Mass Spectrometry Imaging for the Classification of Tumor Tissue

ISBN/EAN: 978-90-77209-97-4

The work described in this thesis was performed at the FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands.

This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). This work is supported by the Dutch national program COMMIT.

An electronic copy of this thesis is available at www.amolf.nl/publications.

Mass Spectrometry Imaging for the Classification of Tumor Tissue

Massaspectrometrie Imaging
voor de Classificatie van Tumorweefsel
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op woensdag 20 januari 2016 des middags te 12.45 uur

door

Nadine Evertine Mascini

geboren op 31 mei 1987 te Utrecht

Promotor: Prof. dr. R.M.A. Heeren

Table of Contents

1

Introduction

1.1 Tumor heterogeneity

Tumors are biologically heterogeneous. Considerable differences are observed between individuals with the same tumor type, due to variation in genetic subtypes and non-genetic factors[1]. Heterogeneity also exists within individual tumors[2], as depicted in Figure 1.1. Saunders *et al.* defines intratumor heterogeneity as 'the variation in genome, epigenome, proteome and cell and tissue behavior that is found within an individual tumor and its stromal constituents'[3].

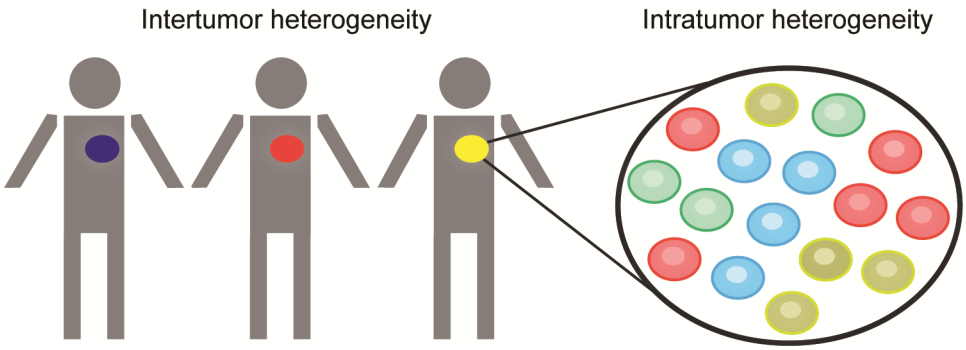

Figure 1.1 Schematic representation of inter- and intratumor heterogeneity.

Tumor cells are embedded in surrounding tissue or stroma, which consists of various cell types, including fibroblasts, immune cells and extracellular matrix components. Taken together they comprise the tumor microenvironment. In addition, variations in

blood vessel architecture are accompanied by regional differences in oxygen and nutrient supply, pH, cell death, and drug uptake.

These local tumor characteristics influence treatment response. For example, tumor cells with an inadequate oxygen supply (hypoxic cells) are known to be more resistant to radiotherapy and chemotherapy than cells with a sufficient oxygen supply[4, 5]. The earliest observations of the insensitivity of tumors to radiotherapy under reduced oxygen conditions date back to the 1930s[6].

Recent evidence suggests that genetic intratumor heterogeneity can also contribute to treatment failure and drug resistance[7, 8]. These reports show that some tumors consist of genetically different cancer cells that vary in their sensitivity to chemotherapeutic drugs. Despite advances in the characterization of intratumor heterogeneity, especially on the genetic level, the question remains how to translate this knowledge into clinical practice[9, 10].

Given the importance of the tumor microenvironment, tumor heterogeneity is also studied within a tissue context. Magnetic resonance imaging and positron emission tomography approaches are able to characterize and quantify this phenomenon[11, 12]. In this thesis the capabilities of mass spectrometry imaging are explored to investigate tumor heterogeneity.

1.2 Mass Spectrometry Imaging

Mass spectrometry imaging (MSI) is based on a surface sampling process in which mass spectrometric data is acquired in a spatially resolved manner. With this technique, multiple analytes can be analyzed simultaneously from a surface: for each detected ion species an image can be generated that represents the distribution of this analyte on the surface.

The earliest examples of mass spectrometry imaging date back almost fifty years, when secondary ion mass spectrometry was used to analyze inorganic surfaces[13]. However, the analysis of biological samples only became possible with the development of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), so-called 'soft' ionization techniques. The application of MALDI-MSI to study biomolecules directly from tissue sections was first demonstrated by Caprioli and coworkers in the late 1990s[14]. Since then, the technique has been established as a useful tool in biomedical and clinical research. The number of applications is still increasing.

1.3 Application of MSI in cancer research

A wide variety of biomolecules can be studied with mass spectrometry imaging. So far, most applications use MALDI-MSI to analyze proteins and proteolytic peptides. However, also lipids, metabolites and pharmaceuticals are studied. Other ionization techniques used for biomedical and clinical imaging are for example desorption electrospray ionization (DESI) and secondary ion mass spectrometry (SIMS). More recently developed methods include laser ablation electrospray ionization (LAESI) and rapid evaporative ionization mass spectrometry (REIMS). Table 1.1 briefly summarizes the main characteristics of these techniques.

Table 1.1 Techniques used for biomedical and clinical MSI. *For MALDI the sample preparation can also include tissue washing, derivatization and proteolytic digestion. In general, the preparation depends on the tissue type, the targeted molecular class and the required spatial resolution.

	Sample preparation	**Ionization**	**Spatial resolution**	**Molecular class**
MALDI	Matrix coating*	Laser ablation and desorption/ionization	10-200 µm	Lipids, peptides, proteins, metabolites
DESI	None	Desorption by solvent droplets and ESI-type ionization	>100 µm	Mostly lipids and metabolites
SIMS	None or matrix/metal coating	Sputtering with a primary ion beam	<1 µm dynamic mode, >1 µm static mode	Elements, fatty acids and lipids
LAESI	None	Laser ablation combined with ESI	>200 µm	Lipids, peptides, proteins, metabolites
REIMS	None	Thermal evaporation	>500 µm	Phospholipids

MSI has two features that make it well suited for cancer research: it requires no target-specific labeling and can thus be used as a discovery tool where the targets are unknown prior to analysis. In addition, MSI analysis leaves the tissue intact. Molecular distributions, as measured by MSI, can be directly compared with the microscopic structure (i.e. histology) of the tissue. Combined, these features enable two types of approaches:

1) Histology-driven approach. Mass spectra are extracted from specific tissue regions and used for tissue classification or biomarker (pattern) discovery. The studies described in Chapter 3 and 4 are examples of this approach.
2) MSI-driven approach. Tissue is annotated based on the mass spectral profiles generated by MSI. Regions with distinct MS profiles are then compared to the tissue's histology.

The information obtained with either of the approaches can be correlated with biomedical or clinical data. In clinical MSI, research is mostly concerned with biomarker discovery. Biomarker discovery aims to find measurable indicators of disease state. These indicators can then be used as surrogate markers of disease to assist in diagnosis, predict disease progression or aid in treatment decisions.

The cellular specificity of MSI might partly explain its ability to find biomarker (patterns), even though only abundant tissue analytes can be detected. A tissue area of 100 x 100 µm, which is a routinely used pixel size for MALDI-MSI, contains around 25 cells, assuming an average cell size of 20 µm. In comparison, approaches that use laser capture microdissection need at least 500 cells, and more often populations of thousands of cells, to study cell-type specific proteomes[15, 16].

Small tissue pieces such as needle core biopsies from patients can be easily analyzed with MALDI-MSI. Tissue microarrays (TMAs) consist of arrays of tissue cores with a typical diameter of 0.6-2 mm. These cores are selected to enable comparison between patients. Using TMAs, hundreds of patient samples can be studied within a single experiment and thus under highly similar conditions. Moreover, TMAs are often constructed from well-documented patient cohorts, for which follow-up data has been sometimes collected over many years. These features make TMAs attractive samples for MALDI-MSI analysis.

Most MALDI-MSI studies use fresh frozen tissue. However, tissue microarrays contain formalin-fixed paraffin-embedded (FFPE) tissue. FFPE tissue is conserved by dehydration and cross-linking of the proteins with formalin. After formalin fixation, the tissue is embedded in paraffin to preserve its morphology and allow thin sectioning. FFPE tissue is widely used for clinical applications, due to easy storage and handling. MALDI-MSI analysis of FFPE tissue became possible with the adaptation of antigen retrieval protocols used for immunohistochemistry[17, 18]. The preparation of FFPE tissue is more time-consuming as compared to the preparation of fresh frozen

tissue, and includes steps to remove paraffin and reverse part of the cross-linking to make the proteins amenable to MALDI-MSI analysis.

1.4 Clinical and biomedical questions

A central theme in today's MSI field is whether we can use the molecular profiles generated by MSI to aid medical doctors, specifically pathologists, in clinical decision making. Can MSI data provide information to diagnose disease, predict disease progression or aid in treatment choice?

Information to guide treatment would be helpful in cases where only a subset of patients benefits from a treatment. MALDI-MSI was for example successfully used in the differentiation of Spitz nevus (benign) from Spitzoid malignant melanoma[19]. For pathologists it is sometimes impossible to distinguish between the two lesions, but ideally none of the Spitz nevus patients is unnecessarily subjected to surgery and chemotherapy. Other examples in which MSI could potentially aid in treatment choice are described in Chapter 3 and 4.

MSI can also be used for more fundamental cancer-related research. The investigation of phenotypic intratumor heterogeneity and the identification of lipids co-localizing with hypoxic tumor regions are notable examples[20, 21]. In general, the molecular profiles generated by MSI also provide information to further investigate underlying biological processes with other bioanalytical techniques.

1.5 Practical aspects of data analysis

Data analysis is an important part of MSI projects. The generated data sets are large (ranging from a couple to hundreds of gigabytes) and require extensive processing before they can be used for hypothesis testing[22, 23].

In general, MALDI-MSI based biomarker discovery aims to find differences between classes (e.g. healthy and diseased) that have predictive power. For this purpose a statistically significant difference between the classes is not sufficient; the classification accuracy depends on the overlap in distributions between the classes. Single biomarkers are typically not sensitive and specific enough; more often combinations of markers are used. Different algorithms can be employed to build a classifier, as reviewed by Hilario *et al.*[22]. In many cases, MALDI-MSI data is high-dimensional. This means that the number of measured *m/z* values is much larger than

the number of measured samples. Classifying high-dimensional data poses some challenges, which are described in detail in Chapter 4.

The choice for a data analysis strategy depends not only on the aim of the study, but also on the type and number of samples available. For studies with human samples a much larger number of samples is needed to account for variability introduced by genetic and environmental influences (e.g. diet, age) as compared to studies using animal models. Moreover, careful matching of human samples is required to minimize the effect of these other sources of variability in the data. Patient-derived xenograft models are human tumors implanted and propagated in immunodeficient mice. They retain the heterogeneity of human tumors, but are grown in a controlled laboratory environment and thus exhibit less variability as compared to clinical samples.

Finally, biomarkers require validation using an independent sample set and preferably also another technique, as for example immunohistochemistry[24, 25]. The interested reader is referred to Jones *et al.* and Schwamborn *et al.* for a detailed overview of study setup and data analysis methods for biomedical and clinical MSI[23, 26].

1.6 Scope of the thesis

The research presented in this thesis was mainly carried out in the Biomolecular Imaging Mass Spectrometry group at the FOM Institute AMOLF in Amsterdam. At the time the thesis was written, the research group moved to the University of Maastricht. The division of Imaging Mass Spectrometry is now part of the Maastricht Multimodal Molecular Imaging institute. The research still focuses on the development and application of analytical tools for MSI. This thesis presents examples of MSI-based methods to study tumor tissue.

The research was carried out in collaboration with research groups from the Netherlands Cancer Institute (Chapter 3), the University Medical Center Utrecht (Chapter 4), and Johns Hopkins University School of Medicine (Chapter 5).

Chapter 2 describes one of the bottlenecks in biomolecular MSI: identification of the analyte(s) of interest. Typically only a few of the hundreds of observed biomolecular signals in a MSI spectrum can easily be identified. This chapter presents some of the current strategies for protein identification with a focus on advances in instrumentation, experimental workflow and bioinformatic tools that improve the

number and confidence of protein identifications. It also includes a description of MALDI-MSI.

Chapter 3 presents an approach that combines MALDI-MSI on tissue microarrays with principal component analysis and linear discriminant analysis (PCA-LDA) to predict treatment response. The feasibility of this approach was evaluated on a set of patient-derived xenograft models of triple-negative breast cancer. We also investigate to which extent heterogeneity between the tissue cores from a particular tumor model compromises response prediction.

Chapter 4 further investigates the use of tissue microarrays for MALDI-MSI. It compares different multivariate data analysis methods to predict lymph node metastasis and disease-free survival for a set of head and neck cancer patient samples. In addition, it discusses the challenges associated with MALDI-MSI based biomarker discovery.

Chapter 5 focuses on intratumor heterogeneity. This chapter demonstrates the detection of the exogenous hypoxia marker pimonidazole by MALDI-MSI directly from tissue sections. It shows the visualization of hypoxic regions in a breast tumor xenograft model with this approach. In addition, it presents the identification of endogenous hypoxia-associated molecules.

Chapter 6 presents the conclusions and provides an outlook on future research.

2

Protein identification in mass spectrometry imaging

With a rapidly growing number of biomedical applications of mass spectrometry imaging (MSI) and expansion of the technique into the clinic, spectrum annotation is an increasingly pressing issue in MSI. Although identification of the species of interest is the key to answering biomedical research questions, only a few of the hundreds of observed biomolecular signals in each MSI spectrum can easily be identified or interpreted. So far no standardized protocols resolve this issue.

Present strategies for protein identification in MSI, their limitations and future developments are the scope of this chapter. We discuss advances in MSI technology, workflows and bioinformatic tools to improve the confidence and number of protein identifications within MSI studies.

2.1 Introduction

In recent years, technological and methodological advances have brought mass spectrometry imaging (MSI) to the biomedical field. MSI allows for the analysis and visualization of peptides, proteins, lipids, metabolites and pharmaceuticals directly from biological tissue and cell samples[27, 28]. The technique uses a surface sampling process in which mass spectra are collected at discrete locations according to a predefined Cartesian grid. In this way, the distribution of ions of interest can be mapped.

MSI has several advantages compared to other imaging techniques such as immunohistochemistry (IHC) or positron emission tomography. It has the capability to detect hundreds of (unknown) compounds simultaneously in one molecular imaging experiment, allowing multiplexed analysis and discovery-based research. As MSI requires no target-specific labeling, unmodified species can be studied. Importantly, in contrast to standard mass spectrometric analysis, which requires tissue homogenization, MSI leaves the molecular distribution in the tissue intact, so it can be utilized to assess molecular differences between specific cellular regions within a tissue.

An increasing number of studies report on applications of MSI in the biomedical field. MSI is used in distribution studies of pharmaceutical compounds and their metabolic products for drug evaluation[29, 30] and in (clinical) proteomics applications[31, 32]. MSI has already been employed to assist in diagnosis, prognosis and biomarker discovery. The technique is used to construct protein profiles that predict disease status or progression, to identify molecular patterns for disease prognosis and to assess molecular markers in treatment response studies[33-35]. Not only can a better fundamental understanding of the molecular processes underlying disease be acquired using MSI, but this knowledge can also aid in the development of new drugs and treatments. The study of the molecular basis of intratumor heterogeneity, for example, is not only expected to lead to improved understanding of tumor biology, but also fits in the trend towards personalized medicine[36, 37].

The most widely used ionization technique for MSI is matrix-assisted laser desorption/ionization (MALDI)[38]. MALDI-MSI was introduced in 1997 by Caprioli and coworkers and uses a matrix, typically an acidic aromatic compound[14]. As the matrix compound absorbs energy at the wavelength of the laser, exposure of the crystals to laser pulses results in desorption and ionization of the sample. Ions are separated

based on their mass-to-charge (*m/z*) ratio, usually by a time-of-flight (TOF) mass analyzer, which is high-throughput and sensitive, and has a broad mass range[39].

Despite the fast developments in MSI technology and workflows, several challenges still need to be addressed for MSI to become an established tool in biomedical and clinical research. Apart from the need for improved mass resolution, spatial resolution and sensitivity of the instruments used for MSI, an important limitation is that only a few of the hundreds of observed signals in each mass spectrum can be easily identified or interpreted. Annotation of ions of interest requires an additional step in the experimental workflow and so far no standardized protocols exist that solve this issue. Identification might be hampered by (unknown) modifications, even when the compound class is known. This holds for example for ions derived from proteins, where posttranslational modifications (PTMs), protein isoforms and chemical modifications resulting from sample preparation or proteolysis can hinder interpretation.

However, from the point of view of a biomedical researcher identification of the species of interest is an essential step to solve a biomedical research question. Although recently studies have been published in which statistical data analysis tools were used to annotate tissue sections solely based on their mass spectrometric profiles[36, 37, 40], MSI data needs to be complemented with information on the nature of the biomolecular species to access the full potential of MSI[41].

With a rapidly growing number of biomedical applications and expansion of MSI into the clinic, spectrum annotation is an increasingly pressing issue in MSI. Present strategies to provide annotation of MSI spectra, their limitations, and newly developed identification strategies are the scope of this chapter. As proteins are the biomolecules most often probed by MSI in a biomedical context, this chapter focuses on protein identification in MSI. However, confident chemical assignment of any biomolecular species in MSI spectra faces similar challenges and some of the approaches described here could also be employed in that context.

2.2 MS-based protein identification methods

Mass spectrometry (MS) is an established analytical technique for protein characterization both at both species level and at proteome level. Numerous, often very sophisticated, methods of MS-based protein identification have been developed[42, 43].

Current MSI has implemented MS-based protein identification methods in its work-flows according to the needs and constraints posed by the technique.

In general, two approaches exist for MS-based protein identification:

1) **In a top-down experiment**, identification is performed through intact mass measurement followed by tandem MS (MS/MS) analysis. Sequence-specific fragmentation patterns are used for identification through database searching, in which the experimentally obtained fragments are compared with theoretical fragments. A top-down approach in MALDI-MSI works for small-to-medium-sized proteins up to 7-10 kDa, because large singly charged molecules will not easily dissociate.
2) **In a bottom-up experiment**, a protein or protein mixture is first enzymatically digested. The resulting proteolytic peptides are analyzed by MS (so called peptide mass fingerprinting), and MS/MS in the case of a protein mixture. In a bottom-up imaging approach, multiple peptide matches per protein are required for confident identification of the protein.

An ideal MSI experiment consists of automatically triggered MS/MS experiments on proteins or peptides directly from tissue, thereby combining the localization of species with their identification within a single experiment. However, low sensitivity seriously hampers the identification as compared to standard identification approaches using protein extraction followed by gel-based separation or liquid chromatography (LC) coupled to electrospray ionization (ESI) MS/MS. The low sensitivity is caused by ion suppression effects due to the complex molecular composition of tissue. In addition, ions generated by MALDI typically have only unit charge. The resulting inefficient ion activation of larger ions renders intact proteins too big for direct identification through fragmentation. As a result, efficient MS/MS can only be performed in a mass range of 500-3500 Da on the majority of mass spectrometers used for MSI. Figure 2.1 summarizes the protein identification workflows used in MSI.

Top-down approaches in MSI

Few examples of a top-down approach used in MALDI-MSI can be found in literature. Minerva and coworkers identified several endogenous peptides up to 3.5 kDa using MALDI-TOF/TOF directly on mouse pancreatic tissue[44].

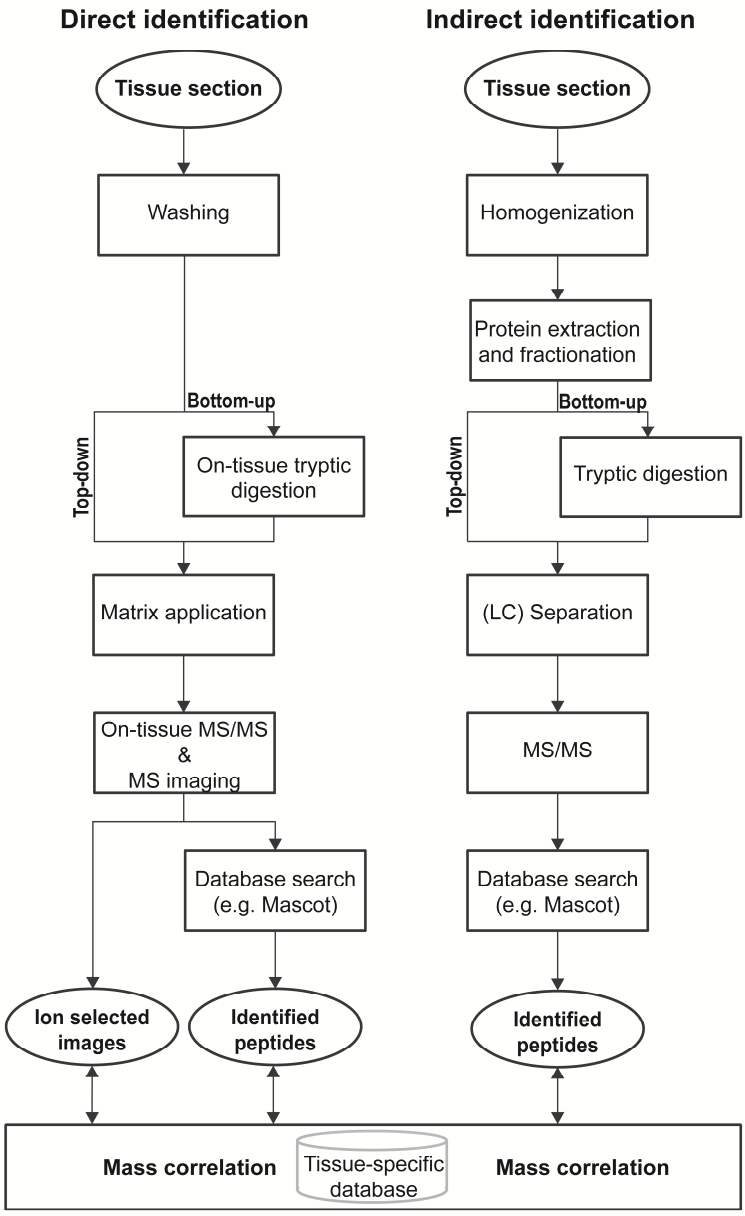

Figure 2.1 Workflow for MS-based protein identification. Identification can be performed within the MSI experiment itself (direct identification) or by using independent MS/MS data followed by mass correlation (indirect identification).

Alternative identification strategies are employed to annotate larger masses, which combine fractionation of tissue extracts by LC, MALDI-MS for fraction selection, followed by ESI-MS/MS[45-47]. In this way, identification of the 8.4 kDa cysteine-rich intestinal protein 1 in breast cancer tissue was demonstrated[45]. Top-down analysis of the 14+ charge state resulted in identification of the protein, which was found to be correlated with human epidermal growth factor receptor 2, an important marker for treatment response prediction. These classical identification strategies are labor intensive and require extraction of the protein of interest, so they remain limited to only a few identifications per study.

An alternative approach to standard fragmentation techniques in MSI, such as TOF/TOF and collision induced dissociation (CID), is in-source decay, where ions are fragmented in the source region before extraction[48]. However, this technique suffers from the lack of precursor ion selection, which makes the mass spectra hard to interpret. In-source decay is therefore only rarely used for protein identification. A recent trend is the development of electron-based MS/MS techniques. Electron-induced dissociation of singly-charged peptides has been demonstrated[49].

Despite the limited utility of a top-down approach due to technical and practical constraints, it should be kept in mind that by studying intact proteins, not only information on the complete amino acid sequence is retained, which allows for high-confidence protein assignment, but also on the protein state. Cazares and coworkers for example, identified specifically a fragment of the MEKK2 protein to discriminate tumor from normal tissue[50]. This type of information typically cannot be obtained using IHC or a bottom-up approach (See 'Indirect identification approaches').

Bottom-up approaches in MSI

In a bottom-up approach, proteins are digested on-tissue while their spatial distribution is preserved. Trypsin is the enzyme of choice for digestion and can be applied by automated spotting devices. These devices deposit picoliter (pL) droplets in an array with a spot size of 100-200 μm. After incubation, matrix can be deposited onto the tissue using the same device[51]. The resulting tryptic peptides are subjected to MS/MS directly on-tissue[52, 53]. This *in situ* digestion approach is often considered the method of preference for MSI studies, because it facilitates on-tissue fragmentation, hence peptide identification within the imaging experiment itself. An additional advantage is that on-tissue digestion can be used to 'unlock' proteins from the

formalin-fixed paraffin-embedded tissue widely used in bio(medical) research, the proteome of which would otherwise remain inaccessible for MSI analysis[17, 54-56].

However, the on-tissue digestion approach also suffers from considerable background signal from the tissue. This results in identification of mainly highly abundant proteins and only a limited number of peptides per identified protein. The limited amino acid sequence coverage per protein may also result in loss of information on protein state, for example on type and location of PTMs.

Several improvements in sample preparation and instrumental set-up have found implementation in bottom-up MALDI-MSI workflows. The addition of the detergent n-octylglucoside to the trypsin buffer solution was found to increase the number of peptide signals and their signal intensities and led to enhanced detection of lipophilic proteins[53]. Furthermore, on-tissue chemical peptide derivatization strategies were developed for enhanced identification. Franck and coworkers showed that on-tissue derivatization of tryptic peptides is compatible with an *in situ* digestion approach[57]. MS/MS spectra recorded on conventional MALDI-TOF instruments are often difficult to interpret due to the different types of ion series generated. This results in only small sequence tags being available for identification. The addition of a N-terminal negative charge by derivatization with sulfonation agents generated (almost) complete y-ion series and even allowed for *de novo* sequencing (i.e. without the help of a protein database) of tryptic peptides.

A recent advancement is the combination of ion mobility spectrometry (IMS) with MSI[51, 53, 55, 58, 59]. IMS is a gas-phase chromatographic technique, which separates ions based on their collision cross-section (i.e. size and shape). The extra dimension provided by the post-ionization ion mobility separation allows separate inspection of isobaric contributions to a spectrum. This is especially useful for complex spectra resulting from the analysis of *in situ* digested tissue, which show unresolved peaks from overlapping species, as for example isotopic distributions of peptide species, lipids and matrix ions. In this way, the complexity of MS/MS spectra is reduced. Figure 2.2 shows how ion mobility separation prior to fragmentation of two singly charged tryptic peptide ions that both have a molecular weight of *m/z* 1039 resulted in their identification (from tubulin and ubiquitin, respectively)[59]. An MS/MS database search using the Mascot engine without ion mobility separation resulted in a score which was too low for confident identification of either of the two tryptic peptides and the proteins they originate from.

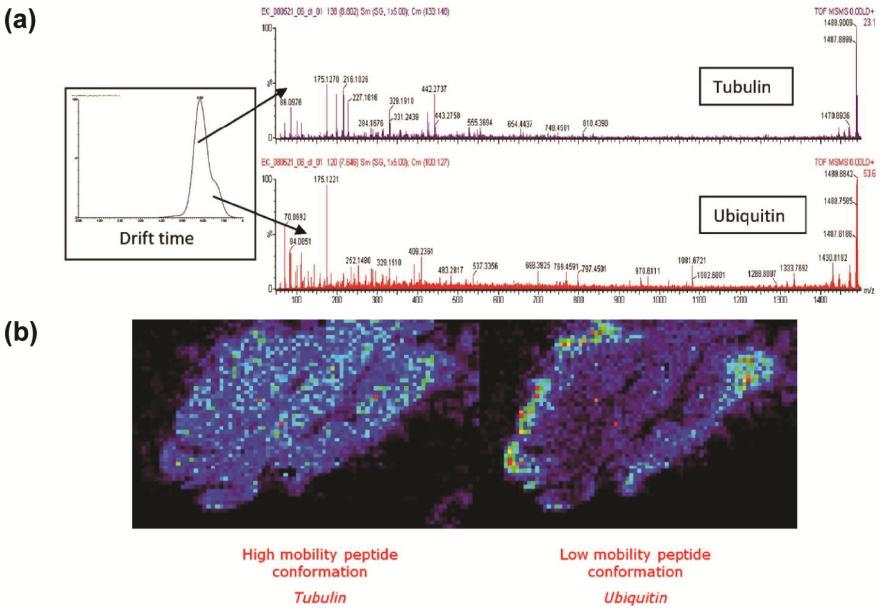

Figure 2.2 Ion mobility separation combined with MSI. (a) On-tissue MS/MS spectra of ion mobility separated tryptic peptides (*m/z* 1039), identified as tubulin and ubiquitin fragments. (b) Ion images of separated tubulin and ubiquitin fragments. Without ion mobility separation (no drift time selection), the ion image would have corresponded to the superposition of the two images (Adapted with permission[59]).

Indirect identification approaches

Indirect approaches for protein identification are often used to avoid sensitivity issues with on-tissue fragmentation. The aim of these strategies is to eliminate ion suppression effects, which introduce ionization bias in the MSI analysis, and to increase the dynamic range of the analysis.

In short, MSI data are matched with data generated using complementary methods that include a fractionation step (mostly LC-MS)[60]. This is not a trivial undertaking as ESI, the most commonly used ionization technique for LC-MS due to the simplicity of the interface, favors the ionization of different peptides as compared to MALDI. LC-MALDI is used less frequently[61], and LC coupled to secondary ion MS still has to prove its utility[62]. In an indirect approach, independent experimental data serves as a tissue-specific reference database, which can be searched to identify peptides in the MSI data (Figure 2.1).

The capability of MSI to measure complex samples might be further enhanced by targeting specific cell populations from tissue by laser capture microdissection (LCM)[63], which is an especially useful enrichment technique for tissue showing a high degree of heterogeneity (e.g. breast cancer tissue)[64].

Mass correlation between MSI data and data from independent experiments requires extensive (manual) data interpretation, often combined with prior knowledge of the species of interest. Wide mass tolerance windows for mass matching of up to ±2 Da are reported[65]. Masses with tolerance windows of this size theoretically match thousands of possible peptides. In these cases, additional validation is an absolute necessity to prevent erroneous protein identification.

An MSI study of tumor margins in renal cell carcinoma reported on the use of an additional peptide characteristic to eliminate false positive protein identifications[66]. Tryptic peptides from tissue extract were isoelectrically focused using an immobilized pH gradient strip to provide additional information on the peptide's isoelectric point to match the experimental with theoretical peptides.

