

Aus dem Institut für Medizintechnologie
der Universität Heidelberg und der Hochschule Mannheim
(Geschäftsführender Direktor: Prof. Dr. med. Norbert Gretz)

Multimodal FTIR Microscopy-guided Acquisition and Interpretation of
MALDI Mass Spectrometry Imaging Data

Inauguraldissertation
zur Erlangung des akademischen Grades Doctor scientiarum humanarum
(Dr. sc. hum.)
der
Medizinischen Fakultät Mannheim
der Ruprecht-Karls-Universität
zu
Heidelberg

vorgelegt von
Jan-Hinrich Rabe

aus
Holzminden

2017

Dekan: Prof. Dr. med. Sergij Goerd
Referent: Prof. Dr. rer. nat. Carsten Hopf



*Für meine Familie,
ihre beispiellose Liebe und Unterstützung....*

ABBREVIATIONS

ABA	Allen Mouse Brain Reference Atlas
ACN	Acetonitrile
AUC	Area Under the Curve
CCD	Charge-Coupled Device
CMC	Carboxymethyl Cellulose
CPU	Central Processing Unit
CT	Computed Tomography
DESI	Desorption Electrospray Ionization
ddH ₂ O	Ultra-Pure Water
DMSO	Dimethyl sulfoxide
DSC	Dice Similarity Coefficient
EPD	Edge-Preserving Denoising
ESI	Electrospray Ionization
EtOH	Ethanol
FTICR	Fourier Transform Ion Cyclotron Resonance
FTIR	Fourier Transform Infrared Spectroscopy
GIST	Gastrointestinal Stromal Tumors
H&E	Hematoxylin & Eosin
ITO	Indium Tin Oxide
LCMD	Laser Capture Microdissection
LC-MS/MS	Liquid Chromatography – Tandem Mass Spectrometry
<i>m/z</i>	Mass-to-Charge Ratio
MALDI-TOF-MS	Matrix Assisted Laser Desorption/ Ionization-Time of Flight Mass Spectrometry
MeOH	Methanol
MIR	Mid-Infrared
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
MSI	Mass Spectrometry Imaging
Nd:YAG	Neodymium-Doped Yttrium Aluminium Garnet
NIR	Near-Infrared
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PhCCAA	Phenyl- α -Cyanocinnamic Acid Amide
REIMS	Rapid Evaporation Ionization Mass Spectrometry
ROI	Region of Interest
RT	Room temperature
S/N	Signal-to-Noise Ratio
SNV	Standard Normal Variate
SPECT	Single-Photon Emission Computed Tomography
TFA	Trifluoroacetic acid
TIC	Total Ion Current
t-SNE	t-Distributed Stochastic Neighbor Embedding
U.S. FDA	United States Food and Drug Administration

TABLE OF CONTENTS

	Page
1 INTRODUCTION	1
1.1 Multimodality imaging	1
1.2 FTIR and Raman microscopy	2
1.2.1 History of vibrational spectroscopy in the field of biomedicine	2
1.2.2 Basic principles of vibrational spectroscopy	3
1.2.3 Diagnostic applications of vibrational spectroscopy	7
1.3 Mass spectrometry	8
1.3.1 History of mass spectrometry	8
1.3.2 Basic principles of mass spectrometry used in biomedical science	9
1.3.3 Mass spectrometry imaging	12
1.3.4 Diagnostic and pharmacological applications of MSI	13
1.4 Neuronal disorders and cancer	14
1.4.1 Gastrointestinal stromal tumors	15
1.4.2 Glioblastoma multiforme	16
1.4.3 Niemann-Pick disease, type C	16
2 AIMS	18
3 DATA ANALYSIS BACKGROUND	19
3.1 Dimensionality reduction	19
3.1.1 Principal Component Analysis	19
3.1.2 t-distributed stochastic neighbor embedding	21
3.2 Information extraction	22
3.2.1 k-means++ segmentation	22
3.2.2 Feature extraction	24
3.3 Quality measures	24
3.3.1 Cluster evaluation	24
3.3.2 Measure of data dispersion	25

3.3.3	Sample correlation and dependence	26
3.4	Image registration and evaluation.....	26
4	MATERIALS AND METHODS	28
4.1	Materials.....	28
4.1.1	Reagents	28
4.1.2	Solutions.....	29
4.1.3	Consumables.....	29
4.1.4	Equipment	30
4.1.5	Software	31
4.2	Methods.....	32
4.2.1	Animal studies	32
4.2.2	Human tissue specimen	32
4.2.3	Tissue section preparation.....	33
4.2.4	Multimodality imaging	34
4.2.5	Image processing and multivariate statistical evaluation	37
5	RESULTS.....	41
5.1	Exploration of single modality capabilities	42
5.1.1	The influence of object slide coating on FTIR spectroscopy.....	42
5.1.2	The influence of object slide coating on MS quality	47
5.1.3	Tissue segmentation by FTIR and MS imaging	50
5.1.4	Feasibility of FTIR imaging on FFPE tissues	53
5.1.5	Reproducibility of FTIR image segmentation	55
5.2	Development of an automated multimodal imaging workflow	58
5.2.1	Image registration to enable FTIR-MS workflow development	58
5.2.2	Evaluation of FTIR-to-MSI registration.....	61
5.3	Analytic applications of multimodality FTIR-MS imaging	64
5.3.1	Automated FTIR-guided tumor-targeting and mass feature extraction 64	
5.3.2	Automated MS marker signature retrieval for defined tissue morphologies.....	69
5.3.3	Identification of GIST by means of FTIR cluster analysis	73
5.4	Same-slice targeted FTIR-guided FTICR-MS acquisition.....	77

6 DISCUSSION	81
6.1 Technical evaluation of FTIR and MS imaging	81
6.2 Development of FTIR-guided MSI	83
6.3 Tumor targeting and spatially restrictive mass feature identification.....	84
6.4 Automated screening for local mass biomarker signatures	85
6.5 Assessment of histological heterogeneity.....	86
6.6 Targeted spatially-restrictive high-resolution MSI acquisition	88
6.7 Different cluster approaches.....	90
7 SUMMARY AND OUTLOOK	92
8 REFERENCES	95
9 SUPPLEMENT	106
10 <i>CURRICULUM VITAE</i>	107
11 ACKNOWLEDGEMENT	108

1 INTRODUCTION

1.1 Multimodality imaging

In the last decade, the field of multimodality imaging that combines the varying capabilities of multiple sensor types has become of crucial significance to clinical *in vivo* research and practice [1]. Commonly used anatomical imaging technologies such as computed tomography (CT) and magnetic resonance imaging (MRI) provide structural detail in the order of 0.5-1 mm [2]; whereas functional modalities such as positron emission tomography (PET) and single-photon emission computed tomography (SPECT) provide insight into the biochemical and physiological functions (i.e. metabolism and perfusion) of an organ or tissue. Consequentially, spatial co-localization of complementary information (i.e. structure and function) provided by two or more modalities ensures a better elucidation of physiological mechanisms at molecular and cellular levels. The benefits of integrating different sensory capabilities such as contrast, sensitivity, spatial and temporal resolution in relationship to diseases while overcoming individual limitations led to the development of many instrumentations, such as PET-MRI, SPECT-CT or optical PET. The original concept of scanning an object from many angles was presented for the first time by Allan Cormack in 1963 [3], and a first prototype of diagnostic CT-SPECT imaging was developed in 1966 [4]. However, the need to combine functional and anatomical imaging was overlooked until Hasegawa *et al.* placed a clinical CT in tandem with a SPECT camera in the early 1990s [5, 6]. At the same time, a similar concept of combining PET and CT was suggested in 1991 [7], although a prototype was not completed until 1998. Following several years of obvious concerns with operational issues, cost, complexity and reliability, the first PET-CT designs appeared in medical centers by early 2001, followed by SPECT-CT designs in 2004 [8]. Since then, their use has grown exponentially.

In a similar fashion, multimodal *ex vivo* tissue analyses that utilize technologies of high chemical specificity such as mass spectrometry (MS) for a more detailed characterization of the molecular makeup have recently been presented [9-13]. Mass spectrometry imaging (MSI) in particular has proven its clinical value, as it can record the spatial distribution of hundreds of biomolecules without the need of prior labeling.

The acquisition of proteins, metabolites, lipids, peptides and pharmaceuticals directly from tissues has been used to examine, among other things, disease-related alterations in multiple neurological disorders (i.e. Alzheimer) and cancer types. Although the usage of hyperspectral yet slow sensors impedes synchronous acquisition as performed during clinical diagnosis, it provides structural and chemical information of unparalleled detail in the order of 10-50 μm lateral resolution. Concomitantly, asynchronous multimodal tissue diagnoses of unprecedented data volumes and acquisition times have been fueling the need for complex post-processing transformations to ensure proper image alignment [14-17]. MS, while providing chemical-rich information is insufficient to reveal the microscopic structures that are observed in established histopathological examinations at a manageable data level. Conversely, microscopy-based tissue histology provides precise structural information but lacks chemical specificity. Hence, opposing to the conventional approach of fusing complementary information into a single hyperspectral plane, an alternate usage of multimodality capabilities has emerged. Thereby, spatial dependencies derived from one technology are utilized to guide data acquisition and interpretation of one or more follow-up modalities [18-20]. Although this extension has seen rapid progress, with its positive impact on data load and acquisition time having been demonstrated in numerous studies [21], the clinical transfer is still in development.

1.2 FTIR and Raman microscopy

Vibrational micro-spectroscopic imaging techniques have gained acceptance in many fields such as cell biology and nanoscience [22]. The two most common techniques to date are imaging Fourier transform infrared spectroscopy (FTIR) and Raman spectroscopy. In both of these techniques, the inherent vibrational fingerprint spectra of biomolecules are observed.

1.2.1 History of vibrational spectroscopy in the field of biomedicine

Although infrared light was discovered as early as 1800 by Sir Frederick William Herschel [23], its analytical worth was considered to be irrelevant for many decades. In 1928, Sir Chandrasekhara Venkata Raman firstly observed the inelastic scattering of a photon by molecules which are excited to higher energy [24]. The effect was named after him and Raman spectroscopy was later used to provide the first catalog

of molecular vibrational frequencies. However, when IR spectroscopy was revived in the 1940s and commercial IR-spectrophotometers became available, Raman spectroscopy lost relevance as it needed much higher sample concentration and volume in order to maximize sensitivity. The first spectroscopic efforts to analyze disease related effects in tissues were described by Blout and Mellors in 1949 [25], and Woernley in 1952 [26], who utilized single beam IR spectroscopy for the analysis of tissue homogenates. Due to instrumental deficiencies and a lack of methods for the interpretation of vibrational spectra at that time, the field was not pursued until the 1960s. In that time, the advent of laser technology resulted in simplified and more sensitive Raman instruments and thousands of studies regarding structure-spectra correlation of biomolecules were reported for both infrared and Raman spectroscopy [27, 28]. In the 1980s, the acceptance of Fourier transform methodology into the field of IR spectroscopy fueled the investigation of bacterial and fungal pathogens [29]. Since then, vibrational spectroscopy research aimed at detecting diseases in human cells and tissues on the basis of cell pellet experiments until the development of micro-spectroscopic methods utilizing a microscope in the 1990s. Nowadays, both techniques have gained recognition, among other things, as diagnostic tools revealing histopathological features in various disease types [30-34].

1.2.2 Basic principles of vibrational spectroscopy

Imaging FTIR and Raman microscopy are analytical methods that determine the radiation response of matter in a spatially resolved fashion. Whereas FTIR spectroscopy measures the absorption of energy at different wavelengths, Raman spectroscopy measures the exchange of energy with electromagnetic radiation of a particular wavelength, usually mitigated from a near-infrared (NIR) range laser. From the exchange in electromagnetic energy a measureable shift in the wavelength of incident laser light occurs. This inelastic light scattering effect, known as Raman scattering, is complementary to the absorption of infrared radiation.

In imaging FTIR, a screening process is applied in order to construct a spectral map that consists of spectra collected from multiple pixels. At each (x,y)-coordinate, multiple beams that each consist of complex light frequencies are pointed onto the observed specimen and the wavelength-specific absorbance/transmission is recorded. The

beams originate from a broadband light source (containing the full measured wavelength range) whose light is transmitted into a Michelson interferometer [35]. The interferometer consists of a configuration of two mirrors of which one is constantly moving and one is fixed, and a beam splitter (**Fig 1**). The incoming light beam is split and in equal parts forwarded to the fixed and moving mirror. Depending on the distance between beam splitter and the moving mirror, each wavelength of the reflected and recombined light beam is periodically blocked and transmitted due to wave interference. The rate at which wavelengths are modulated differs, so that at each moment a different excitation spectrum is applied onto the sample. The resulting absorption values at many discrete positions of the moving mirror form an interferogram that is converted by the eponymous Fourier transform algorithm to receive the actual spectrum.

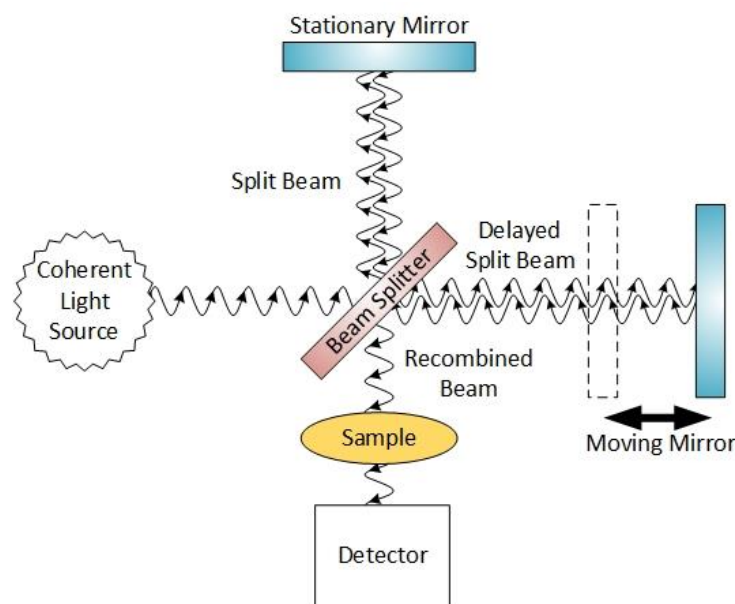


Figure 1 Optical arrangement and light path of a Michelson interferometer Light of a broadband light source is split and partially blocked by a constantly moving mirror to periodically transmit a recombined beam consisting of different wavelengths onto a sample. In FTIR, the resulting interferogram is converted by Fourier transform.

The recorded absorption pattern allows for conclusions to be made on the specimen's chemical composition as molecules absorb frequencies characteristic of their structure. Molecular bonds with an electric dipole moment that can change by atomic displacement owing to natural vibrations are IR-active [36]. An IR-active molecule's absorption behavior can differ which results in different vibrational modes (**Fig 2**). For molecules with N atoms, linear molecules feature $3*N-5$ different vibrational modes

and nonlinear molecules have $3 \cdot N - 6$. The resonant frequencies of each bond are related to the masses of the atoms as well as the strength of the bond. Lighter atoms and stronger bonds result in higher frequencies, when compared to heavier atoms and weaker bonds. Simple diatomic molecules have only one bond and vibrational band whereas larger molecules are correspondingly more complex.

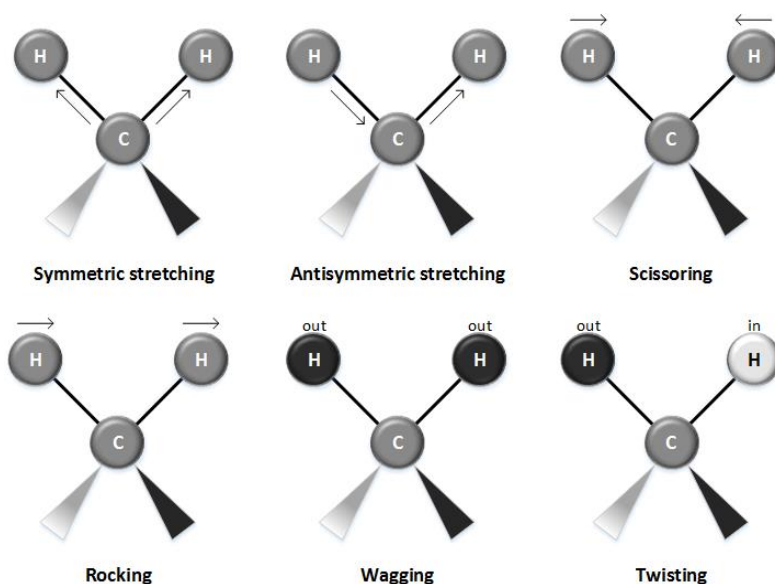


Figure 2 Vibrational modes of a methylene (-CH₂) group. In order for a vibrational mode to be “IR-active”, it must be associated with changes in the dipole moment. Six vibrations caused by the CH₂ moiety, commonly found in organic compounds, are displayed (Figure adapted from [37]).

Similar to imaging FTIR, Raman micro-spectroscopy can be applied in a screening process. A monochromatic laser is utilized in order to apply light of a specific wavelength onto the examined specimen. In biomedical applications, a AlGaAs diode with a wavelength of 785 nm is typically utilized, as the use of a lower energy wavelengths reduces the risk of damaging the sample [34]. The laser light leads to an excitation of (bio)molecular bonds resulting in a red-shift (Raman shift) of some of the applied initial radiation that forms a unique fingerprint of peaks. In order to assess the Raman shift, the electromagnetic radiation from the illuminated spot is passed through a spectrometer, which disperses the light into a spectrum that can be recorded by a cooled charge-coupled device (CCD)-camera.

When examining biological materials, imaging vibrational spectroscopy provides the ability to visualize the distribution of molecular bonds that are characteristic to biochemical classes. Qualitative acquisition of a specimen’s composition occurs in a

rapid, label-free and non-destructive manner which makes vibrational spectroscopy a powerful technique for *in situ* analyses. FTIR examination of biological materials typically focusses on measuring stretching vibrations such as C-H, S-H, N-H and O-H ($2550\text{-}3500\text{ cm}^{-1}$), of which for example C-H₂ is associated with the fatty acid chains of lipids. Other regions of interests are the amide I&II region ($1500\text{-}1700\text{ cm}^{-1}$) which comprises of protein-associated bond vibrations and the fingerprint region ($600\text{-}1450\text{ cm}^{-1}$) which comprises of nucleic acid-characteristic P-O and carbohydrate-associated stretching vibrations [38]. Together, these regions comprise a biochemical signature of interrogated cellular specimen (**Fig. 3**).

