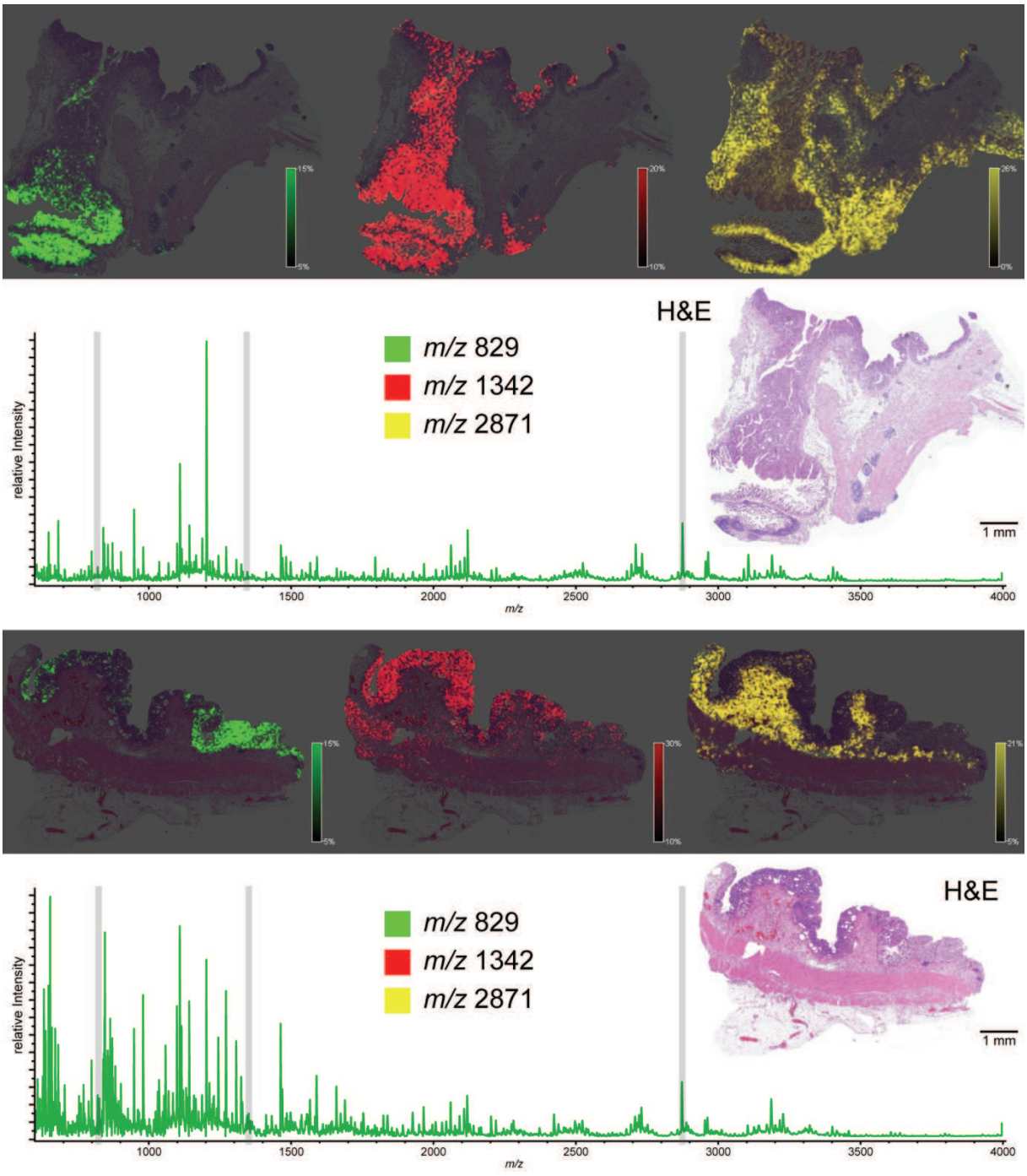


Figure 21 – MALDI Imaging of two representative, formalin-fixed colon cancer samples (#1 and #3, **Table 16**). For both samples, three masses are visualised which are expressed predominately in the mucosa (green, *m/z* 829), in the tumour (red, *m/z* 1342), or in the submucosa (yellow, *m/z* 2871). Below, the average spectra, which have the three visualised *m/z* species highlighted in grey, and the haematoxylin and eosin staining of the tissues are depicted.

The spectra contained between 151 – 188 *m/z* species (s/n > 3, **Table 16**). Increasing the signal to noise level reduced the number of detected peaks in all analysed samples. A signal to noise level of 10 resulted in 16 – 38 detected *m/z* species. No significant inter-sample variation of detectable *m/z* species could be observed. Slight variations were present, but there

was no indication that individual samples were better or less suited for tryptic digest. Many of the *m/z* species displayed a differential expression, which resulted in a tissue type specific revisualisation (**Figure 21**). For further analysis of the data quality, the data sets were subjected to a cluster analysis. Due to the data size and its complexity this could not be done at the Institute of Pathology. Instead, it was done in collaboration with Dr. Theodore Alexandrov and Prof. Dr. Peter Maass, Centre for Industrial Mathematics, University of Bremen. There, an algorithm for processing large MALDI Imaging data sets had been developed previously.[146] Since MALDI Imaging data have a low signal to noise ratio this algorithm relied on an edge-preserving denoising of the mass spectra data which prevented the loss of signal information. After this denoising a high dimensional discriminant clustering was applied to the MALDI Imaging data set.[146] This algorithm was employed on several sample data sets of the colon cancer cohort (**Figure 22** and **Table 16**).

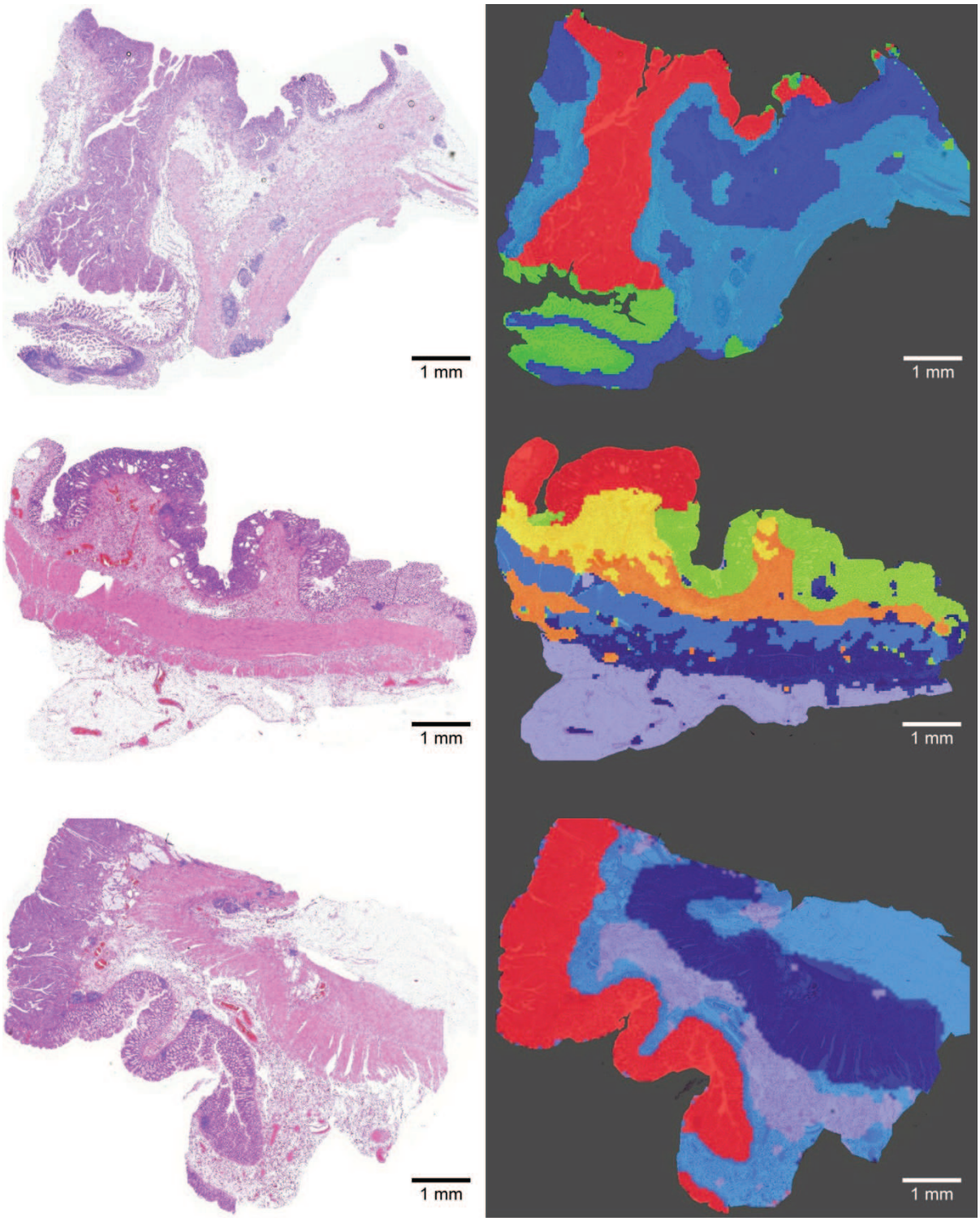


Figure 22 – Haematoxylin and eosin staining (left) and clustering (right) of three representative, formalin-fixed colon cancer samples (#1, #3 and #10, **Table 16**). The number of clusters was selected according to the morphology of the respective tissue. For the topmost sample four clusters, which correspond with the mucosa (green), the tumour (red) and the submucosa and muscle layer (light and dark blue), were chosen. For the middle sample seven clusters, which correspond with the mucosa (green), the tumour (red), the submucosa (yellow and orange), the muscle layer (middle and dark blue) and the subserosa (light blue), were chosen. For the lowermost sample four clusters, which correspond with the tumour and mucosa (red), the submucosa and subserosa (light and middle blue) and the muscle layer (dark blue), were chosen.

This clustering algorithm allowed a good discrimination of many morphological features of the tissue sections (**Figure 22**). Especially the tumour and mucosa regions could be discriminated well from the other morphological features of the tissues. Statistical comparison of the mucosa, adenoma and carcinoma regions of the whole sample cohort was performed, but did not result in statistically significant, discriminating *m/z* species.

The MALDI Imaging data sets of the formalin-fixed colon cancer samples were additionally used for improving the previously mentioned clustering algorithm.[61] Spatial resolution is still limited in MALDI Imaging for technical and preparatory reasons. In order to cope with this limitation, a computational approach for improving the spatial resolution of the clustering data of MALDI Imaging data sets was developed.[61] This so-called super-resolution algorithm relies on a sophisticated smoothing of the boundaries of the clusters within the clustering, also termed segmentation maps. This led to an improved correlation of the clustering data with the histology of the tissue (**Figure 23** and **24**).