A recent paper by Schober and coworkers described an improved indirect strategy, which combined MALDI-MSI with complementary off-line LC coupled to ESI-MS/MS[67, 68]. All results were based on accurate mass measurements recorded on Fourier transform MS instruments, which allowed for improved quality and quantity of peptide identifications. Fourier transform ion cyclotron (FT-ICR) and Orbitrap mass spectrometers have mass accuracies in the low to sub ppm range instead of the at most 10-50 ppm mass accuracy obtained by using TOF systems. The high mass accuracy and mass resolving power of these instruments provide enhanced means to resolve the complexity of biological samples, but the use of these mass analyzers for protein identification in MALDI-MSI is still limited to only a few examples due the limited sensitivity of FT-ICR, the limited mass range of Orbitrap, and the relatively long measurement time needed to obtain high accuracy[69, 70]. Moreover, an indirect accurate mass approach only works if both the MALDI-MSI data and the LC-ESI data are recorded with high mass accuracy.

2.3 The importance of mass accuracy for protein identification

Mass accuracy can be defined as the degree of similarity between a measured value and its theoretical value. If multiple species are assessed, usually the root mean square or mass measurement error is used. Precision is defined as the degree to which a measured value is similar during a (series of) experiment(s). In MS-based protein and proteomics research, statistical tools have already been widely implemented to assess accuracy, precision and hence confidence of identification[71, 72].

MS-based identification can be extremely accurate and precise because it uses the intrinsic property of a species (i.e. its mass), which can be measured by MS in an unbiased way. A highly accurate monoisotopic mass (sub ppm for a peptide of 1 kDa) provides information on mass defect and isotopic distribution, and can even specify the elemental composition. Accurate mass measurements lead therefore to improved confidence in protein assignment.

In addition, an important parameter in peptide identification, directly related to mass accuracy, is the threshold value (or mass tolerance window) used in a database search. The set threshold is a tradeoff between maximum specificity and maximum sensitivity. At strict thresholds, true positives are potentially rejected, whereas, at less strict thresholds, the mass resolution of the recorded data might not be fully used.

Until now, little emphasis has been placed on assessment of peptide annotation reliability in MSI studies. Instead, orthogonal validation methods such as IHC are employed. However, recent high mass resolution MSI studies show that results can largely vary depending on the mass bin width used for ion selected images, which directly depends on the mass accuracy and mass resolution of the data, as exemplified in Figure 2.3[73].

A 2011 study described the assessment of the mass accuracy of MALDI-MSI data[74]. In this paper, MSI data were linked to MS/MS data from independent experiments by employing an intermediate step using accurate mass data from FT-MS measurements. The mass accuracy of the recorded MALDI-MSI data was found to decrease with increasing mass range and the applied mass tolerance window for mass correlation was adjusted accordingly. These examples demonstrate the trend to use accurate mass data to improve the number and the confidence of peptide annotations.

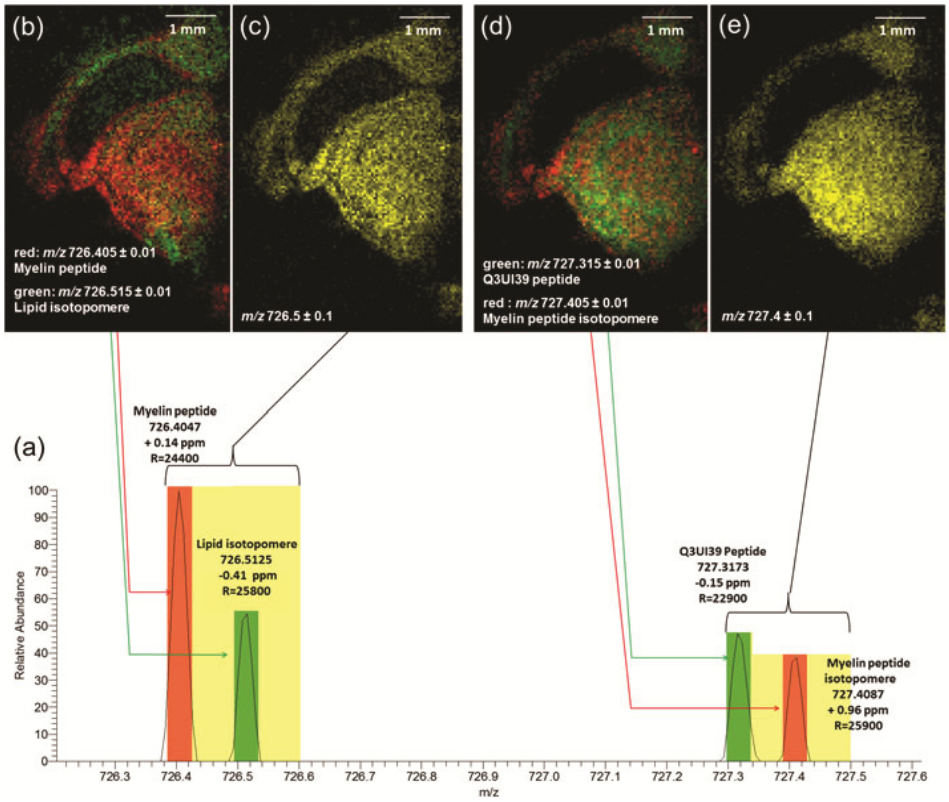

Figure 2.3 High mass resolution MSI not only improves the reliability of peptide assignment, but also the spatial distribution information. (a) Zoomed Orbitrap mass spectrum of a MALDI-MSI analysis of a mouse brain. (b) An overlay of ion images generated with an *m/z* bin width of ±0.01 shows the different spatial distributions of a myelin tryptic peptide (red) and a phospholipid (green) at *m/z* 726.405 and *m/z* 726.515, respectively. (c) An ion image generated with a larger bin width of ±0.1 leads to a superposition of the two images, hence the spatial distribution of the two ions is not resolved. (d) An overlay of ion images of a tryptic peptide of SNAP-91 (green) and a myelin peptide isotopomeric peak (red) at *m/z* 727.315±0.01 and *m/z* 727.405±0.01 and (e) an ion image at *m/z* 727.4±0.1 show the same effect (Reprinted with permission[73]).

2.4 Data analysis

Protein identification also heavily depends on the data processing and mining strategy chosen, the quality of the protein database and the database searching algorithms used. As MSI data analysis uses bioinformatic tools and databases developed for MS-based protein and proteomics research, the challenges in data analysis show large

overlap. An introduction to the already well-described problems associated with protein identification from MS/MS spectra can be found in reference[75]. However, existing data analysis tools are often specifically designed for (ESI)-MS data, so they might not perform optimally on MSI data. Occasionally, in-house developed algorithms are reported[68, 74]. In a MALDI-MSI study of the obese mouse pancreas, for example, MS/MS data sets were clustered to allow identification of structurally related peptides[74].

Importantly, the large data sets generated in MSI experiments create new bioinformatic challenges. Data are processed and mined to reduce the influence of technical and analytical variation and extract information relevant to the biological problem, with either standard software (e.g. Biomap, ClinProTools[76]) or in-house developed algorithms[77]. The interested reader is referred to www.maldi-msi.org that provides a concise overview of available MSI software.

Key here is identification and extraction of relevant spatial and chemical features. Mass spectra generated from different locations on a tissue probe differences in molecular make-up of that tissue, so they can be used for clustering or classification. Data mining methods allow identification of signature masses for specific tissue regions or tissue states, which can be assigned using either protein extraction or *in situ* digestion approaches. In this way, MSI enables the targeted analysis of relevant species for biomarker discovery.

Supervised methods make use of prior knowledge about the tissue, and typically use histological images to define different regions of interest (ROIs)[17, 50, 78-80]. These approaches are referred to as 'histology-directed' MSI[78, 81]. Differentially expressed peaks between the ROIs are identified using statistical tests and used to generate classification models. A recent report showed that histology-directed classification of MALDI-MSI data led to the identification of differentially expressed modified protein species in skin cancer (e.g. multiply acetylated forms of histone H4 and H2A)[78]. As PTMs reflect the actual biological state of proteins, they can be highly relevant for biomarker discovery.

Unsupervised methods, including multivariate methods such as Principal Component Analysis and Hierarchical Clustering, can reveal histology-independent regions[37, 82]. Combinations of (un)supervised methods and newly developed strategies have also been reported[36, 83, 84]. The use of multiple multivariate techniques on one

MSI data set can provide a more accurate description of regions with distinct MS profiles, but this type of study is still at a developmental stage[85].

Furthermore, the large data sets generated by MSI have a big impact on the computational infrastructure necessary for data processing and analysis[77]. Further advancements in high throughput analysis are crucial to improve speed and reliability of protein annotation.

2.5 Validation

In MSI, the use of additional validation methods is a necessity. As explained in Section 2.3, one should be careful about inferring identities between (imaging) data sets. Once MSI has pinpointed towards interesting species, they are typically further investigated using techniques from biochemistry, such as IHC[55, 79] and *in situ* hybridization[80]. The use of standards, common practice in drug distribution studies, is uncommon for independent validation of peptides or proteins, because the preparation of (synthetic) isotope-labeled peptides or protein standards is often far from straightforward.

Also targeted chemical labeling of proteins for direct protein identification from tissue is reported. The addition of a tag allows for enhanced detection of specific species in an MSI experiment, but requires prior knowledge of the protein of interest. The advantage of such an approach is that it allows for multiplexed analysis of preselected proteins, usually a problem when using IHC. Moreover, it enables the analysis of low-abundance and high-mass proteins, which are hard to probe by MALDI-MSI.

Thiery and coworkers showed multiplex immunolabeling of proteins, named TAMSIM for TArgeted multiplex MS IMaging[86]. In this strategy, proteins are linked to an antibody with a mass tag, which is released upon laser irradiation and subsequently detected by MSI without the need for matrix addition. Lemaire and coworkers showed the similar concept of 'Tag-Mass': the addition of a probe with a photocleavable tag of known mass linked to mRNA or protein[87, 88]. Proof of principle was shown for the 180 kDa carboxypeptidase D membrane protein from rat brain tissue (Figure 2.4). As antibodies can show high specificity for their corresponding antigen, this method allows for specific protein identification in an MSI experiment and can be used for validation, as shown for a new potential biomarker for ovary cancer[25]. The Tag-Mass technology is now patented for use in quantitative diagnostic assays.

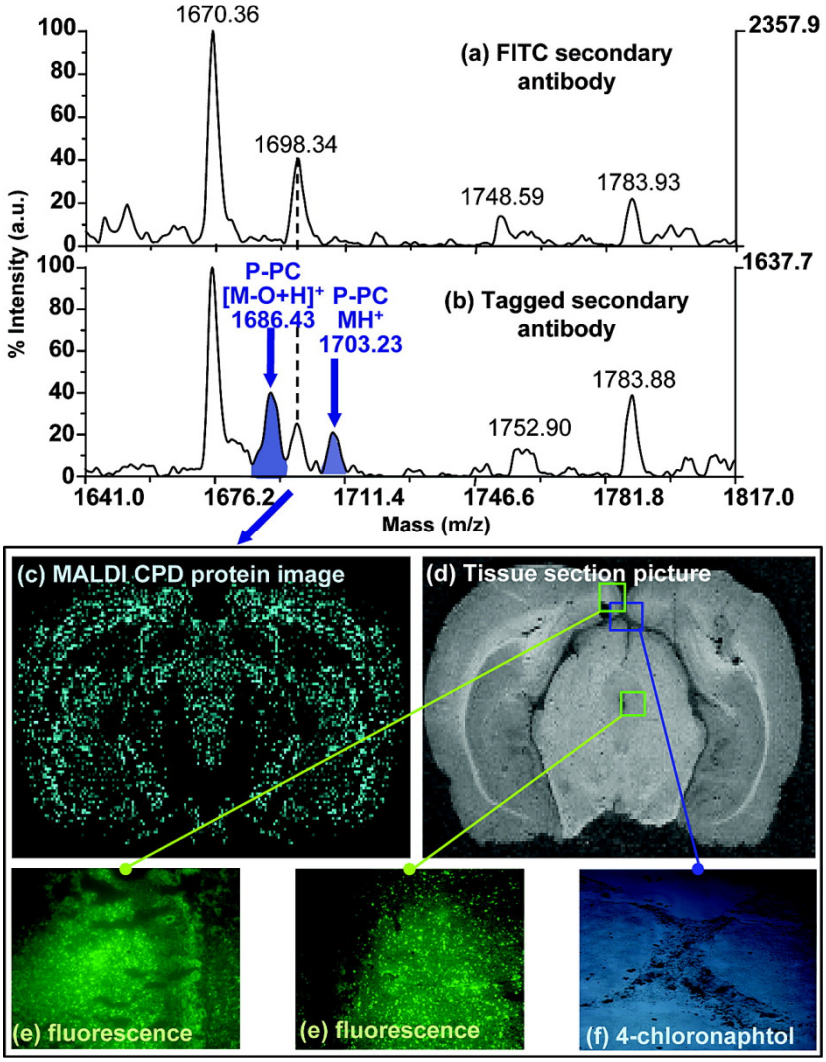

Figure 2.4 Example of the Tag-Mass concept. MALDI mass spectra from adjacent rat brain tissue sections after IHC against carboxypeptidase D (CPD), with (a) untagged and (b) tagged secondary antibody. Two characteristic signals for the Mass-Tag (P-PC) were observed (*m/z* 1686.43 and *m/z* 1703.23). (c) Corresponding ion image at *m/z* 1686.43. (d) Rat brain tissue before analysis. Similar results were obtained with secondary antibody detection with (e) fluorescence or (f) peroxidase staining using 4-chloronaphtol (Reprinted with permission[87]).

More recently, multiplexed detection of proteins with imaging mass cytometry was reported[89]. Imaging mass cytometry is an approach that combines immunohistochemistry with lanthanide-labeled antibodies and laser-ablation inductively-coupled-plasma MS (LA-ICP-MS). Giesen and coworkers showed the simultaneous imaging of 32 proteins at a 1 µm spatial resolution. However, all validation methods have one disadvantage in common, namely that they only allow the validation of a limited number of proteins per study.

An emerging approach for validation is to study the same sample with different MS or spectroscopic techniques as magnetic resonance spectroscopic imaging[90, 91]. In addition to its use for independent verification, a multimodal approach can provide a more comprehensive view of the sample being investigated.

2.6 Conclusions and future perspectives

The greatest strength of MSI is that it can provide relatively unbiased molecular information in an anatomical context. Used as a discovery tool, MSI can highlight interesting species to be further investigated. Alternatively, MSI can be used to visualize a biomedical hypothesis. Although MSI has proved its capability in biomedical research, it is not yet widely adopted. This is partly due to the still existing gap between technique-based method development of MSI and the demands from the biomedical research community.

This chapter addressed one of the main bottlenecks, namely protein identification in an MSI experiment. At the moment, spectral annotation is a laborious, often complicated task for each new set of samples within a laboratory. The number of annotations therefore remains limited to at most tens per study, while the studied spectra contain easily 10-fold more signals.

An obvious way to improve annotation is to merge MSI data with complementary data, as exemplified by the indirect approaches described. However, this will require a substantial effort in bioinformatics. Algorithms need to be developed to improve data correlation and to allow for smarter and faster annotation workflows. Furthermore, the implementation of statistical evaluation methods commonly used in MS-based protein identification is expected to enhance the reliability of an MSI study.

Recently, a case was made for improved identification through community annotation[41]. An online data repository for published MSI data sets should allow for data re-mining. In this way, one can benefit from annotations by laboratories with

other expertise. This initiative exemplifies the need for improved annotation within the MSI community, but will face several challenges, not least standardization of data formats.

Databanks with patient data combined with biomaterials already exist to facilitate (bio)medical research. Although the set-up and use of these databases is governed by strict guidelines from medical ethics, we are convinced that MSI data integration with these 'biobanks', but also with imaging data from other imaging modalities and '-omics' data will greatly improve our capability to mine and annotate MSI data.

Recently, the technique of laserspray ionization (LSI) was applied to protein analysis directly from tissue sections[92]. Although this technique is not yet suitable for imaging experiments, the generation of multiply charged ions directly from tissue, combined with high mass resolution and mass accuracy, might facilitate protein imaging and identification in the future. The generation of multiply charged species in LSI makes possible electron transfer dissociation (ETD) fragmentation on tissue. Formation of *c*-ions and *z*-ions during ETD provides complementary fragmentation information and also suggests the possibility to study posttranslational modifications retained during ETD fragmentation but lost during the CID typically used in MSI.

For MSI to become a standard technique in (bio)medical research, it is of the utmost importance that MSI workflows for protein identification are further developed to provide useful information to the (bio)medical research community. As researchers extend MSI technology to study more complex biological problems, there will be an increasing need for (bioinformatic) tools that improve the confidence and number of identified proteins within these studies. Ongoing efforts to embed MSI into the interdisciplinary world of life sciences will move the field into the next decade.

3

The use of mass spectrometry imaging to predict treatment response of patient-derived xenograft models of triple-negative breast cancer

In recent years, mass spectrometry imaging (MSI) has emerged as a promising technique in oncology. However, the effective application of MSI is hampered by the complexity of the generated data. Bioinformatic approaches that reduce the complexity of these data are needed for the effective use in a (bio)medical setting. This holds especially for the analysis of tissue microarrays (TMAs), which consist of hundreds of small tissue cores.

Here we present an approach that combines MSI on tissue microarrays with Principal Component Analysis and Linear Discriminant Analysis (PCA-LDA) to predict treatment response. The feasibility of such an approach was evaluated on a set of patient-derived xenograft models of triple-negative breast cancer (TNBC). PCA-LDA was used to classify TNBC tumor tissue based on the proteomic information obtained with matrix-assisted laser desorption/ionization (MALDI) MSI from the TMA surface. Classifiers based on two different tissue microarrays from the same tumor models showed overall classification accuracies between 59 and 77%, as determined by cross-validation. Reproducibility tests revealed that the two models were similar. A clear effect of intratumor heterogeneity of the classification scores was observed. These results demonstrate that the analysis of MALDI-MSI data by PCA-LDA is a valuable approach for the classification of treatment response and tumor heterogeneity in breast cancer.

3.1 Introduction

In cancer treatment, there is a great need to develop tools that can predict response to treatment. Mass spectrometry imaging (MSI) is a powerful analytical technique that provides complex molecular information to meet this objective. In particular, matrix-assisted laser desorption/ionization (MALDI) MSI has shown its applicability to cancer research: it can probe intratumor heterogeneity[36, 37, 85], and it can be used for tissue classification[17, 45, 93, 94], disease prognosis and prediction of treatment response[95-97]. High-throughput analysis of clinical samples has been made possible by the establishment of protocols for the MALDI-MSI analysis of tissue microarrays (TMAs)[17, 18]. This has enabled the large-scale analysis of heterogeneous samples of limited quantity.

The analysis of TMAs with MSI easily generates thousands of spectra from hundreds of different tissue cores. In addition, each spectrum consists of hundreds of different molecular ions. Bioinformatic approaches that reduce this complexity are required to exploit the full potential of MSI. Up to now, a small number of studies have reported the use of PCA in combination with LDA or related statistical methods for the classification of MSI data[94, 98-100]. PCA is used as dimensionality and noise reduction method, followed by LDA to build a classification model. Efficient separation of tissue type based on lipid profiles has been shown[99-101]. Also disease-specific peptides and proteins could be identified in osteoarthritis and pancreatic cancer by this method[94, 98].

Here we report on the use of MALDI-MSI in combination with PCA-LDA to study the proteomic content of triple-negative breast cancer (TNBC) patient-derived xenograft (PDX) tumors. TNBCs account for approximately 15% of breast cancers[102]. TNBC is characterized by the lack of expression of the estrogen receptor, progesterone receptor, and the human epidermal growth factor receptor type 2 (HER2). Therefore, it is considered difficult to treat because no targeted treatment is available yet for this subtype of breast cancer, and resistance to conventional toxic chemotherapy frequently develops[103, 104].

Proteomic profiling of breast cancers has shown its usefulness for response prediction and the selection of more effective treatment strategies[95, 96]. In current practice, no reliable predictor of treatment response in TNBC before systemic treatment starts is available. The analysis of xenograft models by MALDI-MSI enabled us to study the proteomic content of these tumors under controlled conditions. For each tumor model multiple tissue cores were analyzed with MALDI-MSI and used to

predict response to the chemotherapeutic drug cisplatin. Here we determine the predictive strength of MALDI-MSI data for treatment response using PCA-LDA. We also establish to which extent heterogeneity between the tissue cores from a particular tumor model compromises response prediction.

3.2 Experimental section

Tissue microarrays

The triple-negative PDX models had been specifically generated to study the mechanisms involved in chemotherapy response and acquired resistance. In a separate study, the models were treated with cisplatin, as clinical studies have shown a good response of TNBC to this cytotoxic drug[105]. The measured initial response of the models was categorized based on tumor size (Figure 3.1): nine models responded well to the cisplatin treatment, resulting in a reduction of the tumor size ('good response'). Three models did not shrink or grow ('stable disease'), and seven showed reduced growth when compared to the control tumors ('progression'). Also, three models did not respond to the treatment ('no response').

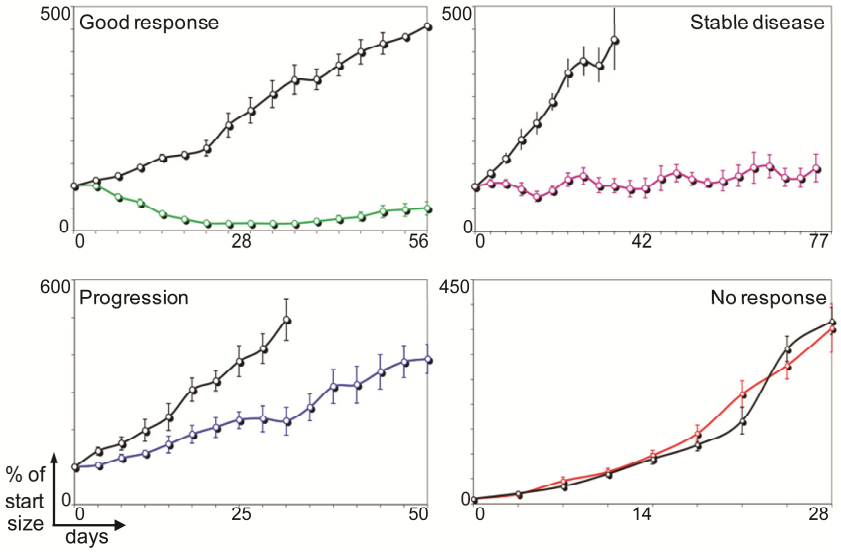

Figure 3.1 Response to cisplatin. Black lines show the growth of untreated controls. In all graphs, the y-axis depicts tumor size (percentage of start size), and the x-axis depicts the days after start of treatment. Representative examples for each response are shown.

Two TMAs (called hereafter TMA1 and TMA2) were constructed, both containing cores (triplicate 0.6 mm cores) from all 22 PDX breast cancer models. All tumor models were poorly differentiated (grade III) TNBCs as assessed by immunohisto-chemistry. The TMAs were constructed from treatment-naive samples. The TMA2 contained tissue cores from the same tissue blocks as TMA1, plus extra tissue cores from different tumors of the same PDX models. In this way, the reproducibility of the method and the tumor-to-tumor variation could be determined. Figure 3.2 shows hematoxylin and eosin (H&E) stained adjacent sections from the TMAs.

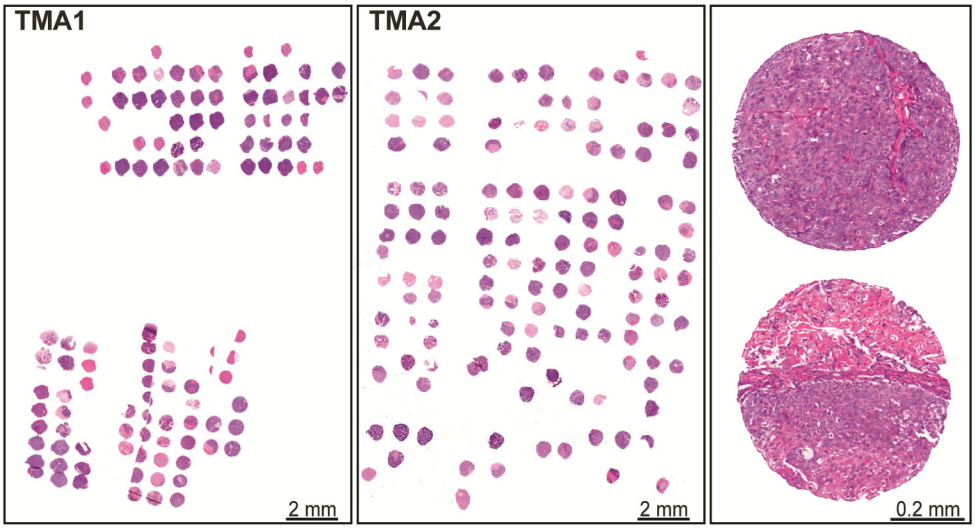

Figure 3.2 Optical images of H&E stained TMA sections. All cores on the TMAs are shown, including control cores and cores from non-TNBC PDX models, which were not included in this study. Right hand panel: a relatively homogeneous core with a tumor cell content of 90% (top) and a heterogeneous core of the same tumor containing stromal (pink) regions (bottom).

Tissue preparation

Serial 5 µm sections were cut from the TMA blocks and mounted onto indium tin oxide coated glass slides. The tissue cores were deparaffinized using xylene washes (100%, twice for 5 min) and rehydrated using graded ethanol washes (100% twice and 95%, 80% and 70%, all 5 min) followed by water washes (twice, 3 min) to make the TMAs amenable to MALDI-MSI analysis. Antigen retrieval was performed by heating the slides in a 10 mM Tris buffer (pH 9.0) at 95 ^{0}C for 20 min. The slides were

allowed to cool down to room temperature, rinsed with water and dried in a desiccator.

Local, on-tissue digestion was performed with trypsin, thereby preserving the spatial localization of the proteolytic peptides. A trypsin solution of 0.05 µg/µL was spotted in an automated manner (CHIP 1000, Shimadzu). A total of 5 nL was deposited per spot with a raster size of 200 x 200 µm. Trypsin spots measured approximately 100 µm in diameter. The sections were incubated overnight at 37 °C.

Finally, α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution was prepared at a concentration of 10 mg/mL in 50% acetonitrile (vol/vol) and 0.1% trifluoroacetic acid (vol/vol) in water and was sprayed onto the sections by a vibrational sprayer (ImagePrep, Bruker Daltonics).

MALDI-MSI experiments

MALDI-MSI analyses were performed using a MALDI quadrupole time-of-flight SYNAPT HDMS mass spectrometer (Waters Corporation). The mass spectrometer was operated in TOF mode optimized for positively charged ions. Data were acquired in the range of *m/z* 200-3500 at a raster size of 150 µm. On average 16 spatially resolved spectra were recorded for each core.

Data processing

Tissue core-specific spectra were extracted for data processing and subsequent statistical analysis. The spectra were subjected to peak detection using an in-house developed algorithm[106]. The ChemomeTricks toolbox for MATLAB was used for further pre-processing and analyses[106]. All spectra per core were averaged to create one representative spectrum per core. Averaged spectra per core were used to reduce the influence of outliers in the data and improve the signal-to-noise-ratio. This approach improved the stability of the multivariate analysis results. A similar observation was reported by Gerbig *et al.*[99]. Histological assessment of the tissue cores revealed that they were highly heterogeneous. Spectra were selected only from tissue regions with at least 80% tumor cells to reduce the variability caused by the presence of mostly stroma, but also some necrotic regions. On average 13 spectra were selected per tissue core and used for subsequent analyses. This selection was compared with the full data set (on average 16 spectra per core) to determine the influence of the introduction of additional variability.

Multivariate statistical approach

Data sets

We performed multivariate statistical analyses on the data sets of both TMAs (TMA1 and TMA2). The aim was to identify a proteomic signature that could differentiate between the tumor models that did respond to the cisplatin treatment (responders) and the models that did not or hardly respond to the treatment (non-responders). For this purpose, tumor models that had experimentally shown to have a 'good response' or 'stable disease' were categorized as responders. Tumor models that had shown 'no response' or 'progression' were categorized as non-responders.

Table 3.1 Samples used for construction of the classifiers based on TMA1 and TMA2.

Data set	Response class	No. of tumor models	No. of tissue cores
TMA1	Responder	12	70
	Non-responder	10	51
TMA2	Responder	12	113
	Non-responder	10	67

The average tissue core spectra were either assigned to the responder (n = 12) or non-responder (n = 10) class (Table 3.1). On average six and eight tissue cores per tumor model were present in TMA1 and TMA2, respectively. Measurements of consecutive TMA sections were used to build the classifier. The data analysis workflow is summarized in Figure 3.3.

Principal Component Analysis

The spectra were normalized to their total ion count and the mass intensities were standardized to zero-mean and unit variance prior to PCA. PCA was performed on the average tissue core spectra with 3524 variables (mass intensities) each. PCA performs a linear transformation of the data in the direction of the largest variance. It defines new variables consisting of linear combinations of the original ones, so-called Principal Components (PCs). The first purpose of PCA was to reduce the dimensionality of the data. Second, PCA was employed to discard noise. It is important to note here that by using PCA it is assumed that the differences between the treatment response classes are one of the main sources of variation in the data and are thus described by the PCs. Otherwise these differences are lost in the PCA-based data reduction.

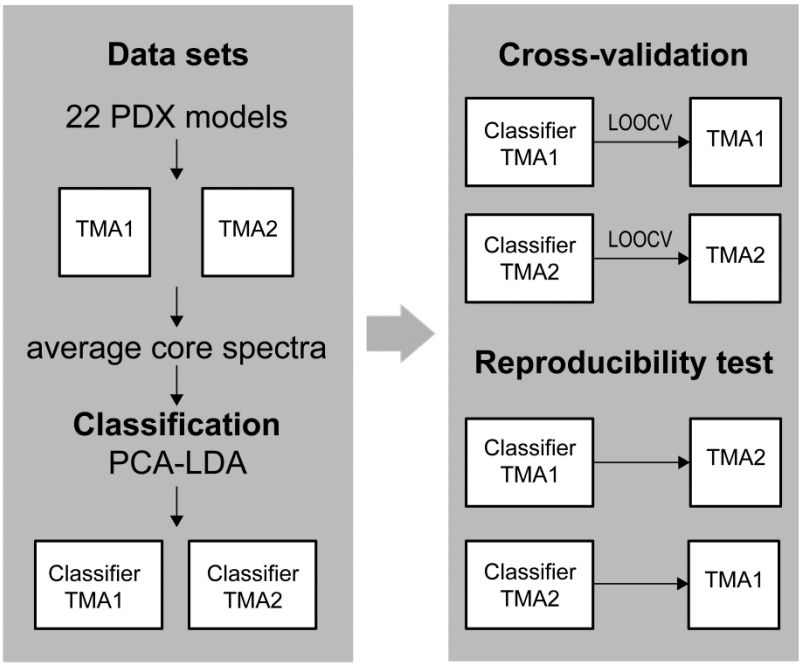

Figure 3.3 Summary of the data analysis workflow. The classifiers based on TMA1 and TMA2 are used to predict treatment response on the tumor, tissue core and pixel (single spectrum) level.