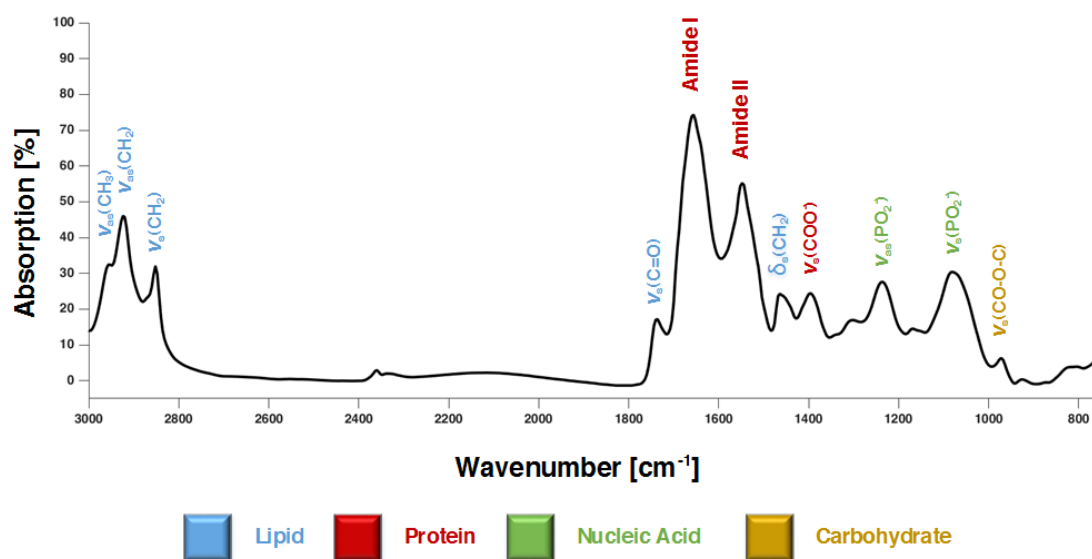


Figure 3 Biological infrared fingerprint spectrum showing biomolecular peak assignment. Transmission-spectrum of a cryosectioned ductal breast carcinoma *in situ*. Typical peaks are highlighted, where ν = stretching vibrations, δ = bending vibrations, s = symmetric vibrations and as = asymmetric vibrations. (Figure adapted from [39]).

Similarly, the Raman spectrum also comprise of vibrational bonds that are associated to biochemical classes, as Raman scattering and IR absorption are complementary effects [40]. Typical regions of interest are associated to proteins ($1500\text{-}1700\text{ cm}^{-1}$), carbohydrates ($470\text{-}1200\text{ cm}^{-1}$) and phosphate groups of DNA ($980, 1080$ and 1240 cm^{-1}). At higher wavenumbers ($2700\text{-}3500\text{ cm}^{-1}$), higher-frequency C-H, N-H and O-H stretching bonds that are associated with lipids and proteins can be recorded [41].

1.2.3 Diagnostic applications of vibrational spectroscopy

One excellent application field of vibrational spectroscopy imaging is the discrimination of specific cell types on the basis of their protein, lipid and carbohydrate profile. Since only light is used to record the necessary data, both technologies are entirely non-perturbing and can be performed complementary to the current practice in histopathology. In comparison, each technology has its distinct advantages. Whereas, mid-infrared spectroscopy is very sensitive to water interferences that strongly absorb radiation, Raman spectroscopy is mostly unaffected, as water is concurrently a weak scatterer [42]. For clinical applications, this enables the direct collection *in vivo* of Raman spectra. Moreover, Raman spectroscopy can achieve sub-cellular lateral resolution (~300 nm) that is superior compared to FTIR imaging (~10 μm). However, the collection of Raman spectra is comparatively more time consuming (1-10 s per spectrum), as the Raman effect is relatively weak and only one in $\sim 10^8$ photons typically undergoes inelastic light scattering events [34]. For this reason, it is often necessary to average a large number of acquired spectra to keep the acquisition time manageable. In addition, Raman spectroscopy is very sensitive to fluorescence mitigated from biological samples depending on the chosen excitation wavelength.

In the past decade FTIR imaging has been reported as a promising tool for the diagnosis of several diseases [43-45], especially for the classification of cancer in various tissues (i.e. breast [46-48], colon [49-51], lung [52-54] and prostate [55, 56]). Furthermore, the possibility to monitor susceptibility to therapy was recently demonstrated in a study from Zawlik *et al.* who reported discriminatory FTIR features for triple-negative breast cancer patients undergoing chemotherapy [57]. In a similar fashion, Raman spectroscopy has been used to *in vivo* investigate cancerous and normal cells of skin [58], brain [59] and other tissues (reviewed in [60, 61]).

In addition to its identification and classification capabilities, the non-perturbing nature of vibrational imaging spectroscopy makes it particularly interesting for multimodality imaging purposes. In that regard, Jermyn *et al.* recently reported *in situ* cancer detection in brain, lung and colon samples by combining Raman spectroscopy, diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy [62]. Moreover, multimodality imaging of frozen head and neck cancer sections by combined Raman, fluorescence and second-harmonic generation microscopy imaging was presented by

Heuke *et al.* [63]. In order to assess molecular cancer heterogeneity of a larynx carcinoma sample, Bocklitz *et al.* demonstrated the positive impact of combining vibrational Raman spectroscopy and MALDI-MSI [12].

In the case of FTIR, only few multimodal studies have been performed to date. Le Naour *et al.* reported the combination of synchrotron FTIR and synchrotron X-ray fluorescence micro-spectroscopy for *in situ* analysis of a single tissue slide [64]. Furthermore, Passot *et al.* investigated cryotolerance of lactic acid bacteria by combined synchrotron FTIR and synchrotron fluorescence microscopy [65]. Following the recent trend of predictive multimodality imaging, Großerueschkamp *et al.* presented FTIR-guided laser capture microdissection (LCMD) for subsequent proteomic analysis using liquid chromatography – tandem mass spectrometry (LC-MS/MS) [20]. This study reflects the high potential and positive impact of guided multimodality imaging.

1.3 Mass spectrometry

In mass spectrometry (MS), the mass of pure chemical species as well as complex mixtures is determined by ionization and separate recording of the created ions based on their mass-to-charge ratio (m/z). To date, a multitude of techniques for the generation of detectable ions as well as for the disaggregation and the measurement of m/z -values exist. The most prominent technique for the ionization of biomolecules is matrix-assisted laser desorption/ionization (MALDI) that uses laser energy absorbing reagents in order to create stable ions from large molecules. The created ion's m/z -values are commonly determined by time-of-flight mass spectrometry (TOF-MS) or Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) that provides substantially improved mass resolving power for ultrahigh-resolution MS.

1.3.1 History of mass spectrometry

The technique was first presented in the early 1910s by Joseph John Thomson and Francis William Aston [66, 67] who discovered the existence of isotopes by means of elemental MS. In 1919, Aston also constructed the first velocity focusing mass spectrometer with a resolving power of 130 [68]. Due to the high energy ionization,

initial MS experiments were restricted to elemental and small molecule analyses until the development of soft ionization techniques. First attempts to enhance the detectable mass range were reported by William Stephens in 1946 who demonstrated pulsed ionization which formed the basis for the development of TOF-MS in 1948 [69, 70]. Decades later, in 1974, FTICR-MS with superior mass resolution was demonstrated by Melvin Comisarow [71]. However, a major breakthrough for the analysis of larger molecules was the development of chemical ionization in the 1960s [72]. Thereby, ionization was achieved by means of ion-transfer reactions of a sample's molecules and reagent ions. This less energetic procedure was proven to reduce fragmentation and resulted in more stable ions when compared to electron ionization methods. To date, the most frequently used soft ionization technique used for the analysis of biomolecules (i.e. proteins, peptides, lipids, sugars) is MALDI, which was first reported by Franz Hillenkamp and Michael Karas in 1985 [73, 74]. By using tryptophan for ionization and a pulsed 266nm laser they demonstrated the ionization of peptide molecules up to 2843 Da. Desorption ionization of larger molecules (up to 34472 Da) by combining 30 nm cobalt particles in glycerol and a 337nm nitrogen laser was reported by Koichi Tanaka in 1987 [75]. The low cost of 337nm nitrogen lasers led to the first commercially available MALDI-MS instruments in the early 1990s. Nowadays, 355nm neodymium-doped yttrium aluminium garnet (Nd:YAG) laser are commonly used.

1.3.2 Basic principles of mass spectrometry used in biomedical science

A mass spectrometry experiment can be divided into two major steps for which multiple techniques exist. Firstly, the elements or molecules of interest are ionized. Secondly, the generated ions are separated based on their m/z -value and the intensity pattern is measured. In biomedical science, MALDI is the most commonly used ionization techniques to date. During MALDI-MS, a layer of matrix solution that typically consists of a small organic acid is either applied onto a tissue or mixed with extracted cells. The matrix incorporates the sample's analytes into matrix crystals formed upon solvent evaporation. Molecular ions are then generated upon radiation by a pulsed and focused laser aimed at the sample surface (**Fig. 4**).

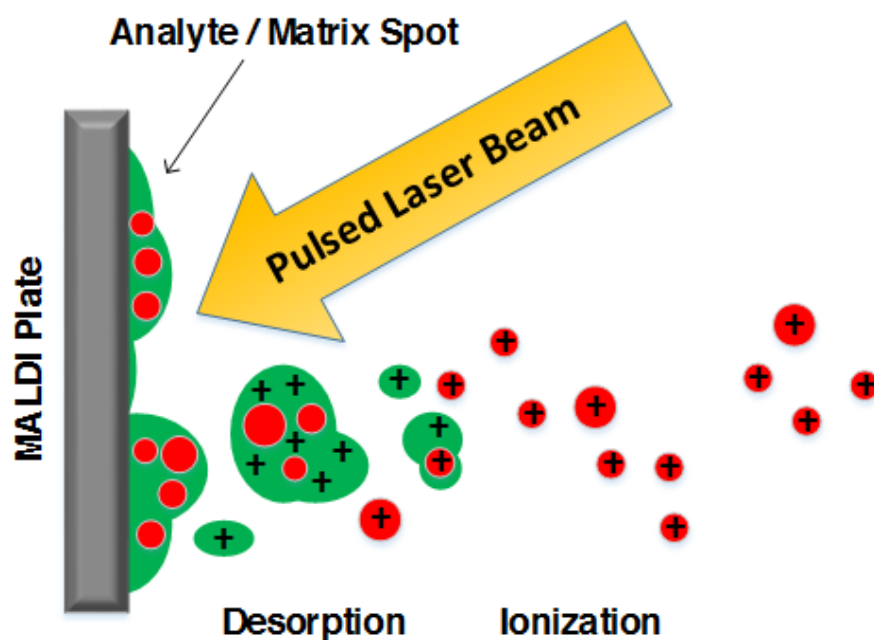


Figure 4 Schematic illustration of MALDI-MS soft ionization. Samples are mounted on conductive slides and matrix is applied across the specimen to incorporate analytes into crystals. Upon radiation by a pulsed laser beam, molecular ions are generated by means of ion-transfer reactions between analyte and matrix. Mass spectra are generated by separating the ions based on their m/z -value.

The precise and standardized application of matrix is of crucial importance to ensure optimal extraction and to avoid diffusion and delocalization effects [76]. Moreover, the choice of matrix is of importance, as several different matrixes allow for an array of biomolecular classes to be analyzed [77]. For example, super-2,5-dihydrobenzoic acid (sDHB), a mixture that consists of 2,5-Dihydrobenzoic Acid (DHB) and 2-Hydroxy-5-Methoxybenzoic Acid is highly effective for the analysis of proteins, whereas 4-Phenyl- α -Cyanocinnamic acid amide (PhCCA) is better suited to analyze (negatively) charged peptides [78]. In order to avoid potential artifacts that may arise due to matrix application, it is important to apply the substance homogeneously [79]. For the reproducible production of analyte-ion signals, automated application systems that spray-coat or sublime matrix are often utilized in modern MALDI-MSI. Apart from MALDI, alternative ionization techniques like desorption electrospray ionization (DESI) and rapid evaporative ionization mass spectrometry (REIMS) have recently shown their potential to classify clinical tissues [80, 81]. DESI enables ambient ion generation by employing a fast-moving charged solvent stream onto the sample surface. After ionization, the ions travel through air into an atmospheric pressure interface that is connected to a mass spectrometer. In contrast, REIMS is a unique technique that can be utilized intra-operatively. A specialized tissue-cutting tool (i.e. a surgical laser) leads

to the formation of an aerosol that contains tissue-specific ionized cell constructs. This flue gas is then analyzed by means of mass spectrometry. However, MALDI-MSI is to date the most popular method for clinical applications [82].

Just like with the generation of ions, a multitude of techniques to disaggregate and determine the ion's m/z ratio are available. Commonly used techniques are TOF-MS and FTICR-MS for ultrahigh-resolution MS. In TOF-MS, ions are accelerated by an electric field of known strength and directed through an evacuated flight tube. Velocity and consequently drift time t of every ion depends on its m/z ratio ($t^2 \sim m/z$), as heavier ions are slower than lighter ions and a higher charge results in increased velocity. Mass resolution and accuracy in TOF-MS can be improved by an ion mirror (reflectron) at the opposing end of the flight tube that reverts the ions travel direction (**Fig. 5A**). Thereby, the spread in kinetic energy of ions with the same m/z at the exit from the ion source gets corrected [83]. The separated masses of ions in the reflectron can range from a few Daltons to a few million Daltons.

In FTICR, the m/z of an ion is determined based on its cyclotron frequency in a fixed magnetic field. Generated ions are retained in a magnetic field with electric trapping plates (Penning trap) where they are excited by an oscillating electric field orthogonal to the magnetic field. After excitation, the ions rotate at their cyclotron frequency and induce a charge when passing a pair of electrodes (**Fig. 5B**). The resulting interferogram can be converted to give a mass spectrum by means of Fourier transform calculation.

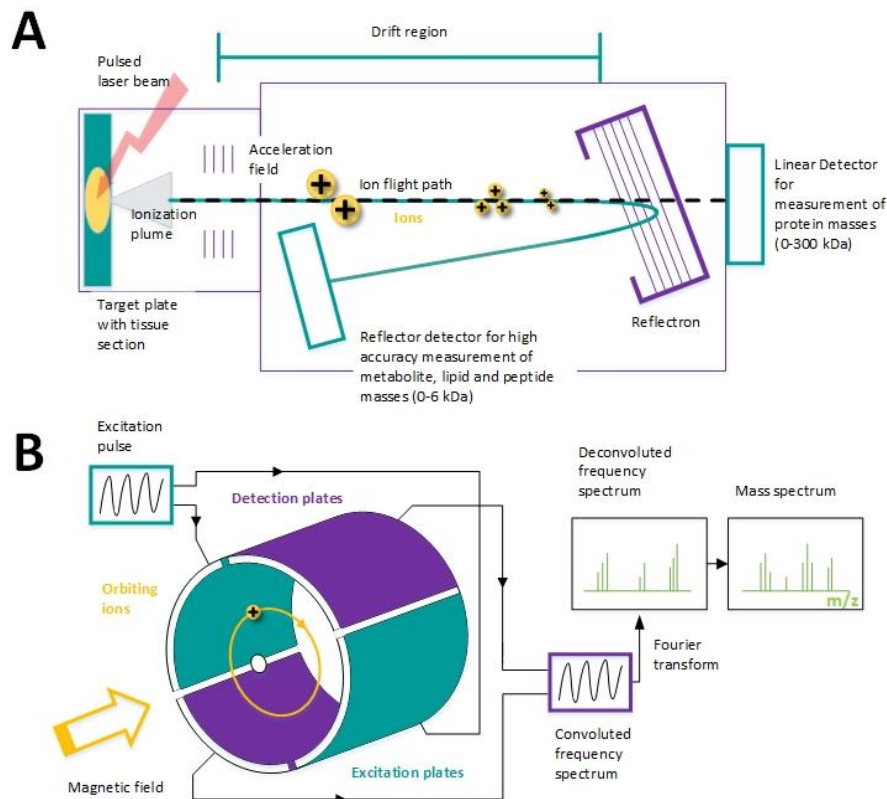


Figure 5 Schematic outline of a time-of-flight (TOF) and Fourier transform ion cyclotron resonance (FTICR) mass analyzer. (A) During a MALDI-TOF-MS measurement, the molecules of the examined sample are ionized, vaporized and directed through an evacuated flight tube. Ions are accelerated by an electric field depending on their m/z -value. Low mass molecules reach the detector in a shorter time than heavier ones and are thus separated. A reflectron may correct the spread in kinetic energy of ions with the same m/z to improve mass resolution and accuracy for smaller molecules. (B) During a FTICR-MS measurement, the generated ions enter a Penning trap in which they are excited at their resonant cyclotron frequencies orthogonal to the magnetic field. Each ion cycles at characteristic radius and frequency and induces an electric signal as it passes the detection plates. The resulting signal is converted to a time-domain frequency spectrum that converts into a mass spectrum by Fourier transform.

1.3.3 Mass spectrometry imaging

Mass spectrometry imaging (MSI) is an *ex vivo* imaging technology that extends conventional MS with a screening process. It enables the assessment of molecular information in a spatially resolved area and was first described by Helmut Liebl in 1967 [84]. The first imaging experiments that utilized MALDI were performed by Richard Caprioli in the late 1990s [85]. In a spatially resolved screening process, ions are created at any given (x/y)-coordinate of a matrix-coated specimen placed on a conductive object carrier (i.e. organ or biopsy sections) by applying laser radiation. Thereby, the precise and standardized preparation of tissue sections as well as a

homogeneous application of matrix substance are of crucial importance. Subsequently to ion production, a spectrum containing the abundance of molecules within a given m/z -range is recorded. The entirety of recorded spectra is used to construct a hyperspectral map in which the prevalence of each acquired m/z -value can be translated into a false-color distribution image (**Fig. 6**).