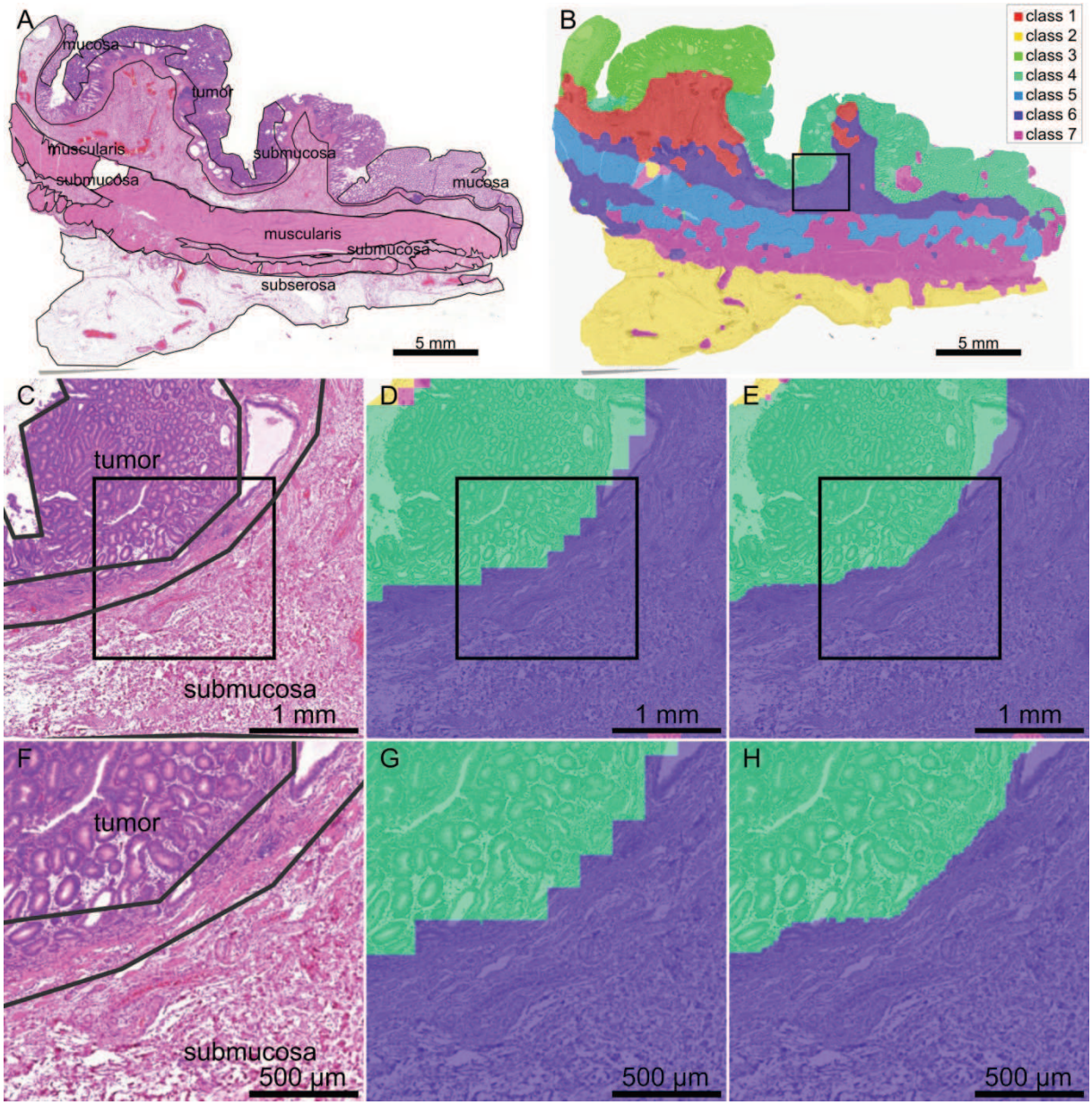


Figure 23 – Super-resolution segmentation of a formalin-fixed colon cancer sample. At the top, the haematoxylin and eosin staining (A) and the full segmentation map (B) of the sample are given. In the middle and at the bottom, haematoxylin and eosin staining (C and F), a segmentation map without super-resolution (D and G) and a segmentation map with super-resolution (E and H) are depicted.[61]

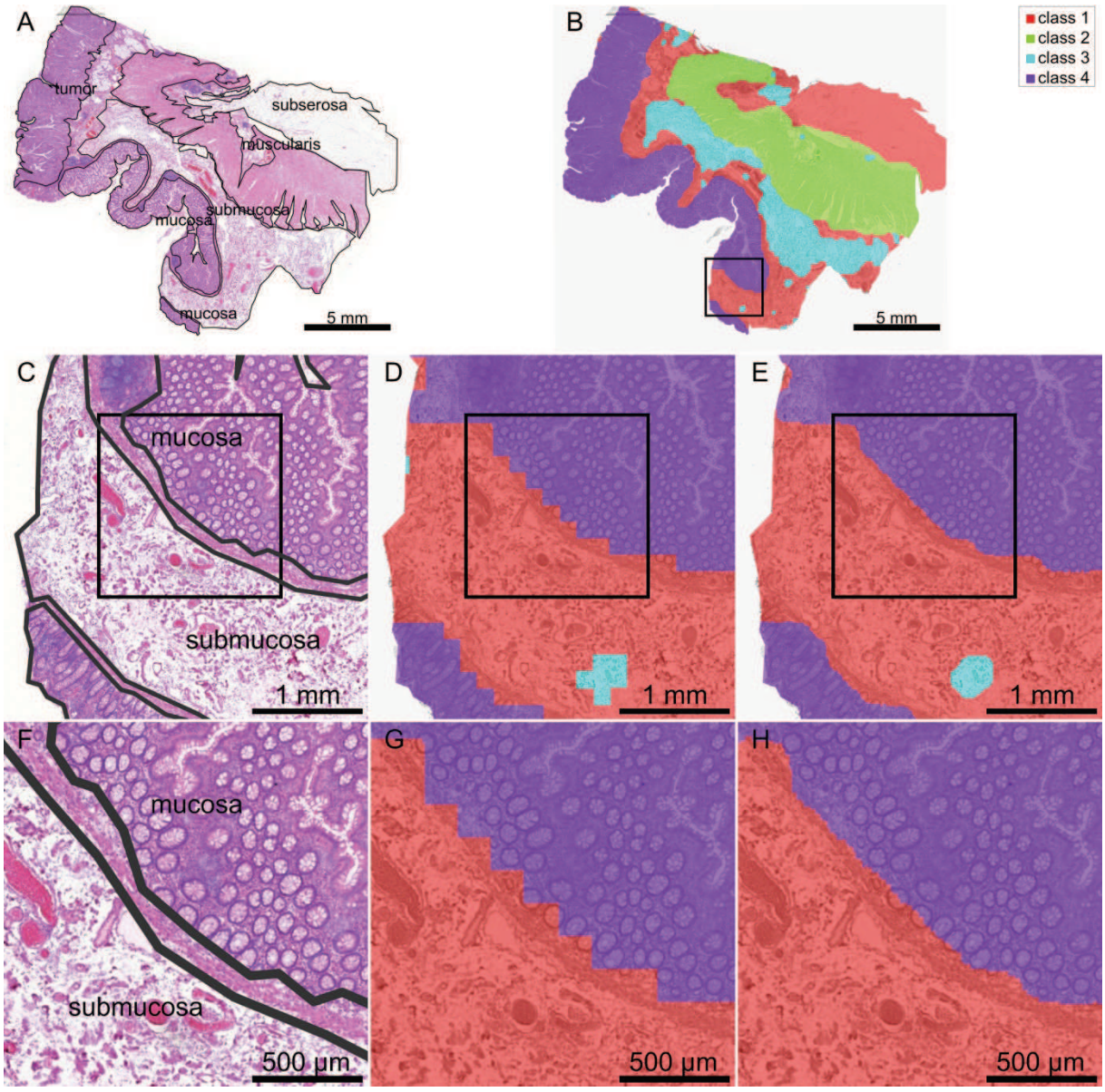


Figure 24 – Super-resolution segmentation of a formalin-fixed colon cancer sample. At the top, the haematoxylin and eosin staining (A) and the full segmentation map (B) of the sample are given. In the middle and at the bottom, haematoxylin and eosin staining (C and F), a segmentation map without super-resolution (D and G) and a segmentation map with super-resolution (E and H) are depicted.[61]

In **Figures 23** and **24** the effect of super-resolution is shown. This algorithm enhances the data resolution after data acquisition. Due to this, the correlation of the individual clusters of the segmentation maps with the tissue morphology is improved which makes data interpretation easier.[61]

5.4.2 MALDI Imaging on alcohol-fixed paraffin-embedded tissues – Bypassing formalin-fixation

A different approach to tackle the formalin "problem" is to circumvent it by using a different fixative that does not cause covalent cross-linking of the proteins. Of course, this can only be done in prospective clinical studies or in studies with animal models. Herein, the feasibility of MALDI Imaging on tissues fixed with a commercial, alcohol-based reagent (PAXgene) is reported.[56] This was done in collaboration with Dr. Bilge Ergin (now: Reischauer) and Prof. Dr. Karl-Friedrich Becker, Institute of Pathology, Technische Universität München. Human pancreatic tissues were used for comparison of the PAXgene fixative with two standard methods of tissue fixation: formalin fixation and cryo preservation. PFPE (PAXgene-fixed paraffin-embedded) and FFPE tissue sections were deparaffinised and spray coated with sinapinic acid matrix. No antigen retrieval was performed. Cryo preserved tissue sections were washed with ethanol and spray coated with sinapinic acid matrix. MALDI Imaging was performed in linear mode and with a lateral resolution of 70 µm. The different fixation methods resulted in spectra of different quality (**Figure 25**).

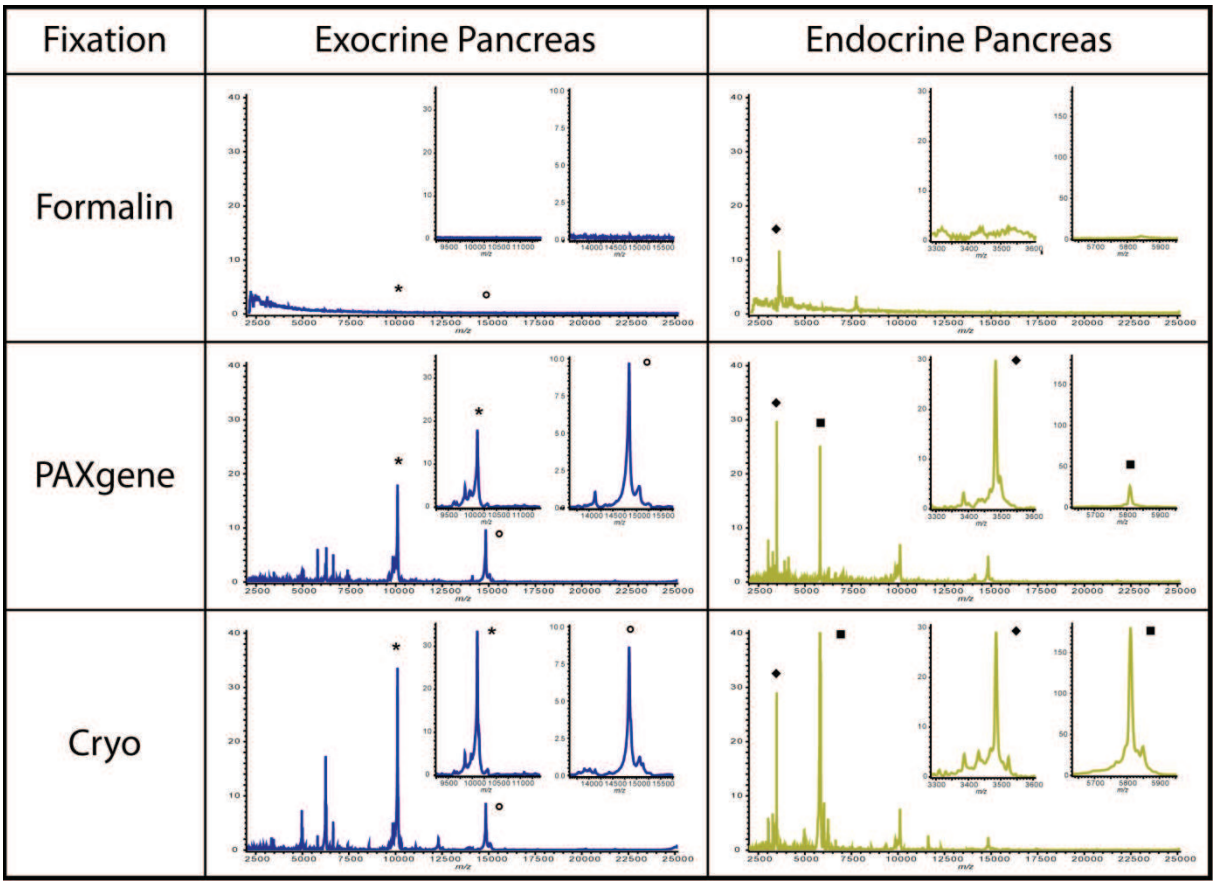


Figure 25 – Average spectra of the exo- (left) and endocrine pancreas (right) of human pancreatic tissue measured by MALDI Imaging. The effect of formalin fixation (top), PAXgene fixation (middle) and cryo preservation (bottom) on MALDI Imaging mass spectra is displayed. The marked *m/z* species are enlarged.[56]

While formalin fixation resulted in a nearly complete absence of *m/z* species in MALDI Imaging, PAXgene fixation and cryo preservation resulted in spectra of good quality. Formalin fixation resulted in no detectable *m/z* species in the endocrine pancreas (**Table 17**). The two visible peaks in the spectrum of the endocrine pancreas (**Figure 25**) have a signal to noise ratio below 5 (s/n < 5) and are therefore not detected as *m/z* signals by the applied software algorithm. MALDI Imaging of PAXgene-fixed and cryo preserved pancreatic tissue sections resulted in similar spectra and a similar number of *m/z* species in the exocrine and the endocrine pancreas, respectively (**Table 17**). PAXgene-fixed tissues displayed slightly lower signal intensities, but signal intensity was high enough for high quality MALDI Imaging mass spectra.[56]

Fixation method	Number of *m/z* species in the exocrine pancreas	Number of *m/z* species in the endocrine pancreas
Formalin	2	0
PAXgene	87	68
Cryo	96	73

Table 17 – Number of *m/z* species detected by MALDI Imaging in formalin-fixed, PAXgene-fixed and cryo preserved pancreatic tissue samples (s/n > 5).