Linear Discriminant Analysis

Two individual classifiers were constructed based on TMA1 and TMA2, respectively. For this purpose, we used a two-step supervised classification method using a combination of PCA and LDA[107]. The PCs were used as input variables for the LDA. LDA calculates a linear combination of variables, in this case the PCs, that maximizes the ratio of the between-class variance and the within-class variance (Fisher's criterion). In other words, it finds the combination of PCs that leads to small discriminant score distances in LDA space within each class and large score distances between the classes. A tumor model is assigned to class *i* if the mean discriminant score of the tissue core spectra of this model is closest to the mean discriminant score of class *i*.

Double cross-validation

Classifiers built on highly dimensional data sets are prone to overfitting. It is, therefore, important to evaluate whether the classifier has been built with random fluctuations in the data or has predictive power. Classifiers are typically validated on independent samples. However, only 22 PDX models with tested initial treatment response were available. Instead, a leave-one-out cross-validation procedure was used to estimate the error-rates of the classifiers. Leave-one-out cross-validation is an accepted validation method when the size of the data set is small[108]. Before anything else, the number of input PCs for the PCA-LDA needed to be estimated. This estimation was incorporated in the cross-validation by using a double leave-one-out cross-validation procedure to avoid the introduction of bias, as previously described[109].

The double leave-one-out cross-validation was performed as follows: all spectra from one tumor model were set apart as test spectra. Next, the optimal number of PCs was determined based on leave-one-out cross-validation using the spectra from the remaining 21 tumor models. The number of PCs to use was chosen based on optimal classification performance of the classifier, using the least number of PCs. The optimal number of PCs was 34 for the classifier based on TMA1, and 25 for the classifier based on TMA2. Then, the separate test tumor model was classified using the number of PCs as determined independently from the test tumor model. This procedure was repeated for all tumor models. Combined, the total number of misclassified tumor models gave an estimate of the error rate of the classifier.

Reproducibility tests

The reproducibility of the method was evaluated by testing the classifiers using the alternate TMA. A classifier was trained on TMA1 and tested on TMA2 and vice versa. The same number of PCs was used as previously determined. The day-to-day variability was corrected using PCA-LDA as follows: each TMA data set was assigned a class. PCA-LDA was performed and the variance described by the resulting discriminant function was excluded from both data sets. The corrected data sets were used for the reproducibility tests.

3.3 Results and discussion

Predictions

The classifier based on TMA1 correctly predicted the treatment response for 17 out of 22 tumors, as determined by double cross-validation (Table 3.2). The classification model based on TMA2 had a classification accuracy of 13 out of 22 tumor models, of which 11 were also correctly classified based on the TMA1 classifier. The tissue core spectra classification scores for the cross-validation of TMA1 and TMA2 can be found in Figure 3.4. The correlation between the two models was quantified by calculating the correlation coefficients of the resulting loading plots and the average discriminant score per tumor model. The correlation coefficients were 0.73 and 0.82, respectively, indicating the similarity between the two models. Figure 3.5a shows the PCA-LDA loading plots of the TMA1 and TMA2 classifiers colored according to the extent of peptide peak contributions to the models. The images that depict the spatial distribution of exemplary peptide peaks with high loadings in both models are shown in Figure 3.5b.

Table 3.2 Classification results for the double cross-validations and the reproducibility tests using the alternate TMA as training set.

Data set		**Cross-validation**		**Reproducibility test**	
TMA1	Tumor model	17/22	(77%)	15/22	(68%)
	Tissue core	78/121	(64%)	76/121	(63%)
TMA2	Tumor model	13/22	(59%)	18/22	(82%)
	Tissue core	127/180	(71%)	125/180	(69%)

Duplicate measurements of consecutive TMA sections resulted in very similar classification scores for each tumor model in the cross-validation (data not shown). Thus, the difference between the performances of the classifiers might be due to the different composition of the tissue microarrays of TMA1 and TMA2; only 40% of the tissue cores of TMA2 originated from the same tissue pieces as the tissue cores of TMA1. The remaining cores originated from different tumor pieces of the same PDX models.

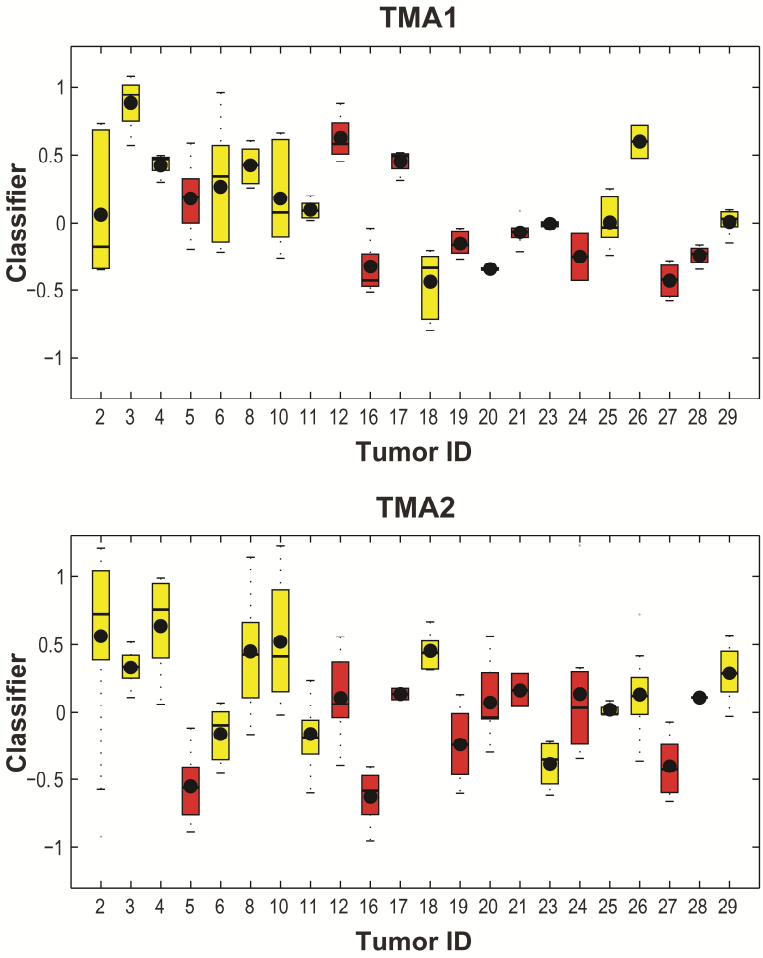

Figure 3.4 Box plots for the cross-validation of TMA1 and TMA2. The classification (discriminant function 1) scores are shown for the responder models (yellow) and the non-responder models (red). The classifier was constructed using the average core spectra from 21 tumors (training set), followed by classification of the spectra from the 22nd tumor (test set). Responder spectra are assigned positive values and non-responder spectra are assigned negative values in the classification models. The box plots represent the lower quartile, median (stripe), mean (dot) and upper quartile of the rescaled classification scores.

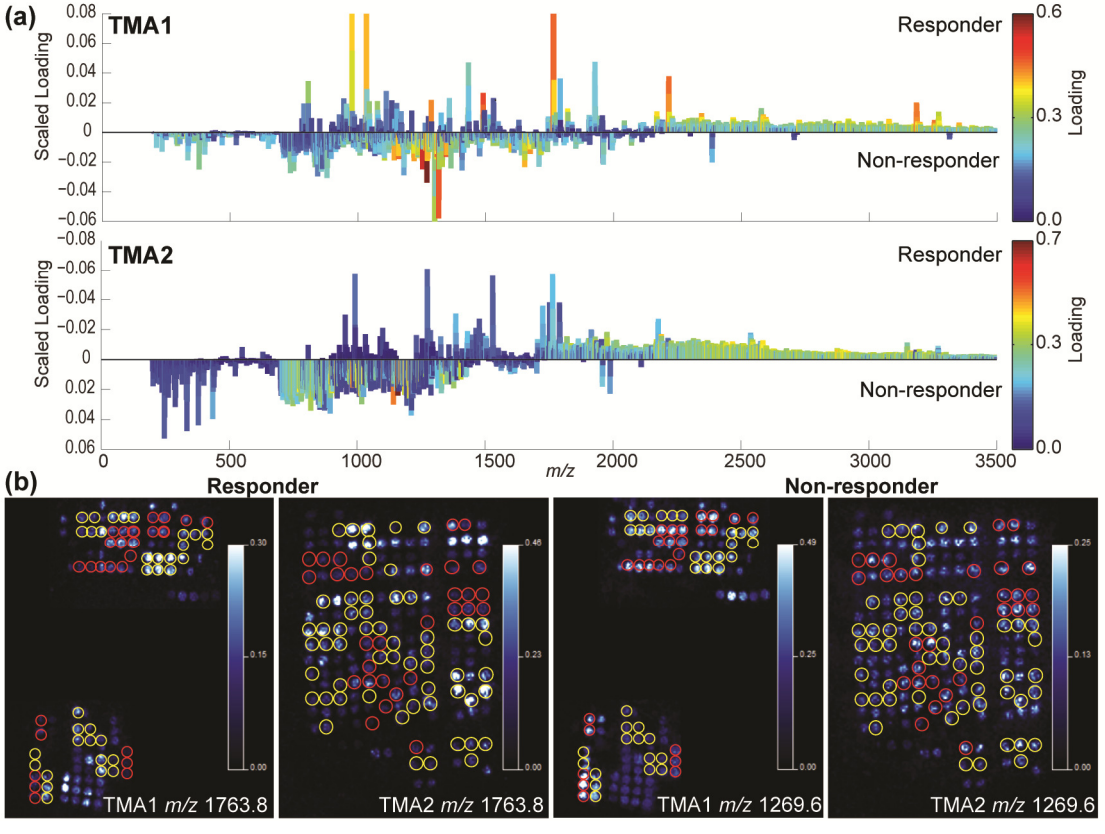

Figure 3.5 (a) PCA-LDA scaled loading plots of TMA1 and TMA2 colored as a function of variable (peptide peak) contribution in the projection. Scaling is applied by multiplying with the standard deviation of the original variables. (b) Selected ion images of variables with high loadings.

Projection of the predictions on the pixel level

The results can be visualized as a class image, because spatial information is retained and remains associated with the spectra. The classification score for each pixel is plotted using a color code. Figure 3.6b shows the classification images for TMA1 and TMA2. The observed similarity of the scores for each tissue core gives an indication of the stability of the classifier. Tumor cores with ambiguous classification scores on the core level typically exhibit a mixture of yellow and red pixels, indicative for the heterogeneity of the tissue. H&E stained adjacent sections from the entire TMAs can be found in Figure 3.2.

Comparison of the classifiers

On the tissue core level

The reproducibility of the method was evaluated by testing the classifiers, which were both based on all 22 PDX models, using TMA1 as training set and TMA2 for validation and vice versa. It should be noted here that the two TMAs were analyzed independently from each other at more than half a year interval. In addition, the tissue core layout of TMA2 was completely randomized as compared with TMA1 to avoid bias introduced by the position of the tissue cores in the TMA. The reproducibility tests revealed 4 wrong predictions out of 22 for the TMA1 classifier and 7 out of 22 for the TMA2 classifier (Table 3.2). These results can be explained by the performance of the classifiers: 8 out of the 11 misclassifications in the reproducibility tests were also misclassified in the cross-validations. This means that these tumor models are not well described by the classifiers. It should be noted that the tumor models were assigned to two discrete classes only for the purpose of classification. It was expected that the intermediate response PDX models would be more difficult to classify than the ‘good response’ or ‘no response’ PDX models. Unsurprisingly, the four most often misclassified tumor models were models with an intermediate experimental response, that is, ‘progression’ or ‘stable disease’. Figure 3.7 shows the number of misclassifications per tumor model.

One would expect that variation between different tumors from the same PDX model might result in a poorer performance of the classifiers. However, a similar percentage of misclassifications was found for duplicate cores from the same tumor piece as for cores from different tumors of the same PDX model.

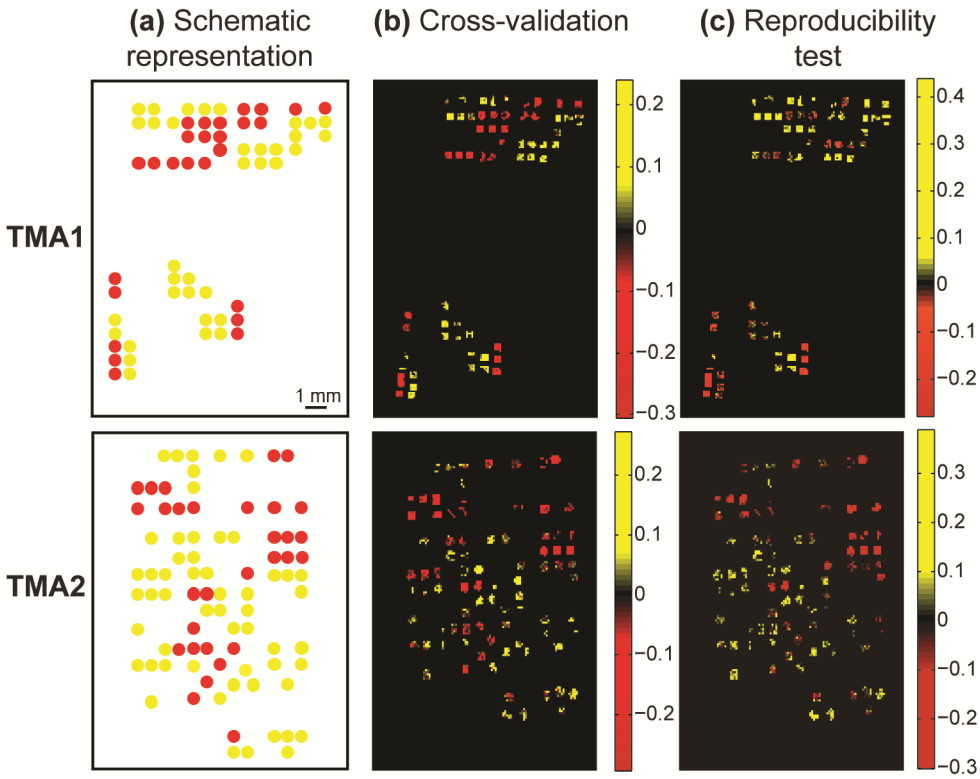

Figure 3.6 Classification results for TMA1 and TMA2 on the pixel level. The classification score for each spectrum is projected on its pixel location. (a) Schematic representation of the TMAs with the responders (yellow) and non-responders (red). (b) Cross-validation results. A clear difference between the responders and non-responders is observed. (c) Reproducibility test results for TMA1 and TMA2 on the pixel level. The TMA1 classifier was used to test the spectra from TMA2 and vice versa. Overall agreement between the schematic representation of the tissue microarrays and the reproducibility test results is observed. Also mixed color tissue cores are present. These cores show a heterogeneous classification: responder (yellow) pixels and non-responder (red) pixels are observed within one core.

A duplicate tissue microarray for TMA1 was measured, at a year interval, to further test the reproducibility of the method. This new data set is predicted, using the classifier based on TMA1, with three misclassifications (8 of the 56 cores were misclassified). In line with our previous findings, those three tumor models were already determined to not fit well in the classification model based on TMA1. Proteomic differences between the duplicate tissue microarray and TMA1 cannot be excluded because the duplicate originated from another part of the TMA block.

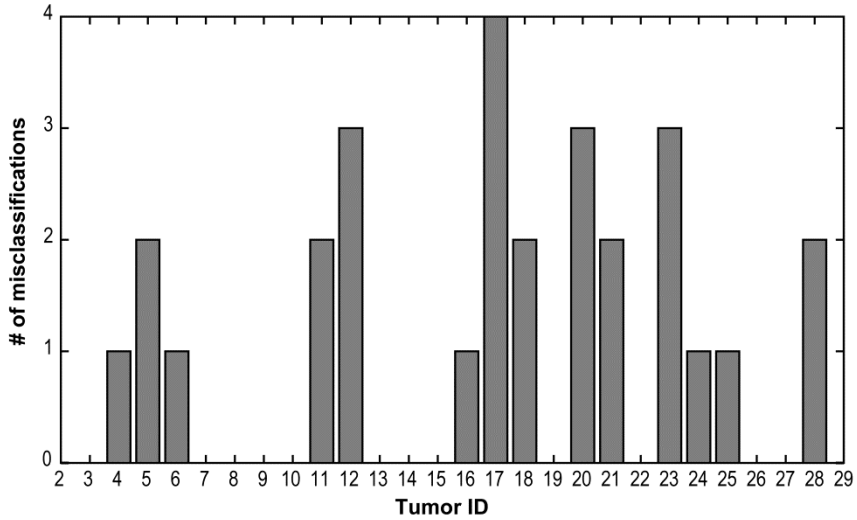

Figure 3.7 The total number of misclassifications per PDX model. For each PDX model, the misclassifications of the cross-validations and the reproducibility tests are summed. The four most often misclassified models (no. 12, 17, 20 and 23) had all shown an intermediate response phenotype in the treatment response experiments.

On the pixel level

Figure 3.6c shows the classification maps for the reproducibility tests. The reproducibility test predictions are overall in agreement with the treatment responses. Similar predictions on the pixel level are observed for most tissue cores. However, also some heterogeneous predictions for single tissue cores are present, represented by mixed colors in the class images.

The extent of the observed heterogeneity in predictions was quantified as follows: for each PDX model the percentage of correctly classified pixels was determined, and this percentage was averaged over all data sets. Overall, 11 models show limited variation in classification score, defined as >70% or <30% correctly classified pixels. The other models show a larger spread in classification scores. MALDI-MSI spectra are known to exhibit variability due to technical issues, for example noise, the probing of mixtures of cells at the used spatial resolution and matrix preparation effects[110]; therefore improved reliability of classification can be obtained at the core or tumor level. A higher percentage of tissue cores with mixed prediction scores are present in TMA2 than in TMA1, reflecting the poorer separation achieved by the classifier based on TMA2. One would expect, from a histological perspective, that biomolecular heterogeneity gives rise to spectra with high variance. This heterogeneity might

contribute to the observed distribution of classification scores. The classifiers could be tested on larger, homogeneous tissue sections to determine to which extent the spread in classification scores is caused by real tissue heterogeneity and not by technical variability.

In general, technical and biological variability between experiments limited the classification accuracies that we could obtain, especially because the study was based on a small number of tumors. The small sample set was a reason for using LDA instead of a more complicated model with multiple free parameters. For example, genetic algorithms would not be very useful, because they are prone to overfitting. An advantage of PCA-LDA is the possibility to evaluate the contribution of individual variables to the model. One would expect that different tumor classes have different peptide profiles that show up in the measurements. A follow-up study with clinical samples is required to externally validate these results. Moreover, the statistical approach might benefit from improved feature extraction. It is assumed in PCA-based feature extraction, as previously mentioned, that treatment response is one of the main sources of variation in the data. However, variance due to small differences between the treatment response classes might be poorly described by the PCs. These subtle changes, as for example differences in proteomic content in this study, can provide valuable information.

Taking into account tissue heterogeneity

All results shown so far have been obtained with the spectra from regions with high tumor cell content. However, it is generally accepted that the tumor microenvironment has an impact on treatment response[111]. New classifiers were built based on TMA1 and TMA2 using all spectra per core. Highly similar classification accuracies were obtained as compared to the accuracies previously reported. The new classifiers based on TMA1 and TMA2 correctly predicted the treatment response for 17 out of 22 and 14 out of 22 tumors, respectively. Interestingly, the reproducibility was increased. In particular, the classifier based on TMA2 showed a higher reproducibility (~20% higher). Histological analysis showed that the tumor cell content was on average the same in both treatment response classes (data not shown); therefore an effect caused by an uneven distribution of tumor tissue can be excluded. Although it is difficult to find a biological interpretation of this result due to the 'black box' nature of the experiment, it is clear that the included heterogeneity has a positive impact on the classification in this study.

3.4 Conclusions

In this chapter an approach is presented for the prediction of treatment response of PDX models of TNBC on the basis of MALDI-MSI data of TMAs. The results show its potential as a tool to study and predict treatment response in a high-throughput way; hundreds of cores can be analyzed in a single measurement. In addition, the method described here permits the classification of treatment response with direct correlation to histologically defined regions of interest. We have described how multiple tumors from the same PDX model could be used to assess the reproducibility of the method, showing both technical and biological variability. Further development of multivariate statistical approaches will bring MSI closer to clinical application.

4

Predicting head and neck cancer metastasis and disease-specific survival from MALDI-MSI data: feasible or not?

The presence of lymph node metastasis is a major predictor for prognosis in head and neck cancer. However, with current diagnostic imaging techniques around 30% of lymph node metastases are not detected. Here, we investigate whether lymph node metastasis and disease-specific survival can be predicted from matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) data. We perform MALDI-MSI measurements on tissue microarrays (TMAs) that contain 240 primary tumor samples.

We describe how MALDI-MSI spectra have been acquired, and discuss how this data is processed so that it can be used in five different classifiers. We do not observe any predictive power for lymph node metastasis; for disease-specific survival some of the classifiers show a small predictive power. We demonstrate that our method is sensitive to intensity differences of 50% in typical peptide peaks. The observed broad peak distributions make it difficult to detect differences between the classes.

With MALDI-MSI a large number of biomolecules can be studied simultaneously from hundreds of samples, with spatially resolved profiles. This makes MALDI-MSI a potentially attractive technique for biomarker discovery. There are drawbacks however: when a large number of biomolecules is measured, it can be hard to discern signal from noise. Spectra can also be hard to reproduce. For the formalin-fixed paraffin-embedded (FFPE) tissue that was used in this study, these drawbacks hampered biomarker discovery - we expected to find only small differences between the classes, in line with previous gene expression studies. Future efforts to improve the spectral quality of MALDI-MSI on FFPE tissue would therefore be valuable.

4.1 Introduction

For the treatment of cancer an accurate prediction of the course of disease can be used to provide tailored treatment. However, in many cases no means exist to predict disease progression, or the existing methods lack sensitivity. Biomarker discovery aims to find measurable indicators of disease state. These indicators can then be used to assist in diagnosis, predict disease progression or aid in treatment decisions. With matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI), biomolecules such as proteins, peptides, lipids and metabolites can be studied directly from tissue sections. Up to thousands of biomolecular species can be studied simultaneously. Tissues with different disease states can be compared to find differences in the expression of biomolecules. In this way, biomarkers or biomarker patterns might be identified that are associated with specific disease states.

For the discovery of biomarkers, a study should fulfill several requirements[26, 112]. First of all, data on the clinical parameter of interest, such as the survival status of the patient, should be available for each patient sample. Secondly, a sample set is needed that is well-matched, which means that the classes or groups should contain a similar distribution of patients. For example, an unequal distribution of males and females could lead to the discovery of `false' biomarkers that are based on the gender of the patient. Furthermore, the sample set should also be sufficiently large to account for expected biological variability.

In recent years it has become possible to study so-called tissue microarray (TMA) samples with MALDI-MSI. These samples consist of arrays of small tissue pieces from different patients. Using TMAs, one can measure a large sample set under highly similar experimental conditions, and correlate the acquired data with clinical data. These properties make TMAs well suited for biomarker discovery studies. Up to a thousand patient samples per study can now be analyzed in a single experiment[113].

There have been several studies that correlate MALDI-MSI data with disease progression[78, 114-119]. In [113, 120], TMAs with samples from more than 100 patients were used. A variety of statistical methods is employed in these studies. Typically, the first step is to select features (*m/z* values) using for example a univariate statistical test or Principal Component Analysis (PCA). There are several ways to investigate the predictive power of the selected features. Perhaps the simplest way is to correlate individual features to the parameter of interest[113, 120]. It is also possible to combine features to increase their predictive power, for example by Hierarchical Clustering[114,

[116, 119]. Alternatively, a classifier can be trained to predict the parameter of interest from the features[114, 115, 118]. Regardless of the approach chosen, the number of features associated with a prognostic parameter is typically less than 20. There is always the danger of *overfitting* in these high-dimensional data sets. Techniques such as cross-validation can reduce this, and in some studies independent validation of the identified proteins (typically not all discriminatory *m/z* values) is performed by immunohistochemistry.

Head and neck cancer is the world's sixth most common cancer. The tumors are biologically highly heterogeneous. Despite advances in diagnostics and treatment strategies, survival rates have not improved over the last decades and remain poor, with a 5-year survival of approximately 50%[121]. The presence of lymph node metastasis at the time of diagnosis is a major predictor for prognosis. Unfortunately, current diagnostic imaging techniques lack sensitivity. In around 30% of patients existing lymph node metastases are not detected. Depending on the chosen treatment strategy, this results in a large number of patients receiving over- or undertreatment. Tumor profiling with biomarkers has shown promising results[122, 123].

A recently validated gene expression profile accurately predicted the absence of nodal metastasis in 89% of the patients. However, the use of this gene profile would lead to a large number of patients undergoing unneccesary treatment[124]. Also MS-based proteomic methods have been applied to identify markers associated with tumor aggressiveness and metastasis in oral cancer. Polachini *et al*. found 155 differentially expressed proteins favoring metastasis[125]. More recently, a proteomic analysis by Harris *et al*. revealed 72 peptide features associated with disease-specific death, metastasis and recurrence[126]. However, to our knowledge no proteomic profile with predictive capability has been generated yet.

Here, we investigate whether disease-specific survival and lymph node metastasis can be predicted from MALDI-MSI data of head and neck cancer tumors. First, the sample preparation and measurement technique are discussed, then we describe the data processing and classification approach. After the results have been presented, we discuss some of the challenges in biomarker discovery with MALDI-MSI.

4.2 Methods

Patient samples

For this study, 240 patient samples were available: 212 cases of oral squamous cell carcinoma (OSCC) and 28 cases of oropharyngeal squamous cell carcinoma (OPSCC). These samples came from patients with histologically proven oral or oropharyngeal squamous cell carcinoma that underwent surgery between 1996 and 2005 at the University Medical Center Utrecht in The Netherlands. Patients diagnosed with synchronous primary tumors or previous malignancies in the head and neck region were not included in this cohort[127].

Table 4.1 The 178 patient samples used in this study were assigned to two binary classes, one for lymph node metastasis and one for disease-specific survival.

Class description	Yes	No
Lymph node metastasis	75	103
Disease-specific survival	115	63

In our study, we use two TMAs that contain tissue cores excised from patient samples. Per patient, three cores (0.6 mm in diameter) of the central part of the primary tumor were present, see Figure 4.1b. In Table 4.1, the number of patients which showed lymph node metastasis or disease-specific survival is listed. Negative disease-specific survival implies recurrence or death due to disease within the 5-year follow-up period, which occurred in 63 of the 178 patients. Note that not all 240 patients are included in this table, because some of them are filtered out during the pre-processing of the data.

Sample preparation

The TMAs contain formalin-fixed paraffin-embedded (FFPE) tissue. This tissue is conserved by dehydration and cross-linking of the proteins with formalin. After formalin fixation, the tissue is embedded in paraffin to preserve tissue morphology and allow thin sectioning of the tissue. FFPE tissue is widely used for clinical applications, due to easy storage and handling. Most MALDI-MSI studies so far have used fresh frozen tissue instead of FFPE tissue.

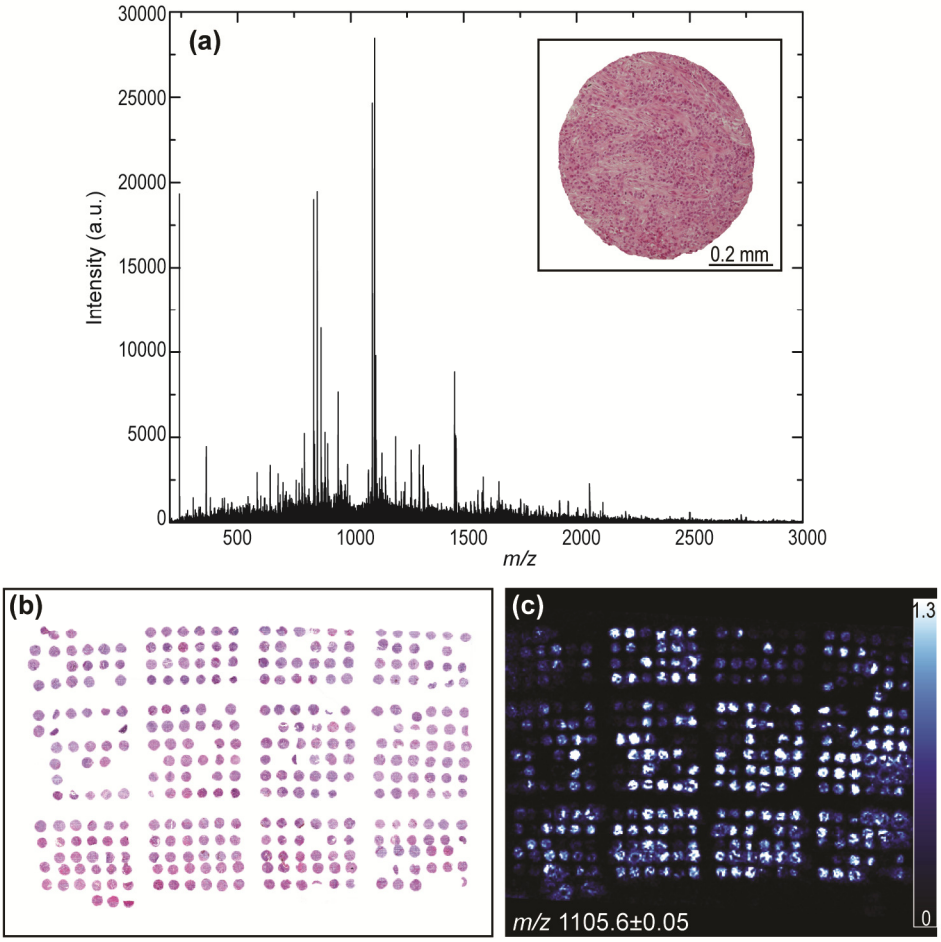

Figure 4.1 Example of a head and neck cancer TMA (a) Unprocessed MALDI-MSI spectrum from a tissue core. Inset: Hematoxylin and eosin (H&E) stained tissue core with 80% tumor cells. (b) H&E stained TMA of 120 patients. (c) Selected ion image shows the distribution of a typical peptide peak.