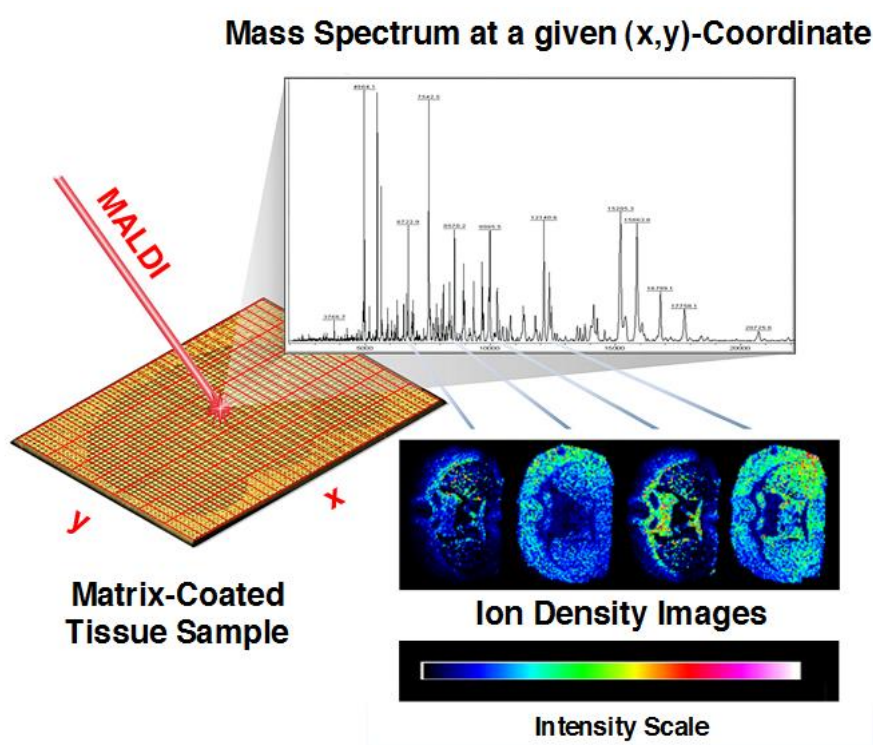


Figure 6 Principle of MALDI-MSI In a screening process, mass spectra are recorded after laser irradiation in a grid-like manner for multiple (x,y) -coordinates. At each spot, m/z -values of abundant ions are recorded and stored within a given range. The prevalence of each m/z can be presented as a false-color distribution map and registered to the tissue's histology. (Figure source [86]).

1.3.4 Diagnostic and pharmacological applications of MSI

The ability of MSI to localize chemical compounds like pharmaceuticals as well as uncover disease-related biomolecular changes has led to a rapid and substantial impact on clinical and pharmacological research (reviewed in [87]). In comparison to conventional *in vivo* imaging techniques and stains, MSI provides unparalleled label-free chemical insight into the distribution pattern of hundreds of biomolecules. Several key studies have demonstrated its ability to uncover disease-related biomolecular shifts, even if highly localized or invisible to established histopathology. For this reason, one of the most prevalent applications for MSI is the identification and localization of disease-associated biomarker profiles that have been demonstrated to

complement classical histopathology (reviewed in [88]). In addition, MSI also plays a vital role in drug discovery as it provides imaging of pharmaceuticals and their metabolites in two-dimensional sections [89] and 3D tissue models that combine multiple subsequently acquired images [90]. In recent publications, MALDI-MSI-derived mass profiles were used to distinguish between transformed cells in formalin-fixed paraffin-embedded (FFPE) and fresh frozen cancer tissues [91, 92]. Moreover, marker proteins identified by means of MALDI-MSI were used to predict survival rates in gastric cancer [93] as well as therapy response in breast cancer [94] and esophageal carcinoma [95]. These studies reflect the high potential of MALDI-MSI in identifying and visualizing transformed cells in tissues as well as the molecular changes linked to their aberration [96]. However, cell heterogeneity impedes the identification of true molecular biomarkers, as it becomes uncertain which mass signature belongs to which subpopulation. For this, the prior selection of “pure” cell groups of similar molecular makeup is necessary, which emphasizes the need for prior histological specification. In this regard, the ability to apply tissue staining after MSI acquisition [97] without the loss of histopathological features has fueled the integration of MSI and classical histology [21, 98]. However, particular limitations still impede the applicability of MALDI-MSI in clinical practice, as the acquisition time needed per section makes especially ultrahigh-resolution MSI impractical for the analysis of larger tissue cohorts. The high amount of created data and the lack of standardized workflow represents additional challenges. In order to contradict these limitations, first attempts to utilize guiding modalities like FTIR or polarimetry that provide fast in-depth characterization of a tissue’s structural composition were recently reported [18, 20].

1.4 Neuronal disorders and cancer

MSI experiments provide spectral information that further elucidates the molecular shifts that differentiate between tissue types in different disease states. It has therefore been the analytical technique of choice to distinguish between alternated cell populations and retrieve potential biomarkers in many cancer studies [99, 100]. Cancer remains the major cause of death in economically developed countries. In 2015, a study [101] reported that in 2012 about 14.1 million new cases occurred worldwide and 8.2 million deaths were caused by the disease. Two cancer types of particular interest, namely gastrointestinal stromal tumors (GIST) and glioblastoma, will be described in

greater detail. Apart from cancer, MALDI-MSI's capabilities to spatially resolve the distribution of proteins, metabolites, lipids, peptides and pharmaceuticals have made the technique especially useful to examine structural organs. In this sense, MSI has extensively been used to investigate biomolecular shifts in neurological disorders. The potential to differentiate between biomolecule species makes it especially useful for the examination of brain lipid accumulations caused by lysosomal storage diseases, i.e. Niemann-Pick type C.

1.4.1 Gastrointestinal stromal tumors

Among the various phenotypes of cancer, sarcoma that arise from transformed cells of mesenchymal origin are quite rare and account for less than 1% of all adult solid malignant tumors [102]. About 90% of diagnosed sarcomas are soft tissue sarcomas, among which gastrointestinal stromal tumors are the most common. Unlike most gastrointestinal tumors, GIST arises from connective tissue instead of epithelial tissue. The aberrant cells typically originate from the stomach (70%), but can later on disseminate to the liver, omentum and peritoneal cavity. GIST is thought to arise from dispersed interstitial Cajal-like cells that have the potential for bidirectional differentiation towards neural lines and myoid cells. In most cases, the cellular aberration is driven by a mutation of the KIT or PDGFR α gene (80-90%), but wild-types that feature multiple or deviating mutations have also been reported [103]. For this reason, the genetic characterization of GIST is important as it denotes the effectiveness of targeted drug therapies. In cases for which resection is anatomically and physiologically feasible, complete surgical excision of the localized tumor region is the treatment of choice [104]. For the treatment of metastatic and unresectable GIST, Imatinib (Gleevec®) and other tyrosine kinase inhibitors (Sunitinib, Regorafenib) are approved by the United States Food and Drug Administration (U.S. FDA) [105]. Imatinib inhibits, among other targets, c-KIT and PDGFR tyrosine kinase activity by binding close to its ATP binding site, thus locking it in a closed or self-inhibited conformation and ultimately inducing cell death. Although Imatinib also binds to non-cancer cells, these cells typically possess additional redundant tyrosine kinases that secure their function. Imatinib is also considered for adjuvant treatment in order to minimize the risk of tumor resurgence. However, the drug's efficacy highly depends on the respective genotype, as multiple forms of c-KIT- and PDGFR α - mutation-negative

GIST as well as the PDGFR α D842V mutant subtype are considered to be Imatinib-resistant.

1.4.2 Glioblastoma multiforme

Glioblastoma multiforme is the most common and aggressive primary brain cancer found in humans. It is usually rapidly fatal, with a short median overall survival rate of 15 months only [106], although in 3-5% of cases, long-term survival that exceeds three years has been reported [107]. The cellular origin of glioblastoma is unknown, but most likely includes glial type cells, astrocytes, oligodendrocytes progenitor cells and neural stem cells [108]. In contrast to other astrocytoma grades, glioblastoma multiforme is characterized by the presence of necrotic lesions that are surrounded by anaplastic cells and hyperplastic blood vessels. For diagnosis, a combination of CT, MRI and tissue biopsy taking is typically utilized. The standard treatment, if feasible, is gross resection, followed by radiotherapy and administration of Temozolamide, a chemotherapy drug [109]. Temozolamide is an alkylating agent that is toxic to cancer cells due to inhibition of tumor cell DNA replication. Alkylation typically occurs at the N-7 or O-6 position of DNA guanine residues.

1.4.3 Niemann-Pick disease, type C

The group of Niemann-Pick diseases includes multiple severe metabolic disorders that result in sphingomyelin accumulation. Unlike type A and B, for which the cause of disease is complete or partial enzymatic deficiency of acid sphingomyelinase, type C disease is associated with mutations in the NPC1 and/or NPC2 gene that encode endosomal-lysosomal transporter proteins. The mutations lead to impaired intracellular lipid trafficking and accumulation of unesterified cholesterol and glycosphingolipids in the brain and other tissues [110]. The age at onset may vary from the perinatal period (< 3 months) to adult age. The disease has manifold symptoms, including liver and spleen enlargement as well as a variety of neurological disorders, such as epilepsy, cerebellar ataxia, dysarthria, dysphagia, spasticity, psychosis and progressive dementia. To date, treatment options are limited and no known cure exists. Since glycolipid storage appears to contribute to the neuropathological features of the disease [111], Miglustat (Zavesca®), an iminosugar inhibitor, has been approved in

Europe, Japan and Canada for treating progressive neurological complications in patients that suffer from Niemann-Pick disease, type C. Miglustat is a glucose-analogue that irreversibly inhibits the activity of the glycosylceramide synthase enzyme, thus reducing the production and accumulation of glycosylceramide in cells.

2 AIMS

MSI, encompassing all modes and methods can bring informative in-depth molecular insight into the pathology and chemical biology of (pre)clinical tissue samples. The implementation of MSI into the (pre)clinical setting holds great promise in the enhancement of personalized medicine and biomarker detection. However, key limitations of MALDI-MSI are sample throughput, data load, the availability of validated/standardized technical routine's and the integration with conventional pathology or other medical imaging techniques. The analysis of large tissue series is impractical, as established approaches acquire MSI data of whole tissue sections prior to histopathological annotation. In contrast, optical imaging techniques like vibrational spectroscopy can assess the structural composition of tissues with relative ease, but lack the chemical specificity needed for biomarker retrieval. A multimodality approach that combines both technologies capabilities might have substantial impact on clinical practice.

For this reason, the study at hand focused on the following objectives:

- 1) Determination of each modality's capabilities to segment and classify divergent cell populations. In this context, the influence of different slide-coating materials was investigated.
- 2) Development and technical evaluation of an automated and robust workflow that enables combined, multimodal imaging of tissues using FTIR and MALDI-MS imaging techniques.
- 3) Evaluation of the routine's capability to correctly distinguish between normal and diseased cell populations in tissues. For this task, various mouse models as well as human cancer tissues were examined.
- 4) Evaluation of the routine's capability to enable unbiased targeting for MALDI-FTICR-MS, thus addressing computational and practical challenges that restrict the transition of ultrahigh-resolution MS to clinical practice.

3 DATA ANALYSIS BACKGROUND

Extracting significant features out of complex data sets is a major challenge in biomedical imaging. In that regard, the high dimensionality of i.e. FTIR and particularly MS images introduces profound restrictions. Apart from introducing practical and computational limitations regarding data processing, the concept of distance between data points becomes very imprecise as dimensionality grows [112]. This problem referred to as the “Curse of Dimensionality” degrades the predictive power of classification models and the efficacy of segmentation. Multimodality imaging introduces additional challenges as joint interpretation and visualization are impossible without alignment of all images into a common coordinate system. In this section, multivariate statistical tools that tackle the challenges associated with multimodality imaging and high dimensionality will be examined. Moreover, methods that extract spatial and spectral features will be demonstrated.

3.1 Dimensionality reduction

Collection of FTIR and MS spectra results in high-dimensional data, with FTICR spectra typically featuring intensities of around one million m/z -values. This introduces some practical and computational restrictions, as the produced image size becomes very large and the computation time to process the data is drastically increased [113]. Especially for MSI, the collected data is very sparse and contains many dimensions more than it contains information. Dimensionality reduction aims to project the data onto a smaller subspace that preserves its information by combining redundant measures and discarding unsubstantial dimensions. Prominent examples of dimensionality reduction tools are principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) [114, 115].

3.1.1 Principal Component Analysis

PCA is a widespread factorization method that reduces dimensionality while retaining most of the data variance. The algorithm reorients the original data onto a new space that is defined by a number of orthogonal axes (principal components) equivalent to the original data’s dimensionality. The principal components are sorted in descending

order by the amount of variance they preserve. For that reason, the first few principal components (PC) describe the majority of data variance, thus dimensionality reduction can be achieved by discarding later components (**Fig. 7**). Typically, the subset of retained PCs is chosen so that at least 95% of the original data's variance are preserved.

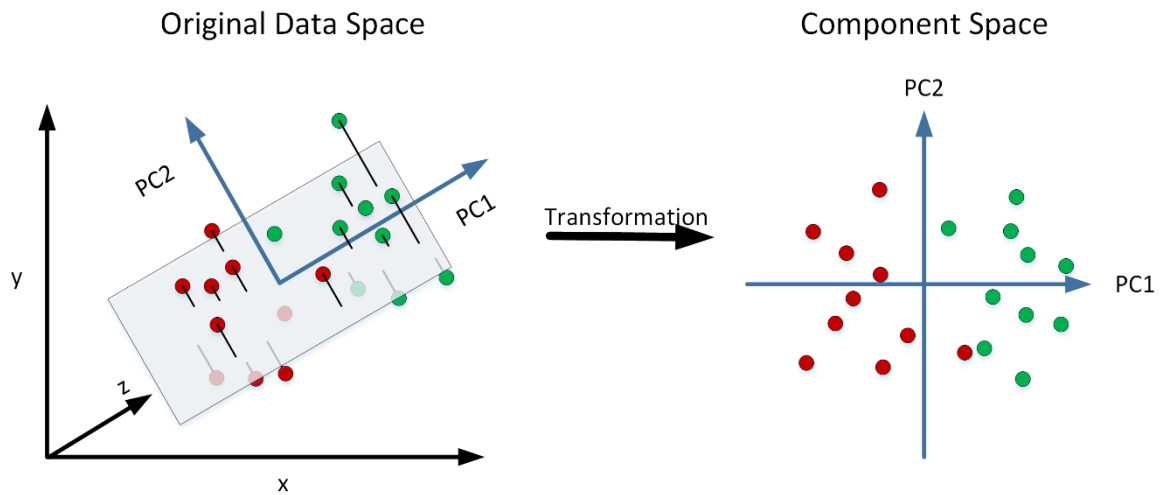


Figure 7 Dimensionality reductions by means of principal component analysis. A reduced two-dimensional component space is defined by the first two principal components (PC1, PC2) and used for transformation of the whole data set (red and green data points). Thereby, the PCs preserve the original data variance in descending order. Transformation of data into a smaller component space allows for dimensionality reduction while preserving most of the original data's variance.

PCs can be located through singular value decomposition (1) in which the mean centered data \tilde{X} gives [116]:

$$\tilde{X} = U\Sigma W^T \quad (1)$$

where Σ is an n-by-p diagonal matrix of the singular values of \tilde{X} , U is an n-by-n matrix of orthogonal unit vectors of length n (the left singular vectors of \tilde{X}) and W is a p-by-p matrix of orthogonal unit vectors of length p (the right singular vectors of \tilde{X}). The right singular vectors W are equivalent to the eigenvectors of $X^T X$, while the singular values of Σ are equal to the square roots of the eigenvalues λ . Using singular value decomposition, the new projection of data T_L (called scores) that considers only the first L largest singular values and their singular vectors is given by (2):

$$T_L = U_L \Sigma_L = \tilde{X} W_L \quad (2)$$

3.1.2 t-distributed stochastic neighbor embedding

PCA is a linear dimensionality reduction technique and thus highlights the major variance while minimizing subtle changes. For this reason, the similarities of highly-dimensional data can often not be faithfully represented. Recently emerging nonlinear methods like t-SNE have proven their capability to preserve local detail and global data structure in life-science applications by emphasizing similarities between data points [117, 118]. The algorithm starts by calculating the distance $d(x_i, x_j)$ between each pair of data points x_i and x_j of the input data X consisting of N objects. Different metrics can be applied for distance calculation, i. e. Euclidean metric (3) that defines $d(x_i, x_j)$ as follows:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{i,k} - x_{j,k})^2} \quad (3)$$

Then, t-SNE constructs a probability distribution over pairs of x_i and x_j in terms of a Gaussian distribution, so that similar objects have a high conditional probability $p_{j/i}$ (4) of being picked as neighbors, whereas this probability to pick dissimilar points is extremely small.

$$p_{j/i} = \frac{\exp\left(-d(x_i, x_j)^2 / (2\sigma_i^2)\right)}{\sum_{k \neq i} \exp\left(-d(x_i, x_k)^2 / (2\sigma_i^2)\right)} \quad (4)$$

$$p_{i/i} = 0$$

Based on this probability, t-SNE then aims to learn a d -dimensional map that reflects the joint probability p_{ij} (5) as precise as possible.

$$p_{ij} = \frac{p_{j/i} + p_{i/j}}{2N} \quad (5)$$

This minimization procedure begins with an initial set Y of random Gaussian-distributed points for which the probability model q_{ij} of the distribution of the distances between points y_i and y_j is calculated (6).

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_k \sum_{l \neq k} (1 + \|y_k - y_l\|^2)^{-1}} \quad (6)$$

$$q_{ii} = 0$$

This initial model is then iteratively updated to minimize the Kullback-Leibler divergence $KL(P \parallel Q)$ between a Gaussian distribution in the high-dimensional space P and a t-distribution in the low-dimensional space Q [119]. For optimization, the algorithm uses a modified gradient descent procedure to reach a good local minimum (7).

$$KL(P \parallel Q) = \sum_j \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (7)$$

3.2 Information extraction

Common biomedical experiments focus either on the identification of discriminatory features used to differentiate between disease states, or segmentation of the image into sub-regions of spectral similarity used for exploratory analysis [120]. In that context, information extraction aims to present the user with spectra or images that represent important features of the data.

3.2.1 k-means++ segmentation

The k-means++ algorithm aims to partition the data into k subgroups of greatest internal similarity [121]. Based on a specific similarity measure, for example the L1-

distance which is more suitable in high-dimensional space and defines the distance between two p -dimensional observations (i.e. spectra) x and y as follows **(8)**, each spectrum is assigned to the closest of k centroids.

$$d(x, y) = \sum_{i=1}^p |x_i - y_i| \quad (8)$$

In contrast to the classic k -means approach (Lloyd's algorithm) in which k initial centroids are placed randomly within the data space and assignment of spectra is achieved in an iterative refinement loop [122], k -means++ uses a heuristic process to improve run time and classification efficacy [121]. Thereby, a random observation of the data set X is chosen as the first centroid c_1 and the distances $d(X_m, c_1)$ of all other observations $m \in n$ are computed. Thereafter, the next centroid c_2 is selected at random with probability defined as follows **(9)**.

$$\frac{d^2(X_m, c_1)}{\sum_{j=1}^n d^2(X_j, c_1)} \quad (9)$$

Center j is chosen by computing the distances from each observation to each centroid and subsequent assignment of each observation to its closest centroid. Then, for $m = 1, \dots, n$ and $q = 1, \dots, j-1$, the next centroid j is selected at random with probability defined as follows **(10)**, with \hat{C}_p being the set of all observations closest to the centroid c_p and X_m belonging to \hat{C}_p .

$$\frac{d^2(X_m, c_p)}{\sum_{\{h; X_h \in \hat{C}_p\}} d^2(X_h, c_p)} \quad (10)$$

Distance calculation, assignment and centroid selection are repeated until k centroids are chosen. For segmentation purposes, it is possible to color-code each pixel of an image based on its allocation to a given centroid.