The next step in assessing the suitablity of PAXgene fixation for MALDI Imaging was to visualise the expression patterns of *m/z* species in the tissue samples and to check if a delocalisation of the *m/z* species had occured. Two proteins of the endocrine pancreas are ideally suited for this task. Insulin (*m/z* 3485) and glucagon (*m/z* 5808) are strongly expressed in the endocrine pancreas while being absent in the exocrine pancreas. Visualisation of these two *m/z* species showed that they were predominantly expressed in the endocrine pancreas, while being nearly absent in the exocrine pancreas (**Figure 26**). So, no delocalisation of these two *m/z* species could be observed. No significant differences in the visualisation quality could be determined between PAXgene-fixed and cryo preserved pancreatic tissues, too.[56]

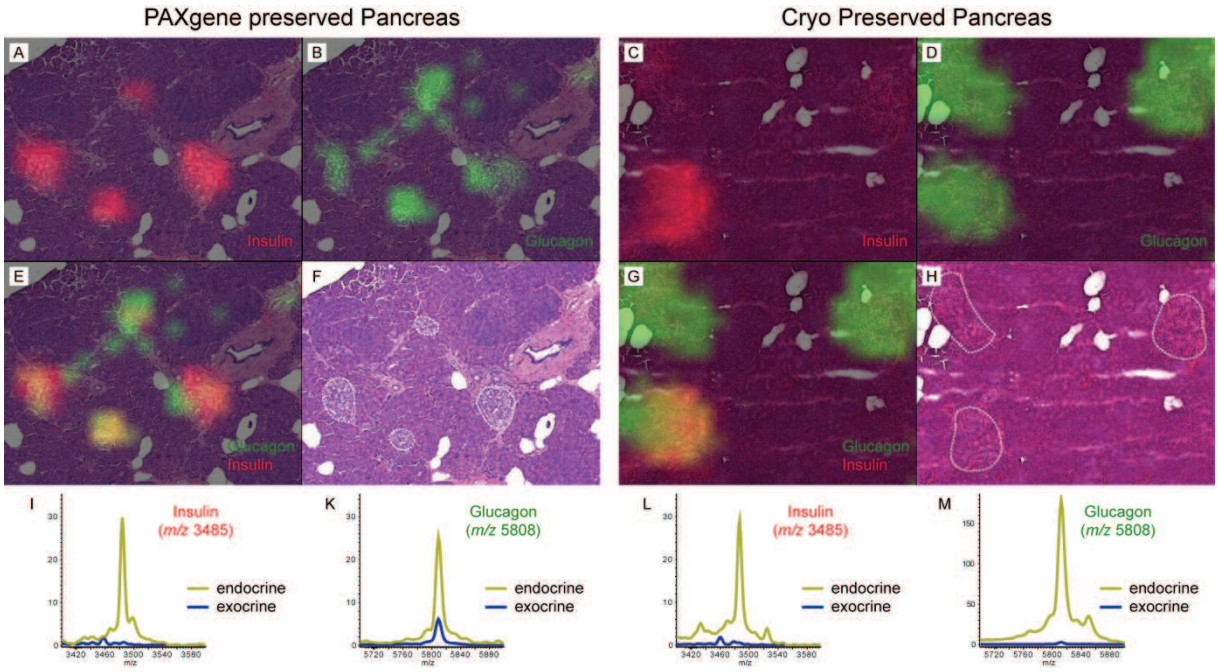


Figure 26 – Expression of insulin (*m/z* 3485) and glucagon (*m/z* 5808) in PAXgene-fixed (left) and cryo preserved (right) pancreatic tissue sections analysed by MALDI Imaging. On the image tiles insulin is visualised in red (A, C), glucagon in green (B, D), insulin and glucagon are co-visualised (E, G), and the haematoxylin and eosin staining of the pancreatic tissue sections is shown (F, H). In the haematoxylin and eosin stained pancreatic sections the endocrine pancreas (Islets of Langerhans) are encircled. In the bottom row spectra displaying the expression levels of insulin (I, L) and glucagon (K, M) in the endocrine (yellow) and exocrine (blue) pancreas are given.[56]

After proving that PAXgene fixation is well suited for MALDI Imaging, the fixation method was used in a proteomic study: *in situ* analysis of the mammalian retina on a proteomic level. The light and dark adaptation of the mammalian retina is a complex process that is only partially understood. Therefore, it was analysed *in situ* by MALDI Imaging. Freshly resected porcine eyecups were procured from the local abattoir. The eyecups were PAXgene-fixed instead of cryo preserved for two reasons. Firstly, correct and reproducible orientation of the sample material during sectioning could only be ascertained with paraffin-embedded material. The eyecups were coronally sectioned. The posterior halves of the eyecups were sagitally oriented while paraffin embedding. Secondly, tissue morphology had to be fully retained (**Figure 27**).

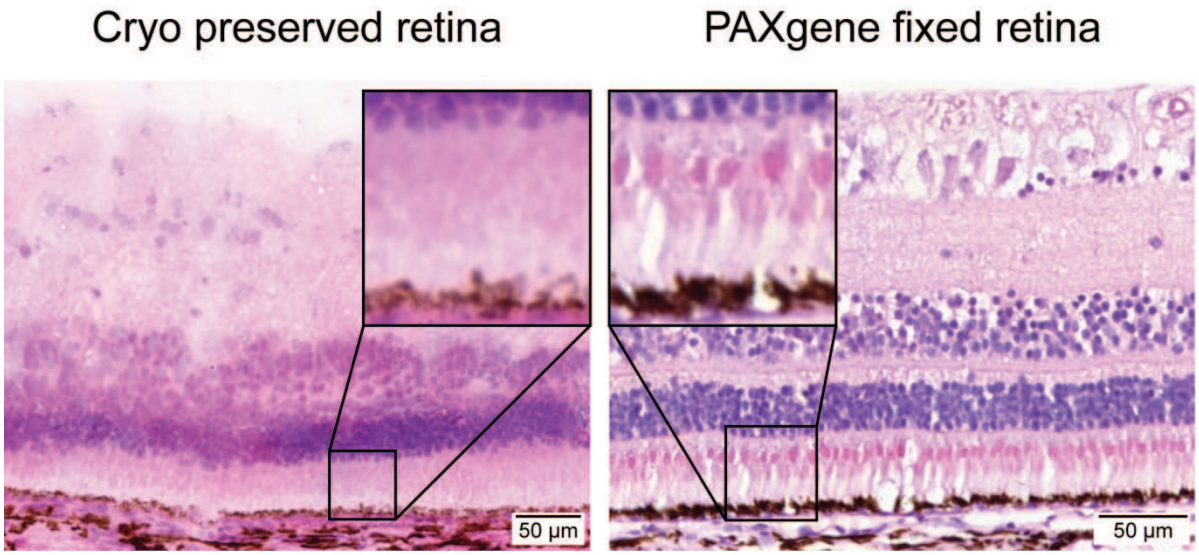


Figure 27 – Comparison of cryo preserved and PAXgene-fixed retina tissue sections. On the left side, a haematoxylin and eosin stained section of a cryo preserved retina sample is displayed. Severe distortions of the tissue morphology are visible. On the right side, a haematoxylin and eosin stained section of a PAXgene fixed retina sample is displayed. The tissue morphology is fully retained. A representative area of the cones- and rods-layer of the two sections is enlarged. In the section of the cryo preserved tissue the cones and rods are no longer visible while in the section of the PAXgene fixed tissue they are fully intact.

The retina is a very fragile, multilayered tissue. Cryo preservation led to distortions of the tissue morphology while PAXgene fixation resulted in a very good preservation of the tissue integrity (**Figure 27**). Thus, PAXgene fixation was better suited for retina preparation than cryo preservation. For analysing light and dark adaptation of the retina, the eyecups were either illuminated for 30 min or kept in the dark, before they were PAXgene fixed as described above. Then, tissue sections were cut, mounted onto MALDI Imaging glass slides, deparaffinised, and spray coated with sinapinic acid matrix. MALDI Imaging was performed in linear mode and with a lateral resolution of 50 µm. MALDI Imaging of light- and dark-adapted retina tissue sections resulted in differentially expressed *m/z* species (**Figure 28**).

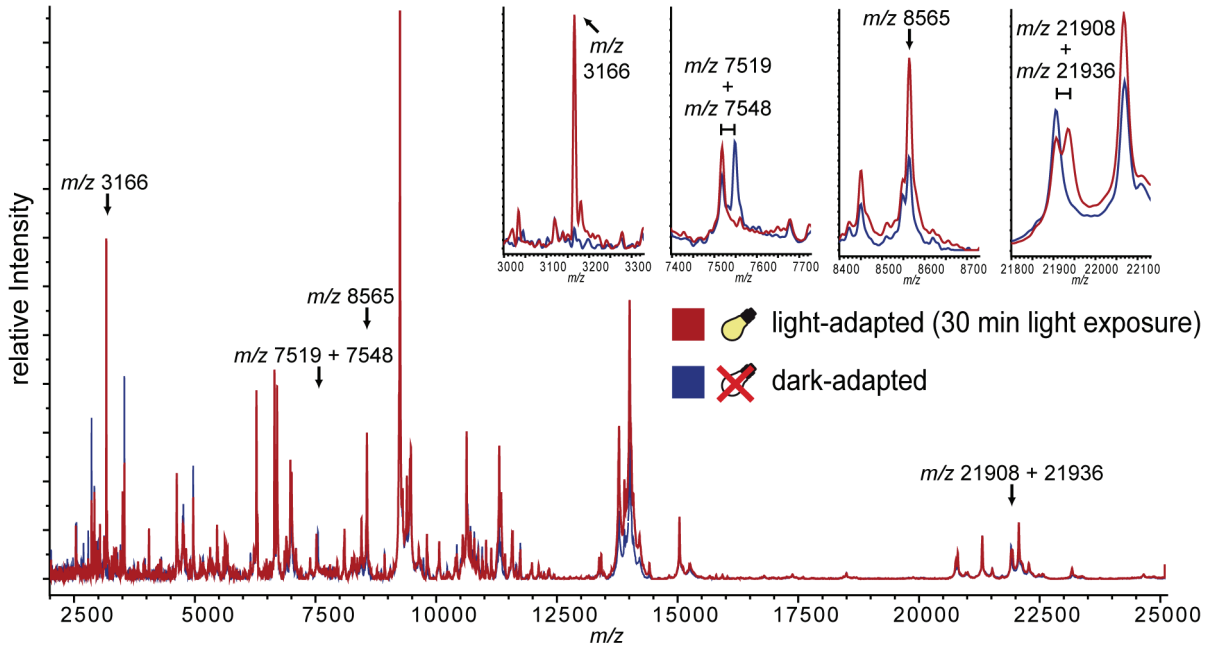


Figure 28 – MALDI Imaging of light- and dark-adapted retina. The average spectra of light- and dark-adapted retina analysed by MALDI Imaging are displayed. The four *m/z* species which show the most prominent differences between light-adapted and dark-adapted state are enlarged. *M/z* 3166 and *m/z* 8565 show an increased expression in light-adapted state. The signal pairs of *m/z* 7519 / 7548 and *m/z* 21908 / 21936 show a different pattern. In either dark-adapted or light-adapted state, a double peak is visible which does not exist in the other respective state. The appearance of a double peak with a size difference of 29 / 28 Da upon adaptation can be best explained by the occurrence of posttranslational protein modifications. A double N-methylation would be the most likely modification.

Several *m/z* species were differentially expressed. The four most prominent ones are depicted in **Figure 28** and are listed in **Table 18**.

***M/z* species**	**Expression pattern**	**Putative PTM**	**Putative protein name**
3166	Upregulated in light-adapted state	-	-
7519 / 7548	Shift of +29 Da in dark-adapted state	Mono- → Trimethylation or Dimethylation	-
8565	Upregulated in light-adapted state	-	Ubiquitin
21908 / 21936	Shift of +28 Da in light-adapted state	Mono- → Trimethylation or Dimethylation	-

Table 18 – Differential *m/z* species between light-adapted and dark-adapted state of the mammalian retina identified by MALDI Imaging. (PTM = posttranslational modification)

Two different ways of signal changes could be observed. In case of *m/z* 3166 and *m/z* 8565 changes in expression intensities were observed (**Figure 28**). Both *m/z* species were

upregulated in the light-adapted state. The signal pairs *m/z* 7519 / 7548 and *m/z* 21908 / 21936 showed a different pattern. A shift of the signal intensities could be observed. In the case of the signal pair *m/z* 7519 / 7548 an additional signal with a shift of 29 Da occurred in the dark-adapted state while the intensity of the *m/z* 7548 signal was lowered in the dark-adapted state. The signal pair *m/z* 21908 / 21936 was present in the light-adapted state. In the dark-adapted state only the *m/z* 21908 signal was present. The 29 / 28 Da shift of *m/z* signals was in both cases probably caused by adaptation induced posttranslational modifications (**Figure 28**). The most reasonable explanation is a change in the methylation state of the measured protein. For example, a newly introduced double N-methylation or a change from single N-methylation to triple N-methylation (**Table 18**). Further experiments for protein identification and validation will be done in collaboration with Dr. Detlev Suckau, Bruker Daltonics, Bremen and with Dr. Alice Ly, Dr. Stefanie Hauck, and Prof. Dr. Marius Ueffing, Research Unit Protein Science, Helmholtz Zentrum München.