The TMAs were prepared for MALDI-MSI analysis as previously described[18]. We summarize the important steps below. First, serial 5 µm tick tissue sections were cut from the TMA blocks and mounted onto conductive indium tin oxide (ITO) coated glass slides (Delta Technologies, USA). Paraffin was removed using xylene washes (twice, 5 min). Paraffin needs to be removed, as it causes ion suppression during mass spectrometric analysis.

The goal of the next steps in the sample preparation is to make the proteins amenable to MSI analysis. First, rehydration is performed using graded ethanol washes (100%, 100%, 95%, 75% and 30%, all 5 min) and water washes (twice, 3 min). Then, (part of) the cross-linking is reversed (most likely through heat-induced hydrolysis[128]) by incubation of the sample in a buffer at high temperature. The samples were incubated at 95 °C in a 10 mM Tris buffer pH 9 for 20 minutes, and allowed to cool down to room temperature before briefly rinsing them with water. Afterwards, the samples were dried in a desiccator.

Trypsin was dissolved in 50 mM ammoniumbicarbonate plus 25 μM octylglucoside at a final concentration of 0.05 μg/μL. On-tissue digestion was performed using the Suncollect automatic sprayer (SunChrom, Germany). Eight layers were applied with a flow rate of 7.5 μL/min. The quick movement of the localized spray over the tissue ensures that the spatial information is retained. The samples were incubated at 37 °C overnight in a humid environment (50% methanol in deionized water). Proteins in FFPE tissue are typically digested with trypsin to free them from remaining cross-links and aggregation which hinder their detection.

A matrix solution of 5 mg/mL alpha-cyano-4-hydroxycinnamic acid (CHCA) in 1:1 ACN:H_2O with 0.1% trifluoroacetic acid was used. Matrix was applied with the Suncollect sprayer. Eight successive coats were applied with an increasing flow rate of 7.5-20 μL/min.

MALDI-MSI measurements

The experiments were performed with a MALDI-QTOF instrument (Synapt G2Si HDMS, Waters, UK) in positive mode and an *m/z* range of 200-3500. Spectra were acquired with a stage step size of 100 μm and a laser frequency of 1000 Hz, with the quadrupole set to have optimal transmission in the peptide region of the mass spectrum (above *m/z* 900 for 90% of the scan time). The instrument was operated in sensitivity mode during all experiments, and on average 34 spatially resolved spectra were recorded for each core.

MALDI-MSI spectra of FFPE tissue consist mainly of tryptic peptide and matrix peaks. An exemplary unprocessed mass spectrum can be found in Figure 4.1. MALDI-MSI studies of TMAs report the detection of around 500 peptide peaks[94, 129]. This is in line with our observation of on average 700 tissue-related peaks per patient in the range of *m/z* 700-3500.

As was previously observed for FFPE tissue[54], the baseline of the spectrum has a hill shape. This feature becomes more pronounced after peak-picking of the spectra, see the next section and Figure 4.2a. The elevated baseline is most likely caused by the rich mixture of molecules that is desorbed and ionized from the digested tissue surface. Formalin fixation-induced adducts might further increase the number of different ionized species[130]. Unresolved peaks lead to a loss in spectral resolution and a decrease in signal-to-noise ratio. A shorter digestion time of two hours did not improve the quality of the spectra. Replacing the Tris buffer (pH 9) with an acidic buffer (10 mM citric acid at pH 6)[65], did also not improve the signal-to-noise ratio.

The samples were collected over a period of ten years. When comparing older with younger samples, there was generally a decrease of peak intensity in the mass region above *m/z* 1300 and an increase in peak intensity below *m/z* 1300. Because the samples were ordered by their age on the TMAs, we cannot exclude that this difference was caused by the sample preparation or a slight bias in the measurements. Compared to the spectral variability of the samples, the effect was relatively minor.

Pre-processing of the spectra

After MALDI-MSI analysis, the samples were stained with hematoxylin and eosin (H&E) using a standard protocol, see Figure 4.1. The percentage of tumor cells per tissue core was determined by a dedicated head and neck pathologist. To reduce the biological variability in our data, cores with a high percentage of tumor stroma or muscle tissue were excluded. Only tissue cores with more than 50% tumor cells were used. Furthermore, low intensity tissue cores (less than 30% of average core intensity) were also excluded. After this procedure there were still spectra from 178 patients, with an average of 80 spectra per patient. Clinical follow-up data were available for all patients (see Appendix).

Spectra from the cores were extracted for data processing and subsequent statistical analysis using an in-house developed software tool. This tool co-registers the MSI data and the H&E scan of the same sample to accurately extract 'on-tissue' spectra. The extracted spectra were subjected to peak detection using the PEAPI algorithm[106].

Classification procedure

Below, our approach for tumor classification is described. First, the spectra that we have obtained (see the previous section) are further processed. Then a `feature selection' step is performed, because the classifiers that we use cannot operate on high-dimensional data. These classifiers are briefly described, as well as the (standard) cross-validation procedure. For the data processing and classification, we have used routines from SciPy[131] and Scikit-learn[132]. The numerical infrastructure provided by these Python libraries allowed us to quickly test different feature selection methods and classifiers.

Processing the spectra

To improve the sensitivity of our classification approach, several operations are performed on the measured spectra:

1) Combine the measurements per patient (by summation)
2) Remove matrix-related peaks
3) Remove the baseline per patient
4) Normalize the spectra

The goal of these steps, which are discussed below, is to reduce the variance that is not caused by protein-differences.

Combining measurements

For each patient, we have performed measurements on all available tumor cores (maximum three), and for each core, spectra were obtained at multiple locations or *pixels*. After peak-picking, we combine the measurements per patient by summation. In Figure 4.2a, an example of such a spectrum is shown.

Removing matrix-related peaks

The m/z values that we remove are indicated by red dots in Figure 4.2a. All peaks below $m/z = 710$ are filtered out, because most of them do not correspond to peptides. Peaks related to the matrix are located at the bottom of the spectrum. To remove them, we apply the following procedure twice:

1) Compute the averaged spectrum over all the patients $\bar{S}$
2) Apply a second order Savitzky-Golay filter with a width of 125 *m/z* values to $\bar{S}$ to get $\bar{S}_f$
3) Remove peaks for which $\bar{S} < 0.75\ \bar{S}_f$

Baseline removal

The spectra have different baselines, as can be seen in Figure 4.2b. We perform baseline correction in the following way. For each patient, we first estimate the baseline with a median filter. Such a filter computes the median over a window of width *w*; we use *w* = 125 *m/z* values. This estimate of the baseline is still rather noisy, so we smooth it with a Savitzky-Golay filter of order two and the same width. The smoothed baseline is then subtracted from the data, see Figure 4.3a. Note that some of the peaks are located below the baseline, and will therefore become negative.

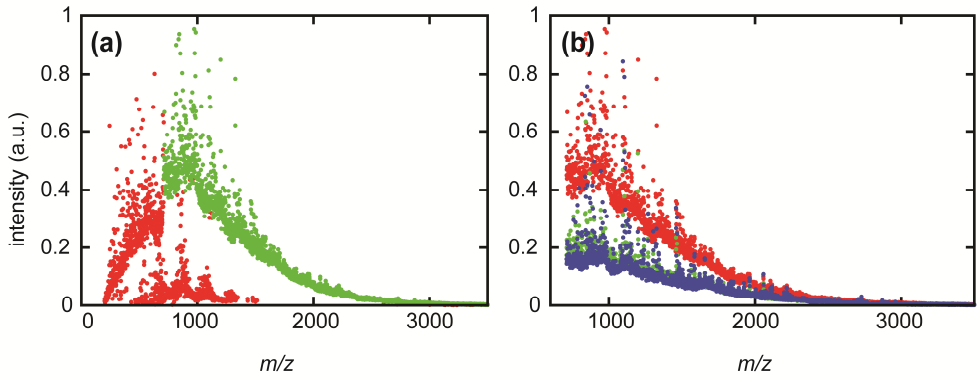

Figure 4.2 (a) Example of a peak-picked spectrum, obtained by combining data from all tumor cores of the same patient. The *m/z* values indicated by red dots are filtered out during the pre-processing. (b) Three filtered spectra from different patients.

Normalizing the total ion count

There are still significant differences in the total intensity of the spectra at this point. As these differences are mostly related to sample preparation and experimental conditions, they are filtered out by normalizing the spectra by their total ion count. In Figure 4.3b, examples of the resulting spectra are shown.

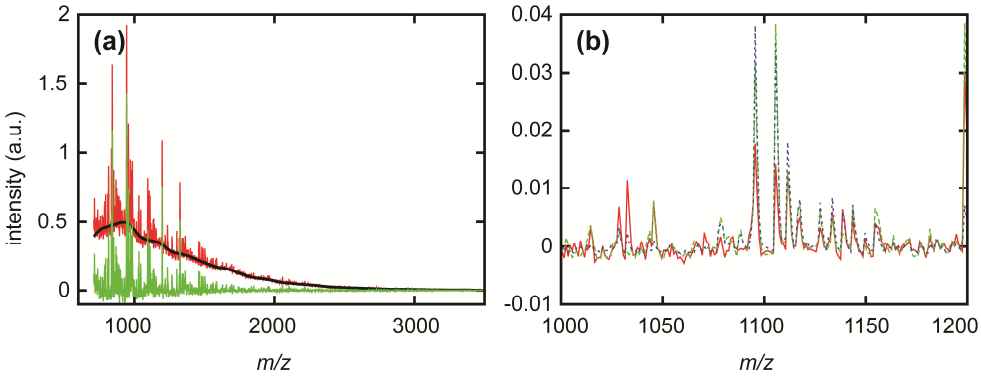

Figure 4.3 (a) Example of a peak-picked spectrum with its baseline (black solid curve). The spectrum after baseline subtraction is also shown (green spectrum). (b) Peak-picked spectra of three patients, after pre-processing (zoom). Peak-picked spectra are shown as continuous graphs for better readability.

Feature selection

Most classification methods can only be trained on data for which $D < N$: the dimension of an observation should be (significantly) smaller than number of observations. We therefore have to reduce the dimensionality of our data. There are several ways to do this, see for example the discussion by Hilario *et al.*[22].

We use univariate feature selection, which means that the features are ranked individually, without taking into account the possibly complex relations between them. Features that are discriminatory only in combination with other features are therefore unlikely to be selected, which makes univariate feature selection unsuitable for detecting patterns involving small changes in a large number of features. However, here we assume that class differences are expressed in a limited number of features. Even if these features are correlated, there will often be significant univariate differences. After such differences have been detected, one can use a suitable classifier to search for complex patterns in the selected features. Another argument for using a univariate approach is that with more multivariate methods, *overfitting* is more likely to occur. The reason for this is that the number of possible patterns rapidly increases with the complexity of the pattern. For example, in a data set with 500 features, there are already about 2.6×10^{11} ways to select five of them. The problem of overfitting is especially relevant because we have a limited sample size and data with large intrinsic variability.

Given a set of training data for two classes, the features (peaks) are ranked according to two tests: the Wilcoxon rank-sum test and the Kolmogorov-Smirnov test. We then retain the features with the smallest *p*-values. These *p*-values indicate the probability of getting a difference at least as big as the observed difference, assuming that the *null hypothesis* is true. The Wilcoxon rank-sum test is closely related to the AUC (area-under-curve) statistic for ROC (receiver operator characteristic) curves. Its null-hypothesis is that $P(x > y) = P(y > x) = 1/2$ for samples x and y taken from two classes. The Kolmogorov-Smirnov test is based on the distance between the estimated cumulative distribution functions of x and y. Its null hypothesis is that these distribution functions are the same. This makes the Kolmogorov-Smirnov test one of the most generally applicable statistical tests.

One of the classifiers that we will use is a Support Vector Machine (SVM). For SVMs, normalization of the data is important, and therefore the selected features are normalized to the range [0, 1] when they are used with the SVM.

Classifiers used

We use five different classifiers, as implemented in the Scikit-Learn library[132]:

- LDA: Linear Discriminant Analysis
- QDA: Quadratic Discriminant Analysis
- NBC: Naive Bayes Classifier
- DTC: Decision Tree Classifier, we use a maximum depth of three
- SVM: Support Vector Machine, the parameters C and γ were both set to one

These classifiers were selected for a number of reasons. LDA is one of the canonical classifiers, which should work well for linearly separable data. QDA is a bit more general and allows for a `quadratic' decision boundary between classes. The Naive Bayes Classifier works well for data in which each single feature independently has (some) predictive power, whereas the Decision Tree classifier can handle complex relations between the features. Support Vector Machines are also quite flexible, and can be used to find complicated decision boundaries between the classes.

Cross-validation procedure

To test the performance of the different classifiers, we use so-called k-fold cross-validation, with $k = 10$. This means that the data is randomly partitioned into k sub-

samples of nearly equal size. Then the training and testing of the classifiers is performed k times. Each time one subsample is used for testing and the rest of the data for feature selection and training. Data from a patient is thus either used for testing or for training, but never for both. After cross-validation, each patient has been part of the test-group exactly once, so a full set of predicted class labels is obtained.

4.3 Results

To investigate whether we can predict the occurrence of lymph node metastasis or disease-specific survival from MALDI-MSI spectra, the patients were assigned to binary classes. The class counts can be found in Table 4.1.

After the cross-validation procedure, the predicted classes are compared with the actual classes. Then a *confusion matrix M* can be constructed, indicating the number of true positives (*TP*), true negatives (*TN*), false positives (*FP*) and false negatives (*FN*):

$$M = \begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix}$$

We express the accuracy of the classifiers as a single number, for which we use Matthews correlation coefficient (MCC), also known as the φ coefficient:

$$MCC = \frac{TP \text{ x } TN - FP \text{ x } FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

This coefficient lies between -1 and 1. A value of 1 indicates that all predictions are correct, 0 means that there is no predictive power, and -1 means that all predictions are wrong. If there are for example two equal size classes, for both of which 75% of the predictions is correct, then we have MCC = 0.5.

Results with artificially modified data

To get an idea of the sensitivity of the classification approach, we have performed tests with modified data. One representative peak with an *m/z* value of 1325.7 was selected. This peak was increased with a certain percentage in one of the classes *after* pre-processing, but before feature selection (see 'The complexity of discriminatory

patterns'). (If we would keep the baseline before increasing the peaks, then a much smaller increase would be needed.)

Classification results for different increases are shown in Figure 4.4a, using the LDA classifier and five selected features. The other classifiers give quite similar results, which are therefore not shown. Data from 20 runs with randomized class labels was used to generate this figure. The feature selection methods perform about the same. With an increase of about 50%, the MCC score is about 0.2 to 0.3, and it increases to about 0.4 for an increase of 80%. Using the method and data presented in this chapter, we should thus be able to detect differences of about 50% in the intensity of typical peaks.

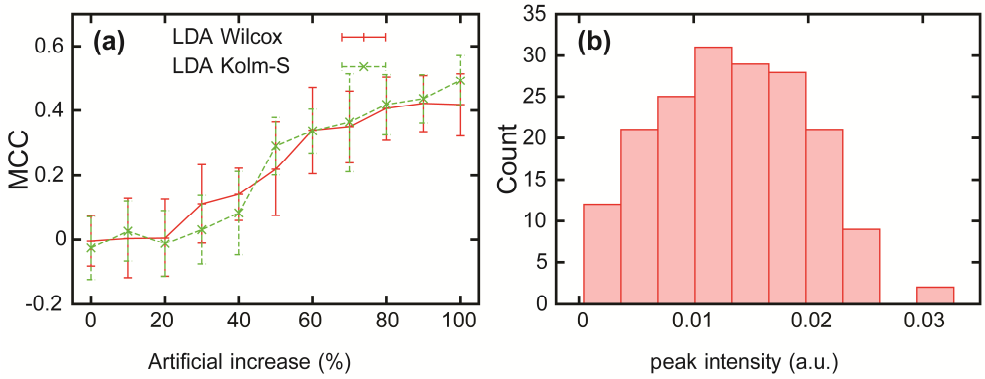

Figure 4.4 (a) Test of the feature selection methods (Wilcoxon rank-sum test and Kolmogorov-Smirnov test) using *artificial* data. A typical peptide peak at *m/z* 1325.7 has been increased by a certain percentage in one of the classes. The error bars indicate plus and minus one standard deviation. Data was collected from 20 runs with randomized class labels. (b) Histogram of the peptide peak at *m/z* 1325.7 after processing.

These tests with modified data give an indication of the quality of the spectra. Suppose that an *m/z* value has a different distribution in classes A and B. How well we are able to discriminate between A and B depends on the overlap between the distributions. For a narrow distribution, a small shift in the mean can already result in a small overlap, but for a broader distribution a much larger shift is required. In Figure 4.4b, the intensity distribution for the *m/z* value of 1325.7 is shown. This is clearly a quite broad distribution, which explains why we have to increase the intensity of this peak with 80% to obtain an MCC score of 0.4. The broad distribution makes it harder to detect differences between the classes. Ideally, peaks would have a

much narrower distribution. A more quantitative way of saying this is that peaks should have a small relative standard deviation (standard deviation over mean), of course *after* baseline removal.

The error bars in Figure 4.4a indicate the standard deviation in the MCC score using randomized class labels. When there is no increase (0%), this standard deviation is about 0.1. This means that a classifier on the real data should show MCC scores significantly larger than 0.1 in order to be successful.

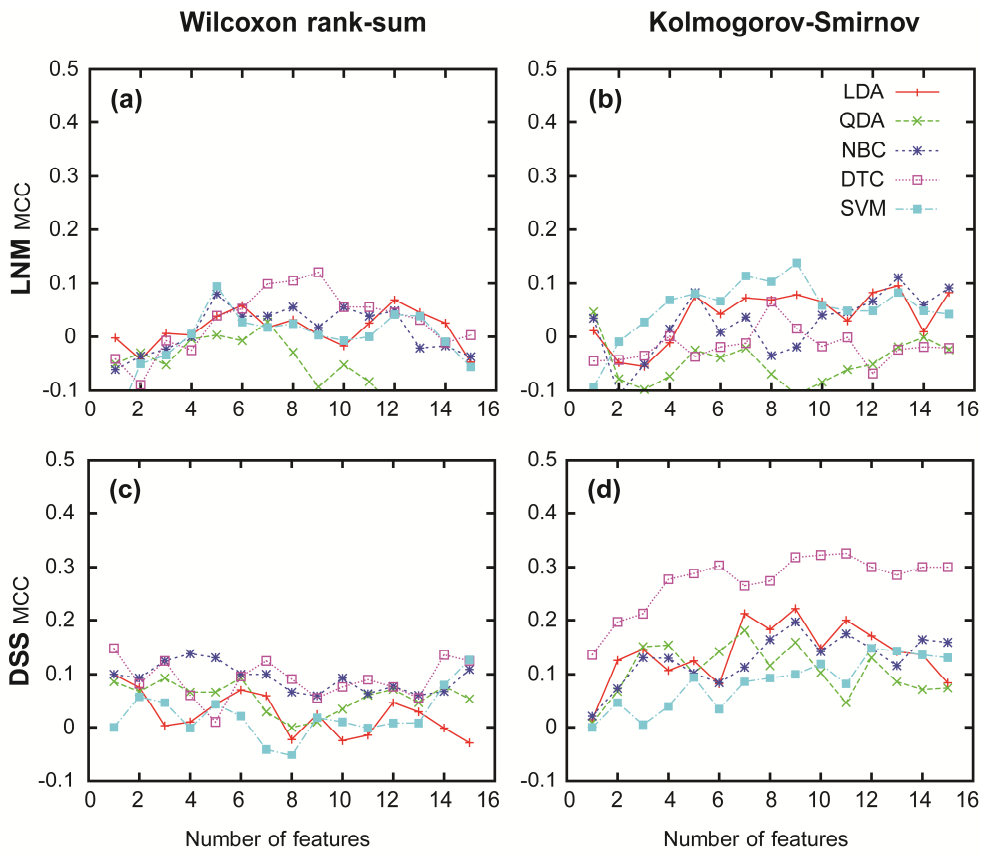

Figure 4.5 Classification results for lymph node metastasis (LNM, a,b) and disease-specific survival (DSS, c,d). The left hand figures show results for feature selection with the Wilcoxon rank-sum test. The right hand figures show results using the Kolmogorov-Smirnov test.

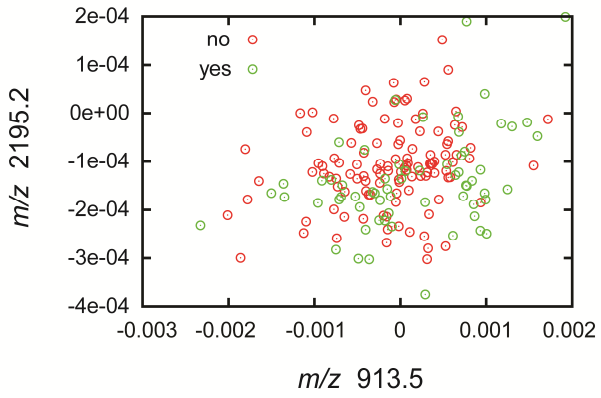

Figure 4.6 The classes visualized according to two discriminatory *m/z* values for disease-specific survival.

Classification of lymph node metastasis

In Figure 4.5a,b classification results are presented for lymph node metastasis. The MCC scores have a maximum of about 0.1, which means that no significant predictive power is observed for any classifier, regardless of the feature selection method and the number of features.

Classification of disease-specific survival

In Figure 4.5c,d classification results are presented for disease-specific survival. With the Wilcoxon rank-sum test, the best MCC scores are about 0.1, which approximately equals one standard deviation. With the Kolmogorov-Smirnov test, the results look significantly better. Using five features, the two most frequently selected *m/z* values are: 913.5 (100%) and 2195.2 (80%). The decision tree classifier gives an MCC score of 0.3, and the confusion matrix is

$$M_{DTC} = \begin{pmatrix} 90 & 31 \\ 25 & 32 \end{pmatrix}$$

The best MCC scores for disease-specific survival correspond to two or three standard deviations, which could indicate that there is some predictive power in the spectra. However, we should emphasize that our results are not significant enough to really make such a claim. This is also illustrated by Figure 4.6, in which the two most discriminatory features are used to display samples from the two classes. No clear

pattern seems to be present, and both peaks are often below the baseline (indicated by negative values).

A small predictive power can appear to be present for a number of reasons, for example:

- There can be patterns in the data that are just random fluctuations. The probability of finding such a pattern grows with the size of the `search'. Here we consider five classifiers, one to fifteen features, two feature selection methods and two classes, which amounts to 300 cases. On the other hand, positive MCC scores are observed for most classifiers, and the scores are not very sensitive to the number of features.
- There can be small differences between the patient distributions in the classes.
- An uneven distribution of the samples over the tissue microarray or inhomogeneities in the sample preparation can lead to `artificial' differences between the classes.

These causes can only be ruled out by independent validation studies with different samples.

4.4 Discussion

Challenges in MALDI-MSI based biomarker discovery

Why are we not able to (clearly) predict metastasis or disease-specific survival from the MALDI-MSI spectra? Are there simply no discriminating features in the tissue, or have we just been unable to detect them? Below, we discuss some of the challenges in MALDI-MSI based biomarker discovery, which could explain our result. We also discuss how our results could be improved in the future.

A large, well-matched sample set is required

For biomarker discovery, one needs a sample set that is both well-matched (to prevent finding false biomarkers) and large (to account for the intrinsic heterogeneity of patient samples). The signal-to-noise ratio increases with the size of the sample set, and has to be large enough so that the biomarker can be detected. Unfortunately, one

can only tell how large a sample set has to be after a biomarker has actually been discovered.

For the study we perform here, there is an additional complication: cancer tissue is quite heterogeneous. It has been shown that interaction of tumor cells with the microenvironment, such as stroma and immune response factors, is important in the progression of the disease[133]. Moreover, comprehensive analyses of cancer genomes have found a great deal of heterogeneity within cancers of a single type, and even within one tumor[112]. These heterogeneous cell populations might differ in malignancy. In this regard, MALDI-MSI has one advantage: histology-directed analysis is possible, using for example tissue regions selected by a pathologist. Only a small tissue region per patient is selected for use in the TMAs used in this study. This small piece of tissue does not have to be representative for the tumor.

Sample preparation and data acquisition

For biomarker discovery with MALDI-MSI, we have the following requirements:

- A large number of *m/z* values should be detected, to maximize the information gained about the tissue
- Peaks should be detected with a high signal-to-noise ratio
- Measurements have to be reproducible

However, some of these requirements are competing. For example, the more analytes are desorbed from the tissue surface, the more they compete for ionization, hampering each others detection. In addition, the more *m/z* values are measured, the more peaks will overlap, reducing the signal-to-noise ratio. Small variations in the mixture of analytes, either tissue-based or introduced during the sample preparation (e.g. proteolytic digestion or matrix application), therefore greatly affect the reproducibility of the measurements.

To improve reproducibility, analytes can be separated before mass spectrometric analysis. Chromatographic separation is not compatible with MSI, which requires preservation of analyte distribution in the tissue section. Instead, ion mobility separation can be used. Additionally, overlapping peaks can sometimes be separated with higher resolution measurements.

High-dimensionality of the data

In many cases, MALDI-MSI data is *high-dimensional*. This means that $D \gg N$, where D is the dimension of an observation (the number of m/z values) and N is the number of observations available (the number of samples). Classifying high-dimensional data poses some challenges, which is sometimes referred to as the high-dimensionality-small-sample (HDSS) problem[22].

A simple example can demonstrate one of problems with high-dimensional data. Suppose we have a set of D+1 coins, which can be flipped by person A or person B. There is one special coin: if A flips it, it has chance β (> 0.5) of heads, and if B flips it, it has a chance β of tails. The other D coins are fair. Given a list of coin outcomes, we can tell whether person A or B flipped them with accuracy β, by looking at the special coin.

Now suppose we have to build a classifier, without knowledge of the special coin, from N labeled observations (coin outcomes), $N/2$ from A and $N/2$ from B. If N = 80 and β = 0.7, then the probability that a fair coin is at least as discriminating as the special one is about 0.56%. For different values of D (the number of fair coins) we can compute the probability that the special coin gives the best results on the training data. For D = 10, this probability is about 95%, for D = 100 it is 57% and for D = 1000 it is only 0.36%.

When D is increased, we effectively add more noise to the observations, whereas the amount of *signal* stays the same, because there is just one special coin. This example demonstrates a *fundamental* problem: when the signal-to-noise ratio gets too low, a good classifier cannot be built. Note that no `smart' algorithm can help here, the only solution is to have more training data available.

The complexity of discriminatory patterns

In previous studies, it was found that single biomarkers usually have limited predictive power[134, 135]. The classifiers that we use can combine the information of several features in different, non-linear ways. However, the `right' features first have to be selected. Since we use a univariate feature selection method, the detection of biomarkers consisting of many features is unlikely, especially if univariate differences are small.

There is a general problem in finding such biomarkers: since the number of possible patterns increases rapidly with the complexity of the pattern, large data sets are required to prevent overfitting.

Outlook

The quality of the spectra is an important factor in biomarker discovery. Some parameters in the sample preparation can still be optimized, for example the tissue thickness or the digestion enzymes. A recent paper by Heijs *et al*. describes the use of Lys-C, Arg-C and r-Lys-N for the proteolytic digestion of fresh frozen tissue[136]. Compared to the use of trypsin, more high intensity *m/z* values were observed. This came, however, at the cost of the detection of fewer *m/z* values.

We expected to find only small differences between the classes. A recently validated gene signature that can predict lymph node metastasis consists of no less than 732 probes[124]. No predictive power was observed for lymph node metastasis and only limited predictive power for disease-specific survival. No clear differences were detected between the classes on the level of the proteome, as probed by MALDI-MSI. A more in-depth analysis, for example by liquid chromatography MS, might be used to investigate changes in the low-abundant proteins.

4.5 Conclusion

We have investigated whether lymph node metastasis and disease-specific survival for head and neck cancer can be predicted from MALDI-MSI data. Measurements were performed on two tissue microarrays, which contained tumor cores from 240 patients. Our sample preparation and measurement technique have been described in the first part of the chapter. The processing of these data so that it could be used in different classifiers was described in the second part of the chapter.

Using five classifiers and two feature selection methods, we did not observe predictive power for lymph node metastasis. For disease-specific survival, some of the classifiers showed a small predictive power. We have shown that our method is sensitive to intensity differences of 50% in typical peptide peaks, using artificially modified data. The proteomic differences associated with lymph node metastasis or disease-specific survival thus have to be smaller, if they exist.

5

Mass spectrometry imaging of the hypoxia marker pimonidazole in a breast tumor model

Although tumor hypoxia is associated with tumor aggressiveness and resistance to cancer treatment, many details of hypoxia-induced changes in tumors remain to be elucidated. Mass spectrometry imaging (MSI) is a technique that is well suited to study the biomolecular composition of specific tissue regions, such as hypoxic tumor regions.

Here, we investigate the use of pimonidazole as exogenous hypoxia marker for matrix-assisted laser desorption/ionization (MALDI) MSI. In hypoxic cells, pimonidazole is reduced and forms reactive products that bind to thiol groups in proteins, peptides and amino acids. We show that a reductively activated pimonidazole metabolite can be imaged by MALDI-MSI in a breast tumor xenograft model. Immunohistochemical detection of pimonidazole adducts on adjacent tissue sections confirmed that this metabolite is localized to hypoxic tissue regions.

We used this metabolite to image hypoxic tissue regions and their associated lipid and small molecule distributions with MALDI-MSI. We identified a heterogeneous distribution of 1-methylnicotinamide and acetylcarnitine, which mostly co-localized with hypoxic tumor regions.

As pimonidazole is a widely used immunohistochemical marker of tissue hypoxia, it is likely that the presented direct MALDI-MSI approach is also applicable to other tissues from pimonidazole-injected animals or humans.

5.1 Introduction

Tumor hypoxia, caused by abnormal tumor vasculature, is associated with tumor aggressiveness and resistance to cancer treatment[137]. Tissue regions with a partial oxygen pressure (pO_2) below 10 mmHg are typically considered hypoxic, but the degree of hypoxia can vary considerably inside and between different tumors[138]. Hypoxia is found in many solid tumors and triggers a complex response that involves many different molecular pathways. These pathways influence cellular processes such as apoptosis, angiogenesis, proliferation and anaerobic metabolism. Understanding the hypoxia-induced changes in tumors is essential for the development of more effective cancer treatment. This requires the development of innovative techniques that can image hypoxia and its associated biomolecular changes.