3.2.2 Feature extraction

MS and FTIR spectra, though rich in biomolecular information contain high levels of redundancy, as many acquired m/z and cm^{-1} intensities either mitigate from correlated sources (i.e. isotopic peaks in MS) or refer to signal-free noise. By extracting informative features (i.e. peaks in MS or wavenumber ranges in FTIR) and discarding the rest, a dataset's size and complexity can be reduced substantially [123]. Feature extraction can also be applied to screen for an optimal sub-feature space that better represents the differences between a pre-defined training set. A suitable technique for the selection of discriminatory features with respect to tissue classification is the receiver operating characteristic curve (ROC) criterion [124]. Thereby, the overlap between the probability density functions describing each feature's data distribution in two classes is calculated as a measure of discriminatory power. For complete overlap the area under the ROC curve (AUC) and the random classifier slope equals zero, whereas an $\text{AUC} = 0.5$ indicates complete separation. The monitored m/z - and cm^{-1} -values obtained from tissues of different morphological makeup are then sorted in descending order based on their AUC.

3.3 Quality measures

Dimensionality reduction and information extraction techniques aim to present manageable and substantial data by reducing redundancy and noise. In order to evaluate the quality of the obtained results several techniques exist.

3.3.1 Cluster evaluation

The number of clusters (k) assumed during k-means++ segmentation is a freely chosen input parameter and a major drawback to the technique, as inappropriate k -values may yield poor results. For this reason, several criteria that serve as a diagnostic check in order to determine the number of anticipated groups in a data set consisting of N observations exist. One of the most common cluster evaluation criteria is the Calinski-Harabasz index published in 1974 [125]. For a given value of k , the Calinski-Harabasz criterion CH_c is a measure of the between-cluster variance SS_B to within-cluster variance SS_W ratio **(11)**.

$$CH_c = \frac{SS_B}{SS_W} \times \frac{(N - k)}{(k - 1)} \quad (11)$$

Thereby, the overall between-cluster variance SS_B is defined as follows (12), with m being the mean of all sample data, m_i being the centroid of cluster i , and $\|m_i - m\|$ being the L^2 norm between both vectors.

$$SS_B = \sum_{i=1}^k n_i \|m_i - m\|^2 \quad (12)$$

In contrast, the overall within-cluster variance is defined as follows (13), with x being a data point, c_i being the i -th cluster and $\|x - m_i\|$ being the L^2 norm between the two vectors.

$$SS_W = \sum_{i=1}^k \sum_{x \in c_i} \|x - m_i\|^2 \quad (13)$$

Well-defined clusters have a large between-cluster variance and a small within-cluster variance, so a larger CH_c resembles better data partition. In order to determine the optimal number of clusters k , the Calinski-Harabasz index can be calculated for multiple k -values. The highest computed CH_c is then assumed to correspond to the optimal number of clusters.

3.3.2 Measure of data dispersion

Several measures can be applied in order to assess the consistency of dimensionality-reduced spectra of similar origin. In that regard, the mean absolute deviation (MAD) is a suitable technique to evaluate stability and technical reproducibility [126]. The MAD of a data set X consisting of x_1, x_2, \dots, x_n is defined as follows (14):

$$\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (14)$$

Thereby, \bar{x} denotes the mean of x . In terms of technical reproducibility, the calculated MAD-value should be as small as possible, as a high value indicates considerable data dispersion.

3.3.3 Sample correlation and dependence

One of the most common measures to determine similarity in scientific research is the sample Pearson correlation coefficient R , which expresses the degree of linear correlation between observations as a value between -1 and +1. Thereby, +1 and -1 refer to a total positive or negative correlation, whereas $R = 0$ refers to no linear correlation. The sample correlation coefficient was developed by Karl Pearson in 1895 and determines the dependence between two series of n observations X and Y [127], where X consists of x_1, x_2, \dots, x_n and Y consists of y_1, y_2, \dots, y_n , as follows (15):

$$\frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right) \quad (15)$$

Thereby, \bar{x} and σ_x denote the mean and standard deviation of X , and \bar{y} and σ_y denote the mean and standard deviation of Y .

3.4 Image registration and evaluation

Image registration is an indispensable tool for the integration of multimodality data sets, especially in medical and biological applications [128, 129]. The process aims to spatially align two or more images of the same object. Typically, one image is defined to be fixed and serves as a reference I_f to align the other moving image/images I_m . In an optimization process, the moving image is then geometrically transformed to be spatially aligned with the fixed image. The optimized transformation parameters μ are

calculated according to the standard registration optimization problem (16), in which the cost function C is minimized with respect to the transformation model T .

$$\hat{\mu} = \arg \min_{\mu} C[I_f, T(I_m)] \quad (16)$$

The accuracy of transformation can be evaluated by calculating the Dice similarity coefficient (DSC) [130]. The DSC defines similarity as the intersection $|X \cap Y|$ between I_m and I_f divided by the numbers of elements (pixels) $|I_m|$ and $|I_f|$ in the two images (17). If $DSC = 1$, a perfect match between transformed and fixed image could be achieved, whereas $DSC = 0$ indicates two completely separated images.

$$DSC = \frac{2|I_m \cap I_f|}{|I_m| + |I_f|} \quad (17)$$

4 MATERIALS AND METHODS

4.1 Materials

4.1.1 Reagents

NAME	CATALOGUE	COMPANY	LOCATION
Acetic Acid	6755.2	Merck	Darmstadt, GER
Acetone	20067.320	VWR Intern.	Darmstadt, GER
Acetonitrile	83640.320	VWR Intern.	Darmstadt, GER
Carboxymethyl Cellulose	6190.1	Carl Roth GmbH + Co. KG	Karlsruhe, GER
Chloroform	83626.320	VWR Intern.	Darmstadt, GER
2,5-Dihydroxybenzoic Acid	8201346	Bruker Daltonics	Bremen, GER
Ethanol absolute	8.18760.2500	AppliChem	Darmstadt, GER
Eosin Y	X883.2	Merck	Darmstadt, GER
Hematoxylin	1.09249.2500	Merck	Darmstadt, GER
n-Hexane	83991.320	VWR Intern.	Darmstadt, GER
Magnesium sulfate	A4837.1000	AppliChem	Darmstadt, GER
Methanol	83638.320	VWR Intern.	Darmstadt, GER
4-Phenyl- α -Cyanocinnamic Acid Amide	SC-1400	SiChem	Bremen, GER
sDHB MALDI Matrix (Mixture of 2,5-Dihydrobenzoic Acid and 2-Hydroxy- 5-Methoxybenzoic Acid	8209813	Bruker Daltonics	Bremen, GER
Sodium hydrogen carbonate	1.06392.0500	Merck	Darmstadt, GER
Trifluoroacetic Acid	1082620100	Merck	Darmstadt, GER

4.1.2 Solutions

NAME		COMPOSITION
Blueing Solution	2 g	NaHCO ₃
	20 g	MgSO ₄
	Add to 1 L with ddH ₂ O	
Carnoy's Fluid	60 mL	EtOH
	30 mL	CHCl ₃
	10 mL	CH ₃ COOH

4.1.3 Consumables

NAME	COMPANY	LOCATION
Centrifuge Tubes 0.5/1/2 mL	Eppendorf	Hamburg, GER
Centrifuge Tubes 15/50 mL	VWR Intern.	Darmstadt, GER
Gold-coated Slides	Science Services	München, GER
MALDI ITO glass slides	Bruker Daltonics	Bremen, GER
MirrIR slides	Kevley Technologies	Chesterland, USA
Nitril examination gloves	VWR Intern.	Darmstadt, GER
Peptide calibration standard II	Bruker Daltonics	Bremen, GER
Pipet Tips 10/200/1000 µL	VWR Intern.	Darmstadt, GER
Protein calibration standard I	Bruker Daltonics	Bremen, GER
Starfrost adhesive slides	R. Langenbrinck GmbH	Emmendingen, GER

4.1.4 Equipment

NAME	COMPANY	LOCATION
Aperio CS2 Digital Pathology Slide Scanner	Leica Biosystems	Nussloch, GER
Autoflex Speed MALDI-TOF/TOF Mass Spectrometer	Bruker Daltonics	Bremen, GER
CanoScan 9000F Mark II optical scanner	Canon Deutschland GmbH	Stuttgart, GER
Cryostat Leica CM 1950	Leica Biosystems	Nussloch, GER
Nalgene™ Polycarbonate Desiccator	Thermo Fisher Scientific	Darmstadt, GER
Frontier MIR system	Perkin Elmer Germany	Rodgau, GER
Matrix application device Sun Collect	SunChrom	Friedrichsdorf, GER
solariX XR 7T MALDI-FTICR Mass Spectrometer	Bruker Daltonics	Bremen, GER
Sonic Bath	VWR Intern	Darmstadt, GER
Spotlight 400 Infrared Microscope	Perkin Elmer Germany	Rodgau, GER
Ultraflex MALDI-TOF/TOF Mass Spectrometer	Bruker Daltonics	Bremen, GER
Vacuum Pump N726 FT.18	KNF Neuberger GmbH	Freiburg im Breisgau, GER

4.1.5 Software

Adobe Photoshop CS6	Adobe	San José, USA
flexControl	Bruker Daltonics	Bremen, GER
flexImaging 4.1	Bruker Daltonics	Bremen, GER
MATLAB R2017a	The Mathworks	Natick, USA
R	R Core Team (https://www.r-project.org)	Vienna, Austria
Tableau	Tableau Software Inc.	Seattle, USA

4.2 Methods

4.2.1 Animal studies

Animal studies on engrafted CD1 *nu/nu* mice were conducted at the German Cancer Research Center (DKFZ, Heidelberg) and supervised by institutional animal protection officials in accordance with the National Institute of Health's guidelines *Guide for the Care and Use of Laboratory Animals*. The experiments were approved by governmental authorities (Regierungspräsidium Karlsruhe, Germany). Female 6-12 week old mice were inoculated with 10^5 human U87-MG glioblastoma cells and sacrificed when the first mouse became symptomatic (5-6 weeks after inoculation). In addition, C57BL/6 mouse brain samples were provided for reference purposes.

Brain samples of Niemann-Pick-C1 I1061T knock-in-mice and C57BL/6 control mice were provided by the Ory laboratory at Washington University in St. Louis, USA [131]. Animals were kept in a controlled animal facility and given standard chow and water ad libitum. Weaning occurred after 3-4 weeks. Experimental procedures were approved by the Washington University Animal Studies Committees and were conducted in accordance with the USDA Animal Welfare Act and the Public Health Service Policy for the Human Care and Use of Laboratory Animals.

All provided animal organs and xenograft tissues were excised and immediately snap-frozen in liquid nitrogen. Until further processing, samples were stored in a freezer at -80 °C.

4.2.2 Human tissue specimen

Patient-derived GIST and prostate cancer samples for the analysis of mid-infrared (MIR) and mass signatures were provided by the University Medical Centre Mannheim. Informed consent was obtained in all cases from patients whose tissues were used in this dissertation. FTIR and MS acquisition were performed in accordance with applicable laws and regulations, good clinical practices and approved by an independent ethics committee of Heidelberg University ("Medizinische Ethik-Kommision II"; No. 2012-293N-MA).

4.2.3 Tissue section preparation

Frozen organs and tissue specimen were cut using a CM1950 cryostat (Leica Biosystems, Nussloch). For all experiments, the cryostat chamber temperature was set to -20°C. First of all, the samples were mounted onto a cryostat metal plate by applying a drop of 2% (w/v) carboxymethyl cellulose (Merck, Darmstadt). Subsequently, the tissue was cut into sections of 8 µm thickness for an improved permeation of mid-infrared radiation [49]. Tissue slides designated for multimodal measurements on single sections were placed on gold-coated slides (Science Services, München). Sections designated for exclusive MALDI-MSI were mounted on indium tin oxide (ITO)-coated glass slides (Bruker Daltonics, Bremen). Sections designated for exclusive FTIR measurements were mounted on MirrIR object slides (Kevley Technologies, Chesterland). Sections prepared for pathological examination and annotation at the Pathology Institute of the University Medical Center Mannheim (UMM) were placed on Starfrost adhesive microscope slides (R. Langenbrinck GmbH, Emmendingen). After cutting, all slides were dried at room temperature (RT) in a Nalgene™ Polycarbonate Desiccator (Thermo Fisher Scientific, Darmstadt) connected to a vacuum pump N726 FT.18 (KNF Neuberger GmbH, Freiburg im Breisgau) for at least 3h. The samples were either measured immediately or stored at -80°C for later image acquisition.

For protein MSI, a multi-step washing protocol was carried out on ice to minimize lipid interferences [132]. The protocol comprises the application of **(1)** 70% ethanol (EtOH; AppliChem, Darmstadt) for 30 sec, **(2)** 100% EtOH for 30 sec, **(3)** Carnoy's fluid, a mixture consisting of 60% EtOH, 30% chloroform (VWR Intern., Darmstadt) and 10% acetic acid (Merck) for 2min [133, 134], **(4)** 100% EtOH for 30 sec, **(5)** purified water (ddH₂O) for 30 sec and **(6)** 100% EtOH for 30 sec. Lipid washing was followed by an overnight drying step in the desiccator at RT.

For the measurement of FFPE tissue samples, a de-waxing protocol was carried out by Malgorzata Muhm. The protocol comprises the application of **(1)** two times n-hexane (VWR Intern.) for 5min, **(2)** 100% EtOH for 5min and **(3)** ddH₂O for 1min.

4.2.4 Multimodality imaging

4.2.4.1 FTIR imaging

FTIR experiments were performed using a Spotlight 400 FTIR Imaging System (Perkin Elmer Germany, Rodgau). Before each image acquisition, the device was precooled by liquid nitrogen and detector check and aperture correction were performed using the Spectrum IMAGE (Perkin Elmer Germany) software. After calibration, tissue-samples on metal-coated slides were placed onto the slide holder. Images used for virtual dissection experiments and feature extraction were taken in the range of 4000-650 cm^{-1} with a spectral resolution of 8 cm^{-1} and a 25x25 μm pixel size. Images recorded for FTIR-guided, exclusive MALDI-FTICR MSI acquisition were recorded in the range of 3200-750 cm^{-1} with a spectral resolution of 12 cm^{-1} and spatial resolution of 6.25 μm . For all measurements, the reflection mode acquisition was specified in the Spectrum IMAGE software. During image recording, mirror velocity was set to 2.2 cm/s in order to reduce acquisition time. Furthermore, it was selected that the stored mean spectrum per pixel would be calculated based on two scans. For each experiment, a tissue-free reference point on the metal-coated slide surface was measured at identical parameters prior to FTIR image acquisition in order to enable automated background subtraction performed in Spectrum IMAGE.

4.2.4.2 Matrix application for MSI

MALDI matrix was deposited on all tissue slides mounted on gold- or ITO-coated slides using the SunCollect automatic sprayer (SunChrom, Friedrichsdorf). Air pressure was set to 2.5 bar with a distance of 25.3 mm between spray head and tissue surface. Matrix substance and application workflow were adjusted with respect to the molecular features of interest during subsequent MS acquisition:

Slides designated for **reflector positive** MS measurement of lipids and small molecules were coated with five layers of 60 mg/ml 2,5-dihydroxybenzoic acid (DHB; Bruker Daltonics) in acetonitrile (ACN; VWR Intern.)/ ddH₂O/ trifluoroacetic acid (TFA; Merck) solution (50/49.5/0.5, v/v/v). The initial layer was applied at 10 $\mu\text{L}/\text{min}$, followed by a second layer at 15 $\mu\text{L}/\text{min}$ and three layers at 20 $\mu\text{L}/\text{min}$. All layers were deposited using the SunChrom slow spraying option (400 mm/min).

Slides designated for **reflector negative** MS measurement of lipids and small molecules were coated with nine layers of 5 mg/ml phenyl- α -cyanocinnamic acid amide (PhCCAA; SiChem, Bremen) in 90% acetone (VWR intern.). The first three layers were applied at 10/15/20 μ L/min, followed by six layers at 25 μ L/min. Matrix coating was performed using the fast spraying option (800 mm/min).

Slides designated for subsequent **linear detector positive** MS measurement of protein signals were sprayed with five layers of 60mg/ml sDHB MALDI matrix (Bruker Daltonics) in ACN/ ddH₂O/ TFA solution (40/59.5/0.5, v/v/v). The initial two layers were applied using a flow-rate of 10 μ L/min followed by three layers at 15 μ L/min. A drying step of 3 min at 37°C was introduced between each spraying step. Layers were spotted using the slow spraying option.

Prior to matrix application, 1 mL of the Bruker peptide calibration standard I was pipetted onto a free slide-spot to enable mass axis calibration during subsequent **reflector** MALDI-MS acquisition of lipids and smaller molecules. Similarly, 1 mL of the Bruker protein calibration standard II was applied prior to the subsequent analysis of **linear detector** MALDI-MS acquisition.

4.2.4.3 MALDI-TOF-MSI

MALDI-MSI experiments were performed on an ultraflex MALDI-TOF/TOF mass spectrometer mass spectrometer (Bruker Daltonics). Data acquisition parameters, i.e. laser shots per pixel, pixel size and acquired mass range, were specified in the flexControl software (Bruker Daltonics) and varied between the experiments. For all experiments, an optical image of the tissue-bearing object carrier was recorded before placement into the mass spectrometer by using a CanoScan 9000F Mark II optical scanner (Canon Deutschland GmbH, Stuttgart). Reflector negative images of NPC1 I1061T knock-in mice brain sections were measured in the range of 100-2000 m/z using 500 laser shots per pixel and a laser raster width of 50 μ m. Single measurements on the previously applied Bruker peptide calibration standard I spot were used to enable mass axis calibration by auto-correlation to a pre-defined list of corresponding peaks. Thereby, the quadratic mass axis correction option was chosen in flexControl.