6 Discussion

6.1 Proteomic markers for regional lymph node metastasis in colon cancer primary tumours

Regional lymph node metastasis drastically reduces the survival rate in colon cancer[14,16,17] and strongly affects the therapeutic approach.[19] Therefore, a better understanding of the underlying molecular mechanisms causing lymphoid metastasis and reliable proteomic markers for clinical diagnostics are needed. We were able to correlate 38 proteins (identified or *m/z* species) which were differentially expressed in the primary tumours with respect to the nodal status of the patients. Of the 10 *m/z* species discovered by MALDI Imaging two were identified by name. Both have been described in the context of cancer but none of them has yet been shown to correlate with regional lymph node metastasis in colon cancer (**Table 6**). By label-free quantitative proteomics a total of 28 discriminating proteins were identified. For reducing the rate of false positives, only proteins were considered differentially expressed whose identification and quantification relied on 2 or more peptides. In tissue-based studies, 11 of them have never been associated with cancer, 22 have never been associated with colon cancer, and only two of them have been previously described to be involved in metastasis (**Table 7**).[115,125] Thus, by combining two tissue-based proteomic methods we found a large panel of novel proteins that could lead to a better understanding of the molecular processes underlying regional lymph node metastasis. Three of them, FXYD3, S100A11 and GSTM3, were confirmed by immunohistochemical staining on an independent tissue cohort (**Figure 12** and **Table 8**). Their association with regional lymph node metastasis in colon cancer has not been described so far. FXYD3 and GSTM3 have not been associated with regional lymph node metastasis, irrespective of the tumour type, while S100A11 has been associated with regional lymph node metastasis in lung cancer.[147]

FXYD3 is a modulator of Na^+/K^+-ATPase activity and has been reported to be expressed in several healthy human tissues.[148] Several studies have reported its dysregulation in tumour cell lines and tissues.[94,95,97,149-153] FXYD3 was found upregulated in a number of adenocarcinomas when compared to the respective normal tissue[95,97,151-153] and downregulated in lung cancer.[94] It was observed that FXYD3 expression correlated with unfavourable prognosis in rectal cancer patients with neoadjuvant treatment.[96] In colon cancer, correlation of reduced FXYD3 expression with cancer progression (Dukes staging) was shown[97] but no association with regional lymph node metastasis. In our study, we could clearly show an association of lowered FXYD3 expression with the presence of lymph node metastasis. In univariate analysis, the nodal status was the strongest correlating factor ($p = 0.0110$) compared to tumour depth and tumour grading which both displayed a less prominent but still significant correlation (**Table 8**). In multivariate analysis, FXYD3 was correlated by trend with the nodal status ($p = 0.0683$). Therefore, it seems that loss of FXYD3 expression is necessary for tumour progression and for the development of regional lymph node metastasis.

S100A11, also known as Calgizzarin or S100C, is a calcium binding intracellular regulatory protein.[154] In many studies it has been associated with cancer, mostly with progression or patient survival. In gastric cancer,[89] pancreatic cancer[155] and colorectal cancer[91-93,156] it was shown to be overexpressed in comparison with the respective healthy tissues. While upregulation of S100A11 expression was mostly associated with tumour progression, it was also found that tumour progression led to a reduced S100A11 expression.[155,157] In bladder cancer loss of S100A11 expression resulted in unfavourable prognosis,[157] while in breast cancer increased expression of S110A11 resulted in unfavourable prognosis.[88] In lung cancer it was shown that S100A11 overexpression was correlated with regional lymph node metastasis and tumour progression.[147] Thus, S100A11 seems to be associated with tumour progression in adenocarcinomas but its actual role remains ambiguous and varying between different tumour entities. In our study, we found a correlation of reduced S100A11 expression

and the regional lymph node metastasis in colon cancer with two independent methods on two independent patient cohorts. The observed association might indicate that the mechanisms leading to regional lymph node metastasis are different in different tumour types, such as lung and colon cancer.

Glutathione S-transferase μ (Mu) 3 (GSTM3) is a cytosolic enzyme involved in prostaglandin and leukotrienes synthesis and in metabolisation of toxic compounds, such as chemotherapeutic drugs, insecticides, herbicides, carcinogens and by-products of oxidative stress.[158] GSTM3 has been associated with cancer risk on a genomic level where an association between a certain polymorphism and increased cancer risk was found in oesophageal cancer,[159] hepatocellular carcinoma,[160] lung cancer[161] and colorectal cancer.[101,162] Promoter hypermethylation of GSTM3 was found in Barrett's adenocarcinoma.[163] On a proteomic level GSTM3 was shown to be downregulated in ovarian cancer.[100] Correlations with clinical endpoints have rarely been undertaken. In urinary bladder cancer low GSTM3 expression was associated with better survival.[99] In our study, we could show a strong association of GSTM3 expression with lymph node metastasis on the proteomic level both in univariate ($p = 0.0063$) and multivariate ($p = 0.0173$) analysis. This is accordance with the data of the study on bladder cancer in which upregulation of GSTM3 reduced the survival rate of patients.[99] Therefore, high cellular levels of GSTM3 might support the development of regional lymph node metastasis.

This is the first time that MALDI Imaging and label-free quantitative proteomics were used as complementary methods. By this, we could identify a panel of 38 proteins (identified or *m/z* species) associated with regional lymph node metastasis. Most of them were identified for the first time in the context of metastasis. Three of them were validated by an independent method on an independent tissue cohort. The identification of new proteins associated with regional lymph node metastasis might lead to a better understanding of the molecular processes involved in tumour metastasis. Additionally, proteins correlating with regional

lymph node metastasis could become valuable markers for risk assessment in colon cancer patients. So far, the decision for adjuvant chemotherapy in high-risk UICC II patients is based on macroscopic and morphologic factors. Molecular markers indicating the potential for regional lymph node metastasis could complement the currently used factors leading to a refined decision making. Analysing the potential of the found markers will be the next step. For this, further *in vitro* and *in vivo* studies will be needed.

6.2 Discrimination of tumour entities by proteomic classification

The correct identification of the respective tumour entity is crucial for a personalised, individually tailored treatment regimen. Over the last decades, new molecular methods, such as gene and protein expression analysis, have been established for discriminating different tissue types or tumour entities. Gene expression analyses could provide accurate classification results.[66-72] Proteomic analyses, which first used body fluids, mostly serum, and later turned to tissue samples, were able to discriminate tumour samples from healthy progenitor samples with high accuracy.[73,75,76,78,80] The clinically more relevant and technically more challenging problem of discriminating various tumour entities or molecularly distinct tumour subgroups could also be addressed successfully.[40] Gene expression profiling and proteomic methods which use tissue homogenates face two problems concerning sample material. Often, more sample material would be needed for analysis than can be procured in pretherapeutic diagnostics and the necessary assumption of tissue homogeneity can negatively influence the results. This might be a reason why these methods have not been implemented into diagnostics. In the last years, first mass spectrometric tissue profiling, then MALDI Imaging have emerged in order to cope with these two drawbacks. They require very little tissue material – basically, a single tissue section from an endoscopic biopsy is enough for analysis[43] – and retain the morphology during analysis. By this, the resulting proteomic pattern can be

compared with the histological staining in order to check the cellular composition of the given tissue.[33] For these reasons we chose MALDI Imaging for tumour classification.

The initial classification of six tumour entities located in different organ sites yielded a high accuracy in both training and test set. The training set could be classified nearly perfectly. The classification of the test set yielded a reduced but still high accuracy (82.74% for the Support Vector Machine and 81.18% for the Random Forest algorithm; **Table 11**). Both classifiers yielded comparable results, thus indicating a robustness of the results. Compared with the discrimination of tumour entities by Principal Component Analysis, it is obvious that the discrimination by classifiers is more accurate and thus better suited for reliable tumour classification. The misclassifications are probably not due to tumour depth or grading since the Spearman's rank correlation test of the rate of misclassification with the tumour depth (T) and grading (G) could not determine a significant ($p < 0.05$) correlation (**Tables 19** and **20**).

Tumour depth (T)	**Barrett's cancer**	**Breast cancer**	**Colon cancer**	**Hepat. carcinoma**	**Gastric cancer**	**Thyroid cancer**
p-value (SVM)	0.19	0.82	0.46	0.36	0.05	0.67
p-value (RF)	0.48	0.20	0.31	0.46	0.38	0.25
p-value (combined)	0.13	0.55	0.25	0.36	0.25	0.27

Table 19 – Correlation of misclassification rate with tumour depth (T) in the classification of six tumour entities located in different organ sites.

Grading (G)	**Barrett's cancer**	**Breast cancer**	**Colon cancer**	**Hepat. carcinoma**	**Gastric cancer**	**Thyroid cancer**
p-value (SVM)	0.17	0.33	0.27	0.13	0.13	NA
p-value (RF)	0.82	0.79	0.66	0.87	0.80	NA
p-value (combined)	0.71	0.47	0.39	0.69	0.42	NA

Table 20 – Correlation of misclassification rate with tumour grading (G) in the classification of six tumour entities located in different organ sites.

A reasonable explanation for the misclassifications is the existence of molecular subtypes within tumour entities. In breast cancer five distinct subtypes are recognised which express different molecular features and display a different clinical outcome.[164] Such molecular heterogeneity is likely to exist in all tumour entities.

So far, most proteomic studies were concerned with the discrimination of normal, "healthy" tissue from tumour tissue. However, there are studies which were concerned with the discrimination of tumour entities. Villanueva et al. were able to distinguish three tumour entities, prostate, bladder and breast cancer, using serum samples.[77] Bloom et al. were able to generate a classifier with an overall accuracy of 82% for a patient cohort consisting of six tumour entities: breast, colon, gastric, kidney, lung, and ovary cancer.[74] While these two studies proved that tumour entities can be discriminated by classification our study furthers these results. Apart from also containing six tumour entities it consists of a training set for generation of the classifier and a test set for its validation. This feature is highly relevant for assessing the power of a classification. If a classifier is employed on a new, independent test cohort the risk for data overfitting of the classifier to the initial data set is reduced and thus the classification results become more reliable.[165]

After proving that MALDI Imaging derived proteomic patterns can be used for accurate discrimination of tumour entities a second cohort was analysed which is closer to an important challenge in diagnostics where metastases have to be correctly classified even if the primary tumour is occult. For this, we selected a cohort comprising tumour entities which are either located within the same organ site or which have the same origin. Since colon cancer frequently metastasises to the liver, primary colon cancer, its liver metastasis, and primary hepatocellular carcinoma were selected. Again the classification of the training set was close to perfect. The more interesting classification of the test set yielded also a high accuracy (84.11% for the Support Vector Machine and 82.32% for the Random Forest algorithm; **Table 13**). This result indicates that even such closely related entities as the primary tumour of colon cancer and its liver metastasis could be classified efficiently. As before, the rate of misclassifications displayed no significant ($p < 0.05$) correlation with the tumour depth and grading, which was tested by Spearman's rank correlation test (**Tables 21** and **22**).

Tumour depth (T)	Colon cancer primary tumour	Colon cancer liver metastases	Hepatocellular carcinoma
p-value (SVM)	0.42	NA	0.06
p-value (RF)	0.51	NA	0.39
p-value (combined)	0.35	NA	0.32

Table 21 – Correlation of misclassification rate with tumour depth (T) in the classification of three tumour entities either located within the same organ site (liver) or being of the same origin (colon).

Tumour Grading (G)	Colon cancer primary tumour	Colon cancer liver metastases	Hepatocellular carcinoma
p-value (SVM)	0.12	0.23	0.21
p-value (RF)	0.10	0.25	0.34
p-value (combined)	0.19	0.25	0.44

Table 22 – Correlation of misclassification rate with tumour grading (G) in the classification of three tumour entities either located within the same organ site (liver) or being of the same origin (colon).

The more likely explanation for the misclassifications is again the imminent molecular heterogeneity within tumours of the same entity.

MALDI Imaging opens new fields in tissue sample classification. This proof-of-principle study shows for the first time that proteomic classification of solid tumour entities can be highly accurate while needing only a minimal amount of tissue. Also other applications can be envisioned, for example the classification of lymphomas[79], tumour subtypes[40], or morphologically similar, non-neoplastic diseases such as chronic inflammatory diseases. Thus, MALDI Imaging might become a valuable tool in clinical diagnostics in the future.

6.3 Proteomic markers for relapse in colon cancer primary tumours

A significant proportion of colon cancer patients who have no detectable metastasis at the time of operation develop a relapse after curative resection of the primary tumour.[21-24] These patients would probably benefit from adjuvant chemotherapy.[22,26] However, clear predictive markers for relapse are lacking.[8,21,27] High-risk patients are subjected to adjuvant therapy but the definition for high-risk is still rather vague. According to ESMO (European Society for Medical Oncology) guidelines the general consensus is that patients with stage II are at high

risk if they present at least one of the following characteristics: lymph nodes sampling <12; poorly differentiated tumour; vascular or lymphatic or perineural invasion; tumour presentation with obstruction or tumour perforation and pT4 stage [II].[166] It is obvious that the decision for adjuvant therapy is based rather on macroscopic and morphologic features of the tumour than on markers describing the molecular properties of the cancer. Several molecular markers have been found in clinical studies but none of them is recommended for clinical diagnostics so far.[21,26,27,167] The identified prognostic molecular markers included the expression levels of various microRNAs[23,24] and proteins, such as Bcl-2,[26] Ubiquitin D,[168] S100B,[169] CIAPIN1,[170] LMNA,[171] SPARC and FOXP3.[172] Unfortunately, several of the aforementioned prognostic markers were identified in studies which included UICC stage III patients.[168-170] This makes it difficult to clearly distinguish metastasis induced secondary tumours from tumour relapse. Taken together, no study could produce a predictive marker for relapse in colon cancer that was reliable enough for clinical application.