Given the importance of hypoxia, many techniques have been developed for measuring tumor oxygenation. These techniques include direct pO_2 measurement with polarographic oxygen electrodes or fiber optic probes, magnetic resonance and other imaging techniques[137, 139]. In recent years, endogenous markers such as hypoxia-inducible factor 1 (HIF-1), carbonic anhydrase IX (CAIX), glucose transporter 1 (GLUT1), C-X-C chemokine receptor type 4 (CXCR4), vascular endothelial growth factor (VEGF) and insulin-like growth factor 1 receptor (IGF1R) have been reported for hypoxia imaging[137, 140]. However, the expression of these markers is not directly linked to the oxygenation status of the tissue and they are therefore often referred to as hypoxia-related markers[137].

The exogenous 2-nitroimidazole hypoxia markers were originally designed as radiosensitizers, but the observation that they were activated and retained in viable hypoxic cells with an oxygen dependence similar to that of radioresistance led to the development of this class of molecules as hypoxia markers[141-143]. Several methods exist for the detection of 2-nitroimidazole adducts, including positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetic resonance spectroscopy (MRS) and immunohistochemical assays[144]. Antibody-based detection methods have the advantage that hypoxia can be imaged on a cellular level. The spatial distribution of hypoxia as detected with antibodies recognizing 2-nitroimidazoles can be compared with high spatial detail to other markers of tumor biology[145, 146]. A disadvantage of immunohistochemical assays is that they are targeted assays with limited multiplexing capability, hence only known targets can be studied and only a small number of proteins can be detected at the same time.

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI) can image hundreds of analytes directly from tissue surfaces. It can visualize a wide variety of biomolecules, such as lipids, proteins, peptides and metabolites routinely at 50-100 µm spatial resolution, but recent technical developments have shown that lipids and drug compounds can be imaged with a resolution of 5-10 µm[147-149]. These characteristics make MALDI-MSI well suited to study the biomolecular make-up of hypoxic tissue regions, provided that there is an easy way to discriminate hypoxic from normoxic tissue regions.

Here we present a MALDI-MSI approach that combines the detection of the 2-nitroimidazole hypoxia marker pimonidazole, and the multiplexing capabilities of this technique to image hypoxic regions and their associated biomolecules in a single experiment.

Pimonidazole was shown to be a reliable marker of hypoxia and is approved for clinical use[150, 151]. Importantly, it has favorable chemical properties for mass spectral detection such as a slightly basic piperidine side chain.

Pimonidazole is reduced only under hypoxic conditions to form reactive products that bind to cellular nucleophiles, especially thiol-containing proteins. Reductive activation is inhibited at pO_2 >10 mmHg in solid tissue (half-maximal inhibition at ca. 2 mmHg)[152-154]. Pimonidazole reduction depends upon nitroreductase activity; hence metabolically active cells are required for reductive metabolism of pimonidazole. Limited knowledge exists on the *in vivo* reaction products of pimonidazole[155]. Current knowledge is largely based on *in vitro* and simple *in vivo* experiments that were performed in the 1980's[143, 156-161]. For these experiments a different 2-nitromidazole, misonidazole, was primarily used.

In this chapter, we used MALDI-MSI to study pimonidazole and its metabolism in breast cancer xenograft tissue from pimonidazole-injected mice. We detected the unreacted pimonidazole compound and several pimonidazole metabolites. Accurate mass and product ion measurements with atmospheric pressure scanning microprobe MALDI (AP-SMALDI) MSI and accurate mass liquid chromatography (LC) MS experiments were performed to analyze tumor tissue from pimonidazole-injected mice in detail. Verification was performed by immunohistochemical detection of pimonidazole adducts on adjacent tissue sections. We show that one of the detected pimonidazole metabolites is well suited as marker of hypoxia in MALDI-MSI

experiments. We also present small molecules that co-localize with the hypoxic regions as detected by this pimonidazole metabolite.

5.2 Materials and methods

Chemicals and reagents

We obtained alpha-((2-Nitroimidazol-1-yl)methyl)-1-piperidineethanol (pimonidazole), α-cyano-4-hydroxycinnamic acid (CHCA) and trifluoroacetic acid (TFA) from Sigma-Aldrich (Steinheim, Germany). Acetonitrile (ACN) and methanol (MeOH) were purchased from Biosolve (Valkenswaard, The Netherlands). Cresyl Violet was obtained from Thermo Scientific (cat# 40576, PA, USA). Ponceau S (cat# P3504), Mayer's hematoxylin and eosin were obtained from Sigma-Aldrich (St. Louis, MO, USA). Paraformaldehyde was purchased from Santa Cruz Biotechnology (Dallas, TX, USA.). Other reagents used for immunostaining were purchased from EMD Millipore (MA, USA), unless stated otherwise.

Preparation of breast tumors for stainings and MSI analysis

Triple-negative MDA-MB-231 breast cancer cells, obtained from the American Type Culture Collection (ATCC), were orthotopically inoculated into the mammary fat pad of athymic nude mice. Tumors were grown to 6-8 mm in diameter within about 8 weeks. 400 mg/kg of pimonidazole was injected intravenously into the tail vein. Pimonidazole, primary mouse anti-pimonidazole antibody conjugated with FITC, secondary rabbit anti-FITC antibody, and all immunohistochemical (IHC) staining reagents were purchased as a kit (HP2-100 Kit, Hypoxyprobe, Burlington, MA, USA). After 30 minutes, the mice were sacrificed, tumors were excised and embedded in gelatin. An equal mixture of Cresyl Violet and Ponceau S (0.5 mg/mL of each as final concentration in gelatin), were added as fiducial markers for spatial referencing[91]. The embedded tumors were frozen in liquid nitrogen and stored at -80 °C before sectioning. Four tumors were used in this chapter: three from pimonidazole-injected mice and one from an untreated control mouse.

During cryosectioning, tumors were divided into about ten sets of cryosections. Four 100 µm thick sections were cut for each set and stored in eppendorf tubes. Then, seven 10 µm sections were cut per set. Adjacent sections were mounted on glass slides for hematoxylin and eosin (H&E) and IHC staining and on indium-tin oxide (ITO)

coated slides (Delta Technologies, Stillwater, MN, USA) for MSI analysis. All sections analyzed originated from tissues sets 6-9 as counted from the mouse body wall to the top of the tumor, except for the sections used for metabolite extraction which originated from throughout the tumor. All sections were stored at -80 °C until analysis.

Hematoxylin and eosin (H&E) staining

One frozen slide from each set was thawed and fixed with 4% paraformaldehyde for 30 min, then washed with water. Fresh hematoxylin was applied on top of the tissue sections for 1 min. After washing with distilled water, tissue sections were stained with eosin for 1 min. The slides were then washed with water until there was minimal coloring visible on the gelatin area of the sections. The slides were then mounted with a cover glass using mounting medium (DAKO Faramount aqueous mounting medium, cat# S3025, Carpinteria, CA, USA) and photomicrographs were taken on a Nikon microscope equipped with a CCD camera.

Immunohistochemical (IHC) staining for pimonidazole

One tissue section from each set was thawed, fixed in 4% paraformaldehyde for 30 min and washed three times with TBS/0.1% Tween-20. Peroxidase activity was quenched with 3% H_2O_2 for 10 min, and tissue sections were washed again three times with TBS/0.1% Tween-20. Non-specific binding was blocked with protein blocking reagent (cat# 20773, Millipore, Billerica, MA, USA) and rabbit serum, each for 10 min. Tissue sections were rinsed three times with TBS/0.1% Tween-20. For staining, tissue sections were incubated with primary antibody (primary mouse anti-pimonidazole antibody conjugated with FITC, HP2-100 Kit, Hypoxyprobe, Burlington, MA, USA) diluted to 1:75 with antibody dilution buffer (TBS/0.1% Tween-20) for 30 min. Samples were washed five times with TBS/0.1% Tween-20, followed by incubation with secondary antibody (secondary rabbit anti-FITC antibody, HP2-100 Kit, Hypoxyprobe, Burlington, MA, USA) diluted to 1:75 with antibody dilution buffer for 30 min. Samples were washed five times with TBS/0.1% Tween-20. 3,3'-Diaminobenzidine (DAB) staining was performed for 10 min (DAB Quanta, cat #TA-060-QHDX, Thermo Scientific). The reaction was stopped by washing with water. Samples were washed once with TBS/0.1% Tween-20. Counterstaining was per-formed with hematoxylin for 1 min. Finally, mounting medium was applied and a cover glass attached. Images were acquired using a Nikon microscope equipped with a CCD camera.

MALDI mass spectrometry

CHCA matrix solution was prepared at a concentration of 10 mg/mL in 50% ACN (vol/vol) and 0.2% TFA in water. The pimonidazole standard solution was prepared by mixing a pimonidazole solution of 1 mg/mL in MeOH with CHCA matrix solution in a 1:1 ratio. The sample was spotted on a MALDI target plate at 0.5 µL per spot. We analyzed multiple breast tumor xenograft sections from three pimonidazole-injected mice and from one untreated control mouse. Tissue sections were dried prior to MSI analysis. CHCA matrix solution was sprayed onto the sections by a vibrational sprayer (ImagePrep, Bruker Daltonics, Bremen, Germany).

MALDI mass spectrometric analyses were performed on a MALDI-QTOF instrument (Synapt HDMS, Waters, UK) in positive ion mode. Quadrupole transmission was optimized for detecting pimonidazole-derived ions. MS images were acquired with a laser step size of 100 µm and with a mass range set between *m/z* 75 and 1000. Collision-induced dissociation was performed in the trap cell with a collision energy of 15-35 eV. Fragmentation spectra of pimonidazole-derived *m/z* 223.2 and endogenous tissue species were collected using on-tissue MS/MS. Pimonidazole standard solution spots were used to acquire MS and MS/MS spectra of the compound. Pimonidazole was detectable down to 0.5 pmol (mass accuracy 100 ppm). MALDI MS Images were generated using BioMap 3.8.0.4 software (Novartis, Basel, Switzerland) with $\Delta m/z = \pm 0.02$.

AP-SMALDI mass spectrometry

CHCA matrix solution was prepared at a concentration of 5 mg/mL CHCA in 50% ACN (vol/vol) and 0.2% TFA in water. Matrix solution was sprayed onto a dried tumor tissue section using a Suncollect sprayer (SunChrom, Friedrichsdorf, Germany).

Experiments were performed using an atmospheric pressure scanning microprobe matrix-assisted laser desorption/ionization imaging source (AP-SMALDI10, TransMIT, Giessen, Germany)[162], coupled to an orbital trapping mass spectrometer (Q Exactive, Thermo Fisher Scientific, Bremen, Germany). MS and MS/MS data were acquired in positive ion mode. MS image size was 50 x 50 pixels, with a step size of 100 µm. Internal calibration was achieved using CHCA signals as lock masses.

On-tissue MS/MS of *m/z* 223.2 was performed using higher-energy collisional dissociation (HCD) and a precursor ion isolation window of 1 Da. For HCD the

normalized collision energy was set to 65%. AP-SMALDI MS images were generated with the software package Mirion[163] with $\Delta m/z$ = ±0.004. Spectra were analyzed with XCalibur software (Thermo Scientific).

LC-MS analysis of pimonidazole metabolites

Several 100 µm-tick tumor tissue sections, from one pimonidazole-injected and one untreated mouse, that were adjacent to sections analyzed with MSI, were used for metabolite extraction. The following extraction protocol was used: 25 mg frozen tissue was placed in pre-cooled 1.5 mL tubes, and glass beads (1 mm diameter) and 200 µL pre-cooled MeOH were added. Tissue samples were homogenized six times for 10 s, using a mini-bead beater (Biospec, Bartlesville, OK, USA), with 20 s on ice in between the homogenization rounds to avoid heating of the samples. Next, the samples were centrifuged for 5 min at 10,000g at 4°C. Supernatants were transferred to new tubes and centrifuged again for 5 min at 20,000g at 4°C to remove remaining tissue debris. Samples were stored at -20°C until LC-MS analysis. A pimonidazole solution of 100 ng/mL in MeOH was used as control sample.

High-performance liquid chromatography was performed on a Thermo Scientific Dionex Ultimate 3000 RSLC system equipped with an Accucore C18 column (100 mm x 2.1 mm, particle size 2.6 µm) at 40°C. The injection volume was 2 µL and separation was achieved using a 10 min gradient with a flow rate of 400 µL/min. Mobile phase A consisted of 10 mM ammonium acetate at pH 9.8. Mobile phase B consisted of ACN with 0.1% formic acid. This setup was connected to an Orbitrap Fusion instrument (Thermo Fisher Scientific, Bremen, Germany). This mass spectrometer was operated in positive ion mode with a scan range of *m/z* 80-300. Data analysis was performed with Xcalibur and Compound Discoverer software (Thermo Scientific).

Image co-registration and correlation analysis

Image co-registration and correlation analysis were performed with Matlab software (The Mathworks Inc., Natick, MA, USA) using peak-picked and TIC normalized data[106].

We analyzed three sections from three different breast tumor xenografts from pimonidazole-injected mice that were imaged by the approach described above. Ion images of individual *m/z* values were co-registered to the optical image of the anti-pimonidazole stained tissue samples using the position of fiducial markers and tumor boundary, as previously described[164].

Overlap between IHC and MSI detection of pimonidazole was determined by overlaying individual ion images and the anti-pimonidazole stained images. The overlap images were constructed by using a relative threshold of 0.2 for all ion images, which means that only the 80% of pixels with the highest intensity were plotted as green dots.

Correlation analysis was performed by calculating Pearson's correlation coefficients for *m/z* 223.2 with all other variables (i.e. *m/z* values), for each MSI data set.

Identification of endogenous metabolites

Endogenous metabolites were identified from parent ion masses and fragmentation spectra. Accurate mass data (<1 ppm mass error) were obtained from on-tissue AP-SMALDI MS experiments. AP-SMALDI MS spectra were analyzed with XCalibur software. Fragmentation spectra were obtained from MALDI on-tissue MS/MS product ion acquisition and analyzed with MassLynx software (Waters, UK). LIPID MAPS (www.lipidmaps.org), the Human Metabolome Database (version 3.6, www.hmdb.ca) and MassBank (www.massbank.jp) were used to search for metabolite structures and fragmentation spectra.

5.3 Results and Discussion

Detection of pimonidazole by MALDI-MS

The pimonidazole compound was readily detectable with MALDI as the protonated molecule (*m/z* 255.1). Also several pimonidazole fragments were detected in the MS mode, with the highest intensity fragments at *m/z* 98.1, 124.1, 142.1, 223.2 and 226.2 (Figure 5.1a). The ions at *m/z* 98.1, 124.1 and 142.1 are fragments of the piperidine side chain[150]. Other fragments contain the imidazole ring atoms or parts of the imidazole ring. Only low intensity fragments were observed in an electrospray ionization (ESI) experiment; *m/z* 124.1 and 142.1 were detected with 100x lower intensity than *m/z* 255.1. The observed fragmentation can thus mainly be attributed to the desorption/ionization process used. Misonidazole, a pimonidazole analog, has an absorbance maximum at 325 nm[143], which is close to 355 nm, the wavelength of the laser used. This might explain pimonidazole's easy fragmentation behavior in MALDI.

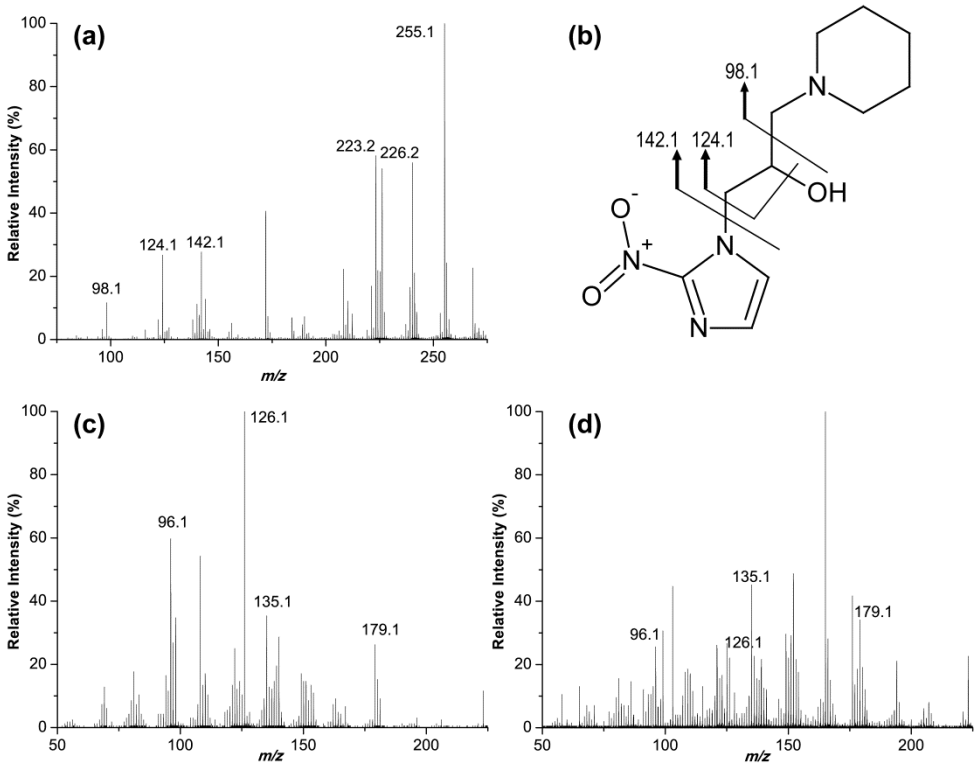

Figure 5.1 MALDI mass spectra of pimonidazole obtained from the pure compound and from tumor tissue from a pimonidazole-injected mouse. (a) MALDI mass spectrum of pimonidazole. Shown are the protonated parent ion at *m/z* 255.1 and its main fragments. (b) Chemical structure of pimonidazole with characteristic fragments as determined by MS/MS. On-tissue MS/MS analysis shows that *m/z* 223.2 is a pimonidazole-derived ion. MS/MS spectra of *m/z* 223.2 from (c) pimonidazole and (d) pimonidazole-treated tissue.

MALDI-MSI of pimonidazole in breast tumor xenografts

We analyzed three breast tumors from pimonidazole-injected mice and one untreated control tumor. Figure 5.2a shows the MS images of three detected pimonidazole-derived ions, *m/z* 124.1, 142.1 and 223.2. The protonated parent ion was also observed at *m/z* 255.1. Overlays of average spectra of pimonidazole-treated and untreated tumor tissue (Figure 5.2b) show the detection of these ions only in the pimonidazole-treated tumor tissue and not in the control tumor tissue.

On-tissue MS/MS fragmentation of *m/z* 223.2 confirmed its pimonidazole-derived nature (Figure 5.1c and d). Given the low molecular weight of pimonidazole, high

mass accuracy measurements with an Orbitrap mass analyzer were performed to confirm the elemental composition of the detected ions. The predicted composition from on-tissue mass measurements matched those of the pimonidazole compound and its derivatives as listed in Table 5.1.

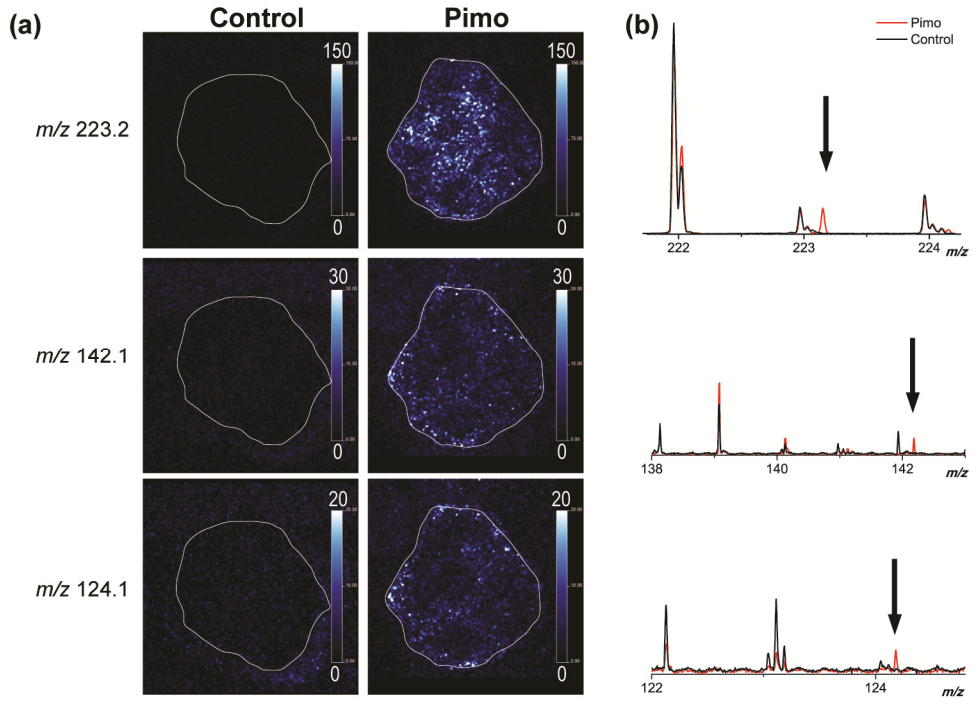

Figure 5.2 MALDI-MSI analysis shows the distribution of pimonidazole-derived ions in tumor tissue. (a) Pimonidazole (Pimo)-derived ions with *m/z* 124.1, 142.1 and 223.2 are solely detected in tumor tissue from pimonidazole-injected mice, and not in control tumor tissue. (b) MALDI-MSI spectra (zoom) from treated (red) and untreated (black) tumor tissue. Presence of pimonidazole-derived ions is indicated with an arrow.

Detection of pimonidazole metabolism

It is well known that 2-nitroimidazoles are heavily metabolized *in vivo*. Under hypoxic conditions, they are reduced, and after a series of steps they finally bind to cellular nucleophiles. It is estimated from *in vitro* experiments that around 20% of reductively activated 2-nitroimidazoles react with thiol containing proteins and small molecules such as glutathione[161]. The remaining 80% is subject to hydrolytic fragmentation.

Several 2-nitroimidazole metabolites are described in the literature, either in *in vivo* or *in vitro* experiments [143, 156-161, 165].

Table 5.1 Monoisotopic mass values for pimonidazole-derived ions from tumor tissue. Assignments are based on high mass accuracy (<3 ppm root-mean-square error) experiments on an AP-SMALDI instrument coupled to an Orbitrap Q Exactive mass spectrometer. *Based on collision energy required to induce fragmentation. **Based on prediction from ESI data of pimonidazole in Metlin database.

Elemental composition	**Adduct**	**Exact mass**	**Measured accurate mass**	**Mass error (ppm)**
$C_{11}H_{19}N_4O_3$	$[M+H]^+$	255.14517	255.14497	-0.8
$C_{11}H_{19}N_4O$	$[M]^{+*}$	223.15534	223.15522	-0.5
$C_8H_{16}NO$	$[M]^{+**}$	142.12264	142.12267	0.2
$C_8H_{14}N$	$[M]^{+**}$	124.11208	124.11222	1.1

We searched for pimonidazole metabolites that are part of the reductive metabolism of pimonidazole. These metabolites are not generated under normoxic conditions, nor in necrotic tissue. Their confinement to hypoxic regions makes them markers of tumor hypoxia.

To characterize the metabolic changes that pimonidazole undergoes in hypoxic regions, metabolites and other small molecules were extracted from treated and untreated tumor tissue, and analyzed by LC-MS. The pimonidazole compound was included as a standard in our analysis to check for mass spectrometry-induced changes to this compound. Several species that are part of the reductive pathway with low or no abundance in the standard and untreated tumor sample were identified.

Pimonidazole hydroxylamine is the four-electron reduction product that is the main reactive species that needs to be formed for thiol-binding to occur[143, 156, 158]. However, this molecule can rearrange to form hydroxy derivatives with the same elemental composition. The detection of *m/z* 241.2 in two major elution peaks points towards the detection of multiple species. The elution of hydroxylamine and hydroxyl derivatives in two chromatographic peaks has been previously reported for the 2-nitroimidazole benznidazole[165, 166].

Further reduction yields the six-electron reduction product of pimonidazole, which is an amine derivative (*m/z* 225.2). Interestingly, also a pimonidazole derivative at *m/z* 223.2 was detected in the tumor tissue samples from pimonidazole-injected mice. An ion with the same elemental composition is postulated to be the nitrenium intermediate that is responsible for binding to thiol groups[158-160]. To our

knowledge, an ion with this elemental composition has not been previously detected for pimonidazole or a 2-nitroimidazole analog.

Table 5.2 Pimonidazole metabolites after *in vivo* reduction of pimonidazole. Assignments are based on high mass accuracy LC-MS experiments (<3 ppm root-mean-square error) using an Orbitrap Fusion mass spectrometer. *Derivative.

RT (min)	Measured accurate mass	Exact mass	Mass error (ppm)	Adduct	Elemental composition	Proposed metabolite
7.7	201.17094	201.17099	-0.2	$[M+H]^+$	$C_9H_{21}N_4O$	Guanidine der.*
2.9	223.15528	223.15534	-0.2	$[M]^+$	$C_{11}H_{19}N_4O$	Nitrenium
7.7	225.17095	225.17099	-0.1	$[M+H]^+$	$C_{11}H_{21}N_4O$	Amine der.
2.8/ 3.9	241.16582	241.16590	-0.3	$[M+H]^+$	$C_{11}H_{21}N_4O_2$	Hydroxylamine Hydroxyl der.
7.7	259.17704	259.17647	2.2	$[M+H]^+$	$C_{11}H_{23}N_4O_3$	Dihydro dihydroxy der.
3.0	271.17639	271.17647	-0.3	$[M+H]^+$	$C_{12}H_{23}N_4O_3$	Methoxy der.
7.7	273.19205	273.19212	-0.2	$[M+H]^+$	$C_{12}H_{25}N_4O_3$	Methoxy der.
2.7	265.62311	265.62321	-0.4	$[M+2H]^{2+}$	$C_{21}H_{37}N_7O_7S$	Glutathione adduct

The binding of reductively activated pimonidazole to a thiol-containing molecule was shown with the detection of an ion at *m/z* 265.6, which was assigned to a doubly charged pimonidazole adduct of glutathione. Its charge state was confirmed by detection of the ^{13}C and ^{34}S isotopic peaks at *m/z* +0.50168 and +0.99790, respectively. This glutathione adduct has previously been reported as a product of the *in vitro* reduction of misonidazole and *in vivo* reduction of benznidazole[157, 158, 165].

As expected, several products of the hydrolytic fragmentation of reductively activated pimonidazole were found as well. Hydrolysis of a hydroxyl derivative yielded a dihydro dihydroxy compound (*m/z* 259.2)[158-160, 166]. This compound can fragment or react with other molecules, releasing glyoxal and a guanidine derivative (*m/z* 201.2)[167-169]. The signals at *m/z* 271.2 and 273.2 were assigned to methoxy derivatives of pimonidazole[165].

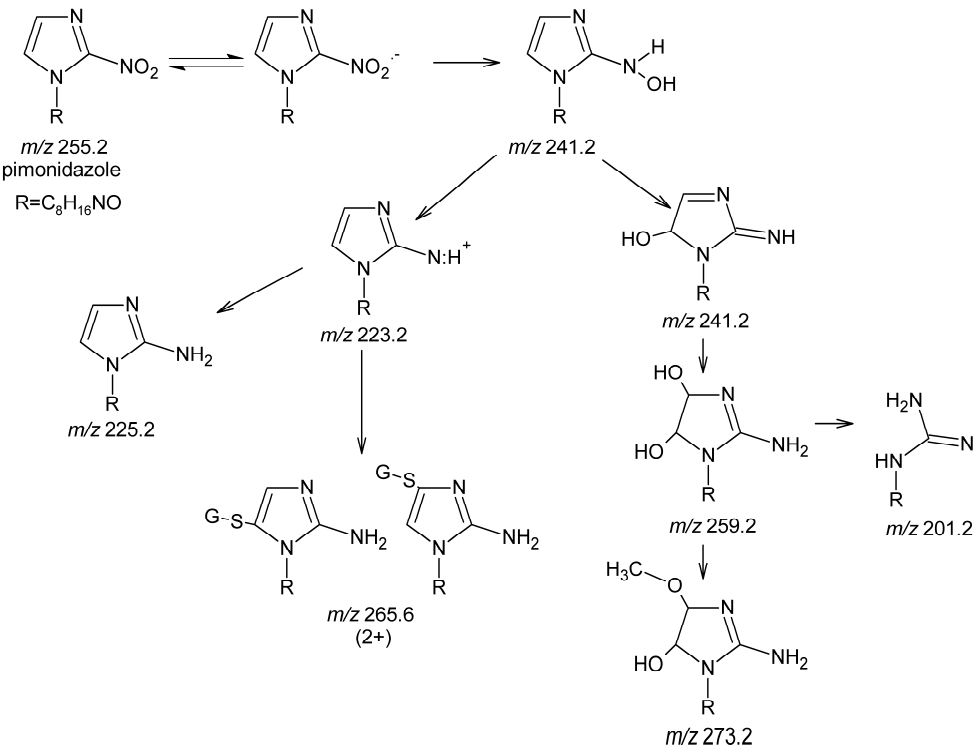

Figure 5.3 Putative chemical structures of the identified pimonidazole-derived metabolites. The indicated pathway is based on existing literature (see 'Detection of pimonidazole metabolism'). For each species the detected *m/z* value is given. R = side chain. G = glutathione.

Pimonidazole metabolites as hypoxia markers in MSI

Most pimonidazole metabolites identified by LC-MS were detected in the AP-SMALDI imaging data, as determined by accurate mass matching. Figure 5.4 shows that a different distribution was observed for *m/z* 201.2 and 223.2 as compared to the distribution for the parent compound at *m/z* 255.1 and the side chain fragment at *m/z* 142.1.