Single measurements were also used to manually define a suitable detector gain voltage and laser attenuation value. MS images of CD1 *nu/nu* mouse xenograft brain tissue sections were kindly measured on an Autoflex Speed MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) by Dr. Sandra Schulz in reflector positive mode. The mass range of 4000-20000 *m/z* was recorded, using 500 laser shots per pixel and a laser raster width of 75 μm . After defining all necessary parameters in flexControl, the previously recorded optical image was used to define measurement areas in the acquisition software flexImaging 4.1 (Bruker Daltonics). Thereby, a point-based registration step, in which three image pixels and their corresponding MS motor axis positions are defined, is performed by the software to enable affine registration. Within the flexImaging software, measurement areas and/or ROI's can then be manually defined.

4.2.4.4 MALDI-FTICR-MSI

Ultrahigh-resolution MS images were acquired on a 7T SolariX XR MALDI-FTICR-MS (Bruker Daltonics) with the kind support of Dr. Bogdan Munteanu who defined all acquisition parameters needed in the ftmsControl software (Bruker Daltonics). The FTICR-MS device is equipped with an Apollo II dual MALDI/electrospray ionization (ESI) ion source and a 2 kHz Smartbeam II laser (Bruker Daltonics). Acquisition parameters were set in order to record in negative-ion mode within the *m/z* range 150-4000 using absorption mode and a 512 kB data point size. Data point lateral resolution was set to 20 μm and 40 laser shots were applied per measurement position. Single measurements on the previously applied Bruker peptide calibration standard I were used to perform mass axis correction and define a suitable laser attenuation value. Similarly to MALDI-TOF-MSI, measurement areas and ROIs were defined by using the flexImaging 4.1 software that performed three-point registration between manually selected motor axis positions and their corresponding pixels in an optical image of the object carrier.

4.2.4.5 H&E staining

For the purpose of correlation between FTIR-segmentation, MS signal distribution and tissue morphology, cryosections were stained in a multi-step workflow with

hematoxylin and eosin (H&E), a standard histological procedure. The staining protocol included 5 major steps. First of all, potential matrix residues were removed by subsequently dipping the slide in (1) 50% methanol (MeOH) for 3 min, (2) 70% MeOH for 3 min, (3) 100% MeOH for 5 min and (4) 100% ACN for an additional 5 min. In a second step, the section got rehydrated by adding (5) tap water for 5 min, followed by (6) ddH₂O for 5 min. After rehydration, cell nuclei were stained as follows: (7) hematoxylin for 2 min, (8) tap water for 3 min, (9) ddH₂O for 10 sec, (10) 0.3% acid alcohol for 30 sec, (11) ddH₂O for 10 sec, (12) blueing solution consisting of sodium hydrogen carbonate (Merck)/ magnesium sulfate (AppliChem)/ ddH₂O (0.2/2/97.8, w/w/v) for 2 min. Eosinophilic compartments were stained with (13) 0.5% eosin Y for 2 min followed by (14) ddH₂O for 1min. In the final step, the cryopreserved section got dehydrated by dipping the slide in (15) 80% EtOH for 2 min, (16) 96% EtOH for 2 min and finally (17) 100% EtOH for 2 min. H&E staining of GIST tissues was carried out by Maria Deligianni at the pathology institute of the University Medical Centre Mannheim . Optical images of H&E stained tissues were taken by using an Aperio CS2 slide scanner (Leica Biosystems).

4.2.5 Image processing and multivariate statistical evaluation

4.2.5.1 FTIR image pre-processing

In accordance with biomedical practice [39, 135], a script consisting of multiple steps was written in MATLAB (The Mathworks) for pre-processing. Initially, spectra were spatially smoothed by means of an edge-preserving denoising (EPD) function that was kindly provided by Dr. Markus Grasmaier [136, 137]. Spectra were then baseline-corrected using a self-written script that performs asymmetric least squares smoothing with a smoothing factor $\lambda = 1000000$ and a weighting factor $p = 0.01$, as suggested by Eilers and Boelens [138, 139]. Corrected spectra were subjected to first derivative calculation using MATLAB's *diff* function and a script that performs standard normal variate (SNV) normalization which consists of mean centering and division by the standard deviation over the spectral intensities, thus giving the resulting spectra a unit standard deviation of one [140]. Spectral windows within the ranges 3100-2900 cm⁻¹ and 1800-950 cm⁻¹ were selected using MATLAB's *find* function and the remaining wavenumbers that don't comprise vibrations most relevant in biomolecules were discarded.

4.2.5.2 MSI pre-processing

Pre-processing of MS data was carried out in R programming language by utilizing a 'MALDIquant' package-based workflow routine that was developed by Denis Abu-Sammour (unpublished) [141]. Spectra were corrected using total ion current (TIC) normalization, which divides the intensity of a spectrum by its TIC (sum of all ion intensities) and then multiplies by the median TIC of all spectra. Normalized spectra were subsequently subjected to Top Hat baseline correction [142], a morphological filter that combines a moving minimum (erosion filter) followed by a moving maximum (dilation filter). To further minimize dimensionality, peak picking using the Friedman's Supersmoothing was applied with a signal-to-noise (S/N) threshold of $S/N > 3$ [143]. In order to make the data accessible to MATLAB, the raw and preprocessed spectral information was converted to text file format (csv).

4.2.5.3 Dimensionality reduction and evaluation of FTIR and MS data

Principal component analysis with the purpose of dimensionality reduction of FTIR and MS imaging data was carried out in MATLAB by utilizing the *pca* function. Thereby, default parameters were used with the exception of *NumComponents*, which was set to three in order to enable 3D-visualization of the transformed spectra.

In a similar fashion, t-distributed stochastic neighbor embedding was performed by utilizing MATLAB's *tsne* function. The function's *Distance* parameter which defines the underlying metric was set to *cityblock* and the *NumDimensions* parameter was set to 3 in order to enable 3D-visualization.

Feature selection and MS peak selection using local maxima identification was achieved by using the *findpeaks* function in MATLAB with default parameters.

The dispersion of transformed data points in tSNE or PCA space was evaluated by calculating the MAD. This was achieved in MATLAB by using the *mad* function and default parameters.

4.2.5.4 Image segmentation and cluster evaluation

FTIR images were segmented in MATLAB using the *kmeans* function with *cityblock* distance metric and 3 *Replicates* using new initial cluster centroid positions [121]. The needed value *k* that denotes the cluster's number of allocation groups was pre-defined based on the examined tissue specimen. Evaluation of FTIR segmentation results for different *k*-values was done by using MATLAB's *evalclusters* function with *criterion* being set to *CalinskiHarabasz*. Thereby, the respective *kmeans* algorithm was specified with a function handle. The highest computed criterion was then assumed to correspond to the optimal cluster number.

Similar to FTIR segmentation, MS images for comparative analyses were dissected in R programming language by Denis Abu-Sammour. In this context, the 'MALDIquant' package was utilized to enable k-means++ segmentation as well as state-of-the-art spatially-aware clustering that enables a more stable segmentation of large and highly-dimensional data sets [144]. MSI-based segmentation results were converted to text file format (csv) in order to make the data accessible to MATLAB.

4.2.5.5 Registration of binary images and efficacy estimation

Binary information depicting the presence and absence of tissue was extracted from images obtained with different sensors for registration purposes. This was achieved by means of k-means++ cluster analysis using the *kmeans* function in MATLAB and the *cityblock* distance metric. Segmentation of spectra obtained on tissue and plain surface by the use of (*k* = 2)-clustering was considered to separate between background and on-tissue signals, as they differentiate the most. Therefore, a binary image was created by setting all pixels belonging to the allocation group with less mean absorption to zero and the remaining pixels with higher mean absorption to one. Subsequently, the geometric transformation matrix needed for registration between the binary images was calculated in MATLAB by utilizing the *imregtform* function with *transformType* being set to *affine*. The function's *optimizer* and *metric* parameter were given by the *imregconfig* function with *modality* being set to *monomodal*. The derived geometric transformation matrix could then be used to register the moving binary image's original sources. This was done by using the *imwarp* function in MATLAB. In order to assess

registration efficacy, a self-written script that calculates the DSC between moving and fixed binary image was utilized.

4.2.5.6 Extraction of discriminatory features

MS and FTIR features that are rich in discriminatory power with respect to tissue classification were extracted by using the *rankfeatures* function in MATLAB. The utilized *Criterion* was set to *roc* in order to enable ROC-based feature extraction. For the remaining parameters of *rankfeatures*, default values were used.

4.2.5.7 Evaluation of spectral similarity

The degree of linear correlation between spectra was assessed by calculating the sample Pearson correlation coefficient. Thereby, the *corrcoef* function in MATLAB was utilized with default parameters.

5 RESULTS

The results of this thesis will be subdivided into four major parts which revolve around its major objectives:

- 1) The first part presents single modality capabilities of FTIR and MALDI-MS imaging with respect to segmentation and classification. In this context, the influence of different slide-coating materials and fixation is investigated.
- 2) In the second part, the technical and computational development of a sensitive, robust and reproducible multimodality imaging workflow that utilizes the concept of guidance in order to combine FTIR micro-spectroscopy and MALDI-MSI is presented.
- 3) The third and major part focusses on evaluation of the developed workflow for discriminatory purposes in animal and human tissue samples. The capability of identifying divergent cell populations (i.e. tumors) as well as extracting morphology-dependent marker signatures, is examined.
- 4) Lastly, the computational and practical challenges related to high mass-resolution MSI are addressed by presenting a more integrated multimodality imaging workflow that enables unbiased targeting of tissue morphologies prior to MSI acquisition.

5.1 Exploration of single modality capabilities

Single modality experiments were carried out in order to evaluate the multimodality capabilities of FTIR micro-spectroscopy and MALDI-MSI. The main focus was to demonstrate modality-dependent advantages of FTIR imaging that would provide the basis for its beneficial integration into a multimodality imaging workflow. In this context, the impact of technical parameters like slide-coating composition, fixation and storage was examined with respect to tissue segmentation and characterization.

5.1.1 The influence of object slide coating on FTIR spectroscopy

The substrate composition on which FTIR measurements are performed has a critical effect on image quality [145] and potentially influences tissue segmentation. In general, metal substrates like gold, aluminum or MirrIR, a reflective silver and tin oxide (Ag/SNO₂) multilayer coating, are used for reflection measurements. In order to determine the optimal optical substrate for FTIR tissue imaging, these three different substrate composition slides were evaluated. Porcine 8 μ m liver tissue sections were chosen due to their high cellular homogeneity and mounted on ITO-, MirrIR- and gold-coated slides. Indium tin oxide slides were examined as they represent the standard material for MALDI-TOF MS [146, 147] that permits visible light transmission and good conductive properties [148]. For each slide's substrate composition, the mean mid-infrared background absorption of 60 scans was recorded (**Fig. 8A**) prior to triplicate measurements (**Fig. 8B**) of rectangular on-tissue areas that consisted of 1600 spectra each. All recorded spectra were computed in PCA space (**Fig. 8C**) in order to visualize each image's dispersion. Thereby, 88.67% of the original data's variance was preserved within the first three PCAs.

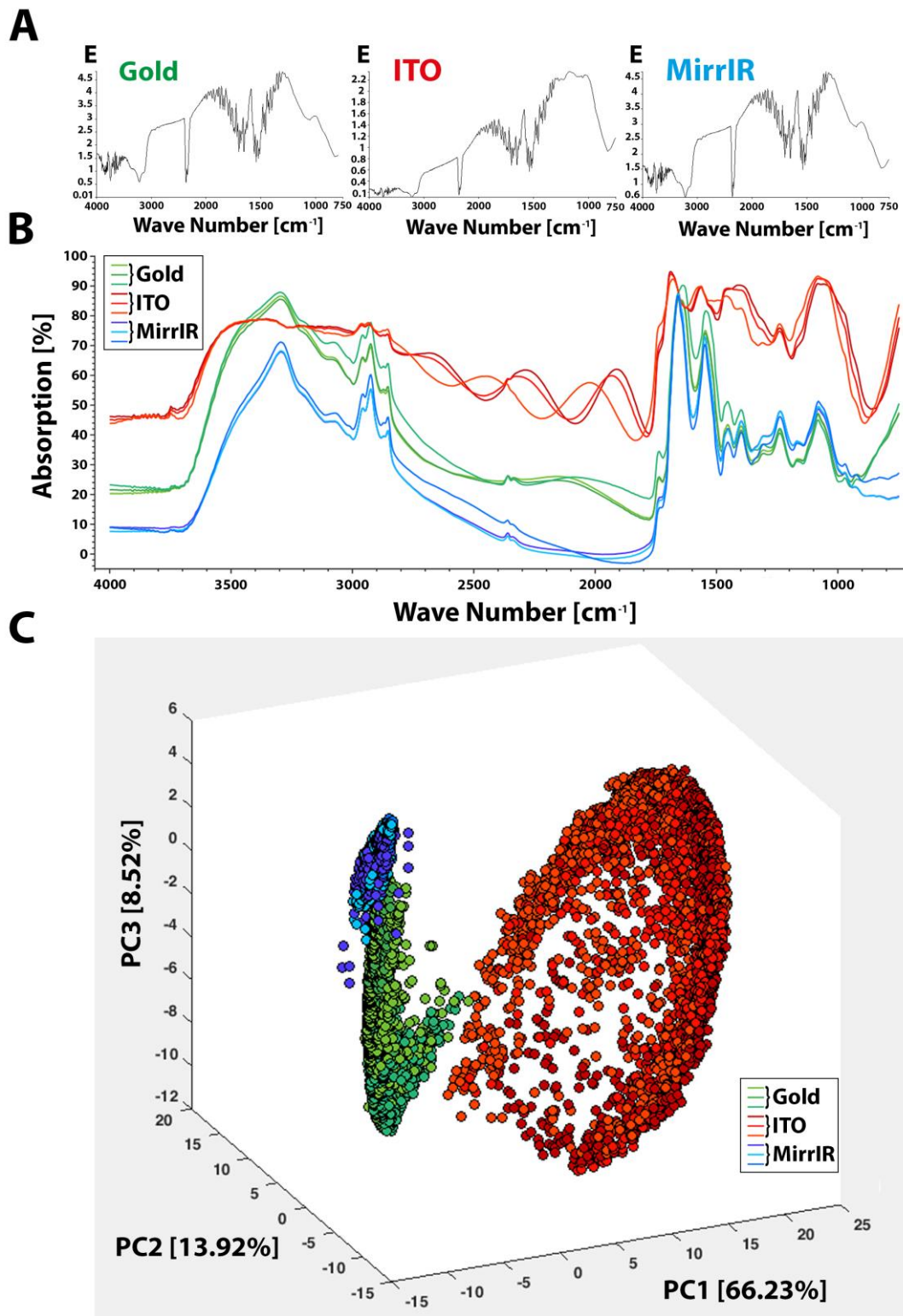


Figure 8 FTIR reflection measurements on different substrate composition slides (A) FTIR background absorption was acquired on gold-, ITO- and MirrIR-coated microscopy slides prior to subsequent measurement of 8 μm porcine liver tissue cryosections. For each liver section, three rectangular measurement areas consisting of 1600 positions with a pixel size of 25x25 μm were defined using the Spectrum IMAGE software. (B) Comparison of each acquisition area's mean spectrum reveals a higher overall absorption (lower S/N) when using ITO-coated slides. The spectra acquired on gold- and MirrIR-coated object carriers appear comparable with a slightly higher mean absorption on gold. (C) By transforming all 14400 spectra into PCA space, the dispersion created by each carrier's coating was visualized. It becomes apparent that FTIR spectra acquired on ITO are less stable when compared to gold or MirrIR.

The monitored background absorption on each coating material appeared comparable, though an increased absorption in the range of 1200-800 cm^{-1} was observed with ITO (**Fig. 8A**). Moreover, an increased mean absorption (65.99%) was observed for tissue spectra acquired on ITO when compared to on-tissue measurements using gold- (41.04%) or MirrIR-coated object slides (27.12%). This difference was found to be mainly caused by a comparatively increased base-intensity. As more than 40% of the incident light was absorbed in the range of 2550-1700 cm^{-1} that is considered devoid of biomolecular vibration signals (**Fig. 8B**), it can be assumed that the coating-material itself absorbs mid-infrared radiation and therefore dilutes spectral information content. The negative influence of ITO coating can further be emphasized by transforming all recorded spectra into three-dimensional PCA space (**Fig. 8C**). Assuming that liver tissue is homogeneous, all recorded spectra should fall together after transformation. However, a considerably increased data spread could be observed for ITO-derived spectra. As a quantitative measure of spectral dispersion, the MAD-value was calculated for each sample set, confirming a higher stability for on-MirrIR infrared acquisition (MAD = 3.51) when compared to ITO (MAD = 7.21) or gold (MAD = 4.32).

However, spectral quality does not necessarily denote segmentation efficiency. In order to investigate the influence of different coating-composition materials on segmentation, subsequent murine C57BL/6 brain tissue cryosections of 8 μm thickness were mounted on gold-, ITO- and MirrIR-coated object slides. The brain was chosen as a suitable test model due to its heterogeneous morphological composition. In order to elucidate the accuracy of image dissection, the Allen Brain Reference Atlas [149], which summarizes publicly available data about the brain's structural composition (**Fig. 9A**) was used as a reference. The derived contours obtained from dissecting each FTIR image into seven segments by means of ($k = 7$) k-means++ cluster analysis were color-coded by assigning a unique color to each pixel of a given index (**Fig. 9B**). Image dissection into seven segments was chosen based on the calculated Calinski-Harabasz criterion CH_c for all k -values in the range of 5-10. The retrieved CH_c suggested an optimal number of seven clusters for the acquired on-gold image as well as the acquired on-MirrIR image and five clusters for the acquired on-ITO image. In order to compare between segmentation results, all images were partitioned based on the highest retrieved cluster optimum and optically compared to

structural brain atlas regions (**Fig. 9A**). Thereby, matching colors were assigned to the pixels of each allocation group. As indices are randomly assigned during cluster analyses, comparable images were manually selected out of all 5040 possible color-coded maps (**Fig. 9B**). All retrieved segmentation images showed a rough division into morphological brain tissue regions, namely hippocampal region, fiber tracts, thalamus, hypothalamus and isocortex. In comparison, dissection of FTIR images recorded on ITO appeared noisy with less defined contours and poor rendering of the thalamus region (**Fig. 9C**). Interestingly, spectral dissection on gold-coated object slides additionally uncovered the granule cell layer of the dentate gyrus and *cornu ammonis* brain region (**Fig. 9D**). This granule cell layer is subject to multiple neurological disease studies i.e. on Alzheimer [150], epilepsy [151] and electroconvulsive therapy [152].