In order to find such predictive markers this study was undertaken. It had several advantages over the previously attempted studies. The large number of samples of patients who were all without regional lymph node or distant metastasis and had a long follow-up of nearly 10 years together with the tissue and morphology based screening method allowed a reliable identification of novel proteomic markers. Applying MALDI Profiling on the tumour tissue cohort resulted in a panel of 27 relapse associated *m/z* species, 8 of which could be identified: S100A9, DEFA1, MRPL14, DEFA2, S100A8, S100P, CRIP1, and HBA1 (**Table 15**). Apart from MRPL14 and HBA1 all of them have been associated with cancer, and none of them has been associated with tumour relapse in colon cancer.

The expression of CRIP1, a cytosolic protein involved in cell proliferation,[45] has been associated with cancer. Most studies were based on expression profiling of cell lines, but several also used tumour tissues for analyses.[45,145,173,174] CRIP1 was reported to be overexpressed in pancreatic cancer,[174] cervical cancer[173] and breast cancer.[145] Additionally,

CRIP1 was part of a set of proteins that could distinguish the HER2 receptor status of breast cancer samples with high accuracy,[45] and it could be used as a prognostic marker in gastric cancer.[48] CRIP1 has never been reported in the context of colon cancer, tumour metastasis, or tumour relapse. In this study, a correlation ($p = 0.0855$) of increased CRIP1 expression with tumour relapse could be observed.

There is increasing evidence suggesting that S100P, an intra- and extracellular signalling molecule, has a significant role in cancer.[175] It was reported to be upregulated in pancreatic cancer.[176] In breast cancer, S100P was reported to be upregulated in cancer cells[177] and associated with HER2 expression.[178] Additionally, overexpression could be associated with unfavourable prognosis and metastasis.[140] In prostate cancer S100P expression was associated with tumour progression and metastasis.[141] In lung cancer S100P expression was shown to be a predictor of distant metastasis and prognosis.[142] Specific expression of S100P in colon cancer cells and positive effects on proliferation and cell migration in colon cancer cell lines were reported.[179] So, association of S100P with prognosis and metastasis has been reported for several cancers. However, little is known about its role in carcinogenesis and tumour progression of colon cancer. This study could for the first time associate reduced S100P levels with relapse in colon cancer ($p = 0.0688$).

S100A8 and S100A9 predominately form a heterodimer under physiological conditions[180] and have intra- and extracellular functions.[181] Their role seems to be ambivalent. Pro- and anti-inflammatory effects are reported.[182] It was shown that S100A8 and S100A9 were expressed in breast cancer, gastric cancer and colorectal cancer tissues,[137-139] and that increased expression levels correlated with reduced tumour differentiation.[131-133] Data from various tumour cancer cell lines and a mouse line suggest a role in tumour cell invasion and migration.[183-185] S100A9 overexpression was associated with unfavourable prognosis in lung cancer.[130] These findings suggest an involvement of S100A8 and S100A9 in cancer

progression. Herein, a correlation of reduced S100A8 ($p = 0.0453$) and S100A9 ($p = 0.0439$) levels with tumour relapse was reported for the first time.

DEFA1 and 2, also known as neutrophil defensin 1 and 2, are antimicrobial and immunomodulatory peptides which are mainly synthesised and secreted by neutrophils, but can also be expressed by tumour cells.[186] Both peptides are overexpressed in gastric cancer,[43] squamous cell carcinoma,[187] bladder cancer,[188] renal cell carcinoma,[189] and colon cancer.[190] They have also been associated with invasiveness in bladder cancer,[188] and with tumour progression.[134,135] In this study, a correlation of reduced DEFA1 ($p = 0.0439$) and DEFA2 ($p = 0.0453$) levels with tumour relapse was reported for the first time.

All six proteins (CRIP1, S100P, S100A8, S100A9, DEFA1 and DEFA2) are potential markers for tumour relapse. Their association with cancer has been reported in several studies. For CRIP1 and S100P an association with prognosis was also reported. In the case of CRIP1, higher CRIP1 expression lead to unfavourable prognosis.[48] The here presented association with tumour relapse is in accordance with this previous result. In the case of S100P, increased S100P levels were also reported to result in unfavourable prognosis.[140,142] This is in contrast to the observation made in this study. Herein, reduced S100P levels correlated with relapse. A possible explanation might me that S100P causes different effects whether it is localised intra- or extracellularly. For S100A8 and A9 same runs true. According to literature they are associated with unfavourable prognosis in lung cancer.[130] This is in contrast with the observed correlation of reduced S100A8 and S100A9 levels with relapse. It was reported that S100A8 and A100A9 are pro- and anti-inflammatory depending on their concentration and localisation.[182] This might explain the difference between the findings of this study and previous ones. There are no reports on an association of DEFA1 and DEFA2 with prognosis. It is known that inflammation and immune response is an important factor in tumour progression. Jensen and co-workers reported that the presence of intratumoural neutrophils was associated with unfavourable patient prognosis in renal cell carcinoma.[191] Since DEFA1

and DEFA2 are predominately expressed by neutrophils one would expect that both proteins are higher expressed in relapse developing patient. In this study the contrary was observed. In this study, the overall levels of DEFA1 and DEFA2 were measured which is not necessarily the same as the number of neutrophils within a tissue, especially since it has been shown that also tumour cells can express DEFA1 and DEFA2.[186] It would also be reasonable if low levels of these inflammatory peptides correspond with an increased risk for tumour relapse.

In conclusion, 6 of the 8 identified proteins which correlate with relapse have already been associated with cancer. Looking at their function and the previously published findings the observed results seem plausible. The next step will be a validation of the results, ideally on an independent cohort. Since such a tissue cohort is hard to obtain, validation will be done on the present cohort using immunohistochemical tissue staining.

6.4 Methodological advances in MALDI Imaging – Opening new fields of research

MALDI Imaging is a relatively novel method and further development of this method is needed. Apart from technical improvements that allow a higher sensitivity and lateral resolution, the tissue preparation protocols will have to be refined. One aspect is most certainly the development of preparation protocols for sample tissues that are not native (cryo preserved).

6.4.1 MALDI Imaging on formalin-fixed paraffin-embedded tissues – Gaining access to the tissue archives of pathology

Human tissue samples, especially those, which are associated with clinical data, are mostly stored as formalin-fixed paraffin-embedded tissue specimens in the archives of pathology departments.[35] Therefore, most histological methods used in diagnostics and research are adapted to this kind of tissue material. Formalin fixation results in covalent cross-linking of proteins and other molecules and therefore, prevents direct mass spectrometric analysis. In order to cope with the covalent cross-linking of proteins, trypsin based protocols for mass

spectrometry on FFPE tissues have been established.[192,193] For MALDI Imaging, several studies have shown that FFPE tissues can be used if *in situ* tryptic digest is done before analysis.[50-54,81-85] In these studies the lateral resolutions reached only 200 µm. For a detailed correlation of MALDI Imaging data sets with the tissue morphology the lateral resolution will have to be improved. With the reported methods of trypsin and matrix application, an improvement of lateral resolution in measurement is technically not possible since they rely on acoustic or mechanic spotting of the trypsin and the matrix solutions. Thus, new tissue preparation protocols will have to be established. Herein, a preparation protocol which used spray application of both trypsin and matrix solutions was reported. Spray application of trypsin resulted in a drastic improvement of the lateral resolution. Now, only the size of the droplets (~20 µm) is limiting.

At first, MALDI Imaging on FFPE tissues was attempted without tryptic digest for preparation. For this an endoscopic biopsy of the distal oesophagus was used. This resulted in a differential expression of several *m/z* species (**Figure 19**) and the hierarchic clustering displayed a high correlation with the tissue morphology (**Figure 20**). Unfortunately, only few signals were specific for Barrett's metaplasia and the mass spectrum displayed broad peaks, which are a strong indication for protein degradation. This led to the conclusion that tryptic digest is indeed necessary for successful MALDI Imaging. Therefore, a preparation protocol was established which included *in situ* tryptic digest by spray application, and a cohort of 14 colon cancer samples was analysed using this protocol (**Table 16**). This preparation method resulted in a multitude of *m/z* species within each spectrum. These signals displayed a differential expression between the different tissue types within each sample (**Figure 21**). Hierarchic clustering was not possible because of the data size. Therefore, clustering was done in collaboration with Dr. Theodore Alexandrov and Prof. Dr. Peter Maass who had developed a clustering algorithm for large MALDI Imaging data sets.[146] This clustering algorithm was applied to MALDI Imaging data sets of 5 colon cancer samples. It resulted in a

good discrimination of different morphological features of the respective tissue samples (**Figure 22**). This indicated that the preparation protocol resulted in peptide spectra that feature tissue type specific proteomic patterns. Comparison of morphological features (mucosa, adenoma or carcinoma) within the sample cohort could not provide significantly differentially expressed *m/z* species. So, it became obvious that the preparation method was successful because it resulted in a specific, differential peptide pattern within each sample measurement. It also became obvious that the preparation method was not robust since it did not provide *m/z* species which were differentially expressed in the present tissue types (mucosa, adenoma, and carcinoma) within the cohort. The reasons for this remain open. It could be speculated that the tryptic digest is inhomogeneous due to slight inter-sample variations of the amount of applied trypsin or due to different enzymatic activity of the applied trypsin caused by the molecular composition of the underlying sample tissue.

The acquired MALDI Imaging data sets of the formalin-fixed colon cancer samples were used for another study which also aimed at improving the lateral resolution in MALDI Imaging.[61] This time it was not attempted to improve the measurement protocol but to increase the resolution of the data set after acquisition by software algorithms. Lateral resolution is still limited in MALDI Imaging. The highest lateral resolution with commercially available mass spectrometers is 20 µm.[38] Normally, MALDI Imaging measurements have a resolution between 50 and 250 µm. In collaboration with Dr. Theodore Alexandrov and Prof. Dr. Peter Maass, a software algorithm that is able to increase the resolution of clustering data (also termed segmentation map) was generated and tested on the formalin-fixed colon cancer data sets (**Figure 23** and **24**). By increasing the resolution of the segmentation maps, a higher correlation of the individual clusters of the segmentation maps with the tissue morphology could be observed. This showed that the generated algorithm is capable to artificially increase the resolution of MALDI Imaging data sets after measurement. Since it is uncertain when and if MALDI Imaging sample preparation and instrumentation will reach the

level of a single cell (1-10 μm), this software algorithm might compensate for the low lateral resolution in the meantime. It could become a valuable tool by increasing the lateral resolution in an indirect way.[61]

6.4.2 MALDI Imaging on alcohol-fixed paraffin-embedded tissues – Bypassing formalin-fixation

As mentioned previously, formalin fixation causes protein cross-linking and therefore, MALDI Imaging on formalin-fixed tissues requires *in situ* tryptic digest for tissue preparation. By this, only peptides, not full-length proteins can be analysed. Alcohol-based tissue fixation is another preservation method well known in pathology. It should also be compatible with MALDI Imaging since the proteins remain intact and unmodified.

Pancreatic tissue specimens were used for testing the suitability of a novel, alcohol-based fixative named PAXgene. Pancreatic tissue consists of two distinct tissue types, the endocrine and the exocrine pancreas. Within the endocrine pancreas, glucagon and insulin are produced and both of them can be measured by MALDI Imaging. The localisation of these two proteins in a small, defined and histologically discernable region is ideal for assessing the quality of the fixative in respect to the preservation of protein localisation. The PAXgene fixation was compared with formalin fixation and cryo preservation. MALDI Imaging was performed on these differently fixed tissues (**Figure 25**). While formalin fixation yielded no signals (since no antigen retrieval or tryptic digest was undertaken), PAXgene fixation and cryo preservation yielded similar spectra. The intensity and the number of the *m/z* species detected in PAXgene-fixed tissues were slightly lower than in cryo preserved tissues (**Table 17**). No delocalisation of insulin and glucagon could be observed (**Figure 26**).[56] These findings are in accordance with other studies. Chaurand and coworkes used 70% ethanol for fixation.[55] Mangé and coworkers used RCL2/CS100 for fixation.[57] Chaurand observed reduced signal intensity. Mangé reported a decrease of signal intensity by a factor of 2. Both saw no negative

effect of alcohol preservation on protein localisation within the measured tissue sections. However, their lateral resolution was only 150 to 200 µm. In this study, the pancreatic sections have been measured with a lateral resolution of 70 µm. The higher lateral resolution together with the histologically distinct regions of exocrine and endocrine pancreas within the pancreas sections allowed a better assessment of the degree of protein (de)localisation within the tissue.[56] Taken together, it could be proven that PAXgene is an appropriate fixative for MALDI Imaging. It could also be shown that PAXgene fixation is compatible with other molecular biological methods.[56] Therefore, PAXgene might become an alternative to formalin for tissue fixation. It combines the advantages of both, formalin fixation and cryo preservation. Due to paraffin embedding the samples can be oriented during preparation resulting in optimal sectioning planes. Additionally, paraffin embedding allows easy long term storage. In the same way as cryo preservation, alcohol preserves fully intact proteins allowing top-down as well as bottom-up proteomics.