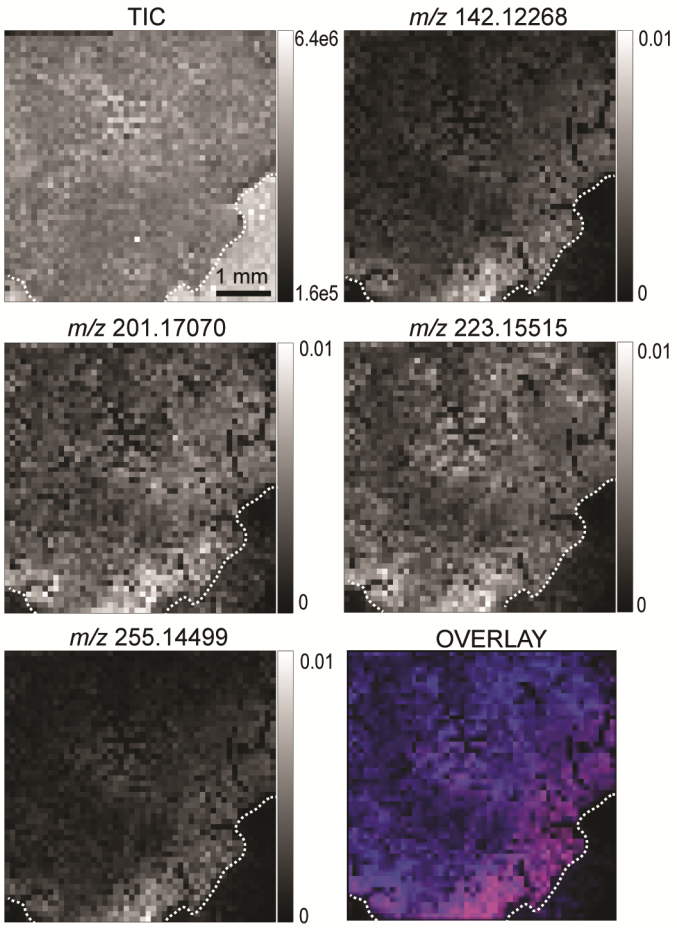

Figure 5.4 AP-SMALDI MSI of a tumor tissue section showing the distribution of pimonidazole-derived ions. A different distribution was observed for *m/z* 201.17070 and 223.15515 as compared to the distribution for the parent compound at *m/z* 255.14499 and the side chain fragment at *m/z* 142.12268. The overlay of *m/z* 223.15515 (blue) and *m/z* 255.14499 (red) shows overlap of the two ions only at the border of the tissue section (pink). The intensity for each ion was normalized to the total ion count (TIC) per pixel.

To determine which pimonidazole metabolites are suitable markers for tumor hypoxia in MALDI-MSI, we compared for each metabolite the signal intensity and the contribution of MALDI fragments to the signal.

For the MALDI-MSI data, only the derivatives at *m/z* 223.2 and 225.2 were detected, due to the lower sensitivity and mass resolution of the QTOF instrument as

compared to the AP-SMALDI instrument. Nevertheless, when comparing *m/z* 223.2 with *m/z* 124.1 or 142.1, a similar difference in distribution was observed as determined by AP-SMALDI (Figure 5.2 and Figure 5.4).

To determine the contribution of MALDI-induced fragmentation to the pimonidazole metabolite signals, 50 pmol pimonidazole was mixed with untreated tissue homogenate and measured with MALDI-MS. Pimonidazole derivative *m/z* 223.2 shows a five times higher detection after *in vivo* metabolism of pimonidazole as compared to the detection from tissue homogenate mixed with unreacted pimonidazole (Figure 5.5). These results suggest that the ion at *m/z* 223.2 observed in tumor tissue from pimonidazole-injected animals can be mainly attributed to the pimonidazole metabolite.

The pimonidazole derivative at *m/z* 223.2 could be easily detected in both the MALDI and AP-SMALDI imaging experiments. We therefore propose to use *m/z* 223.2 as hypoxia marker for mass spectrometric analysis of tumor tissue from pimonidazole-injected animals. All further analyses will focus on this pimonidazole metabolite at *m/z* 223.2.

Verification of pimonidazole distribution with immunohisto-chemistry

Additional verification was performed with immunohistochemistry (IHC). IHC staining against pimonidazole is a widely used method for hypoxia detection[150]. Tissue sections adjacent to the sections used for mass spectrometric analysis were stained. The anti-pimonidazole stained images were co-registered with the MALDI-MSI data. Overlap between IHC and MSI detection of pimonidazole was determined by overlaying individual ion images and the anti-pimonidazole stained images.

Figure 5.6 shows the results of this qualitative overlap analysis for one representative tumor. The darkest anti-pimonidazole stain was observed at the tumor boundary and at the border to necrotic regions. Pimonidazole metabolite *m/z* 223.2 co-localized with the marker detected by IHC mainly around the necrotic tumor core. The larger area stained positive for hypoxia by IHC might be explained by the higher sensitivity of IHC as compared to MSI. The detected species are also different: IHC detects pimonidazole protein adducts and MSI detects unbound pimonidazole metabolites.

Adjacent tissue sections were co-registered for the analysis and did not perfectly overlap. The use of an MSI-based hypoxia marker makes co-registration unnecessary, and thus avoids a potentially error-introducing step.

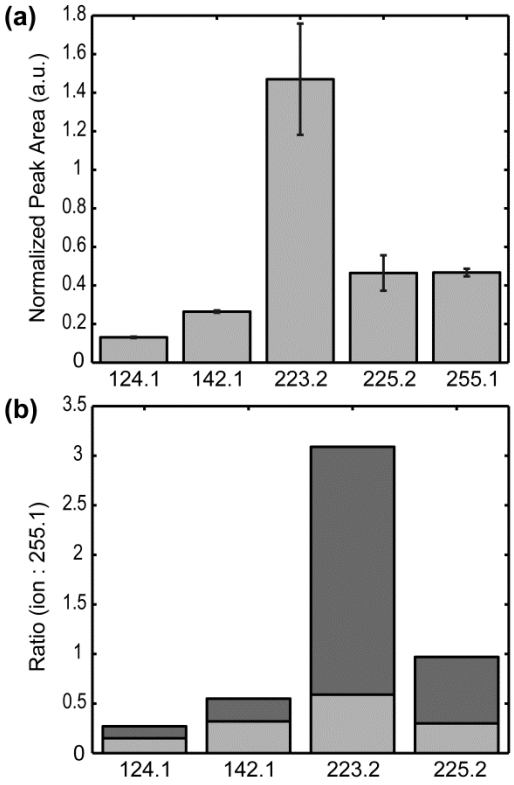

Figure 5.5 Pimonidazole derivative *m/z* 223.2 is the ion detected with the highest intensity from tumor tissue in MALDI-MSI. (a) MALDI-MSI detection of pimonidazole derivatives. Data are averaged for three tumors and shown as average ± standard deviation. (b) Pimonidazole derivative *m/z* 223.2 shows a five times higher detection after *in vivo* metabolism of pimonidazole as compared to detection from tissue homogenate mixed with unreacted pimonidazole (i.e. MALDI-induced fragmentation). Shown is the detection ratio for each ion as compared to *m/z* 255.1. Contribution of MALDI-induced fragmentation to the total signal is shown in light grey.

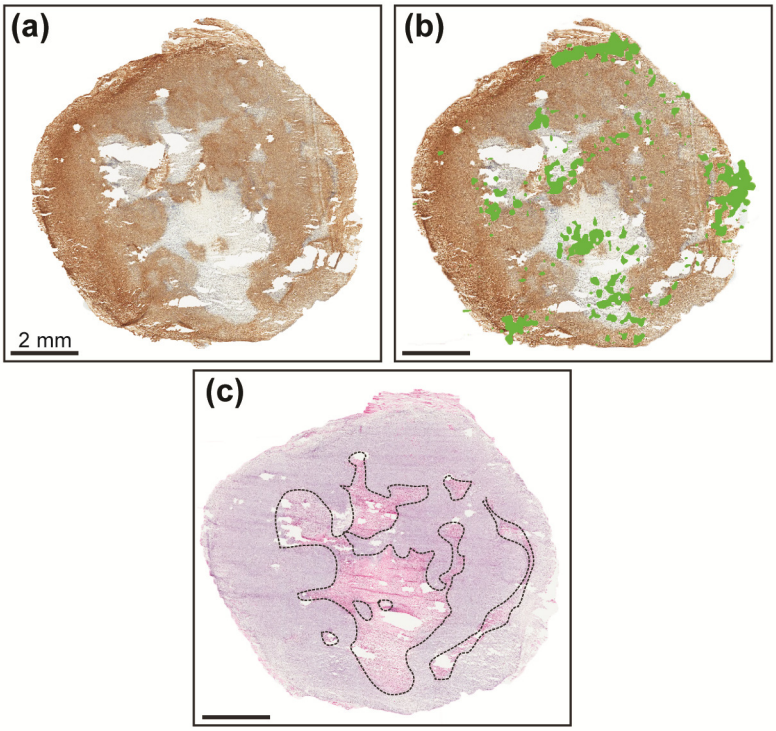

Figure 5.6 IHC detection of pimonidazole. (a) Distribution of hypoxic regions by pimonidazole antibody staining. Brown tissue staining is indicative of pimonidazole binding and is strongest at the tumor boundary and around the necrotic regions. (b) Co-registration of a normalized and thresholded MS image of pimonidazole metabolite *m/z* 223.2 (green) from an adjacent tissue section. (c) H&E stained adjacent tissue section. Necrotic regions are indicated by a dashed line.

Correlation of the pimonidazole metabolite *m/z* 223.2 with endogenous lipids and metabolites

Hypoxic regions in MDA-MB-231 breast tumor xenografts were identified by means of the hypoxic pimonidazole metabolite *m/z* 223.2. We performed correlation analysis to identify biomolecules that are spatially correlated with these hypoxic regions. All three tumors showed a highly similar correlation pattern as determined by Pearson's correlation. The strongest correlations were identified for three low molecular weight ions at *m/z* 137.1, 160.1 and 204.1 (Figure 5.7). These ions show indeed a similar distribution as compared to the pimonidazole metabolite *m/z* 223.2 (insets Figure 5.7). Accurate mass measurements and on-tissue MS/MS fragmentation analyses

identified these ions as 1-methylnicotinamide $[M]^+$ for *m/z* 137.1 and acetylcarnitine $[M+H]^+$ for *m/z* 204.1. The ion at *m/z* 160.1 (elemental composition $C_8H_{18}NO_2$) remains so far unidentified.

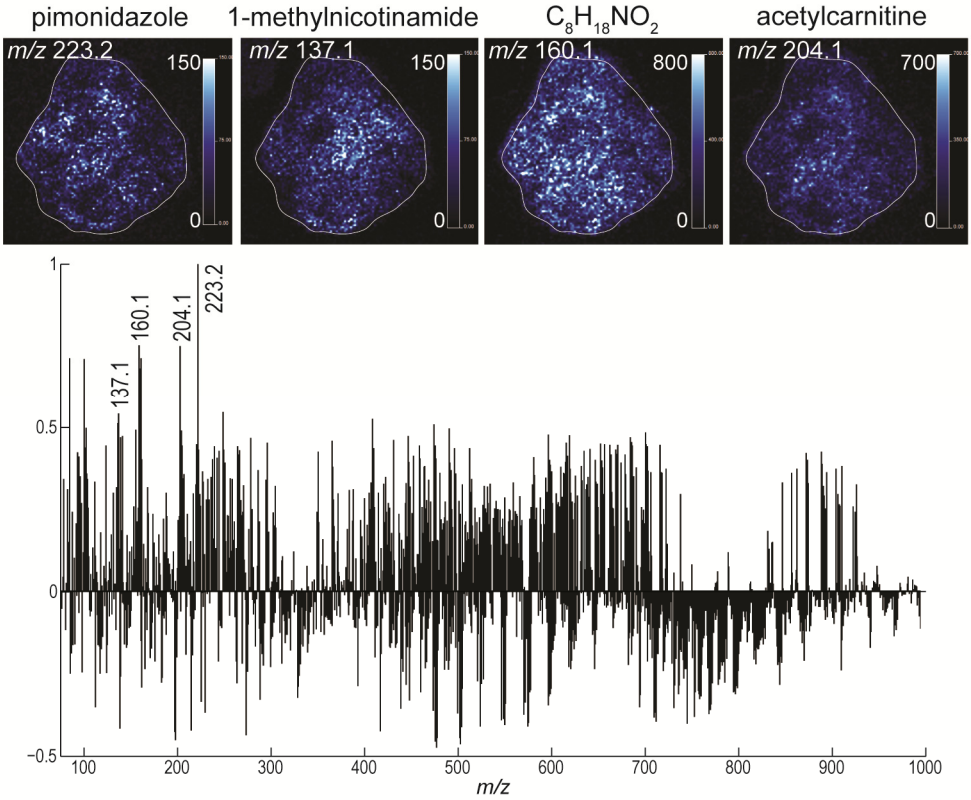

Figure 5.7 Correlation of *m/z* 223.2 with other *m/z* features as determined by Pearson's correlation. 1-methylnicotinamine, $C_8H_{18}NO_2$ and acetylcarnitine were found to be highly correlated with *m/z* 223.2. Data are shown for one representative tumor section.

Elevated levels of 1-methylnicotinamide were previously found in human cell lines overexpressing nicotinamide N-methyltransferase using untargeted LC-MS/MS analysis[170]. Nicotinamide N-methyltransferase catalyzes the transfer of a methyl group from S-adenosyl-L-methionine to nicotinamide, generating S-adenosylhomocysteine and 1-methylnicotinamide. This enzyme is overexpressed in several cancer types and is known to support tumorigenesis[171, 172]. A high expression of nicotinamide N-methyltransferase has been reported for the MDA-MB-231 breast cancer cell line used as

orthotopic xenograft model in this chapter[173]. It was previously reported that nicotinamide N-methyltransferase is involved in the regulation of protein methylation in cancer cells, whereby 1-methylnicotinamide functions as a sink for methylation units[170].

We have identified the ion at *m/z* 204.1 as the protonated ion of acetylcarnitine. A previous MALDI-MSI study using the same breast tumor xenograft model identified two acylcarnitines that localized to hypoxic tumor regions, namely palmitoylcarnitine and stearoylcarnitine[21]. Carnitine and acetylcarnitine are involved in the mitochondrial metabolism of acetyl coenzyme A (acetyl-CoA). Acetyl-CoA is converted to CoA and acetylcarnitine in the presence of carnitine[174]. The free CoA can then be used for fatty acid oxidation and in the citric acid cycle. Hypoxia is associated with a perturbation of CoA homeostasis and an increase in the ratio between acylcarnitines and free carnitine[175].

5.4 Conclusions

With MSI, the distribution of a wide variety of biomolecules can be studied. Ideally, these distributions are directly correlated with tissue regions of interest. However, the markers that delineate these tissue regions can typically only be studied with other imaging techniques, as for example IHC. We have used the exogenous marker pimonidazole for direct detection of hypoxic tissue regions in a breast tumor xenograft model, thereby avoiding the co-registration of MSI data with other imaging data.

A MALDI-MSI approach is presented that combines the detection of pimonidazole and a hypoxic pimonidazole metabolite with the multiplexing capabilities of the technique. We have used this metabolite to image hypoxic tissue regions and their associated biomolecules. Several endogenous species localized to hypoxic tissue regions as defined by the hypoxic pimonidazole metabolite. Interestingly, the identified species are known to be involved in hypoxia or metabolic reprogramming in cancer, although their specific roles remain to be elucidated. Pimonidazole is a widely used marker of tissue hypoxia. We expect that the presented MALDI-MSI approach is also applicable to other tissues from pimonidazole-injected animals or humans.

6

Conclusions and future research

6.1 Conclusions

At the moment, the MSI field is moving from proof-of-principle studies to biomedical and clinical applications. The popularity of the technique is exemplified by the many occurrences of 'MSI' in titles of applied research papers. However, the presence of 'MSI' in a title indicates that the used technique is just as important as the biomedical or clinical research results. As MSI becomes more established, its name will be gradually replaced by these results.

This thesis has explored the use of TMAs for MALDI-MSI. We have investigated the feasibility of using MALDI-MSI data obtained from the analysis of TMAs to predict treatment response and disease progression in cancer. In addition, we have investigated the use of a chemical marker to detect hypoxic tumor regions. For this purpose, we have developed a method to detect the exogenous hypoxia marker pimonidazole directly from tumor tissue using MALDI-MSI.

The use of TMAs for MSI

TMAs are commonly used for high-throughput analysis of cancer tissue. The small size of the tissue cores and array layout enable the analysis of hundreds of cores within a single MALDI-MSI experiment, and thus under highly similar experimental conditions. In addition, clinical follow-up data is typically available for these samples. Combined with MSI, TMAs can be used to find and validate diagnostic or prognostic markers.

In **Chapter 3** and **Chapter 4** we have shown the mass spectra that can be obtained with MALDI-MSI analysis of TMAs. We have used these spectra to predict treatment response and disease progression using different (multivariate) data analysis methods and classifiers. We have observed a small predictive power for some

of the investigated approaches. However, we have also described several issues concerning the MALDI-MSI analysis of TMAs. In general, the detected biological and experimental variability between the tissue cores limited the classification accuracies that we could obtain. The observed variability was larger between the patient samples (**Chapter 4**) than between the patient-derived xenograft (PDX) samples (**Chapter 3**). This can be partially attributed to the controlled laboratory environment in which the PDX tumors were grown and harvested. In Section 6.2, suggestions for improvements are presented.

As a side note, even though clinically useful biomarkers or biomarker patterns are heavily sought after, only a limited number of attempts have been successful. For example, DNA sequencing technology has detected only small differences for the cancer types investigated in this thesis (triple-negative breast cancer and head and neck cancer).

Hypoxia detection using pimonidazole

In **Chapter 5** we have presented a method to detect the hypoxia marker pimonidazole directly from fresh frozen tumor tissue using MALDI-MSI. We have shown that a pimonidazole metabolite can be used to image hypoxic regions and hypoxia-associated lipids and metabolites in a single experiment; no co-registration of MSI data with immunohistochemical (IHC) data is needed.

It has become a standard procedure to stain the tissue section after MSI analysis (typically with hematoxylin and eosin, H&E). The use of one tissue section is more accurate as compared to co-registered adjacent sections. Consecutive sections may have different shapes due to sectioning and preparation of the sections for MSI, histological staining or IHC. Moreover, tumor tissue does consist of easily recognizable structures like brain tissue and is therefore more challenging to co-register. In principle, it is also possible to perform IHC after MSI on the same tissue section. Instead, we have designed a method that omits IHC staining altogether, which is faster, suitable to fragile tissue sections and potentially more accurate.

The presented method facilitates the study of hypoxia-associated lipids and metabolites. These molecular classes play a role in cancer processes, but have not been studied as much as proteins. We have identified a heterogeneous distribution of 1-methylnicotinamide and acetylcarnitine, which mostly co-localized with hypoxic

tumor regions. Ultimately, endogenous molecules might be identified that can serve as hypoxia markers for MSI.

We have provided some insight in the complex *in vivo* metabolism of pimonidazole. To the best of our knowledge this is the first study of pimonidazole metabolism using mass spectrometry.

6.2 Outlook

There are several directions in which the research presented in this thesis can be continued and improved.

Study design and sample preparation

TMAs were originally designed to facilitate the large-scale investigation of diagnostic and prognostic markers using techniques such as IHC and fluorescence *in situ* hybridization. Recently, sample preparation protocols have been developed that make TMAs amenable to MALDI-MSI analysis. However, so far only limited attention has been paid to the influence of sample preparation and storage on the results (see Section 4.2). A thorough investigation of the effects of formalin fixation, paraffin embedding and sample storage would therefore be valuable. Mass spectrometric studies using homogenised samples do not report major effects of storage time, as reviewed by Shi *et al.*[176]. However, the differences in sample preparation between these studies and MSI studies do not exclude a potential effect: sample preparation for MSI typically is less 'harsh' to retain tissue integrity, which might lead to inefficient protein 'unlocking'. Future improvements in sample preparation could focus on improving the reversal of protein aggregation and cross-linking, maybe combined with the testing of other proteolytic enzymes such as Lys-C, Arg-C and r-Lys-N.

For each tumor sample in a TMA only a small tissue piece is measured. This piece might not be representative for the tumor. With the current improvements in measurement speed of mass spectrometers, the analysis of tens of tissue sections (with a diameter of >1 cm, instead of 0.6 mm for TMA tissue cores) can be performed in a relatively small amount of time. The analysis of larger tissue sections as compared to tissue cores also provides information on intratumor heterogeneity.

So far, the studies that were most successful in the identification of prognostic markers performed intact protein analysis on fresh frozen tissue sections[114, 119]. In these studies, TMAs were not used for the discovery set, but for the immunohisto-

chemical validation of the identified markers. A recent study by Dekker *et al.* shows the reproducibility of this type of approach[24]. Two independent studies in different laboratories were able to confirm three out of four peaks associated with tumor-activated stroma in breast cancer.

Data analysis

Clinical samples show a high degree of variability, and therefore clinical studies require large numbers of samples. In these studies, standardization and quality control are important to ensure reproducibility of the results. Several research fields are more advanced on these topics, such as the genomics and proteomics fields[177]. The development of standard software for quality control and data analysis will greatly facilitate clinical MSI studies. The embedding of dedicated (bio)informaticians in research groups is therefore indispensable.

Chemical markers

Chapter 5 has demonstrated the detection of an administered chemical marker to investigate intratumor heterogeneity with MALDI-MSI. Future research might be directed towards the design of a marker with optimal chemical properties for detection by MALDI-MSI. The pimonidazole compound was chosen for this study, because it is an established hypoxia marker. This allowed us to compare detection of this compound by MSI with standard detection by IHC. However, pimonidazole's chemical properties are not ideal for MALDI-MSI detection; it fragments during the MALDI process, which leads to reduced sensitivity and spectral overlap of MALDI-induced fragments and pimonidazole metabolites. The incorporation of an UV-cleavable tag in the side chain, such as used in the 'Tag-Mass' and 'TAMSIM' concepts might improve its detection[86, 87]. The side chain is not involved in adduct formation, but serves as antibody binding site. It can thus be modified provided the modification does not reduce compound uptake in the tumor. The addition of an UV-cleavable tag would allow the detection of protein adducts instead of metabolites. This might have several advantages over the current method, including increased specificity and sensitivity.

General

In recent years, a number of topics have emerged where MSI has the potential to make a contribution, such as prediction of disease progression, the delineation of tumor margins during surgery and the investigation of intratumor heterogeneity. Mass spectrometers are still becoming faster, more sensitive, precise and accurate, which greatly facilitates MSI studies. However, MSI has a few inherent limitations, such as a lack of analytical depth, which also limit its application.

There will be an increasing need for researchers with computing skills and statistical knowledge to fully explore the large and complex MSI data sets. Only close collaborations of researchers with expertise in MSI, histo(patho)logy and data analysis will exploit the full potential of MSI.

References

[1] R.A. Burrell, N. McGranahan, J. Bartek, and C. Swanton, *The causes and consequences of genetic heterogeneity in cancer evolution*, Nature 501 (2013) 338-345.

[2] A. Marusyk, V. Almendro, and K. Polyak, *Intra-tumour heterogeneity: a looking glass for cancer?*, Nature Reviews Cancer 12 (2012) 323-334.

[3] N.A. Saunders, F. Simpson, E.W. Thompson, M.M. Hill, L. Endo-Munoz, G. Leggatt, R.F. Minchin, and A. Guminski, *Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives*, Embo Molecular Medicine 4 (2012) 675-684.

[4] M. Hockel, K. Schlenger, B. Aral, M. Mitze, U. Schaffer, and P. Vaupel, *Association between tumor hypoxia and malignant progression in advanced cancer of the uterine cervix*, Cancer Research 56 (1996) 4509-4515.

[5] C. Grau and J. Overgaard, *Effect of Etoposide, Carmustine, Vincristine, 5-Fluorouracil, or Methotrexate on Radiobiologically Oxic and Hypoxic Cells in a C3h Mouse Mammary-Carcinoma Insitu*, Cancer Chemotherapy and Pharmacology 30 (1992) 277-280.

[6] J.C. Mottram, *A factor of importance in the radiosensitivity of tumours*, Br J of Radiology 9 (1936) 606-614.

[7] C. Ruiz, E. Lenkiewicz, L. Evers, T. Holley, A. Robeson, J. Kiefer, M.J. Demeure, M.A. Hollingsworth, M. Shen, D. Prunkard, P.S. Rabinovitch, T. Zellweger, S. Mousses, J.M. Trent, J.D. Carpten, L. Bubendorf, D. Von Hoff, and M.T. Barrett, *Advancing a clinically relevant perspective of the clonal nature of cancer*, Proceedings of the National Academy of Sciences of the United States of America 108 (2011) 12054-12059.

[8] L. Ding, T.J. Ley, D.E. Larson, C.A. Miller, D.C. Koboldt, J.S. Welch, J.K. Ritchey, M.A. Young, T. Lamprecht, M.D. McLellan, J.F. McMichael, J.W. Wallis, C. Lu, D. Shen, C.C. Harris, D.J. Dooling, R.S. Fulton, L.L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V.J. Magrini, L. Cook, S.D. McGrath, T.L. Vickery, M.C. Wendl, S. Heath, M.A. Watson, D.C. Link, M.H. Tomasson, W.D. Shannon, J.E. Payton, S. Kulkarni, P. Westervelt, M.J. Walter, T.A. Graubert, E.R. Mardis, R.K. Wilson, and J.F. DiPersio, *Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing*, Nature 481 (2012) 506-510.

[9] K. Polyak, *Heterogeneity in breast cancer*, J Clin Invest 121 (2011) 3786-3788.

[10] P.L. Bedard, A.R. Hansen, M.J. Ratain, and L.L. Siu, *Tumour heterogeneity in the clinic*, Nature 501 (2013) 355-364.

[11] M.C. Asselin, J.P. O'Connor, R. Boellaard, N.A. Thacker, and A. Jackson, *Quantifying heterogeneity in human tumours using MRI and PET*, Eur J Cancer 48 (2012) 447-455.

[12] C.C. Cyran, P.M. Paprottka, M. Eisenblatter, D.A. Clevert, C. Rist, K. Nikolaou, K. Lauber, F. Wenz, D. Hausmann, M.F. Reiser, C. Belka, and M. Niyazi, *Visualization, imaging and new preclinical diagnostics in radiation oncology*, Radiat Oncol 9 (2014) 3.

[13] H. Liebl, *Ion Microprobe Mass Analyzer*, J Appl Phys 38 (1967) 5277-5283.

[14] R.M. Caprioli, T.B. Farmer, and J. Gile, *Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS*, Anal Chem 69 (1997) 4751-4760.

[15] J.P. Shapiro, S. Biswas, A.S. Merchant, A. Satoskar, C. Taslim, S.L. Lin, B.H. Rovin, C.K. Sen, S. Roy, and M.A. Freitas, *A quantitative proteomic workflow for characterization of frozen clinical biopsies: Laser capture microdissection coupled with label-free mass spectrometry*, Journal of Proteomics 77 (2012) 433-440.

[16] N.Q. Liu, R.B.H. Braakman, C. Stingl, T.M. Luider, J.W.M. Martens, J.A. Foekens, and A. Umar, *Proteomics Pipeline for Biomarker Discovery of Laser Capture Microdissected Breast Cancer Tissue*, Journal of Mammary Gland Biology and Neoplasia 17 (2012) 155-164.

[17] M.R. Groseclose, P.P. Massion, P. Chaurand, and R.M. Caprioli, *High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry*, Proteomics 8 (2008) 3715-3724.

[18] R. Casadonte and R.M. Caprioli, *Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry*, Nature Protocols 6 (2011) 1695-1709.

[19] R. Lazova, E.H. Seeley, M. Keenan, R. Gueorguieva, and R.M. Caprioli, *Imaging mass spectrometry--a new and promising method to differentiate Spitz nevi from Spitzoid malignant melanomas*, Am J Dermatopathol 34 (2012) 82-90.

[20] B. Balluff, C.K. Frese, S.K. Maier, C. Schone, B. Kuster, M. Schmitt, M. Aubele, H. Hofler, A.M. Deelder, A. Heck, Jr., P.C. Hogendoorn, J. Morreau, A.F. Maarten Altelaar, A. Walch, and L.A. McDonnell, *De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry*, J Pathol 235 (2015) 3-13.

[21] K. Chughtai, L. Jiang, T.R. Greenwood, K. Glunde, and R.M. Heeren, *Mass spectrometry images acylcarnitines, phosphatidylcholines, and sphingomyelin in MDA-MB-231 breast tumor models*, Journal of Lipid Research 54 (2013) 333-344.

[22] M. Hilario, A. Kalousis, C. Pellegrini, and M. Muller, *Processing and classification of protein mass spectra*, Mass Spectrometry Reviews 25 (2006) 409-449.

[23] E.A. Jones, S.O. Deininger, P.C.W. Hogendoorn, A.M. Deelder, and L.A. McDonnell, *Imaging mass spectrometry statistical analysis*, Journal of Proteomics 75 (2012) 4962-4989.

[24] T.J.A. Dekker, B.D. Balluff, E.A. Jones, C.D. Schone, M. Schmitt, M. Aubele, J.R. Kroep, V.T.H.B.M. Smit, R.A.E.M. Tollenaar, W.E. Mesker, A. Walch, and L.A. McDonnell, *Multicenter Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging (MALDI MSI) Identifies Proteomic Differences in Breast-Cancer-Associated Stroma*, Journal of Proteome Research 13 (2014) 4730-4738.

[25] R. Lemaire, S.A. Menguellet, J. Stauber, V. Marchaudon, J.P. Lucot, P. Collinet, M.O. Farine, D. Vinatier, R. Day, P. Ducoroy, M. Salzet, and I. Fournier, *Specific MALDI imaging and profiling for biomarker hunting and validation: Fragment of the 11S proteasome activator complex, reg alpha fragment, is a new potential ovary cancer biomarker*, Journal of Proteome Research 6 (2007) 4127-4134.

[26] K. Schwamborn, *Imaging mass spectrometry in biomarker discovery and validation*, Journal of Proteomics 75 (2012) 4990-4998.

[27] D.S. Cornett, M.L. Reyzer, P. Chaurand, and R.M. Caprioli, *MALDI imaging mass spectrometry: molecular snapshots of biochemical systems*, Nat Methods 4 (2007) 828-833.

[28] K. Chughtai and R.M.A. Heeren, *Mass Spectrometric Imaging for Biomedical Tissue Analysis*, Chemical Reviews 110 (2010) 3237-3277.

[29] D.M. Drexler, T.J. Garrett, J.L. Cantone, R.W. Diters, J.G. Mitroka, M.C. Prieto Conaway, S.P. Adams, R.A. Yost, and M. Sanders, *Utility of imaging mass spectrometry (IMS) by matrix-assisted laser desorption ionization (MALDI) on an ion trap mass spectrometer in the analysis of drugs and metabolites in biological tissues*, J Pharmacol Toxicol Methods 55 (2007) 279-288.

[30] Y. Hsieh, J. Chen, and W.A. Korfmacher, *Mapping pharmaceuticals in tissues using MALDI imaging mass spectrometry*, J Pharmacol Toxicol Methods 55 (2007) 193-200.