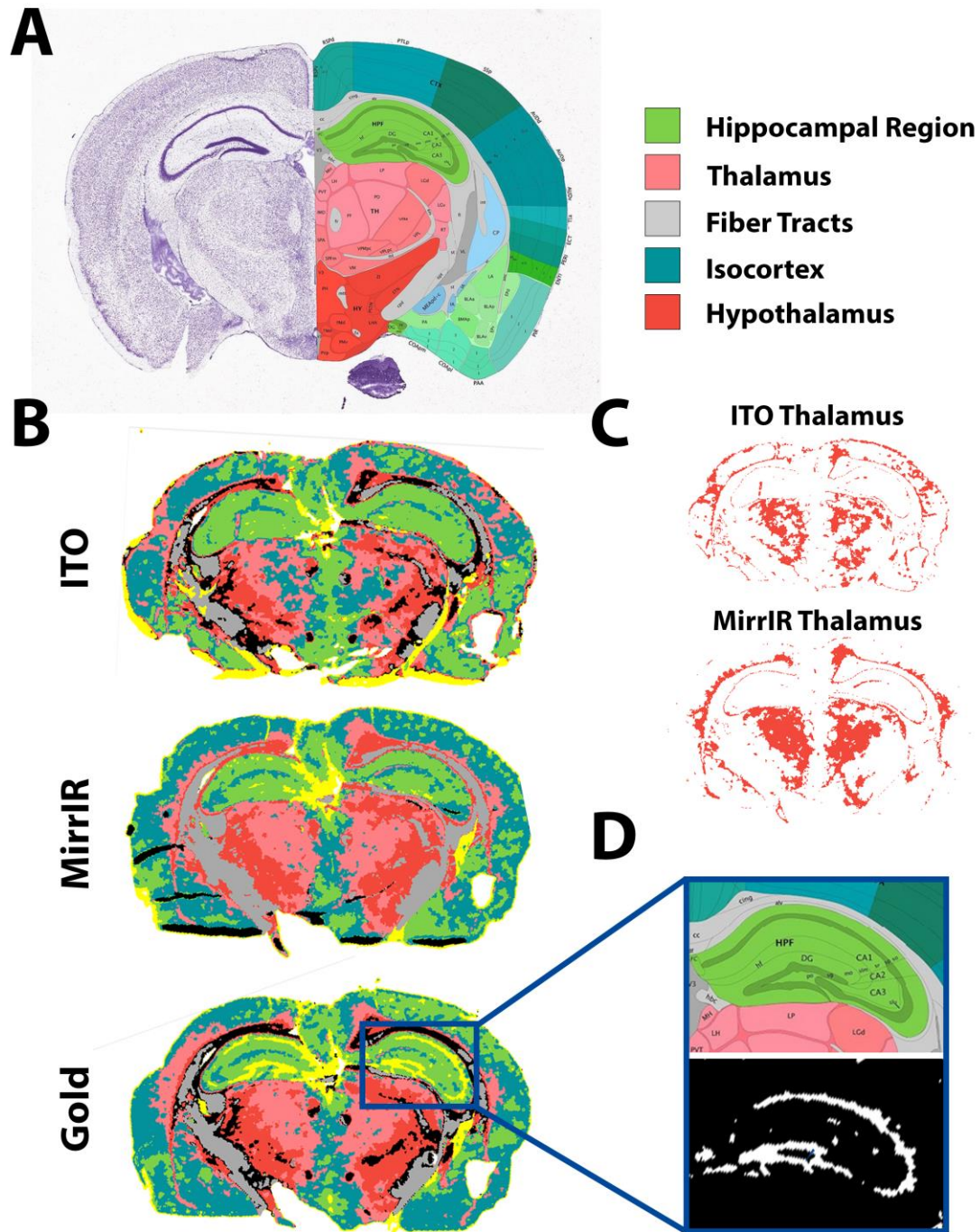


Figure 9 FTIR reflection based brain tissue segmentation on different substrate composition slides
(A) The Allen Brain Reference Atlas (Image credit: Allen Institute Mouse Brain Atlas [149], P56, Coronal, Image 76 of 132) illustrates a mouse brain's structural makeup that enables comparison to **(B)** the spatial contours derived by $(k = 7)$ k-means++ clustering of FTIR images recorded on slides coated with different substrate materials (ITO, MirrIR, gold). Images were recorded on subsequent cryosections of the same C57BL/6 mouse brain sample. The obtained cluster results were used to express spectral grouping in a color-coded image, revealing spectral differences linked to morphological regions, namely fiber tracts, thalamus, hypothalamus, isocortex and hippocampal region. **(C)** Exemplified comparison of the thalamus-section's contours reveals poor rendering of the ITO-retrieved segmentation image. **(D)** Magnification of the hippocampal region and the yellow cluster of the on-gold acquired FTIR image demonstrates the identification of an additional region of interest, namely the granule cell layer of the dentate gyrus and *cornu ammonis* brain region

In summary, FTIR imaging on ITO-coated slides led to poorer spectral quality and less stability. Spatial contours after segmentation were blurry and revealed less morphological structures. The use of MirrIR-coated slides resulted in images of improved spectral quality with less dispersion and a good resolution of the amide I&II region ($1500\text{-}1700\text{ cm}^{-1}$) and the fingerprint region ($600\text{-}1450\text{ cm}^{-1}$). Spectral measurements on gold appeared slightly worse. However, segmentation results were of comparable quality and revealed additional morphological structures.

5.1.2 The influence of object slide coating on MS quality

The conductive properties of different substrate coatings and their interaction with the electronic states of matrix molecules have been proven to influence the ionization process of MALDI- [153], DESI- [154] and other MSI techniques [155]. In order to elucidate the influence of different substrate composition materials on spectral MALDI-MSI quality, subsequent porcine liver sections of $8\text{ }\mu\text{m}$ thickness were placed on ITO- and gold-coated slides. MirrIR-coated slides were not used as they didn't fit into the Bruker target plate holder. Tissues were sprayed with DHB in ACN/ ddH₂O/ TFA (50/49.5/0.5, v/v/v) using the SunChrome spraying system. To ensure comparability, both object slides were jointly put in the same plate holder, and reflector positive MS images were subsequently acquired with identical parameters. For each liver section, three rectangular regions consisting of roughly 50 acquisition spots were defined in flexImaging. The mean spectra of 177 mass spectra acquired on ITO and 169 mass spectra acquired on gold were compared (**Fig. 10**).

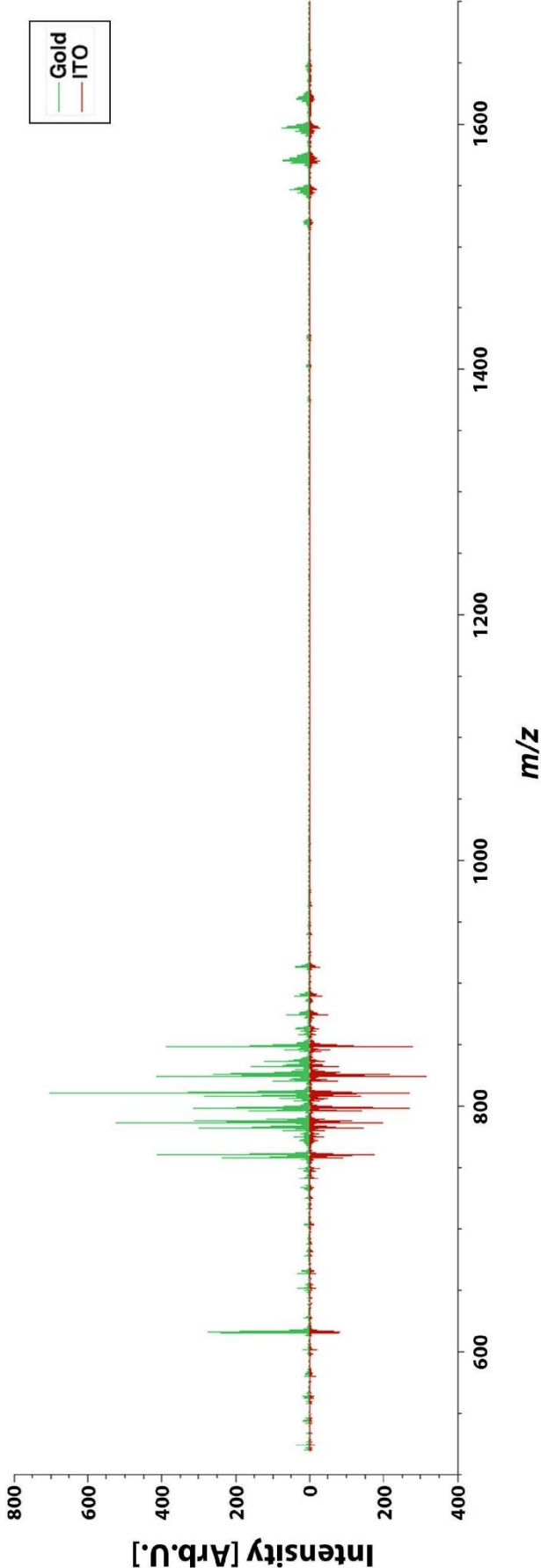


Figure 10 Reflector positive MSI of porcine liver tissue sections on gold- and ITO-coated slides. Porcine liver sections were placed on object carriers coated with different metals. The mean spectra of 177 MS spectra acquired on ITO and 169 MS spectra acquired on gold indicate a substrate composition dependency. Spectra derived from gold-coated slides are of higher intensity, though the overall pattern appears comparable.

The resulting tissue pattern was comparable between spectra acquired on gold and ITO. In general, peak intensities on gold were higher and no negative influence of on-gold MALDI-MSI was observed with respect to m/z -signals. On the contrary, peak selection using local maxima identification in MATLAB identified a higher number of peaks (992) within mass spectrometry data recorded on gold when compared to MS data recorded on ITO (740). However as this experiment was only carried out once, it is difficult to assume better spectral quality for on-gold measurements. In order to compare each data set's dispersion, all recorded mass spectra were transformed by means of PCA and each spectrum's position within the first 3 PC space was visualized (**Fig. 11**). Thereby, 89.86% of the original data's variance was preserved.

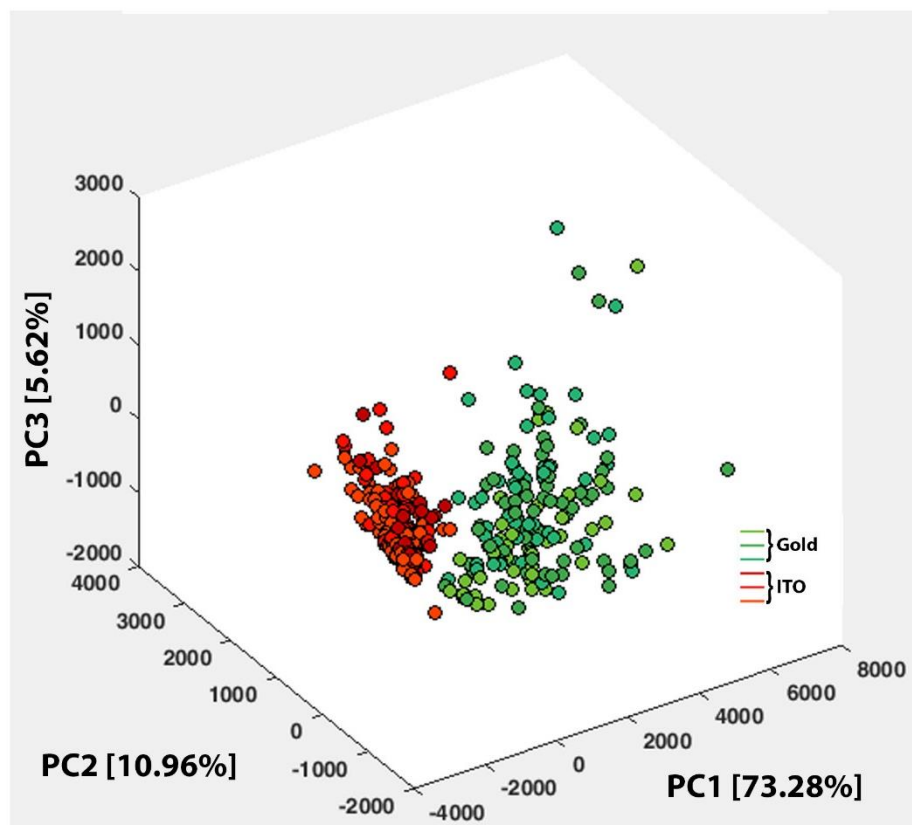


Figure 11 PCA transformation of MS data acquired on gold- and ITO-coated slides. Porcine liver tissue sections were mounted on different metal-coated object carrier and subsequently used for reflector positive MS measurements. Visualization of all 177 spectra acquired on ITO and 169 spectra acquired on gold within the first three PCs demonstrate distinct clustering based on the used coating material.

The transformed mass spectra acquired on gold and ITO formed two distinct clusters which emphasizes the influence of substrate composition on MALDI-MSI. After normalization, the calculated MAD revealed an increased scattering of 6.1% for

on-gold MSI ($MAD = 6.05 \cdot 10^{-4}$) when compared to on-ITO measurements ($MAD = 5.7 \cdot 10^{-4}$). In comparison to FTIR imaging results (**chapter 5.1.1**) for which the dispersion on ITO was increased by 66.9% when compared to gold, this difference in dispersion appears less significant. In summary, the use of gold-coated object slides led to no observable major negative effects, although the obtained spectra were easily distinguishable in PCA space. As this experiment was carried out only once, it remains uncertain whether the distinct clusters indicate improved or worsened spectral quality when measuring on gold-coated object slides. Based on these observations and taking into account the results obtained in chapter 4.1.1, gold was chosen as the material of choice for multimodal acquisition in both systems.

5.1.3 Tissue segmentation by FTIR and MS imaging

The differences between FTIR segmentation for later MSI guidance and direct MS image segmentation were compared. Brain tissue sections of CD1 *nu/nu* mice inoculated with orthotopic U87-MG glioblastoma xenografts were chosen as a suitable test case with rather homogenous lesions. Subsequent sections of 8 μm thickness were mounted on gold- and ITO-coated slides for FTIR spectromicroscopy and MS imaging respectively. H&E stains of on-gold tissues (**Fig. 12A**) were prepared after FTIR image acquisition to serve as a reference for the evaluation of FTIR- and MS image segmentation results. FTIR- images were partitioned by means of k-means++ cluster analysis with an increasing number of centroids ($k = 3$, $k = 4$, $k = 5$). Thereby, the spatial coordinates of all spectra were color-coded based on their allocation to a given centroid. As indices were randomly assigned during each cluster analysis, each segmentation image received an independent color-code. The obtained segmentation images were compared to MSI-based segmentation results using k-means++ clustering as well as state-of-the-art spatially-aware clustering that were provided by Denis Abu-Sammour. In this context (**Fig. 12B**), dissection of MS images using spatially-aware clustering methods (MS_{spk}) was found to be less noisy and featured sharper segment-contours when compared to k-means++ based MS segmentation (MS_{k++}). Despite the lack of histopathological annotation which impedes quantification assessment of segmentation accuracy, the fit between tumor boundaries visible in H&E stains and the FTIR-based k-means++ segmentation image (IR_{k++}) appeared to be significantly better when compared to MS-based segmentation results. Dissection

using five centroids ($k = 5$) added additional morphological information to FTIR segmentation images and revealed to some extent tumor boundaries in MS_{k++} which were not observable for lower k -values. In MS_{spk} images, increasing k added no additional morphological information.

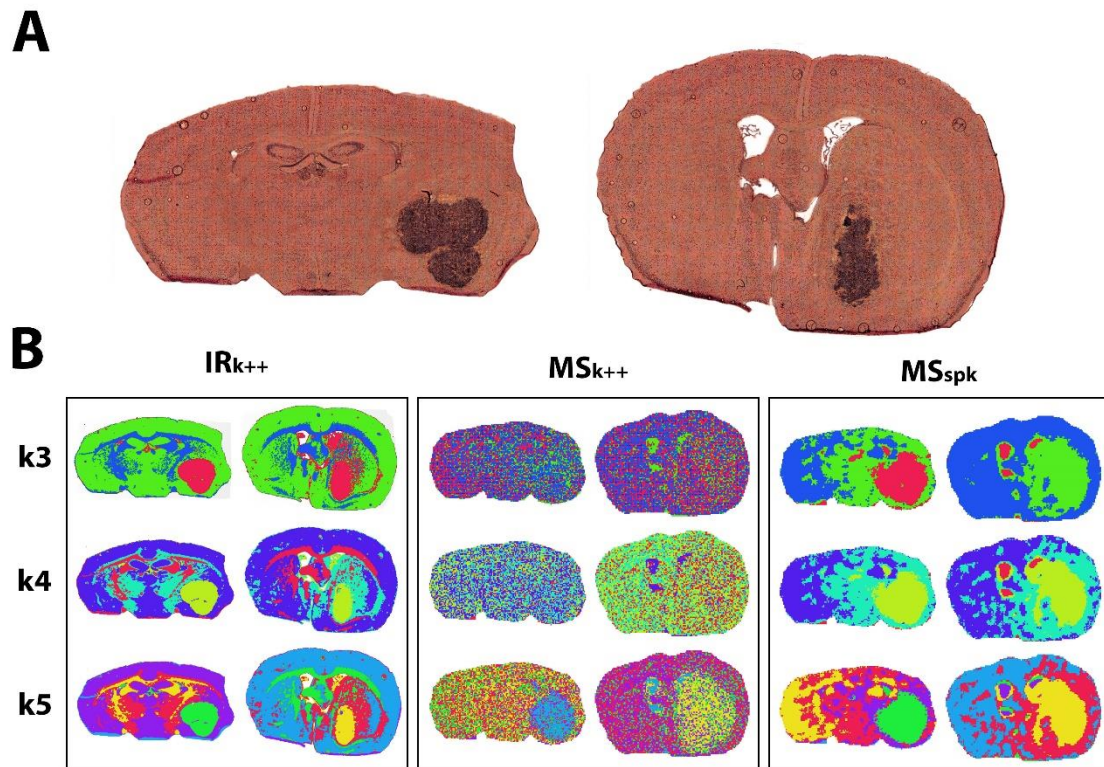


Figure 12 FTIR- and MS-based segmentation of engrafted mouse brain sections. Brain cryosections of CD1 *nu/nu* mice inoculated with human U87-MG glioblastoma cells were mounted on gold-coated slides. (A) After FTIR and MALDI MS image acquisition, H&E stains of the measured sections were prepared to visualize the spatial distribution of tumor cells. Based on this reference, (B) segmentation results based on the previously obtained FTIR- and MS images were evaluated. The recorded images were *in-silico* dissected into multiple areas ($k = 3$, $k = 4$, $k = 5$) of greatest possible data homogeneity. Each spectrum allocation was used to create a color-coded image in which identical colors represent spectral similarity. The resulting segmentation images derived from k -means++ clustering of FTIR- (IR_{k++}) and MS images (MS_{k++}) as well as spatially-aware clustering of MS images (MS_{spk}) are compared.