The PAXgene fixative has been used for the analysis of the mammalian retina. The retina is very fragile. Therefore, cryo preservation of the retina is not possible. The morphology is distorted and most of the cells, especially the cones and rods are lysed. With PAXgene fixation the morphology is fully retained (**Figure 27**). MALDI Imaging on sections which were either adapted to light or dark state yielded differential *m/z* species (**Figure 28**). Looking closer into the differential protein patterns showed that some of the differences within the spectra are probably due to posttranslational modifications, most probably double N-methylations (**Table 18**). This was the first time that molecular modifications induced by light/dark adaptation could be observed *in situ*. This highlights the big advantage of MALDI Imaging and top-down proteomics in general. Since they deal with full-length proteins they also detect the full protein with its posttranslational modifications. Bottom-up proteomics has limitations in the measurement of posttranslational modifications due to its set-up. Peptides are measured and identified and afterwards the corresponding proteins are deduced. This can

cause problems in analysis. For instance, a posttranslational modification can only be detected if the peptide is measured, or a posttranslational modification can prevent the peptide identification if the search algorithm employed for peptide identification cannot cope with it. Regarding the high number of different protein modifications this should happen frequently. Another limitation of bottom-up proteomics in respect to posttranslational modification analysis is the fact that the modifications of a protein, which exists in different states (such as isoforms) with distinct modification patterns, can often not be correctly assigned to the right state. Therefore, a top-down method like MALDI Imaging seems to be best suited for the analysis of dynamic changes of posttranslational modifications *in situ*.

6.5 Conclusion

Within this thesis several relevant clinical problems could be addressed successfully. Novel proteomic markers for regional lymph node metastasis could be identified by MALDI Imaging and label-free quantitative proteomics. Three of them, FXYD3, S100A11 and GSTM3, were additionally validated on an independent cohort. These markers might help to make a better risk assessment for regional lymph metastasis and identify high-risk patients which would benefit from adjuvant chemotherapy.

In a proof-of-principle study it could be shown that MALDI Imaging derived proteomic profiles can be successfully used for highly accurate tumour classification. This might help to establish new diagnostic methods for solving the clinical problem of cancers of unknown primary.

By MALDI Profiling on a patient cohort including relapse patients, 27 m/z species correlating with relapse could be detected. 8 of them were identified by proteomic identification. These markers might help to improve patient treatment and aftercare by identifying patients with a increased risk for a relapse.

Additionally, methodological advances in MALDI Imaging were made. A protocol for MALDI Imaging on formalin-fixed paraffin-embedded tissues and a protocol for MALDI Imaging on alcohol-fixed paraffin-embedded tissues were established. The availability of MALDI Imaging protocols for these kinds of tissue material will give access to a wider range of clinical tissue samples and open new fields of research.

7 References

1. World Health Organization. The global burden of disease : 2004 update. Geneva: World Health Organization; 2008.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
3. Quirke P, Risio M, Lambert R, von Karsa L, Vieth M. Quality assurance in pathology in colorectal cancer screening and diagnosis-European recommendations. Virchows Arch 2011;458:1-19.
4. Bosman FT, Organization WH, Cancer IAfRo. WHO classification of tumours of the digestive system. 4th ed. Lyon: International Agency for Research on Cancer; 2010.
5. Kumar V, Robbins SL. Robbins basic pathology. 8th ed. Philadelphia, PA: Saunders/Elsevier; 2007.
6. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57-70.
7. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759-67.
8. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. N Engl J Med 2009;361:2449-60.
9. Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. Gastroenterology 2010;138:2088-100.
10. Sobin LH, Gospodarowicz MK, Wittekind C, International Union against Cancer., ebrary Inc. TNM classification of malignant tumours. In. 7th ed. Chichester, West Sussex, UK ; Hoboken, NJ: Wiley-Blackwell; 2009.
11. Edge SB, American Joint Committee on Cancer. AJCC cancer staging manual. 7th ed. New York ; London: Springer; 2010.
12. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology 2008;134:1570-95.
13. Lieberman DA. Clinical practice. Screening for colorectal cancer. N Engl J Med 2009;361:1179-87.
14. Weitz J, Koch M, Debus J, Hohler T, Galle PR, Buchler MW. Colorectal cancer. Lancet 2005;365:153-65.
15. Swanson RS, Compton CC, Stewart AK, Bland KI. The prognosis of T3N0 colon cancer is dependent on the number of lymph nodes examined. Ann Surg Oncol 2003;10:65-71.
16. Gunderson LL, Jessup JM, Sargent DJ, Greene FL, Stewart AK. Revised TN categorization for colon cancer based on national survival outcomes data. J Clin Oncol 2010;28:264-71.
17. Compton CC, Greene FL. The staging of colorectal cancer: 2004 and beyond. CA Cancer J Clin 2004;54:295-308.
18. Gill S, Loprinzi CL, Sargent DJ, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? J Clin Oncol 2004;22:1797-806.
19. Meyerhardt JA, Mayer RJ. Systemic therapy for colorectal cancer. N Engl J Med 2005;352:476-87.
20. Van Cutsem E, Oliveira J. Primary colon cancer: ESMO clinical recommendations for diagnosis, adjuvant treatment and follow-up. Ann Oncol 2009;20 Suppl 4:49-50.

21. Hutchins G, Southward K, Handley K, et al. Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. J Clin Oncol 2011;29:1261-70.
22. Gertler R, Rosenberg R, Schuster T, Friess H. Defining a high-risk subgroup with colon cancer stages I and II for possible adjuvant therapy. Eur J Cancer 2009;45:2992-9.
23. Aslam MI, Taylor K, Pringle JH, Jameson JS. MicroRNAs are novel biomarkers of colorectal cancer. Br J Surg 2009;96:702-10.
24. Schepeler T, Reinert JT, Ostenfeld MS, et al. Diagnostic and prognostic microRNAs in stage II colon cancer. Cancer Res 2008;68:6416-24.
25. Pfister DG, Benson AB, 3rd, Somerfield MR. Clinical practice. Surveillance strategies after curative treatment of colorectal cancer. N Engl J Med 2004;350:2375-82.
26. Poincloux L, Durando X, Seitz JF, et al. Loss of Bcl-2 expression in colon cancer: a prognostic factor for recurrence in stage II colon cancer. Surg Oncol 2009;18:357-65.
27. Pritchard CC, Grady WM. Colorectal cancer molecular biology moves into clinical practice. Gut 2011;60:116-29.
28. Van Cutsem E, Nordlinger B, Cervantes A. Advanced colorectal cancer: ESMO Clinical Practice Guidelines for treatment. Ann Oncol 2010;21 Suppl 5:v93-7.
29. Van Cutsem E, Oliveira J. Advanced colorectal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol 2009;20 Suppl 4:61-3.
30. Mulisch M, Romeis B. Romeis Mikroskopische Technik. 18. Aufl. ed. Heidelberg: Spektrum Akademischer Verlag; 2010.
31. Pandey A, Mann M. Proteomics to study genes and genomes. Nature 2000;405:837-46.
32. Walch A, Rauser S, Deininger SO, Höfler H. MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. Histochem Cell Biol 2008;130:421-34.
33. Chaurand P, Sanders ME, Jensen RA, Caprioli RM. Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections. Am J Pathol 2004;165:1057-68.
34. Seeley EH, Caprioli RM. MALDI imaging mass spectrometry of human tissue: method challenges and clinical perspectives. Trends Biotechnol 2011;29:136-43.
35. Rauser S, Deininger SO, Suckau D, Hofler H, Walch A. Approaching MALDI molecular imaging for clinical proteomic research: current state and fields of application. Expert Rev Proteomics 2010;7:927-41.
36. Meding S, Walch A. MALDI Imaging Mass Spectrometry for Direct Tissue Analysis. Methods Mol Biol 2011;in press.
37. Balluff B, Schöne C, Höfler H, Walch A. MALDI imaging mass spectrometry for direct tissue analysis: technological advancements and recent applications. Histochem Cell Biol 2011;136:227-44.
38. Lagarrigue M, Becker M, Lavigne R, et al. Revisiting rat spermatogenesis with MALDI imaging at 20-microm resolution. Mol Cell Proteomics 2011;10:M110 005991.
39. Seeley EH, Caprioli RM. Molecular imaging of proteins in tissues by mass spectrometry. Proc Natl Acad Sci U S A 2008;105:18126-31.
40. Balluff B, Elsner M, Kowarsch A, et al. Classification of HER2/neu status in gastric cancer using a breast-cancer derived proteome classifier. J Proteome Res 2010;9:6317-22.
41. Bauer JA, Chakravarthy AB, Rosenbluth JM, et al. Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. Clin Cancer Res 2010;16:681-90.
42. Cazares LH, Troyer D, Mendrinos S, et al. Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase

kinase kinase 2 discriminates cancer from uninvolved prostate tissue. Clin Cancer Res 2009;15:5541-51.
43. Kim HK, Reyzer ML, Choi IJ, et al. Gastric cancer-specific protein profile identified using endoscopic biopsy samples via MALDI mass spectrometry. Journal of proteome research 2010;9:4123-30.
44. Meistermann H, Norris JL, Aerni HR, et al. Biomarker discovery by imaging mass spectrometry: transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat. Mol Cell Proteomics 2006;5:1876-86.
45. Rauser S, Marquardt C, Balluff B, et al. Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. J Proteome Res 2010;9:1854-63.
46. Reyzer ML, Caldwell RL, Dugger TC, et al. Early changes in protein expression detected by mass spectrometry predict tumor response to molecular therapeutics. Cancer Res 2004;64:9093-100.
47. Yanagisawa K, Shyr Y, Xu BJ, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet 2003;362:433-9.
48. Balluff B, Rauser S, Meding S, et al. MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer. Am J Path 2011;in press.
49. Heeren RM, Smith DF, Stauber J, Kukrer-Kaletas B, MacAleese L. Imaging mass spectrometry: hype or hope? J Am Soc Mass Spectrom 2009;20:1006-14.
50. Djidja MC, Claude E, Snel MF, et al. MALDI-ion mobility separation-mass spectrometry imaging of glucose-regulated protein 78 kDa (Grp78) in human formalin-fixed, paraffin-embedded pancreatic adenocarcinoma tissue sections. J Proteome Res 2009;8:4876-84.
51. Groseclose MR, Massion PP, Chaurand P, Caprioli RM. High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. Proteomics 2008;8:3715-24.
52. Gustafsson JO, Oehler MK, McColl SR, Hoffmann P. Citric acid antigen retrieval (CAAR) for tryptic peptide imaging directly on archived formalin-fixed paraffin-embedded tissue. J Proteome Res 2010;9:4315-28.
53. Lemaire R, Desmons A, Tabet JC, Day R, Salzet M, Fournier I. Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections. J Proteome Res 2007;6:1295-305.
54. Stauber J, MacAleese L, Franck J, et al. On-tissue protein identification and imaging by MALDI-ion mobility mass spectrometry. J Am Soc Mass Spectrom 2010;21:338-47.
55. Chaurand P, Latham JC, Lane KB, et al. Imaging mass spectrometry of intact proteins from alcohol-preserved tissue specimens: bypassing formalin fixation. J Proteome Res 2008;7:3543-55.
56. Ergin B, Meding S, Langer R, et al. Proteomic analysis of PAXgene-fixed tissues. J Proteome Res 2010;9:5188-96.
57. Mange A, Chaurand P, Perrochia H, Roger P, Caprioli RM, Solassol J. Liquid chromatography-tandem and MALDI imaging mass spectrometry analyses of RCL2/CS100-fixed, paraffin-embedded tissues: proteomics evaluation of an alternate fixative for biomarker discovery. J Proteome Res 2009;8:5619-28.
58. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol 2010;2010:840518.
59. Neilson KA, Ali NA, Muralidharan S, et al. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics 2011;11:535-53.
60. Patel VJ, Thalassinos K, Slade SE, et al. A comparison of labeling and label-free mass spectrometry-based proteomics approaches. J Proteome Res 2009;8:3752-9.