[31] S. Rauser, S.O. Deininger, D. Suckau, H. Hofler, and A. Walch, *Approaching MALDI molecular imaging for clinical proteomic research: current state and fields of application*, Expert Rev Proteomics 7 (2010) 927-941.

[32] J. Franck, K. Arafah, M. Elayed, D. Bonnel, D. Vergara, A. Jacquet, D. Vinatier, M. Wisztorski, R. Day, I. Fournier, and M. Salzet, *MALDI imaging mass spectrometry: state of the art technology in clinical proteomics*, Mol Cell Proteomics 8 (2009) 2023-2033.

[33] B. Balluff, C. Schone, H. Hofler, and A. Walch, *MALDI imaging mass spectrometry for direct tissue analysis: technological advancements and recent applications*, Histochem Cell Biol (2011).

[34] L.A. McDonnell, G.L. Corthals, S.M. Willems, A. van Remoortere, R.J. van Zeijl, and A.M. Deelder, *Peptide and protein imaging mass spectrometry in cancer research*, J Proteomics 73 (2010) 1921-1944.

[35] E.H. Seeley and R.M. Caprioli, *MALDI imaging mass spectrometry of human tissue: method challenges and clinical perspectives*, Trends Biotechnol 29 (2011) 136-143.

[36] S.M. Willems, A. van Remoortere, R. van Zeijl, A.M. Deelder, L.A. McDonnell, and P.C.W. Hogendoorn, *Imaging mass spectrometry of myxoid sarcomas identifies proteins and lipids specific to tumour type and grade, and reveals biochemical intratumour heterogeneity*, Journal of Pathology 222 (2010) 400-409.

[37] S.O. Deininger, M.P. Ebert, A. Futterer, M. Gerhard, and C. Rocken, *MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers*, J Proteome Res 7 (2008) 5230-5236.

[38] M. Karas, D. Bachmann, and F. Hillenkamp, *Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules*, Analytical Chemistry 57 (1985) 2935-2939.

[39] P. Chaurand, S.A. Schwartz, M.L. Reyzer, and R.M. Caprioli, *Imaging mass spectrometry: principles and potentials*, Toxicol Pathol 33 (2005) 92-101.

[40] B. Balluff, M. Elsner, A. Kowarsch, S. Rauser, S. Meding, C. Schuhmacher, M. Feith, K. Herrmann, C. Rocken, R.M. Schmid, H. Hofler, A. Walch, and M.P. Ebert, *Classification of HER2/neu status in gastric cancer using a breast-cancer derived proteome classifier*, Journal of Proteome Research 9 (2010) 6317-6322.

[41] J.D. Watrous, T. Alexandrov, and P.C. Dorrestein, *The evolving field of imaging mass spectrometry and its impact on future biological research*, Journal of Mass Spectrometry 46 (2011) 209-222.

[42] B. Domon and R. Aebersold, *Mass spectrometry and protein analysis*, Science 312 (2006) 212-217.

[43] J.R. Yates, C.I. Ruse, and A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications*, Annu Rev Biomed Eng 11 (2009) 49-79.

[44] L. Minerva, S. Clerens, G. Baggerman, and L. Arckens, *Direct profiling and identification of peptide expression differences in the pancreas of control and ob/ob mice by imaging mass spectrometry*, Proteomics 8 (2008) 3763-3774.

[45] S. Rauser, C. Marquardt, B. Balluff, S.O. Deininger, C. Albers, E. Belau, R. Hartmer, D. Suckau, K. Specht, M.P. Ebert, M. Schmitt, M. Aubele, H. Hofler, and A. Walch, *Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry*, J Proteome Res 9 (2010) 1854-1863.

[46] M. Lagarrigue, M. Becker, R. Lavigne, S.O. Deininger, A. Walch, F. Aubry, D. Suckau, and C. Pineau, *Revisiting rat spermatogenesis with MALDI imaging at 20-microm resolution*, Mol Cell Proteomics 10 (2011) M110 005991.

[47] H. Meistermann, J.L. Norris, H.R. Aerni, D.S. Cornett, A. Friedlein, A.R. Erskine, A. Augustin, M.C. De Vera Mudry, S. Ruepp, L. Suter, H. Langen, R.M. Caprioli, and A. Ducret, *Biomarker discovery by imaging mass spectrometry: transthyretin is a biomarker for gentamicin-induced nephrotoxicity in rat*, Mol Cell Proteomics 5 (2006) 1876-1886.

[48] D. Debois, V. Bertrand, L. Quinton, M.C. De Pauw-Gillet, and E. De Pauw, *MALDI-in source decay applied to mass spectrometry imaging: a new tool for protein identification*, Anal Chem 82 (2010) 4036-4045.

[49] Y.M. Fung, C.M. Adams, and R.A. Zubarev, *Electron ionization dissociation of singly and multiply charged peptides*, J Am Chem Soc 131 (2009) 9977-9985.

[50] L.H. Cazares, D. Troyer, S. Mendrinos, R.A. Lance, J.O. Nyalwidhe, H.A. Beydoun, M.A. Clements, R.R. Drake, and O.J. Semmes, *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 15 (2009) 5541-5551.

[51] J.A. McLean, W.B. Ridenour, and R.M. Caprioli, *Profiling and imaging of tissues by imaging ion mobility-mass spectrometry*, Journal of Mass Spectrometry 42 (2007) 1099-1105.

[52] M.R. Groseclose, M. Andersson, W.M. Hardesty, and R.M. Caprioli, *Identification of proteins directly from tissue: in situ tryptic digestions coupled with imaging mass spectrometry*, Journal of Mass Spectrometry 42 (2007) 254-262.

[53] M.C. Djidja, S. Francese, P.M. Loadman, C.W. Sutton, P. Scriven, E. Claude, M.F. Snel, J. Franck, M. Salzet, and M.R. Clench, *Detergent addition to tryptic digests and ion mobility separation prior to MS/MS improves peptide yield and protein identification for in situ proteomic investigation of frozen and formalin-fixed paraffin-embedded adenocarcinoma tissue sections*, Proteomics 9 (2009) 2750-2763.

[54] R. Lemaire, A. Desmons, J.C. Tabet, R. Day, M. Salzet, and I. Fournier, *Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections*, Journal of Proteome Research 6 (2007) 1295-1305.

[55] M.C. Djidja, E. Claude, M.F. Snel, P. Scriven, S. Francese, V. Carolan, and M.R. Clench, *MALDI-ion mobility separation-mass spectrometry imaging of glucose-regulated protein 78 kDa (Grp78) in human formalin-fixed, paraffin-embedded pancreatic adenocarcinoma tissue sections*, J Proteome Res 8 (2009) 4876-4884.

[56] M. Ronci, E. Bonanno, A. Colantoni, L. Pieroni, C. Di Ilio, L.G. Spagnoli, G. Federici, and A. Urbani, *Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations*, Proteomics 8 (2008) 3702-3714.

[57] J. Franck, M. El Ayed, M. Wisztorski, M. Salzet, and I. Fournier, *On-tissue N-terminal peptide derivatizations for enhancing protein identification in MALDI mass spectrometric imaging strategies*, Anal Chem 81 (2009) 8305-8317.

[58] L.M. Cole, M.C. Djidja, J. Bluff, E. Claude, V.A. Carolan, M. Paley, G.M. Tozer, and M.R. Clench, *Investigation of protein induction in tumour vascular targeted strategies by MALDI MSI*, Methods 54 (2011) 442-453.

[59] J. Stauber, L. MacAleese, J. Franck, E. Claude, M. Snel, B.K. Kaletas, I.M. Wiel, M. Wisztorski, I. Fournier, and R.M. Heeren, *On-tissue protein identification and imaging by MALDI-ion mobility mass spectrometry*, J Am Soc Mass Spectrom 21 (2010) 338-347.

[60] J. Stauber, R. Lemaire, J. Franck, D. Bonnel, D. Croix, R. Day, M. Wisztorski, I. Fournier, and M. Salzet, *MALDI imaging of formalin-fixed paraffin-embedded tissues: application to model animals of Parkinson disease for biomarker hunting*, J Proteome Res 7 (2008) 969-978.

[61] E.B. Monroe, S.P. Annangudi, N.G. Hatcher, H.B. Gutstein, S.S. Rubakhin, and J.V. Sweedler, *SIMS and MALDI MS imaging of the spinal cord*, Proteomics 8 (2008) 3746-3754.

[62] L. MacAleese, M.C. Duursma, L.A. Klerk, G. Fisher, and R.M. Heeren, *Protein identification with Liquid Chromatography and Matrix Enhanced Secondary Ion Mass Spectrometry (LC-ME-SIMS)*, J Proteomics 74 (2011) 993-1001.

[63] D. Mustafa, J.M. Kros, and T. Luider, *Combining laser capture microdissection and proteomics techniques*, Methods Mol Biol 428 (2008) 159-178.

[64] R.B. Braakman, M.M. Tilanus-Linthorst, N.Q. Liu, C. Stingl, L.J. Dekker, T.M. Luider, J.W. Martens, J.A. Foekens, and A. Umar, *Optimized nLC-MS workflow for laser capture microdissected breast cancer tissue*, J Proteomics (2012).

[65] J.O.R. Gustafsson, M.K. Oehler, S.R. McColl, and P. Hoffmann, *Citric Acid Antigen Retrieval (CAAR) for Tryptic Peptide Imaging Directly on Archived Formalin-Fixed Paraffin-Embedded Tissue*, Journal of Proteome Research 9 (2010) 4315-4328.

[66] S.R. Oppenheimer, D. Mi, M.E. Sanders, and R.M. Caprioli, *Molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma*, J Proteome Res 9 (2010) 2182-2190.

[67] A.F.M. Altelaar, I.M. Taban, L.A. McDonnell, P.D.E.M. Verhaert, R.P.J. de Lange, R.A.H. Adan, W.J. Mooi, R.M.A. Heeren, and S.R. Piersma, *High-resolution MALDI imaging mass spectrometry allows localization of peptide distributions at cellular length scales in pituitary tissue sections*, International Journal of Mass Spectrometry 260 (2007) 203-211.

[68] Y. Schober, T. Schramm, B. Spengler, and A. Rompp, *Protein identification by accurate mass matrix-assisted laser desorption/ionization imaging of tryptic peptides*, Rapid Commun Mass Spectrom 25 (2011) 2475-2483.

[69] A. Rompp, S. Guenther, Y. Schober, O. Schulz, Z. Takats, W. Kummer, and B. Spengler, *Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging*, Angew Chem Int Ed Engl 49 (2010) 3834-3838.

[70] I.M. Taban, A.F. Altelaar, Y.E. van der Burgt, L.A. McDonnell, R.M. Heeren, J. Fuchser, and G. Baykut, *Imaging of peptides in the rat brain using MALDI-FTICR mass spectrometry*, J Am Soc Mass Spectrom 18 (2007) 145-151.

[71] T. Liu, M.E. Belov, N. Jaitly, W.J. Qian, and R.D. Smith, *Accurate mass measurements in proteomics*, Chemical Reviews 107 (2007) 3621-3653.

[72] R. Zubarev and M. Mann, *On the proper use of mass accuracy in proteomics*, Molecular & Cellular Proteomics 6 (2007) 377-381.

[73] Y. Schober, S. Guenther, B. Spengler, and A. Rompp, *High-resolution matrix-assisted laser desorption/ionization imaging of tryptic peptides from tissue*, Rapid Communications in Mass Spectrometry 26 (2012) 1141-1146.

[74] L. Minerva, K. Boonen, G. Menschaert, B. Landuyt, G. Baggerman, and L. Arckens, *Linking mass spectrometric imaging and traditional peptidomics: a validation in the obese mouse model*, Anal Chem 83 (2011) 7682-7691.

[75] W.S. Noble and M.J. MacCoss, *Computational and statistical analysis of protein mass spectrometry data*, PLoS Comput Biol 8 (2012) e1002296.

[76] R. Ketterlinus, S.Y. Hsieh, S.H. Teng, H. Lee, and W. Pusch, *Fishing for biomarkers: analyzing mass spectrometry data with the new ClinProTools software*, BioTechniques Suppl (2005) 37-40.

[77] L.A. McDonnell, A. van Remoortere, N. de Velde, R.J. van Zeijl, and A.M. Deelder, *Imaging mass spectrometry data reduction: automated feature identification and extraction*, J Am Soc Mass Spectrom 21 (2010) 1969-1978.

[78] W.M. Hardesty, M.C. Kelley, D.M. Mi, R.L. Low, and R.M. Caprioli, *Protein signatures for survival and recurrence in metastatic melanoma*, Journal of Proteomics 74 (2011) 1002-1014.

[79] J. Le Faouder, S. Laouirem, M. Chapelle, M. Albuquerque, J. Belghiti, F. Degos, V. Paradis, J.M. Camadro, and P. Bedossa, *Imaging mass spectrometry provides fingerprints for distinguishing hepatocellular carcinoma from cirrhosis*, J Proteome Res 10 (2011) 3755-3765.

[80] J. Bruand, S. Sistla, C. Meriaux, P.C. Dorrestein, T. Gaasterland, M. Ghassemian, M. Wisztorski, I. Fournier, M. Salzet, E. Macagno, and V. Bafna, *Automated querying and identification of novel peptides using MALDI mass spectrometric imaging*, J Proteome Res 10 (2011) 1915-1928.

[81] D.S. Cornett, J.A. Mobley, E.C. Dias, M. Andersson, C.L. Arteaga, M.E. Sanders, and R.M. Caprioli, *A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer*, Mol Cell Proteomics 5 (2006) 1975-1983.

[82] M. Hanselmann, M. Kirchner, B.Y. Renard, E.R. Amstalden, K. Glunde, R.M.A. Heeren, and F.A. Hamprecht, *Concise Representation of Mass Spectrometry Images by Probabilistic Latent Semantic Analysis*, Anal Chem 80 (2008) 9649-9658.

[83] D. Bonnel, R. Longuespee, J. Franck, M. Roudbaraki, P. Gosset, R. Day, M. Salzet, and I. Fournier, *Multivariate analyses for biomarkers hunting and validation through on-tissue bottom-up or in-source decay in MALDI-MSI: application to prostate cancer*, Analytical and Bioanalytical Chemistry 401 (2011) 149-165.

[84] G. McCombie, D. Staab, M. Stoeckli, and R. Knochenmuss, *Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis*, Anal Chem 77 (2005) 6118-6124.

[85] E.A. Jones, A. van Remoortere, R.J.M. van Zeijl, P.C.W. Hogendoorn, J.V.M.G. Bovee, A.M. Deelder, and L.A. McDonnell, *Multiple Statistical Analysis Techniques Corroborate Intratumor Heterogeneity in Imaging Mass Spectrometry Datasets of Myxofibrosarcoma*, Plos One 6 (2011).

[86] G. Thiery, M.S. Shchepinov, E.M. Southern, A. Audebourg, V. Audard, B. Terris, and L.G. Gut, *Multiplex target protein imaging in tissue sections by mass spectrometry - TAMSIM*, Rapid Communications in Mass Spectrometry 21 (2007) 823-829.

[87] R. Lemaire, J. Stauber, M. Wisztorski, C. Van Camp, A. Desmons, M. Deschamps, G. Proess, I. Rudlof, A.S. Woods, R. Day, M. Salzet, and I. Fournier, *Tag-mass: Specific molecular imaging of transcriptome and proteome by mass spectrometry based on photocleavable tag*, Journal of Proteome Research 6 (2007) 2057-2067.

[88] J. Stauber, M.E. Ayed, M. Wisztorski, M. Salzet, and I. Fournier, *Specific MALDI-MSI: Tag-Mass*, Methods Mol Biol 656 (2010) 339-361.

[89] C. Giesen, H.A.O. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P.J. Schuffler, D. Grolimund, J.M. Buhmann, S. Brandt, Z. Varga, P.J. Wild, D. Gunther, and B. Bodenmillerthat, *Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry*, Nature Methods 11 (2014) 417-422.

[90] E.R. Amstalden van Hove, T.R. Blackwell, I. Klinkert, G.B. Eijkel, R.M. Heeren, and K. Glunde, *Multimodal mass spectrometric imaging of small molecules reveals distinct spatio-molecular signatures in differentially metastatic breast tumor models*, Cancer Res 70 (2010) 9012-9021.

[91] K. Chughtai, L. Jiang, T.R. Greenwood, I. Klinkert, E.R.A. van Hove, R.M.A. Heeren, and K. Gunde, *Fiducial Markers for Combined 3-Dimensional Mass Spectrometric and Optical Tissue Imaging*, Analytical Chemistry 84 (2012) 1817-1823.

[92] E.D. Inutan, A.L. Richards, J. Wager-Miller, K. Mackie, C.N. McEwen, and S. Trimpin, *Laserspray ionization, a new method for protein analysis directly from tissue at atmospheric pressure with ultrahigh mass resolution and electron transfer dissociation*, Mol Cell Proteomics 10 (2011) M110 000760.

[93] S. Meding, U. Nitsche, B. Balluff, M. Elsner, S. Rauser, C. Schone, M. Nipp, M. Maak, M. Feith, M.P. Ebert, H. Friess, R. Langer, H. Hofler, H. Zitzelsberger, R. Rosenberg, and A. Walch, *Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging*, J Proteome Res 11 (2012) 1996-2003.

[94] M.C. Djidja, E. Claude, M.F. Snel, S. Francese, P. Scriven, V. Carolan, and M.R. Clench, *Novel molecular tumour classification using MALDI-mass spectrometry imaging of tissue micro-array*, Analytical and Bioanalytical Chemistry 397 (2010) 587-601.

[95] M.L. Reyzer, R.L. Caldwell, T.C. Dugger, J.T. Forbes, C.A. Ritter, M. Guix, C.L. Arteaga, and R.M. Caprioli, *Early changes in protein expression detected by mass spectrometry predict tumor response to molecular therapeutics*, Cancer Research 64 (2004) 9093-9100.

[96] J.A. Bauer, A.B. Chakravarthy, J.M. Rosenbluth, D. Mi, E.H. Seeley, N. De Matos Granja-Ingram, M.G. Olivares, M.C. Kelley, I.A. Mayer, I.M. Meszoely, J.A. Means-Powell, K.N. Johnson, C.J. Tsai, G.D. Ayers, M.E. Sanders, R.J. Schneider, S.C. Formenti, R.M. Caprioli, and J.A. Pietenpol, *Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation*, Clin Cancer Res 16 (2010) 681-690.

[97] M. Aichler, M. Elsner, N. Ludyga, A. Feuchtinger, V. Zangen, S.K. Maier, B. Balluff, C. Schone, L. Hierber, H. Braselmann, S. Meding, S. Rauser, H. Zischka, M. Aubele, M. Schmitt, M. Feith, S.M. Hauck, M. Ueffing, R. Langer, B. Kuester, H. Zitzelsberger, H. Hofler, and A.K. Walch, *Clinical response to chemotherapy in oesophageal adenocarcinoma patients is linked to defects in mitochondria*, Journal of Pathology (2013).

[98] B. Cillero-Pastor, G.B. Eijkel, A. Kiss, F.J. Blanco, and R.M. Heeren, *Matrix-assisted laser desorption ionization-imaging mass spectrometry: a new methodology to study human osteoarthritic cartilage*, Arthritis Rheum 65 (2013) 710-720.

[99] S. Gerbig, O. Golf, J. Balog, J. Denes, Z. Baranyai, A. Zarand, E. Raso, J. Timar, and Z. Takats, *Analysis of colorectal adenocarcinoma tissue by desorption electrospray ionization mass spectrometric imaging*, Analytical and Bioanalytical Chemistry 403 (2012) 2315-2325.

[100] K.A. Veselkov, R. Mirnezami, N. Strittmatter, R.D. Goldin, J. Kinross, A.V. Speller, T. Abramov, E.A. Jones, A. Darzi, E. Holmes, J.K. Nicholson, and Z. Takats, *Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer*, Proc Natl Acad Sci U S A 111 (2014) 1216-1221.

[101] B. Cillero-Pastor, G. Eijkel, A. Kiss, F.J. Blanco, and R.M. Heeren, *Time-of-flight secondary ion mass spectrometry-based molecular distribution distinguishing healthy and osteoarthritic human cartilage*, Analytical Chemistry 84 (2012) 8909-8916.

[102] K.R. Bauer, M. Brown, R.D. Cress, C.A. Parise, and V. Caggiano, *Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry*, Cancer 109 (2007) 1721-1728.

[103] L.A. Carey, E.C. Dees, L. Sawyer, L. Gatti, D.T. Moore, F. Collichio, D.W. Ollila, C.I. Sartor, M.L. Graham, and C.M. Perou, *The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes*, Clin Cancer Res 13 (2007) 2329-2334.

[104] S. Cleator, W. Heller, and R.C. Coombes, *Triple-negative breast cancer: therapeutic options*, Lancet Oncology 8 (2007) 235-244.

[105] D.P. Silver, A.L. Richardson, A.C. Eklund, Z.C. Wang, Z. Szallasi, Q. Li, N. Juul, C.O. Leong, D. Calogrias, A. Buraimoh, A. Fatima, R.S. Gelman, P.D. Ryan, N.M. Tung, A. De Nicolo, S. Ganesan, A. Miron, C. Colin, D.C. Sgroi, L.W. Ellisen, E.P. Winer, and J.E. Garber, *Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer*, J Clin Oncol 28 (2010) 1145-1153.

[106] G.B. Eijkel, B.K. Kaletas, I.M. van der Wiel, J.M. Kros, T.M. Luider, and R.M.A. Heeren, *Correlating MALDI and SIMS imaging mass spectrometric datasets of biological tissue surfaces*, Surface and Interface Analysis 41 (2009) 675-685.

[107] R. Hoogerbrugge, S.J. Willig, and P.G. Kistemaker, *Discriminant Analysis by Double Stage Principal Component Analysis*, Analytical Chemistry 55 (1983) 1710-1712.

[108] A.W. Simonetti, W.J. Melssen, M. van der Graaf, G.J. Postma, A. Heerschap, and L.M.C. Buydens, *A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy*, Analytical Chemistry 75 (2003) 5352-5361.

[109] M.M.W.B. Hendriks, S. Smit, W.L.M.W. Akkermans, T.H. Reijmers, P.H.C. Eilers, H.C.J. Hoefsloot, C.M. Rubingh, C.G. de Koster, J.M. Aerts, and A.K. Smilde, *How to distinguish healthy from diseased? Classification strategy for mass specitrometry-based clinical proteomics*, Proteomics 7 (2007) 3672-3680.

[110] T. Alexandrov, *MALDI imaging mass spectrometry: statistical data analysis and current computational challenges*, Bmc Bioinformatics 13 (2012).

[111] D.G. DeNardo, D.J. Brennan, E. Rexhepaj, B. Ruffell, S.L. Shiao, S.F. Madden, W.M. Gallagher, N. Wadhwani, S.D. Keil, S.A. Junaid, H.S. Rugo, E.S. Hwang, K. Jirstrom, B.L. West, and L.M. Coussens, *Leukocyte Complexity Predicts Breast Cancer Survival and Functionally Regulates Response to Chemotherapy*, Cancer Discovery 1 (2011) 54-67.

[112] J.D. Brooks, *Translational genomics: The challenge of developing cancer biomarkers*, Genome Research 22 (2012) 183-187.

[113] S. Steurer, C. Borkowski, S. Odinga, M. Buchholz, C. Koop, H. Huland, M. Becker, M. Witt, D. Trede, M. Omidi, O. Kraus, A.S. Bahar, A.S. Seddiqi, J.M. Singer, M. Kwiatkowski, M. Trusch, R. Simon, M. Wurlitzer, S. Minner, T. Schlomm, G. Sauter, and H. Schluter, *MALDI mass spectrometric imaging based identification of clinically relevant signals in prostate cancer using large-scale tissue microarrays*, International Journal of Cancer 133 (2013) 920-928.

[114] B. Balluff, S. Rauser, S. Meding, M. Elsner, C. Schone, A. Feuchtinger, C. Schuhmacher, A. Novotny, U. Jutting, G. Maccarrone, H. Sarioglu, M. Ueffing, H. Braselmann, H. Zitzelsberger, R.M. Schmid, H. Hofler, M.P. Ebert, and A. Walch, *MALDI Imaging Identifies Prognostic Seven-Protein Signature of Novel Tissue Markers in Intestinal-Type Gastric Cancer*, American Journal of Pathology 179 (2011) 2720-2729.

[115] S. Meding, B. Balluff, M. Elsner, C. Schone, S. Rauser, U. Nitsche, M. Maak, A. Schafer, S.M. Hauck, M. Ueffing, R. Langer, H. Hofler, H. Friess, R. Rosenberg, and A. Walch, *Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer*, Journal of Pathology 228 (2012) 459-470.

[116] M. Nipp, M. Elsner, B. Balluff, S. Meding, H. Sarioglu, M. Ueffing, S. Rauser, K. Unger, H. Hofler, A. Walch, and H. Zitzelsberger, *S100-A10, thioredoxin, and S100-A6 as biomarkers of papillary thyroid carcinoma with lymph node metastasis identified by MALDI Imaging*, Journal of Molecular Medicine-Jmm 90 (2012) 163-174.

[117] K. Yanagisawa, Y. Shyr, B.G.J. Xu, P.P. Massion, P.H. Larsen, B.C. White, J.R. Roberts, M. Edgerton, A. Gonzalez, S. Nadaf, J.H. Moore, R.M. Caprioli, and D.P. Carbone, *Proteomic patterns of tumour subsets in non-small-cell lung cancer*, Lancet 362 (2003) 433-439.

[118] S.A. Schwartz, R.J. Weil, R.C. Thompson, Y. Shyr, J.H. Moore, S.A. Toms, M.D. Johnson, and R.M. Caprioli, *Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry*, Cancer Research 65 (2005) 7674-7681.

[119] M. Elsner, S. Rauser, S. Maier, C. Schone, B. Balluff, S. Meding, G. Jung, M. Nipp, H. Sarioglu, G. Maccarrone, M. Aichler, A. Ruchtinger, R. Langer, U. Jutting, M. Feith, B. Kuster, M. Ueffing, H. Zitzelsberger, H. Hofler, and A. Walch, *MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma*, Journal of Proteomics 75 (2012) 4693-4704.

[120] A. Quaas, A.S. Bahar, K. von Loga, A.S. Seddiqi, J.M. Singer, M. Omidi, O. Kraus, M. Kwiatkowski, M. Trusch, S. Minner, E. Burandt, P. Stahl, W. Wilczak, M. Wurlitzer, R. Simon, G. Sauter, A. Marx, and H. Schluter, *MALDI imaging on large-scale tissue microarrays identifies molecular features associated with tumour phenotype in oesophageal cancer*, Histopathology 63 (2013) 455-462.

[121] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, and D. Forman, *Global Cancer Statistics*, Ca-a Cancer Journal for Clinicians 61 (2011) 69-90.

[122] R. de Bree, R.P. Takes, J.A. Castelijns, J.E. Medina, S.J. Stoeckli, A.A. Mancuso, J.L. Hunt, J.P. Rodrigo, A. Triantafyllou, A. Teymoortash, F.J. Civantos, A. Rinaldo, K.T. Pitman, M. Hamoir, K.T. Robbins, C.E. Silver, O.S. Hoekstra, and A. Ferlito, *Advances in diagnostic modalities to detect occult lymph node metastases in head and neck squamous cell carcinoma*, Head Neck (2014).

[123] A. Teymoortash, S. Hoch, B. Eivazi, and J.A. Werner, *Postoperative morbidity after different types of selective neck dissection*, Laryngoscope 120 (2010) 924-929.

[124] S.R. van Hooff, F.K.J. Leusink, P. Roepman, R.J.B. de Jong, E.J.M. Speel, M.W.M. van den Brekel, M.L.F. van Velthuysen, P.J. van Diest, R.J.J. van Es, M.A.W. Merkx, J.A. Kummer, C.R. Leemans, E. Schuuring, J.A. Langendijk, M. Lacko, M.J. De Herdt, J.C. Jansen, R.H. Brakenhoff, P.J. Slootweg, R.P. Takes, and F.C.P. Holstege, *Validation of a Gene Expression Signature for Assessment of Lymph Node Metastasis in Oral Squamous Cell Carcinoma*, Journal of Clinical Oncology 30 (2012) 4104-4110.

[125] G.M. Polachini, L.M. Sobral, A.M. Mercante, A.F. Paes-Leme, F.C. Xavier, T. Henrique, D.M. Guimaraes, A. Vidotto, E.E. Fukuyama, J.F. Gois-Filho, P.M. Cury, O.A. Curioni, P. Michaluart, Jr., A.M. Silva, V. Wunsch-Filho, F.D. Nunes, A.M. Leopoldino, and E.H. Tajara, *Proteomic approaches identify members of cofilin pathway involved in oral tumorigenesis*, PLoS One 7 (2012) e50517.

[126] T.M. Harris, P.C. Du, N. Kawachi, T.J. Belbin, Y.H. Wang, N.F. Schlecht, T.J. Ow, C.E. Keller, G.J. Childs, R.V. Smith, R.H. Angeletti, M.B. Prystowsky, and J. Lim, *Proteomic Analysis of Oral Cavity Squamous Cell Carcinoma Specimens Identifies Patient Outcome-Associated Proteins*, Archives of Pathology & Laboratory Medicine 139 (2015) 494-507.

[127] T.J.W. Klein Nulent, P.J. Van Diest, P. van der Groep, F.K.J. Leusink, C.L.J.J. Kruitwagen, R. Koole, and E.M. Van Cann, *Cannabinoid receptor-2 immunoreactivity is associated with survival in squamous cell carcinoma of the head and neck*, British Journal of Oral & Maxillofacial Surgery 51 (2013) 604-609.

[128] S. Magdeldin and T. Yamamoto, *Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues*, Proteomics 12 (2012) 1045-1058.

[129] T.M. Morgan, E.H. Seeley, O. Fadare, R.M. Caprioli, and P.E. Clark, *Imaging the Clear Cell Renal Cell Carcinoma Proteome*, Journal of Urology 189 (2013) 1097-1103.

[130] V. Redeker, J.Y. Toullec, J. Vinh, J. Rossier, and D. Soyez, *Combination of peptide profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and immunodetection on single glands or cells*, Analytical Chemistry 70 (1998) 1805-1811.

[131] E. Jones, Oliphant, T., Peterson, P., *et al.*, *SciPy: Open source scientific tools for Python. http://www.scipy.org/*, (2001-).

[132] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research 12 (2011) 2825-2830.

[133] D. Hanahan and R.A. Weinberg, *Hallmarks of Cancer: The Next Generation*, Cell 144 (2011) 646-674.