For the purpose of quantifying each modality's potential to monitor tumors in tissues by means of segmentation analyses, images of resected human tissue samples containing GIST were recorded. For reference purposes, adjacent sections were prepared on Starfrost adhesive slides and subsequently H&E-stained (Fig. 13A). Histopathological annotation was provided by the Pathological Institute at the University Medical Center Mannheim (red marks). FTIR-images were dissected into three areas of greatest homogeneity using ($k = 3$) k -means++ cluster analysis and compared to MSI-based k -means++ and spatially-aware image segmentation results

provided by Denis Abu-Sammour. The k -value needed for image partitioning was set to three based on the preceding U87-MG glioblastoma xenograft segmentation results for which $k = 3$ sufficed to visualize tumor contours. The obtained spectral allocation results were then used to form a color-coded segmentation image (**Fig. 13B**). Both the annotated tumor contours as well as the matching colored segments were converted to binary in order to quantify similarities. The segment-retrieved moving black-and-white images were registered to the fixed tumor annotation image by using automated intensity-based image registration tools in MATLAB (**Fig. 13C**). As a measure of precision, the DSC which quantifies the overlap between tumor contours and each modality's segment binary was calculated. Disaggregation based on FTIR data (DSC = 0.881) was proven to better represent tumor contours when compared to MS-based segmentation using spatially-aware clustering (DSC = 0.505) or k -means++ cluster analysis (DSC = 0.383).

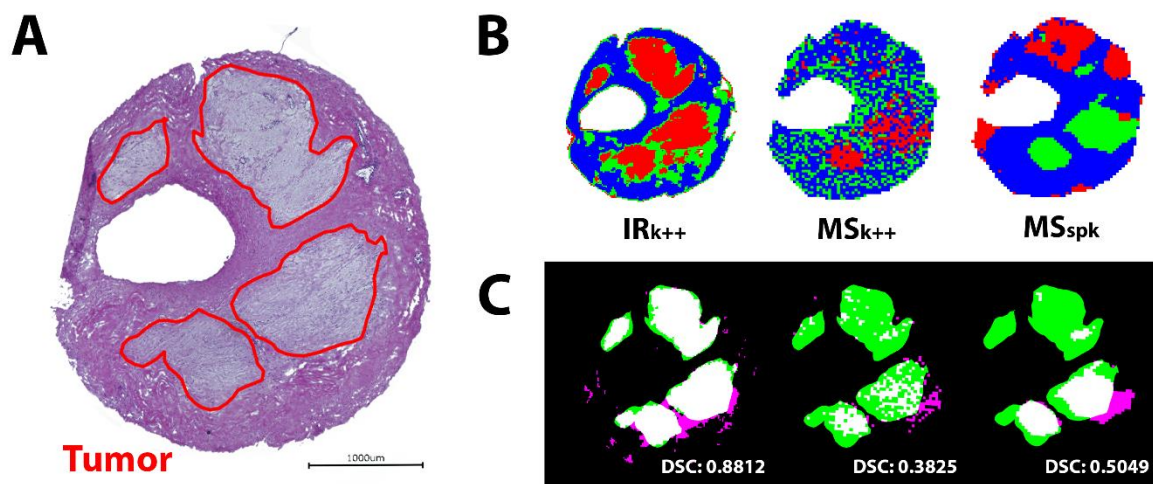


Figure 13 FTIR and MS based segmentation for tumor recognition in human GIST tissue samples. (A) H&E stains of human patient-derived GIST tissues were pathologically annotated to localize tumor areas. (B) IR and MS images recorded on subsequent cryosections were *in-silico* dissected into three areas ($k = 3$) by performing k -means++ clustering of IR (IR_{k++}) and MS data (MS_{k++}). Additionally, MS images were dissected using spatially-aware clustering (MS_{spk}). For visualization, the retrieved segments were color-coded (red, green, blue). (C) Evaluation of each segmentation's efficacy was achieved by converting the pathologically annotated tissue area as well as the matching segment to binary. The retrieved images were overlaid by means of intensity-based registration, showing matching pixels (white) as well as areas unique to the pathological annotation (green) or segment (magenta). Calculation of the DSC demonstrates a more precise tumor identification (DSC = 0.881) for FTIR-based tissue disaggregation when compared to MS-based segmentation. Tissue dissection of MS data by means of spatially-aware clustering proved superior (DSC = 0.505) when compared to MS_{k++} (DSC = 0.383).

In conclusion, segmentation of FTIR images proved to add critical information about a tissue's morphological makeup when compared to MS image clustering. In addition, the technique operates comparatively fast and is less computationally expensive. FTIR

images of 25 μm spatial resolution were recorded in 65 min for both glioblastoma mouse brain sections and 8 min for the presented GIST section. In contrast, MS images of both glioblastoma brain sections featured a lower spatial resolution of 75 μm and recording took 6.1 h on the ultrafleXtreme MS instrument. The presented GIST MS image was recorded in 54 min at 50 μm resolution.

5.1.4 Feasibility of FTIR imaging on FFPE tissues

Tissue embedding is known to have a critical influence on spectral quality in both MS and FTIR imaging. It is therefore a major obstacle, as histopathological examination and storage is typically done using FFPE tissues. Although de-waxing protocols have been proposed for FTIR imaging, little is known about the effect of FFPE residues and de-waxing on spectral quality and segmentation. In order to evaluate FTIR's capability to record and dissect FFPE tissues, embedded, de-waxed and cryopreserved porcine kidney samples were mounted on MirrIR-slides. The derived MIR-spectra recorded on two embedded, two de-waxed and two cryopreserved liver sections were compared (**Fig. 14**). Thereby, calculation of the sample Pearson correlation coefficient as a quantitative measure of spectral similarity revealed that de-waxing led to an increased linear correlation between MIR-spectra acquired on FFPE and cryopreserved tissues ($R = 0.95$) when compared to untreated FFPE tissues ($R = 0.89$). However, an increased mean absorption was observed for MIR-spectra acquired on FFPE (56.2%) and de-waxed (60.8%) tissues in comparison to measurements on cryopreserved material (25.9%).

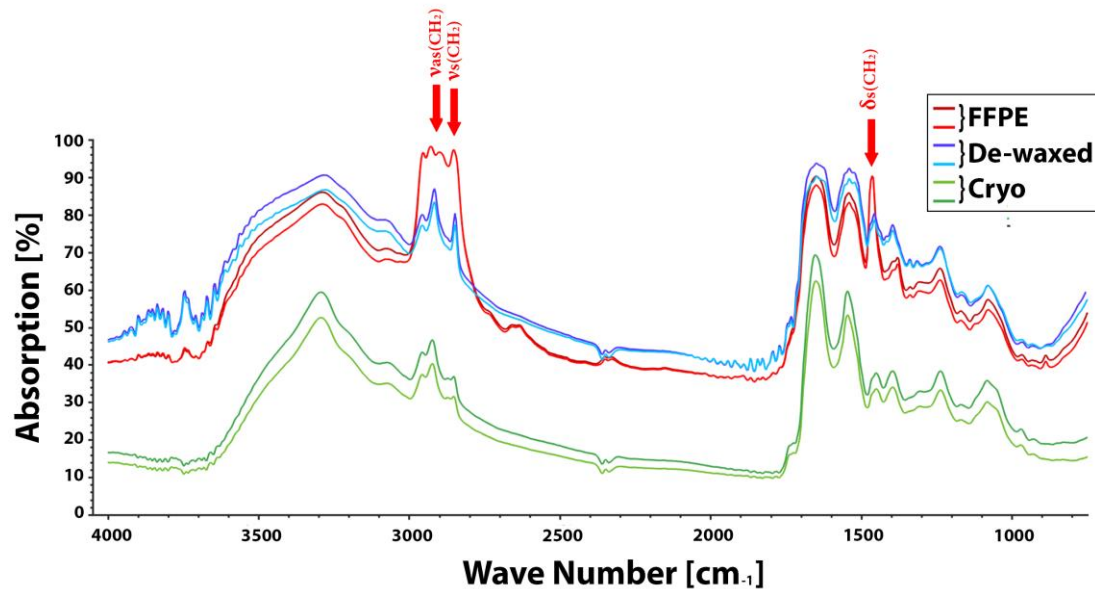


Figure 14 FTIR measurements on formalin-fixed paraffin-embedded tissues. Comparison of MIR-spectra acquired on FFPE and de-waxed tissues reveals a higher overall absorption when compared to cryopreserved materials. Spectral differences are mainly observed in the ranges of 3010-2850 cm⁻¹ and 1460-1470 cm⁻¹ in which CH₂-associated symmetrical (ν_s) and antisymmetrical stretch (ν_{as}) as well as symmetrical bending vibrations (δ_s) occur. Spectral correlation was quantified by calculating the sample Pearson correlation coefficient which confirms a higher correlation of cryopreserved material to de-waxed tissues ($R = 0.95$) than to FFPE tissues ($R = 0.89$).

The calculated differences were found to be caused by a comparatively increased base-intensity (2550-1700 cm⁻¹) and an increase in the wavenumber-ranges 3010-2850 cm⁻¹ and 1460-1470 cm⁻¹ that are associated to CH₂-vibrations. For this, it can be concluded that the typically lipid-associated CH₂-signals are overlaid by the CH₂-bonds of paraffin (C₂₀H₄₂) which impairs spectral quality. Although de-waxed tissue samples also featured an increased base-intensity, no relative increase in the ranges 3010-2850 cm⁻¹ and 1460-1470 cm⁻¹ could be observed. In order to evaluate the segmentation of de-waxed tissue samples, a human derived prostate cancer FFPE-section was mounted on a MirIR-coated tissue slide and de-waxed prior to FTIR image acquisition. The FTIR image was partitioned into five clusters by means of ($k = 5$) k-means++ segmentation in accordance to the optimal cluster number denoted by the Calinski-Harabsz criterion (**Fig. 15**). For comparison, histopathological annotation of a subsequent section was provided by the Pathological Institute at the University Medical Center Mannheim (black mark). In the derived segmentation image, no tumor-describing segment could be observed. Instead disperse clusters were obtained that might refer to other tissue morphologies, namely smooth muscles (blue), gland cells (green) collagen-rich capsule (orange).

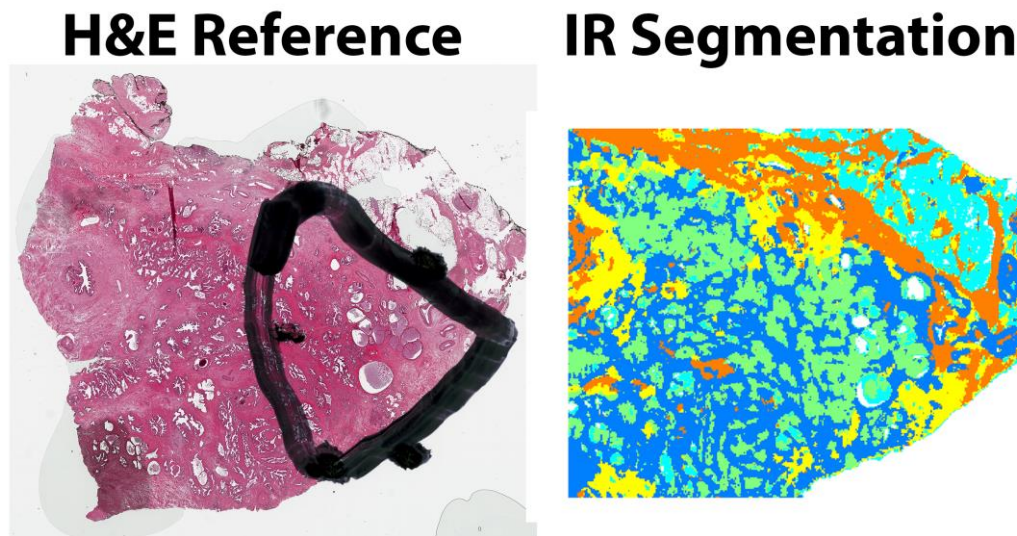


Figure 15 Comparison between H&E-stained human prostate cancer tissue and subsequent FTIR image segmentation. Two adjacent FFPE tissues of a prostate cancer quick section were designated to H&E staining and FTIR image acquisition after de-waxing. The tumor area was highlighted (black mark) after pathological examination at the University Medical Center Mannheim. In comparison to the color-coded image provided by ($k = 5$) k-means++ segmentation of the FTIR image, no segment could be linked to the occurrence of tumor. Instead, disperse segments that might refer to other tissue morphologies were observable.

In summary, no correlation between the occurrence of prostate cancer and IR segmentation was observable. More comprehensive studies of different tissue types would be needed to evaluate the feasibility of on-FFPE measurements for tissue segmentation.

5.1.5 Reproducibility of FTIR image segmentation

Storage conditions could have a striking influence on the obtained infrared signals of tissue materials. To further elucidate the long-term effects of storage, subsequent sections of the murine brain samples with human U87-MG glioblastoma xenografts (presented in **chapter 5.1.4**) were mounted on gold-coated slides and kept in a desiccator at room temperature for three years. FTIR images were recorded and segmentation results of both the stored image and the image that was recorded three years earlier were compared. Thereby, images were partitioned into seven segments by means of ($k = 7$) k-means++ cluster analysis, as denoted by the CHc of previous brain tissue experiments (presented in **chapter 5.1.1**).. Allocation-based color-coding of pixels revealed comparable morphological brain structures as well as the area of

infiltrating glioblastoma cells in both segmentation images (**Fig. 16A**). In order to quantify the structural similarities, both images were registered into a common coordinate system and the DSC was calculated for each corresponding segment-pair, revealing a mean overlap of 58.9% ($DSC = 0.589 \pm 0.211$) between the identified structures. In this context, the individual DSC-values per segment-pair varied greatly, with the isocortex contours detected after three years being the least similar ($DSC = 0.134$) to the original image's contours and the infiltrating tumor zone ($DSC = 0.908$) and fiber tract contours ($DSC = 0.746$) being the most similar. In addition to structural similarities, the respective mean infrared spectra acquired for cerebellar fiber tracts and tumor cells were comparable in both images (**Fig. 16B**). This observation was quantified by calculating the Pearson correlation coefficient [127], as a measure of spectral similarity. Thereby, a high mean spectral correlation ($R = 0.972 \pm 0.004$) between all identified segment-pairs was confirmed.

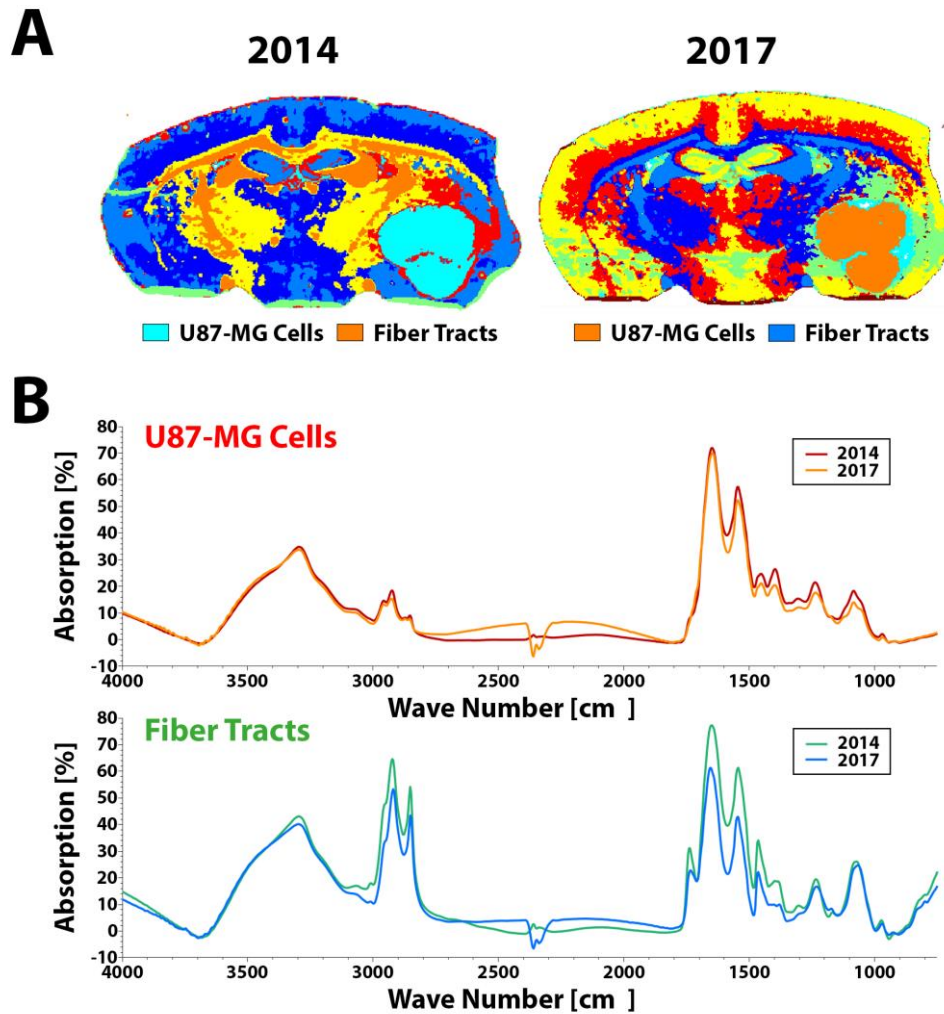


Figure 16 Long term reproducibility of FTIR image segmentation. Brain tissue sections derived from CD1 *nu/nu* mice inoculated with U87-MG glioblastoma cells were stored in a desiccator at room temperature. FTIR images were recorded prior to storage and after three years for comparative cluster analysis. **(A)** Image dissection by means of ($k = 7$) k-means++ analysis leads to comparable color-coded segmentation and reveals the spatial contours. Calculation of the Dice similarity coefficient between matching segments reveals structural similarity of fiber tracts (DSC = 0.746) and infiltrating glioblastoma cells (DSC = 0.908). **(B)** Comparison of the respective mean spectra for both spatial features also demonstrates spectral comparability. This was quantified by calculating the Pearson correlation coefficient which confirms high similarity ($R = 0.972 \pm 0.004$).

In summary, the long-term stability of FTIR image structures appears to vary for different cell populations. Whereas the contours of infiltrating tumor cells and white matter were still observable after three years, grey matter structures, i.e. isocortex, were found to cluster differently. Despite these structural differences, mid-infrared spectra appeared long-term stable as the spectra recorded in 2014 and 2017 showed high linear correlation. More comprehensive studies of different tissue types would be needed to evaluate the influence of storage conditions on segmentation results.