61. Alexandrov T, Meding S, Trede D, et al. Super-resolution segmentation of imaging mass spectrometry data: Solving the issue of low lateral resolution. J Proteome Res 2011.
62. Talmadge JE, Fidler IJ. AACR centennial series: the biology of cancer metastasis: historical perspective. Cancer Res 2010;70:5649-69.
63. Pavlidis N, Fizazi K. Cancer of unknown primary (CUP). Crit Rev Oncol Hematol 2005;54:243-50.
64. Pavlidis N, Fizazi K. Carcinoma of unknown primary (CUP). Crit Rev Oncol Hematol 2009;69:271-8.
65. Bugat R, Bataillard A, Lesimple T, et al. Summary of the Standards, Options and Recommendations for the management of patients with carcinoma of unknown primary site (2002). Br J Cancer 2003;89 Suppl 1:S59-66.
66. Bloom G, Yang IV, Boulware D, et al. Multi-platform, multi-site, microarray-based human tumor classification. Am J Pathol 2004;164:9-16.
67. Ma XJ, Patel R, Wang X, et al. Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay. Arch Pathol Lab Med 2006;130:465-73.
68. Ramaswamy S, Tamayo P, Rifkin R, et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A 2001;98:15149-54.
69. Su AI, Welsh JB, Sapinoso LM, et al. Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res 2001;61:7388-93.
70. Talantov D, Baden J, Jatkoe T, et al. A quantitative reverse transcriptase-polymerase chain reaction assay to identify metastatic carcinoma tissue of origin. J Mol Diagn 2006;8:320-9.
71. Tothill RW, Kowalczyk A, Rischin D, et al. An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. Cancer Res 2005;65:4031-40.
72. Varadhachary GR, Talantov D, Raber MN, et al. Molecular profiling of carcinoma of unknown primary and correlation with clinical evaluation. J Clin Oncol 2008;26:4442-8.
73. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62:3609-14.
74. Bloom GC, Eschrich S, Zhou JX, Coppola D, Yeatman TJ. Elucidation of a protein signature discriminating six common types of adenocarcinoma. Int J Cancer 2007;120:769-75.
75. Lee NP, Chen L, Lin MC, et al. Proteomic expression signature distinguishes cancerous and nonmalignant tissues in hepatocellular carcinoma. J Proteome Res 2009;8:1293-303.
76. Scarlett CJ, Smith RC, Saxby A, et al. Proteomic classification of pancreatic adenocarcinoma tissue using protein chip technology. Gastroenterology 2006;130:1670-8.
77. Villanueva J, Shaffer DR, Philip J, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 2006;116:271-84.
78. Le Faouder J, Laouirem S, Chapelle M, et al. Imaging Mass Spectrometry Provides Fingerprints for Distinguishing Hepatocellular Carcinoma from Cirrhosis. J Proteome Res 2011.
79. Schwamborn K, Krieg RC, Jirak P, et al. Application of MALDI imaging for the diagnosis of classical Hodgkin lymphoma. J Cancer Res Clin Oncol 2010;136:1651-5.

80. Schwartz SA, Weil RJ, Johnson MD, Toms SA, Caprioli RM. Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. Clin Cancer Res 2004;10:981-7.
81. Djidja MC, Francese S, Loadman PM, et al. Detergent addition to tryptic digests and ion mobility separation prior to MS/MS improves peptide yield and protein identification for in situ proteomic investigation of frozen and formalin-fixed paraffin-embedded adenocarcinoma tissue sections. Proteomics 2009;9:2750-63.
82. Morita Y, Ikegami K, Goto-Inoue N, et al. Imaging mass spectrometry of gastric carcinoma in formalin-fixed paraffin-embedded tissue microarray. Cancer Sci 2010;101:267-73.
83. Ronci M, Bonanno E, Colantoni A, et al. Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations. Proteomics 2008;8:3702-14.
84. Seeley EH, Caprioli RM. Molecular imaging of proteins in tissues by mass spectrometry. Proceedings of the National Academy of Sciences 2008;105:18126-31.
85. Stauber J, Lemaire R, Franck J, et al. MALDI imaging of formalin-fixed paraffin-embedded tissues: application to model animals of Parkinson disease for biomarker hunting. J Proteome Res 2008;7:969-78.
86. Hauck SM, Dietter J, Kramer RL, et al. Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. Mol Cell Proteomics 2010;9:2292-305.
87. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9:811-8.
88. McKiernan E, McDermott EW, Evoy D, Crown J, Duffy MJ. The role of S100 genes in breast cancer progression. Tumour Biol 2011;32:441-50.
89. Oue N, Hamai Y, Mitani Y, et al. Gene expression profile of gastric carcinoma: identification of genes and tags potentially involved in invasion, metastasis, and carcinogenesis by serial analysis of gene expression. Cancer Res 2004;64:2397-405.
90. Song HY, Liu YK, Feng JT, et al. Proteomic analysis on metastasis-associated proteins of human hepatocellular carcinoma tissues. J Cancer Res Clin Oncol 2006;132:92-8.
91. Jung Y, Lee S, Choi HS, et al. Clinical validation of colorectal cancer biomarkers identified from bioinformatics analysis of public expression data. Clin Cancer Res 2011;17:700-9.
92. Melle C, Ernst G, Schimmel B, et al. Different expression of calgizzarin (S100A11) in normal colonic epithelium, adenoma and colorectal carcinoma. Int J Oncol 2006;28:195-200.
93. Wang G, Wang X, Wang S, et al. Colorectal cancer progression correlates with upregulation of S100A11 expression in tumor tissues. Int J Colorectal Dis 2008;23:675-82.
94. Okudela K, Yazawa T, Ishii J, et al. Down-regulation of FXYD3 expression in human lung cancers: its mechanism and potential role in carcinogenesis. Am J Pathol 2009;175:2646-56.
95. Kayed H, Kleeff J, Kolb A, et al. FXYD3 is overexpressed in pancreatic ductal adenocarcinoma and influences pancreatic cancer cell growth. Int J Cancer 2006;118:43-54.
96. Loftas P, Onnesjo S, Widegren E, et al. Expression of FXYD-3 is an independent prognostic factor in rectal cancer patients with preoperative radiotherapy. Int J Radiat Oncol Biol Phys 2009;75:137-42.
97. Widegren E, Onnesjo S, Arbman G, et al. Expression of FXYD3 protein in relation to biological and clinicopathological variables in colorectal cancers. Chemotherapy 2009;55:407-13.

98. Oppenheimer SR, Mi D, Sanders M, Caprioli RM. Molecular Analysis of Tumor Margins by MALDI Mass Spectrometry in Renal Carcinoma. J Proteome Res 2010.
99. Mitra AP, Pagliarulo V, Yang D, et al. Generation of a concise gene panel for outcome prediction in urinary bladder cancer. J Clin Oncol 2009;27:3929-37.
100. Lim R, Lappas M, Ahmed N, Permezel M, Quinn MA, Rice GE. 2D-PAGE of ovarian cancer: analysis of soluble and insoluble fractions using medium-range immobilized pH gradients. Biochem Biophys Res Commun 2011;406:408-13.
101. Holley SL, Rajagopal R, Hoban PR, et al. Polymorphisms in the glutathione S-transferase mu cluster are associated with tumour progression and patient outcome in colorectal cancer. Int J Oncol 2006;28:231-6.
102. Hirasawa Y, Arai M, Imazeki F, et al. Methylation status of genes upregulated by demethylating agent 5-aza-2'-deoxycytidine in hepatocellular carcinoma. Oncology 2006;71:77-85.
103. Zhou CZ, Qiu GQ, Wang XL, et al. Screening of tumor suppressor genes on 1q31.1-32.1 in Chinese patients with sporadic colorectal cancer. Chin Med J 2008;121:2479-86.
104. Martens JW, Nimmrich I, Koenig T, et al. Association of DNA methylation of phosphoserine aminotransferase with response to endocrine therapy in patients with recurrent breast cancer. Cancer Res 2005;65:4101-17.
105. Vie N, Copois V, Bascoul-Mollevi C, et al. Overexpression of phosphoserine aminotransferase PSAT1 stimulates cell growth and increases chemoresistance of colon cancer cells. Mol Cancer 2008;7:14.
106. Huang D, Casale GP, Tian J, et al. Udp-glucose dehydrogenase as a novel field-specific candidate biomarker of prostate cancer. International journal of cancer Journal international du cancer 2010;126:315-27.
107. McLaughlin PJ, Stack BC, Jr., Levin RJ, Fedok F, Zagon IS. Defects in the opioid growth factor receptor in human squamous cell carcinoma of the head and neck. Cancer 2003;97:1701-10.
108. Choi JH, Kim TN, Kim S, et al. Overexpression of mitochondrial thioredoxin reductase and peroxiredoxin III in hepatocellular carcinomas. Anticancer Res 2002;22:3331-5.
109. Kim K, Yu M, Han S, et al. Expression of human peroxiredoxin isoforms in response to cervical carcinogenesis. Oncol Rep 2009;21:1391-6.
110. Basu A, Banerjee H, Rojas H, et al. Differential expression of peroxiredoxins in prostate cancer: Consistent upregulation of PRDX3 and PRDX4. Prostate 2010.
111. Holmes JL, Sharp SY, Hobbs S, Workman P. Silencing of HSP90 cochaperone AHA1 expression decreases client protein activation and increases cellular sensitivity to the HSP90 inhibitor 17-allylamino-17-demethoxygeldanamycin. Cancer Res 2008;68:1188-97.
112. Garg M, Kanojia D, Saini S, et al. Germ cell-specific heat shock protein 70-2 is expressed in cervical carcinoma and is involved in the growth, migration, and invasion of cervical cells. Cancer 2010;116:3785-96.
113. Milicevic ZT, Petkovic MZ, Drndarevic NC, Pavlovic MD, Todorovic VN. Expression of heat shock protein 70 (HSP70) in patients with colorectal adenocarcinoma--immunohistochemistry and Western blot analysis. Neoplasma 2007;54:37-45.
114. Pfister K, Radons J, Busch R, et al. Patient survival by Hsp70 membrane phenotype: association with different routes of metastasis. Cancer 2007;110:926-35.
115. Hatakeyama H, Kondo T, Fujii K, et al. Protein clusters associated with carcinogenesis, histological differentiation and nodal metastasis in esophageal cancer. Proteomics 2006;6:6300-16.
116. Tsumuraya M, Kato H, Miyachi K, et al. Comprehensive analysis of genes involved in the malignancy of gastrointestinal stromal tumors. Anticancer Res 2010;30:2705-15.

117. Stewart GD, Skipworth RJ, Pennington CJ, et al. Variation in dermcidin expression in a range of primary human tumours and in hypoxic/oxidatively stressed human cell lines. Br J Cancer 2008;99:126-32.
118. Park HJ, Kim BG, Lee SJ, et al. Proteomic profiling of endothelial cells in human lung cancer. Journal of proteome research 2008;7:1138-50.
119. Schulz DM, Bollner C, Thomas G, et al. Identification of differentially expressed proteins in triple-negative breast carcinomas using DIGE and mass spectrometry. Journal of proteome research 2009;8:3430-8.
120. Trojanowicz B, Sekulla C, Lorenz K, et al. Proteomic approach reveals novel targets for retinoic acid-mediated therapy of thyroid carcinoma. Mol Cell Endocrinol 2010;325:110-7.
121. Wiese AH, Auer J, Lassmann S, et al. Identification of gene signatures for invasive colorectal tumor cells. Cancer Detect Prev 2007;31:282-95.
122. Qi W, Liu X, Qiao D, Martinez JD. Isoform-specific expression of 14-3-3 proteins in human lung cancer tissues. Int J Cancer 2005;113:359-63.
123. Pereira-Faca SR, Kuick R, Puravs E, et al. Identification of 14-3-3 theta as an antigen that induces a humoral response in lung cancer. Cancer Res 2007;67:12000-6.
124. Bai Z, Ye Y, Liang B, et al. Proteomics-based identification of a group of apoptosis-related proteins and biomarkers in gastric cancer. Int J Oncol 2011;38:375-83.
125. Hashimoto Y, Ito T, Inoue H, et al. Prognostic significance of fascin overexpression in human esophageal squamous cell carcinoma. Clin Cancer Res 2005;11:2597-605.
126. Yoder BJ, Tso E, Skacel M, et al. The expression of fascin, an actin-bundling motility protein, correlates with hormone receptor-negative breast cancer and a more aggressive clinical course. Clin Cancer Res 2005;11:186-92.
127. Darnel AD, Behmoaram E, Vollmer RT, et al. Fascin regulates prostate cancer cell invasion and is associated with metastasis and biochemical failure in prostate cancer. Clin Cancer Res 2009;15:1376-83.
128. Ozerhan IH, Ersoz N, Onguru O, Ozturk M, Kurt B, Cetiner S. Fascin expression in colorectal carcinomas. Clinics (Sao Paulo) 2010;65:157-64.
129. Sudarshan S, Shanmugasundaram K, Naylor SL, et al. Reduced expression of fumarate hydratase in clear cell renal cancer mediates HIF-2alpha accumulation and promotes migration and invasion. PLoS One 2011;6:e21037.
130. Kawai H, Minamiya Y, Takahashi N. Prognostic impact of S100A9 overexpression in non-small cell lung cancer. Tumour Biol 2011;32:641-6.
131. Arai K, Teratani T, Nozawa R, Yamada T. Immunohistochemical investigation of S100A9 expression in pulmonary adenocarcinoma: S100A9 expression is associated with tumor differentiation. Oncol Rep 2001;8:591-6.
132. Arai K, Takano S, Teratani T, Ito Y, Yamada T, Nozawa R. S100A8 and S100A9 overexpression is associated with poor pathological parameters in invasive ductal carcinoma of the breast. Curr Cancer Drug Targets 2008;8:243-52.
133. Arai K, Teratani T, Kuruto-Niwa R, Yamada T, Nozawa R. S100A9 expression in invasive ductal carcinoma of the breast: S100A9 expression in adenocarcinoma is closely associated with poor tumour differentiation. Eur J Cancer 2004;40:1179-87.
134. Albrethsen J, Moller CH, Olsen J, Raskov H, Gammeltoft S. Human neutrophil peptides 1, 2 and 3 are biochemical markers for metastatic colorectal cancer. Eur J Cancer 2006;42:3057-64.
135. Mothes H, Melle C, Ernst G, Kaufmann R, von Eggeling F, Settmacher U. Human Neutrophil Peptides 1-3--early markers in development of colorectal adenomas and carcinomas. Dis Markers 2008;25:123-9.