[134] A. de Gramont, S. Watson, L.M. Ellis, J. Rodon, J. Tabernero, A. de Gramont, and S.R. Hamilton, *Pragmatic issues in biomarker evaluation for targeted therapies in cancer*, Nature Reviews Clinical Oncology 12 (2015) 197-212.

[135] R.M. Caprioli, *Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy*, Cancer Research 65 (2005) 10642-10645.
[136] B. Heijs, R.J. Carreira, E.A. Tolner, A.H. de Ru, A.J.M. van den Maagdenberg, P.A. van Veelen, and L.A. McDonnell, *Comprehensive Analysis of the Mouse Brain Proteome Sampled in Mass Spectrometry Imaging*, Analytical Chemistry 87 (2015) 1867-1875.
[137] S.E. Rademakers, P.N. Span, J.H. Kaanders, F.C. Sweep, A.J. van der Kogel, and J. Bussink, *Molecular aspects of tumour hypoxia*, Mol Oncol 2 (2008) 41-53.
[138] M. Hockel and P. Vaupel, *Tumor hypoxia: definitions and current clinical, biologic, and molecular aspects*, J Natl Cancer Inst 93 (2001) 266-276.
[139] K.A. Krohn, J.M. Link, and R.P. Mason, *Molecular imaging of hypoxia*, J Nucl Med 49 Suppl 2 (2008) 129S-148S.
[140] A. Adams, A.S. van Brussel, J.F. Vermeulen, W.P. Mali, E. van der Wall, P.J. van Diest, and S.G. Elias, *The potential of hypoxia markers as target for breast molecular imaging--a systematic review and meta-analysis of human marker expression*, BMC Cancer 13 (2013) 538.
[141] A.J. Varghese, S. Gulyas, and J.K. Mohindra, *Hypoxia-dependent reduction of 1-(2-nitro-1-imidazolyl)-3-methoxy-2-propanol by Chinese hamster ovary cells and KHT tumor cells in vitro and in vivo*, Cancer Research 36 (1976) 3761-3765.
[142] J.D. Chapman, A.J. Franko, and J. Sharplin, *A marker for hypoxic cells in tumours with potential clinical applicability*, Br J Cancer 43 (1981) 546-550.
[143] A.J. Varghese and G.F. Whitmore, *Binding to cellular macromolecules as a possible mechanism for the cytotoxicity of misonidazole*, Cancer Research 40 (1980) 2165-2169.
[144] J.A. Raleigh, G.G. Miller, A.J. Franko, C.J. Koch, A.F. Fuciarelli, and D.A. Kelly, *Fluorescence immunohistochemical detection of hypoxic cells in spheroids and tumours*, Br J Cancer 56 (1987) 395-400.
[145] S. Sobhanifar, C. Aquino-Parsons, E.J. Stanbridge, and P. Olive, *Reduced expression of hypoxia-inducible factor-1alpha in perinecrotic regions of solid tumors*, Cancer Research 65 (2005) 7259-7266.
[146] T. Huang, A.C. Civelek, J. Li, H. Jiang, C.K. Ng, G.C. Postel, B. Shen, and X.F. Li, *Tumor microenvironment-dependent 18F-FDG, 18F-fluorothymidine, and 18F-misonidazole uptake: a pilot study in mouse models of human non-small cell lung cancer*, J Nucl Med 53 (2012) 1262-1268.
[147] A. Rompp, S. Guenther, Z. Takats, and B. Spengler, *Mass spectrometry imaging with high resolution in mass and space (HR(2) MSI) for reliable investigation of drug compound distributions on the cellular level*, Anal Bioanal Chem 401 (2011) 65-73.
[148] A. Thomas, J.L. Charbonneau, E. Fournaise, and P. Chaurand, *Sublimation of new matrix candidates for high spatial resolution imaging mass spectrometry of lipids: enhanced information in both positive and negative polarities after 1,5-diaminonapthalene deposition*, Analytical Chemistry 84 (2012) 2048-2054.

[149] A. Rompp and B. Spengler, *Mass spectrometry imaging with high resolution in mass and space*, Histochem Cell Biol 139 (2013) 759-783.

[150] M.A. Varia, D.P. Calkins-Adams, L.H. Rinker, A.S. Kennedy, D.B. Novotny, W.C. Fowler, and J.A. Raleigh, *Pimonidazole: A novel hypoxia marker for complementary study of tumor hypoxia and cell proliferation in cervical carcinoma*, Gynecologic Oncology 71 (1998) 270-277.

[151] J.A. Raleigh, S.C. Chou, G.E. Arteel, and M.R. Horsman, *Comparisons among pimonidazole binding, oxygen electrode measurements, and radiation response in C3H mouse tumors*, Radiat Res 151 (1999) 580-589.

[152] M.W. Gross, U. Karbach, K. Groebe, A.J. Franko, and W. Mueller-Klieser, *Calibration of misonidazole labeling by simultaneous measurement of oxygen tension and labeling density in multicellular spheroids*, Int J Cancer 61 (1995) 567-573.

[153] G.E. Arteel, R.G. Thurman, J.M. Yates, and J.A. Raleigh, *Evidence That Hypoxia Markers Detect Oxygen Gradients in Liver - Pimonidazole and Retrograde Perfusion of Rat-Liver*, British Journal of Cancer 72 (1995) 889-895.

[154] S.C. Chou, Y. Azuma, M.A. Varia, and J.A. Raleigh, *Evidence that involucrin, a marker for differentiation, is oxygen regulated in human squamous cell carcinomas*, Br J Cancer 90 (2004) 728-735.

[155] G.E. Arteel, R.G. Thurman, and J.A. Raleigh, *Reductive metabolism of the hypoxia marker pimonidazole is regulated by oxygen tension independent of the pyridine nucleotide redox state*, European Journal of Biochemistry 253 (1998) 743-750.

[156] A.J. Varghese and G.F. Whitmore, *Cellular and chemical reduction products of misonidazole*, Chem Biol Interact 36 (1981) 141-151.

[157] A.J. Varghese, *Glutathione conjugates of misonidazole*, Biochem Biophys Res Commun 112 (1983) 1013-1020.

[158] A.J. Varghese and G.F. Whitmore, *Properties of 2-hydroxylaminoimidazoles and their implications for the biological effects of 2-nitroimidazoles*, Chem Biol Interact 56 (1985) 269-287.

[159] R.A. McClelland, R. Panicucci, and A.M. Rauth, *Electrophilic intermediate in the reactions of a 2-(hydroxyamino)imidazole. A model for biological effects of reduced nitroimidazoles*, Journal of the American Chemical Society 107 (1985) 1762-1763.

[160] J.L. Bolton and R.A. Mcclelland, *Kinetics and Mechanism of the Decomposition in Aqueous-Solutions of 2-(Hydroxyamino)Imidazoles*, Journal of the American Chemical Society 111 (1989) 8172-8181.

[161] J.A. Raleigh and C.J. Koch, *Importance of Thiols in the Reductive Binding of 2-Nitroimidazoles to Macromolecules*, Biochemical Pharmacology 40 (1990) 2457-2464.

[162] M. Koestler, D. Kirsch, A. Hester, A. Leisner, S. Guenther, and B. Spengler, *A high-resolution scanning microprobe matrix-assisted laser desorption/ionization ion source for imaging analysis on an ion trap/Fourier transform ion cyclotron resonance mass spectrometer*, Rapid Commun Mass Spectrom 22 (2008) 3275-3285.

[163] C. Paschke, A. Leisner, A. Hester, K. Maass, S. Guenther, W. Bouschen, and B. Spengler, *Mirion--a software package for automatic processing of mass spectrometric images*, J Am Soc Mass Spectrom 24 (2013) 1296-1306.

[164] L. Jiang, T.R. Greenwood, E.R. van Hove, K. Chughtai, V. Raman, P.T. Winnard, Jr., R.M. Heeren, D. Artemov, and K. Glunde, *Combined MR, fluorescence and histology imaging strategy in a human breast tumor xenograft model*, Nmr in Biomedicine 26 (2013) 285-298.

[165] A. Trochine, D.J. Creek, P. Faral-Tello, M.P. Barrett, and C. Robello, *Benznidazole biotransformation and multiple targets in Trypanosoma cruzi revealed by metabolomics*, PLoS Negl Trop Dis 8 (2014) e2844.

[166] B.S. Hall and S.R. Wilkinson, *Activation of benznidazole by trypanosomal type I nitroreductases results in glyoxal formation*, Antimicrob Agents Chemother 56 (2012) 115-123.

[167] A.J. Varghese and G.F. Whitmore, *Modification of guanine derivatives by reduced 2-nitroimidazoles*, Cancer Research 43 (1983) 78-82.

[168] D.C. Heimbrook and A.C. Sartorelli, *Biochemistry of misonidazole reduction by NADPH-cytochrome c (P-450) reductase*, Mol Pharmacol 29 (1986) 168-172.

[169] J.A. Raleigh and S.F. Liu, *Reductive fragmentation of 2-nitroimidazoles in the presence of nitroreductases--glyoxal formation from misonidazole*, Biochemical Pharmacology 32 (1983) 1444-1446.

[170] O.A. Ulanovskaya, A.M. Zuhl, and B.F. Cravatt, *NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink*, Nature Chemical Biology 9 (2013) 300-+.

[171] S.W. Tang, T.C. Yang, W.C. Lin, W.H. Chang, C.C. Wang, M.K. Lai, and J.Y. Lin, *Nicotinamide N-methyltransferase induces cellular invasion through activating matrix metalloproteinase-2 expression in clear cell renal cell carcinoma cells*, Carcinogenesis 32 (2011) 138-145.

[172] Y. Wu, M.S. Siadaty, M.E. Berens, G.M. Hampton, and D. Theodorescu, *Overlapping gene expression profiles of cell migration and tumor invasion in human bladder cancer identify metallothionein 1E and nicotinamide N-methyltransferase as novel regulators of cell migration*, Oncogene 27 (2008) 6679-6689.

[173] J. Zhang, Y.Z. Wang, G.L. Li, H.T. Yu, and X.Y. Xie, *Down-Regulation of Nicotinamide N-methyltransferase Induces Apoptosis in Human Breast Cancer Cells via the Mitochondria-Mediated Pathway*, PLoS One 9 (2014).

[174] J. Kerner and C. Hoppel, *Fatty acid import into mitochondria*, Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids 1486 (2000) 1-17.

[175] M.S. Wainwright, R. Kohli, P.F. Whitington, and D.H. Chace, *Carnitine treatment inhibits increases in cerebral carnitine esters and glutamate detected by mass spectrometry after hypoxia-ischemia in newborn rats*, Stroke 37 (2006) 524-530.

[176] S.R. Shi, C.R. Taylor, C.B. Fowler, and J.T. Mason, *Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies*, Proteomics Clinical Applications 7 (2013) 264-272.

[177] M. Walzer, L.E. Pernas, S. Nasso, W. Bittremieux, S. Nahnsen, P. Kelchtermans, P. Pichler, H.W.P. van den Toorn, A. Staes, J. Vandenbussche, M. Mazanek, T. Taus, R.A. Scheltema, C.D. Kelstrup, L. Gatto, B. van Breukelen, S. Aiche, D. Valkenborg, K. Laukens, K.S. Lilley, J.V. Olsen, A.J.R. Heck, K. Mechtler, R. Aebersold, K. Gevaert, J.A. Vizcaino, H. Hermjakob, O. Kohlbacher, and L. Martens, *qcML: An Exchange Format for Quality Control Metrics from Mass Spectrometry Experiments*, Molecular & Cellular Proteomics 13 (2014) 1905-1913.

Appendix

Table. Clinicopathological parameters of the sample set used in Chapter 4.

Variable	Patients (%)
Sex	
Male	112 (63%)
Female	66 (37%)
Mean age at diagnosis, range (years)	62, 37 – 87
Smoking	
Never	32 (18%)
Stopped > 1 year	27 (15%)
Current smoking or stopped < 1 year	118 (66%)
Missing	1 (1%)
Alcohol	
None	34 (19%)
Occasionally	41 (23%)
1-4 U/day	65 (36%)
>4 U/day	37 (21%)
Missing	1 (1%)
Site	
Oral cavity	156 (88%)
Oropharynx	22 (12%)
Clinical TNM-Stage (AJCC)	
Stage I	28 (16%)
Stage II	46 (26%)
Stage III	32 (18%)
Stage IV	72 (40%)
Histologically nodal metastasis	
No	75 (42%)
Yes	103 (58%)
5-year overall survival	
Alive	90 (51%)
Death	88 (49%)
5-year disease-free survival	
Disease-free	115 (65%)
Recurrence of disease or death	63 (35%)

AJCC, American Joint Committee on Cancer.

Summary

Mass spectrometry imaging (MSI) can detect and identify many different molecules without the need for labeling. In addition, it can provide their spatial distributions as 'molecular maps'. These features make MSI well suited for studying the molecular makeup of tumor tissue.

Currently, there is an interest in using MSI to predict cancer progression and treatment response, which often remains a challenge in today's clinical practice. Experimental evidence shows that tumor heterogeneity (on the inter- and intratumor level) plays an important role in tumor biology and therefore in response to treatment. The molecular profiles generated by MSI reveal part of this heterogeneity.

In this thesis, we have combined matrix-assisted laser desorption/ionization (MALDI) MSI, histological tissue staining and multivariate data analysis to investigate inter- and intratumor heterogeneity. The histopathological analysis of cancer tissue provided information on the distribution of tumor cells, stromal components, and necrotic and hypoxic tissue regions. Spatially resolved molecular profiles were generated with MALDI-MSI. After co-registration of the histological images and MSI data, we used histology-specific molecular profiles to predict disease progression and treatment response. The molecular changes associated with hypoxic tissue regions were also investigated.

In Chapter 2, strategies for protein identification in MSI are discussed, which is one of the current bottlenecks in the field. Identification of the peptide or protein of interest is key to answering biomedical questions, but only few of the molecular signals in MSI spectra can easily be identified. We discuss bottom-up, top-down and indirect identification approaches. The role of mass accuracy in protein identification, and developments in on-tissue chemical labeling of proteins are also discussed.

Chapters 3 and 4 focus on the generation and use of MALDI-MSI data to predict treatment response and disease progression. For prediction purposes a (preferably large) sample set with available biomedical or clinical data is required. We have used arrays of small pieces of formalin-fixed paraffin-embedded (FFPE) tissue, typically referred to as tissue microarrays (TMAs). MALDI-MSI analysis of TMAs enables high-

throughput analysis of tumor tissue that is carefully selected based on its histology. This approach generates thousands of spectra from hundreds of different tissue cores. Moreover, each spectrum consists of hundreds of different molecular ions. Bioinformatic approaches are required to reduce this complexity.

Chapter 3 presents an approach that combines MALDI-MSI on tissue microarrays with Principal Component Analysis and Linear Discriminant Analysis (PCA-LDA) to predict treatment response. The feasibility of this approach is evaluated on a set of 22 patient-derived xenograft models of triple-negative breast cancer. We used PCA-LDA to predict response to the chemotherapeutic drug cisplatin based on the proteomic information obtained with MALDI-MSI. A small predictive power is observed. The comparison of different tissue cores from the same tumor model revealed a clear effect of the biological and experimental variability on classifier performance.

In Chapter 4, the use of a different classification approach is investigated. We tested the feasibility of this approach to predict lymph node metastasis and disease-specific survival from a sample set of 240 head and neck cancers. We discuss how MALDI-MSI data from these samples is processed so that it could be used in five different classifiers: Linear and Quadratic Discriminant Analysis, a Naive Bayes classifier, Decision Tree classifier and Support Vector Machine. Only for disease-specific survival some of the classifiers showed a small predictive power. We demonstrated that our method is sensitive to intensity differences of 50% in typical peptide peaks. The observed broad peak distributions made classification difficult. A small effect of sample age on spectral quality was observed, which might have implications for the comparison of FFPE samples collected over periods of several years. As other proteomic approaches using mass spectrometry do not report such an effect, future efforts to improve protein extraction from FFPE tissue for MALDI-MSI would be valuable.

Chapter 5 focuses on intratumor heterogeneity. It presents a new method to detect and visualize hypoxic tumor regions and hypoxia-associated molecules in a breast tumor xenograft model. We investigated the use of the 2-nitroimidazole pimonidazole as exogenous hypoxia marker for MALDI-MSI. In hypoxic cells, pimonidazole is reduced and forms reactive products that bind to cellular nucleophiles. We demonstrated that a reductively activated pimonidazole metabolite can be imaged by MALDI-MSI. The immunohistochemical detection of pimonidazole adducts on adjacent tissue sections confirmed that this metabolite is localized to

hypoxic tissue regions. We used the metabolite to image hypoxic tissue regions and hypoxia-associated lipids and metabolites. A heterogeneous distribution of 1-methylnicotinamide and acetylcarnitine is identified, which mostly co-localized with hypoxic tumor regions. Thus, the presented approach is capable of imaging hypoxic regions and their associated biomolecules in a single experiment, without the need for co-registration of MSI data with immunohistochemical data. Pimonidazole is a widely used hypoxia marker. We expect that the presented MALDI-MSI approach might be also applicable to other tissues from pimonidazole-injected animals or humans.

In Chapter 6, the overall conclusions are presented. In addition, perspectives for future research are discussed, including further optimization of the sample preparation for FFPE tissue and the use of a modified chemical marker for the detection of hypoxia.

The MSI methods for the analysis of tumor tissue presented in this thesis are examples of the development of the MSI field towards biomedical and clinical applications. We have investigated the feasibility of using TMAs, different data analysis techniques, and the use of an immunohistochemical hypoxia marker for MALDI-MSI. As the focus shifts from MSI technology to application, research projects become increasingly multidisciplinary. This development calls for close collaboration of researchers with expertise in MSI, histo(patho)logy and data analysis.

Samenvatting

Met behulp van massaspectrometrie imaging (MSI) kunnen veel verschillende moleculen worden gedetecteerd en geïdentificeerd zonder deze moleculen te hoeven labelen. Tegelijkertijd kan hun ruimtelijke verdeling in een weefsel in kaart worden gebracht. Deze kenmerken maken MSI zeer geschikt om tumorweefsel op moleculair niveau te bestuderen.

Er is een groeiende interesse in het gebruik van MSI in het onderzoek naar kanker. MSI-datasets worden gebruikt om het verloop van een ziekte en het resultaat van een behandeling te voorspellen. Experimenten laten zien dat heterogeniteit (op het inter- en intratumorniveau) een belangrijke rol speelt in de biologie van een tumor en dus ook in de reactie van een tumor op een behandeling. De moleculaire profielen gegenereerd met MSI leggen een deel van deze heterogeniteit bloot.

We combineerden in dit proefschrift matrixgeassisteerde laser desorptie/ionisatie (MALDI) MSI, histologische kleuring en multivariate data-analyse voor het onderzoeken van inter- en intratumorheterogeniteit. De histopathologische analyse van kankerweefsel gaf informatie over de verdeling van tumorcellen, ondersteunend bindweefsel (stroma), en gebieden met dode (necrotische) en zuurstofarme (hypoxische) cellen in tumorweefsel. Ruimtelijk opgeloste moleculaire profielen werden gegenereerd met MALDI-MSI. Na het co-registreren van de histologische beelden en MSI beelden, gebruikten we histologie-specifieke moleculaire profielen om het resultaat van een cisplatinabehandeling in borstkanker en het verloop van hoofd-halskanker te voorspellen. We onderzochten ook de moleculaire veranderingen in hypoxisch tumorweefsel.

In Hoofdstuk 2 bespraken we strategieën voor eiwitidentificatie met MSI. Hoewel identificatie van het onderzochte peptide of eiwit belangrijk is voor het beantwoorden van een biomedisch vraagstuk, kunnen maar weinig moleculaire signalen in MSI-spectra eenvoudig geïdentificeerd worden. We bespraken 'bottom-up', 'top-down' en indirecte identificatiestrategieën. Het belang van massa-accuraatheid in eiwit-identificatie en de ontwikkelingen in het chemisch labelen van eiwitten in het weefsel werden ook besproken.

In Hoofdstuk 3 en 4 werd gefocust op de generatie en het gebruik van MALDI-MSI data voor het voorspellen van de behandelreactie en het ziekteverloop van twee verschillende soorten kanker. Om een voorspelling te kunnen doen, zijn een (bij voorkeur groot) aantal monsters nodig, waar biomedische of klinische gegevens voor beschikbaar zijn. Wij gebruikten 'arrays' met kleine stukjes formaline-gefixeerd paraffine-ingebed (Engels: FFPE) weefsel, die 'tissue microarrays' (TMAs) worden genoemd. Het gebruik van TMAs maakt de 'high-throughput' analyse van tumorweefsel met MALDI-MSI mogelijk. Deze analyse genereert duizenden spectra van honderden verschillende weefselstukjes. Bovendien bestaat elk spectrum uit honderden verschillende moleculaire ionen. Methoden uit de bioinformatica zijn nodig om deze complexiteit te reduceren.

In Hoofdstuk 3 presenteerden we een strategie die MALDI-MSI analyse van TMAs combineert met Principale Componenten Analyse en Lineaire Discriminant Analyse (PCA-LDA) voor het voorspellen van een behandelreactie. We bepaalden de geschiktheid van deze methode aan de hand van een set van 22 xenograftmodellen (humane tumoren geïmplanteerd in muizen) van drievoudig-negatieve borstkanker. We gebruikten PCA-LDA voor het voorspellen van het effect van een behandeling met het chemotherapeutisch middel cisplatina, gebaseerd op de eiwitinformatie verkregen met MALDI-MSI. We constateerden een klein voorspellend vermogen. De vergelijking van verschillende weefselstukjes van hetzelfde tumormodel liet een duidelijk effect van de biologische en experimentele variabiliteit op het resultaat van de classificatie zien.

In Hoofdstuk 4 beschreven we een andere classificatiemethode. We testten de geschiktheid van deze methode om lymfekliermetastase en ziekte-specifieke overleving te voorspellen van een set van 240 hoofd-halstumoren. We bespraken hoe MALDI-MSI data van deze monsters is verwerkt zodat het gebruikt kon worden in vijf verschillende 'classifiers': Lineare en Quadratische Discriminant Analyse, Naive Bayes classifier, Decision Tree classifier en Support Vector Machine. Alleen voor ziekte-specifieke overleving lieten sommige van de classifiers een klein verspellend vermogen zien. We toonden aan dat onze methode gevoelig is voor intensiteits-verschillen van 50% voor een representatieve peptidepiek. De brede piekverdelingen maakten classificatie moeilijk. We namen een klein effect van de leeftijd van FFPE weefsel op de spectrale kwaliteit waar, wat gevolgen zou kunnen hebben voor de vergelijking van FFPE weefsel dat verzameld is gedurende een periode van meerdere

jaren. Andere 'proteomics'-technieken die gebruikmaken van massaspectrometrie laten dit effect niet zien. Verbetering van de eiwitextractie van FFPE weefsel voor MALDI-MSI zou dus waardevol kunnen zijn.

Hoofdstuk 5 focuste op intratumor heterogeniteit. In dit hoofdstuk presenteerden we een nieuwe methode om hypoxische gebieden in tumoren en hypoxie-geassocieerde moleculen te detecteren en visualiseren in een borsttumormodel. We onderzochten het gebruik van de 2-nitroimidazool pimonidazole als een exogene hypoxiemarker voor MALDI-MSI. Pimonidazole wordt in hypoxische cellen gereduceerd en vormt reactieproducten die aan cellulaire nucleofielen binden. We lieten zien dat een metaboliet van pimonidazole kan worden gedetecteerd en gevisualiseerd met MALDI-MSI. De immunohistochemische detectie van pimonidazole-adducten op opeenvolgende weefselcoupes bevestigde dat dit metaboliet zich bevindt in de hypoxische weefselgebieden. We gebruikten het metaboliet vervolgens voor het detecteren van hypoxie-geassocieerde lipiden en metabolieten. De verdeling van 1-methylnicotinamide en acetylcarnitine overlapte grotendeels met de hypoxische tumorgebieden. Deze methode is dus in staat hypoxische gebieden en de geassocieerde moleculen te detecteren zonder dat co-registratie van MSI beelden met immunohistochemische beelden nodig is. Pimonidazole is een veelgebruikt hypoxiemarker. We verwachten dat de gepresenteerde MALDI-MSI methode ook toepasbaar is op ander weefsel van pimonidazole-geïnjecteerde dieren of mensen.

In Hoofdstuk 6 presenteerden we de algemene conclusies. Daarnaast werden toekomstige onderzoekmogelijkheden besproken, waaronder verdere optimalisatie van de monstervoorbereiding voor FFPE weefsel en het gebruik van een gemodificeerde marker voor hypoxiedetectie.

De methoden voor de analyse van tumorweefsel die zijn gepresenteerd in dit proefschrift zijn voorbeelden van (bio)medische toepassingen van MSI. We hebben het gebruik onderzocht van TMAs, verschillende data-analysetechnieken, en een immunohistochemische hypoxiemarker voor MALDI-MSI. Nu de focus van techniek-ontwikkeling naar toepassing verschuift worden onderzoeksprojecten in toenemende mate multidisciplinair. Deze ontwikkeling maakt hechte samenwerking tussen onderzoekers met expertise in MSI, histo(path)ologie en data-analyse noodzakelijk.

Acknowledgments

Many things have changed during 4.5 years of PhD research. The mass spectrometry imaging field is young, dynamic and multidisciplinary, which is reflected in the developments in my research group. Since I started, the group has not only seen changes in its members (normal for a research group), but also in location, instruments, name, research topics and expertise. To be part of this was an interesting experience. It sometimes felt like I was doing a PhD in chemistry and a PhD in communications at the same time.

First, I would like to thank my promotor Ron Heeren for the opportunity to do a PhD at AMOLF and for initiating the collaborations which were crucial to my PhD.

AMOLF is a great research environment. The people make it a special place. Many thanks to former and current group members for the science and the fun: András Kiss, Julia Jungmann, Don Smith, Berta Cillero Pastor, Kamila Chughtai, Sana Chughtai, Ivo Klinkert, Ronald Buijs, Gert Eijkel, Frans Giskes, Marc Duursma, Piet Kistemaker, Anne Bruinen, Karolina Škrášková, Jo Cappell, Nina Ogrinc Potočnik, Tiffany Porta, Shane Ellis, Mark Jansen, Guzel Kireeva, Klára Ščupáková, Sarfaraz Syed, Bob Hommersom, Florian Marty and Bryn Flinders. Gert, thank you for the many hours you spent helping me with data analysis. Ivo, thank you for sharing the fun task of explaining the link between the mass and identity of a molecule during our labtours at three successive AMOLF Open Days. Bob and Marc, thank you for sharing coffee and lunch time during the last months at AMOLF. Nina, Jo, and Tiff, thank you for making me feel welcome in Maastricht. I wish all members of M4I success in the future.

I would also like to thank my collaborators, without whom my research projects would not have been possible: Jelle Wesseling and Petra ter Brugge from the Netherlands Cancer Institute, Stefan Willems and Rob Noorlag from the University Medical Center Utrecht, and Kristine Glunde, Lu Jiang, Asif Rizwan and Menglin Cheng from Johns Hopkins University. Special thanks to Kristine and her group for the opportunity to visit them in Baltimore.

Acknowledgments

Thanks to all members of the AMOLF works council and the central works council of FOM 2013-2014 for sharing this important task and giving me an insight into the organizational aspects of academia.

I would also like to thank Gilles van Wezel for guiding my first steps in research. You taught me many things. Reza Shariatgorji and Per Andrén, thank you for your introduction in the mass spectrometry imaging field and your hospitality in Uppsala.

Last but not least I would like to thank my family and friends for their support and encouragement. Especially Tessa, Eva and, of course, Jannis. Thank you so much for sharing the ups and downs of a PhD.

List of publications

This thesis is based on the following publications:

N.E. Mascini and R.M.A. Heeren, *Protein identification in mass-spectrometry imaging*, Trac-Trends in Analytical Chemistry 40 (2012) 28-37. (Chapter 2)

N.E. Mascini, G.B. Eijkel, P. ter Brugge, J. Jonkers, J. Wesseling, and R.M. Heeren, *The use of mass spectrometry imaging to predict treatment response of patient-derived xenograft models of triple-negative breast cancer*, Journal of Proteome Research 14 (2015) 1069-1075. (Chapter 3)

N.E. Mascini, J. Teunissen, R. Noorlag, S.M. Willems and R.M.A. Heeren, *Predicting head and neck cancer metastasis and disease-specific survival from MALDI-MSI data: feasible or not?*, submitted to Analytical and Bioanalytical Chemistry. (Chapter 4)

N.E. Mascini, M. Cheng, L. Jiang, A. Rizwan, H. Podmore, A. Römpp, D.R. Bhandari, K. Glunde, R.M.A. Heeren *Mass spectrometry imaging of the hypoxia marker pimonidazole in a breast tumor model*, submitted to Analytical Chemistry. (Chapter 5)

Other:

J.H. Jungmann, N.E. Mascini, A. Kiss, D.F. Smith, I. Klinkert, G.B. Eijkel, M.C. Duursma, B. Cillero Pastor, K. Chughtai, S. Chughtai, and R.M. Heeren, *A MASSive laboratory tour. An interactive mass spectrometry outreach activity for children*, Journal of the American Society for Mass Spectrometry 24 (2013) 979-982.

About the author

Nadine Mascini was born in Utrecht, the Netherlands, in 1987. She attended *Het Nieuwe Lyceum* in Bilthoven from which she obtained her high school diploma in 2005. From 2005, she studied *Life Science and Technology* at Leiden University and Delft University of Technology. After obtaining her bachelor's degree (*cum laude*), she started a research master in *Life Science and Technology* with a focus on molecular biotechnology at Leiden University, but also kept following courses in Delft. Her master research project comprised a study on cell division control in the antibiotic-producing bacterium *Streptomyces coelicolor*. This project was carried out at the Leiden Institute of Chemistry under the supervision of Prof. Dr. Gilles van Wezel. Towards the end of her studies, she completed a second scientific internship at Uppsala University in Sweden. In the *Medical Mass Spectrometry* group of Prof. Dr. Per Andrén she conducted a feasibility study on drug detection in a mouse model of Parkinson's Disease using mass spectrometry imaging. She received her master's degree (*cum laude*) in 2011 and joined the *Biomolecular Imaging Mass Spectrometry* group of Prof. Dr. Ron Heeren at the FOM Institute AMOLF in Amsterdam. The focus of her PhD research is the development of mass spectrometry imaging methods for the classification of tumor tissue, the results of which are presented in this thesis.