136. Kim HK, Reyzer ML, Choi IJ, et al. Gastric cancer-specific protein profile identified using endoscopic biopsy samples via MALDI mass spectrometry. J Proteome Res 2010;9:4123-30.
137. Cross SS, Hamdy FC, Deloulme JC, Rehman I. Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers. Histopathology 2005;46:256-69.
138. El-Rifai W, Moskaluk CA, Abdrabbo MK, et al. Gastric cancers overexpress S100A calcium-binding proteins. Cancer Res 2002;62:6823-6.
139. Stulik J, Koupilova K, Osterreicher J, et al. Protein abundance alterations in matched sets of macroscopically normal colon mucosa and colorectal carcinoma. Electrophoresis 1999;20:3638-46.
140. Wang G, Platt-Higgins A, Carroll J, et al. Induction of metastasis by S100P in a rat mammary model and its association with poor survival of breast cancer patients. Cancer Res 2006;66:1199-207.
141. Mousses S, Bubendorf L, Wagner U, et al. Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays. Cancer Res 2002;62:1256-60.
142. Diederichs S, Bulk E, Steffen B, et al. S100 family members and trypsinogens are predictors of distant metastasis and survival in early-stage non-small cell lung cancer. Cancer Res 2004;64:5564-9.
143. Terris B, Blaveri E, Crnogorac-Jurcevic T, et al. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am J Pathol 2002;160:1745-54.
144. Chen Y, Miller C, Mosher R, et al. Identification of cervical cancer markers by cDNA and tissue microarrays. Cancer Res 2003;63:1927-35.
145. Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A 2003;100:5974-9.
146. Alexandrov T, Becker M, Deininger SO, et al. Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering. J Proteome Res 2010;9:6535-46.
147. Tian T, Hao J, Xu A, et al. Determination of metastasis-associated proteins in non-small cell lung cancer by comparative proteomic analysis. Cancer Sci 2007;98:1265-74.
148. Geering K. FXYD proteins: new regulators of Na-K-ATPase. Am J Physiol Renal Physiol 2006;290:F241-50.
149. Dairkee SH, Sayeed A, Luciani G, et al. Immutable functional attributes of histologic grade revealed by context-independent gene expression in primary breast cancer cells. Cancer Res 2009;69:7826-34.
150. Maxwell PJ, Longley DB, Latif T, et al. Identification of 5-fluorouracil-inducible target genes using cDNA microarray profiling. Cancer Res 2003;63:4602-6.
151. Yamamoto H, Okumura K, Toshima S, et al. FXYD3 protein involved in tumor cell proliferation is overproduced in human breast cancer tissues. Biol Pharm Bull 2009;32:1148-54.
152. Zhang Z, Pang ST, Kasper KA, et al. FXYD3: A Promising Biomarker for Urothelial Carcinoma. Biomark Insights 2011;6:17-26.
153. Zhu ZL, Zhao ZR, Zhang Y, et al. Expression and significance of FXYD-3 protein in gastric adenocarcinoma. Dis Markers 2010;28:63-9.
154. He H, Li J, Weng S, Li M, Yu Y. S100A11: diverse function and pathology corresponding to different target proteins. Cell Biochem Biophys 2009;55:117-26.
155. Ohuchida K, Mizumoto K, Ohhashi S, et al. S100A11, a putative tumor suppressor gene, is overexpressed in pancreatic carcinogenesis. Clin Cancer Res 2006;12:5417-22.

156. Lam F, Jankova L, Dent OF, et al. Identification of distinctive protein expression patterns in colorectal adenoma. Proteomics Clin Appl 2010;4:60-70.
157. Memon AA, Sorensen BS, Meldgaard P, Fokdal L, Thykjaer T, Nexo E. Down-regulation of S100C is associated with bladder cancer progression and poor survival. Clin Cancer Res 2005;11:606-11.
158. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. Annu Rev Pharmacol Toxicol 2005;45:51-88.
159. Jain M, Kumar S, Lal P, Tiwari A, Ghoshal UC, Mittal B. Role of GSTM3 polymorphism in the risk of developing esophageal cancer. Cancer Epidemiol Biomarkers Prev 2007;16:178-81.
160. White DL, Li D, Nurgalieva Z, El-Serag HB. Genetic variants of glutathione S-transferase as possible risk factors for hepatocellular carcinoma: a HuGE systematic review and meta-analysis. Am J Epidemiol 2008;167:377-89.
161. Ye Z, Song H, Higgins JP, Pharoah P, Danesh J. Five glutathione s-transferase gene variants in 23,452 cases of lung cancer and 30,397 controls: meta-analysis of 130 studies. PLoS Med 2006;3:e91.
162. Loktionov A, Watson MA, Gunter M, Stebbings WS, Speakman CT, Bingham SA. Glutathione-S-transferase gene polymorphisms in colorectal cancer patients: interaction between GSTM1 and GSTM3 allele variants as a risk-modulating factor. Carcinogenesis 2001;22:1053-60.
163. Peng DF, Razvi M, Chen H, et al. DNA hypermethylation regulates the expression of members of the Mu-class glutathione S-transferases and glutathione peroxidases in Barrett's adenocarcinoma. Gut 2009;58:5-15.
164. Cianfrocca M, Gradishar W. New molecular classifications of breast cancer. CA Cancer J Clin 2009;59:303-13.
165. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 2004;4:309-14.
166. Labianca R, Nordlinger B, Beretta GD, Brouquet A, Cervantes A. Primary colon cancer: ESMO Clinical Practice Guidelines for diagnosis, adjuvant treatment and follow-up. Ann Oncol 2010;21 Suppl 5:v70-7.
167. Locker GY, Hamilton S, Harris J, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol 2006;24:5313-27.
168. Yan DW, Li DW, Yang YX, et al. Ubiquitin D is correlated with colon cancer progression and predicts recurrence for stage II-III disease after curative surgery. Br J Cancer 2010;103:961-9.
169. Hwang CC, Chai HT, Chen HW, et al. S100B protein expressions as an independent predictor of early relapse in UICC stages II and III colon cancer patients after curative resection. Ann Surg Oncol 2011;18:139-45.
170. Shi H, Zhou Y, Liu H, et al. Expression of CIAPIN1 in human colorectal cancer and its correlation with prognosis. BMC cancer 2010;10:477.
171. Belt EJ, Fijneman RJ, van den Berg EG, et al. Loss of lamin A/C expression in stage II and III colon cancer is associated with disease recurrence. Eur J Cancer 2011;47:1837-45.
172. Chew A, Salama P, Robbshaw A, et al. SPARC, FOXP3, CD8 and CD45 correlation with disease recurrence and long-term disease-free survival in colorectal cancer. PLoS One 2011;6:e22047.
173. Chen Y, Miller C, Mosher R, et al. Identification of cervical cancer markers by cDNA and tissue microarrays. Cancer Res 2003;63:1927-35.
174. Terris B, Blaveri E, Crnogorac-Jurcevic T, et al. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am J Pathol 2002;160:1745-54.

175. Arumugam T, Logsdon CD. S100P: a novel therapeutic target for cancer. Amino Acids 2011;41:893-9.
176. Logsdon CD, Simeone DM, Binkley C, et al. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. Cancer Res 2003;63:2649-57.
177. Guerreiro Da Silva ID, Hu YF, Russo IH, et al. S100P calcium-binding protein overexpression is associated with immortalization of human breast epithelial cells in vitro and early stages of breast cancer development in vivo. Int J Oncol 2000;16:231-40.
178. Mackay A, Jones C, Dexter T, et al. cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. Oncogene 2003;22:2680-8.
179. Fuentes MK, Nigavekar SS, Arumugam T, et al. RAGE activation by S100P in colon cancer stimulates growth, migration, and cell signaling pathways. Dis Colon Rectum 2007;50:1230-40.
180. Vogl T, Leukert N, Barczyk K, Strupat K, Roth J. Biophysical characterization of S100A8 and S100A9 in the absence and presence of bivalent cations. Biochim Biophys Acta 2006;1763:1298-306.
181. Srikrishna G. S100A8 and S100A9: New Insights into Their Roles in Malignancy. J Innate Immun 2011.
182. Goyette J, Geczy CL. Inflammation-associated S100 proteins: new mechanisms that regulate function. Amino Acids 2011;41:821-42.
183. Yong HY, Moon A. Roles of calcium-binding proteins, S100A8 and S100A9, in invasive phenotype of human gastric cancer cells. Arch Pharm Res 2007;30:75-81.
184. Hermani A, De Servi B, Medunjanin S, Tessier PA, Mayer D. S100A8 and S100A9 activate MAP kinase and NF-kappaB signaling pathways and trigger translocation of RAGE in human prostate cancer cells. Exp Cell Res 2006;312:184-97.
185. Ichikawa M, Williams R, Wang L, Vogl T, Srikrishna G. S100A8/A9 activate key genes and pathways in colon tumor progression. Mol Cancer Res 2011;9:133-48.
186. Droin N, Hendra JB, Ducoroy P, Solary E. Human defensins as cancer biomarkers and antitumour molecules. J Proteomics 2009;72:918-27.
187. Lundy FT, Orr DF, Gallagher JR, et al. Identification and overexpression of human neutrophil alpha-defensins (human neutrophil peptides 1, 2 and 3) in squamous cell carcinomas of the human tongue. Oral Oncol 2004;40:139-44.
188. Holterman DA, Diaz JI, Blackmore PF, et al. Overexpression of alpha-defensin is associated with bladder cancer invasiveness. Urol Oncol 2006;24:97-108.
189. Müller CA, Markovic-Lipkovski J, Klatt T, et al. Human alpha-defensins HNPs-1, -2, and -3 in renal cell carcinoma: influences on tumor cell proliferation. Am J Pathol 2002;160:1311-24.
190. Melle C, Ernst G, Schimmel B, et al. Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer. Gastroenterology 2005;129:66-73.
191. Jensen HK, Donskov F, Marcussen N, Nordsmark M, Lundbeck F, von der Maase H. Presence of intratumoral neutrophils is an independent prognostic factor in localized renal cell carcinoma. J Clin Oncol 2009;27:4709-17.
192. Hood BL, Conrads TP, Veenstra TD. Mass spectrometric analysis of formalin-fixed paraffin-embedded tissue: unlocking the proteome within. Proteomics 2006;6:4106-14.
193. Ralton LD, Murray GI. The use of formalin fixed wax embedded tissue for proteomic analysis. J Clin Pathol 2011;64:297-302.

8 Acknowledgements

I would like to thank Prof. Dr. Axel Walch for his excellent supervision.

I would like to thank Prof. Dr. Horst Zitzelsberger for his constant help and support.

I would like to thank Prof. Dr. Jerzy Adamski for being my thesis supervisor.

I would like to thank Prof. Dr. Bernhard Küster and Prof. Dr. Dieter Langosch who have accepted to examine this thesis.

I would like to thank all members of the work group of Axel Walch, for help and fruitful scientific discussions, especially Dr. Sandra Rauser for her supervision and project organisation, Benjamin Balluff for his help with bioformatics and proofreading, and Claudia-Mareike Pflüger and Ulrike Buchholz for their excellent technical assistance.

I would also like to thank my collaboration partners from the Department of Surgery, Klinikum rechts der Isar, Technische Universität München, from the Institute of Pathology, Technische Universität München, from the Research Unit Protein Science, the Research Unit Radiation Cytogenetics, the Institute of Biological and Medical Imaging, the Institute of Biomathematics and Biometry, Helmholtz Zentrum München, from the Chair of Proteomics and Bioanalytics, Technische Universität München, from the Centre for Industrial Mathematics, University Bremen, from the Bruker Daltonik GmbH, and from the Definiens AG. Without them, significant parts of this thesis would have been impossible.