5.2 Development of an automated multimodal imaging workflow

An established routine in clinical MALDI-MSI is the analysis of serial whole-tissue samples. In order to extract valuable analytical information out of the resulting large data volumes, subsequent histological stains of the analyzed specimen are typically co-registered and used to manually define striking morphological regions. The respective mass information is then used to assess additional in-depth molecular characteristics, i.e. tumor-associated mass features. However, as shown in recent publications [21], only a small fraction of the recorded MSI data is used for later analyses. Moreover, the manual definition of striking areas constitutes an increased potential for errors.

FTIR image segmentation can identify tissue substructures of morphological similarity in a more sensitive and robust manner when compared to direct MS clustering. Moreover, the acquisition of FTIR images is comparatively fast, which potentially facilitates their applicability within a clinical setting. Despite these advantages, the obtained MIR-information is insufficient to uncover the molecular coherences involved in disease formation. In order to combine the respective potentials of vibrational spectroscopy and mass spectrometry, a multimodal imaging workflow was developed that utilizes the FTIR-derived structural information in order to guide subsequent MALDI-MSI experiments. In this context, technical and computational developments were made with respect to registration in order to generate templates for MS guidance out of FTIR spectra.

5.2.1 Image registration to enable FTIR-MS workflow development

In order to jointly utilize the spatial information obtained from infrared-based segmentation as well as the molecular information provided by MALDI-MS, a two-step concept was developed (**Fig. 17**), in which FTIR images of tissue cryosections are acquired prior to MSI-data acquisition. The derived FTIR-based segments after image pre-processing and virtual micro-dissection are then used as a template to define distinct, divergent cell populations (i.e. tumors or functionally important tissue morphologies) for MALDI-MSI.

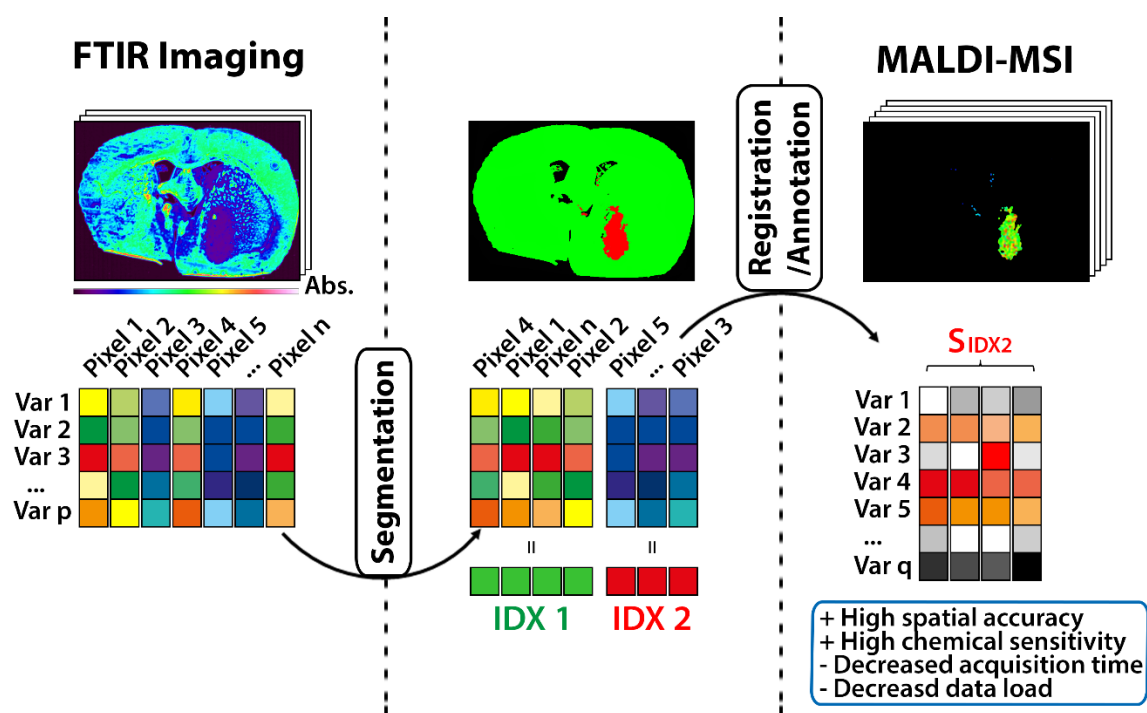


Figure 17 Multimodal FTIR-guided MALDI mass spectrometry imaging. FTIR imaging is utilized in order to record a variable (**Var**) set consisting of p vibrational bands that represent biomolecular features. By applying segmentation, the spectra at any given pixel are assigned to a defined number of subgroups based on similarity. The membership of each pixel is expressed by assigning an exclusive index-value (**IDX**). The spatial properties of a segment that consists of all pixels belonging to a given **IDX** (**S_{IDX}**) are registered to the follow-up modality, MALDI-MSI, thus allowing to target the q -dimensional mass variables for predefined pixels. FTIR and MS imaging measure orthogonal properties, consequently complementing contours of spatial accuracy that exceed MALDI-MSI capabilities with chemically specific mass information.

In order to convert the FTIR image into the same coordinate system as the MALDI-MSI, a routine for automated registration was developed in MATLAB. A multi-step script aligns the recorded moving FTIR images to their corresponding tissue objects within the optical reference image which is used for MSI acquisition (**Fig. 18**). This enables FTIR-based guidance and interpretation of MS images, as the flexImaging software automatically links the position of each mass spectrum to its corresponding optical image pixel by means of affine point set registration. In an initial step, the MATLAB script converts both the FTIR and optical image into binary files by applying ($k = 2$) k -means++ cluster analysis. Thereby, the segment with lower mean intensity is set to zero, as it is considered to resemble background and the remaining tissue-pixel are set to one. The generated binary files are then forwarded to intensity-based automatic image registration which, in an iterative process, estimates the geometric transformation matrix that aligns both images (**Fig. 18**). The process applies affine registration of an internally determined transformation matrix to the moving FTIR-derived binary image with bilinear interpolation. Then, the mean squares image similarity metric is computed between the transformed moving image and the fixed

binary of the optical image. The metric assumes similar image intensities and is generally only suitable for mono-modal registration [156]. However, the previous conversion of both images to binary, bypasses this limitation. Finally, a regular step gradient optimizer [157] adjusts the transformation parameters so that the optimization follows the gradient of the image similarity metric in the direction of the extrema at a constant step length. However, when the gradient changes direction, the step length is halved until either 100 iterations were performed or the step length falls below 10^{-5} .

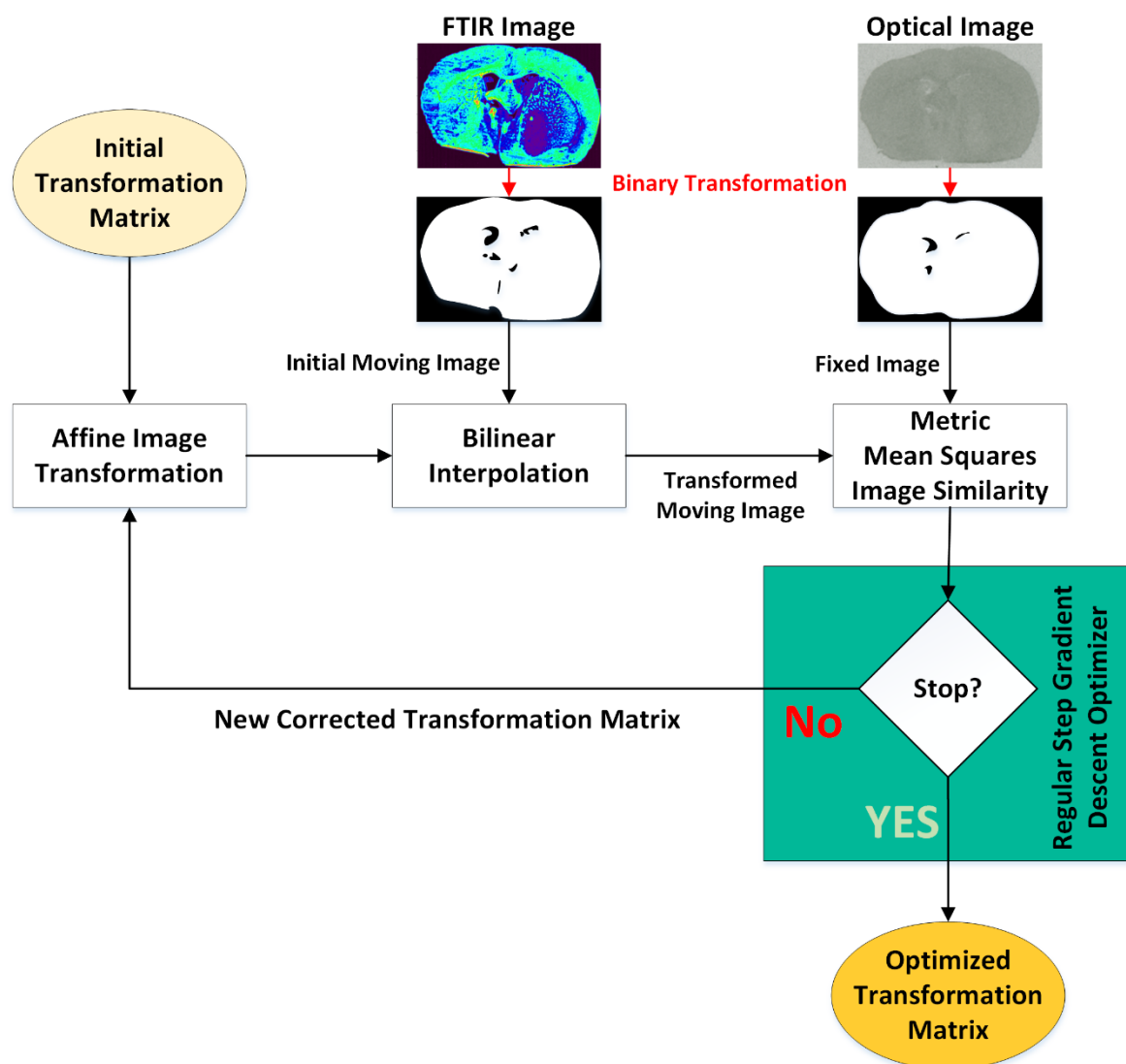


Figure 18: Affine intensity-based automatic registration of FTIR and optical image. The moving binary image originating from a recorded FTIR image is initially transformed with bilinear interpolation. Subsequently, the similarity between the transformed image and a fixed binary image originating from an optical image of the same specimen is calculated by means of mean squares image similarity. If the computed metric doesn't fulfill the condition specified by the regular step gradient descent optimizer, the transformation matrix is adjusted and reapplied to the moving image. If the condition is matched or if a certain count of iterations is reached, the resulting optimized transformation matrix can align the starting FTIR image and optical image.

5.2.2 Evaluation of FTIR-to-MSI registration

For the evaluation of registration efficacy, a murine brain section was prepared on gold and designated to FTIR acquisition (**Fig. 19A**). Four additional sections were mounted on a subsequent object carrier and an optical image was recorded (**Fig. 19B**). Registration was carried out in two subsequent steps. At first, both the FTIR and optical image were converted into binary images (background = 0, on-tissue signal = 1). Secondly, the geometric transformation matrix between the moving FTIR and the fixed optical binary image was calculated by means of intensity-based affine registration. Based on the retrieved geometric transformation matrix, the color-coded FTIR-based segmentation image (**Fig. 19C**) was aligned to match the contours of its subsequent section in the optical image (**Fig. 19D**).

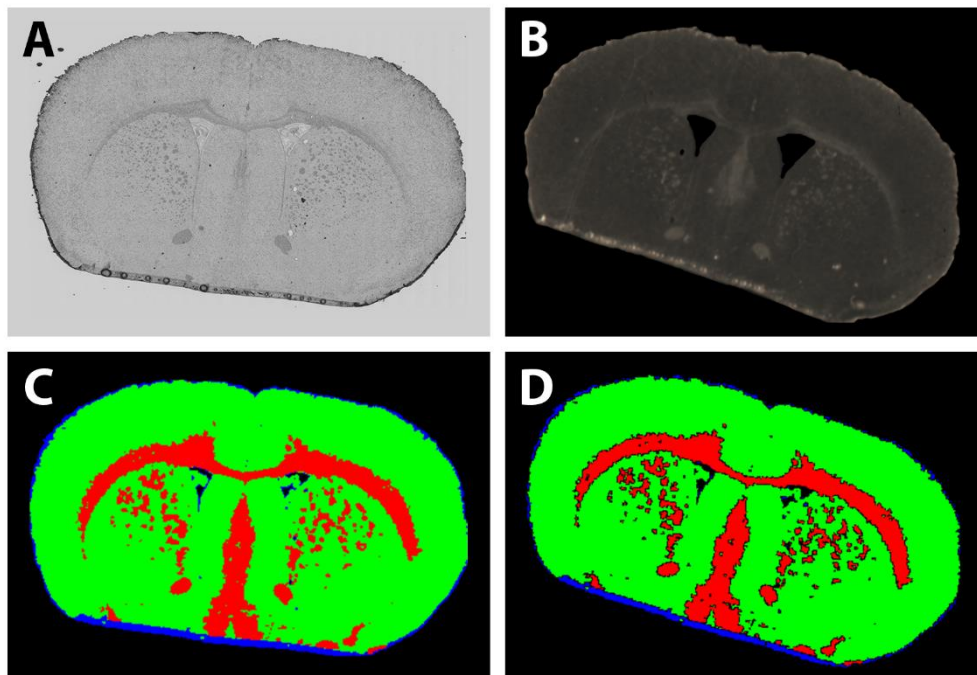


Figure 19 Registration of FTIR-derived segments to multiple subsequent sections used for MALDI-MSI. (A) A brain tissue section designated to FTIR image acquisition will be utilized to denote spatial properties in (B) subsequent sections designated to MALDI-MSI. (C) Morphological information is obtained by k-means++ ($k = 3$) segmentation of the acquired FTIR image. As a result, the image is getting partitioned into multiple color-coded areas of greatest spectral similarity. (D) After registration between the obtained moving FTIR and fixed optical MS image, the derived transformation matrix is used to adjust the segments spatial properties.

For evaluation purposes, the transformation matrix was recalculated for each FTIR-to-subsequent brain tissue section pair, thus allowing the identical FTIR cluster image to be reoriented and rescaled to match all subsequent sections (**Fig. 20A**). The effectiveness of registration was quantified by calculating the DSC, which denotes the

percentage of overlap between the transformed moving FTIR and the fixed optical image (**Fig. 20B**). For this exemplary tissue series, a mean DSC of 0.975 ± 0.006 was calculated, which confirms the precise conversion of FTIR-derived segments.

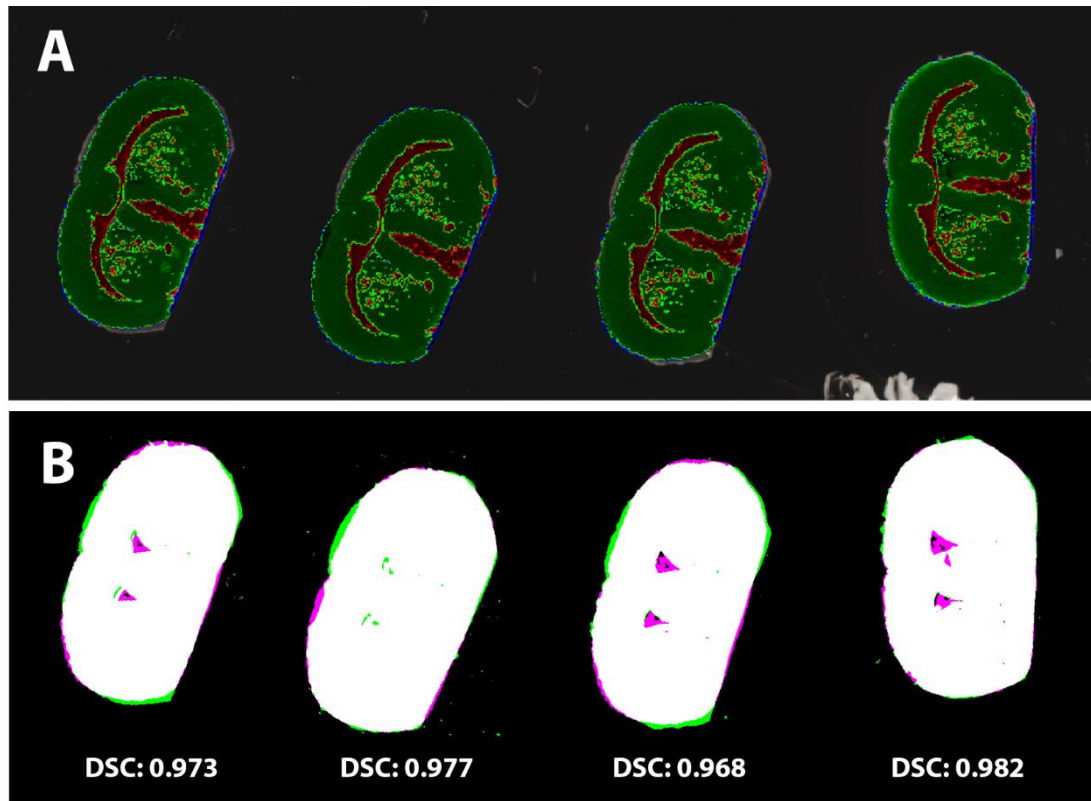


Figure 20 Evaluation of the automatic registration used for FTIR-guided MSI (A) A single segmentation image (red-green) obtained from FTIR mouse brain image dissection by means of ($k = 2$) k -means++ cluster analysis was registered to the optical image of four subsequent tissue sections. (B) Fusion of the binary masks that represent tissue contours in the FTIR- and optical image makes it possible to quantify registration efficiency. Calculation of the Dice similarity coefficient (DSC) is used as a measure of overlap (white) in relation to pixels that are only visible in the optical image binary (green) or FTIR derived binary (green).

The established registration routine was incorporated into a multimodal workflow that makes use of the image acquisition file utilized to denote MALDI-MSI parameters (**Fig. 21**). Within the acquisition file, it becomes possible to define optical image pixel coordinates that frame the desired region of interest (ROI). In order to retrieve the optical pixels that resemble the spatial FTIR-based segmentation contours, the border pixels of each segment were automatically transformed according to the obtained geometrical transformation matrix. In this manner, the FTIR-derived spatial allocations are mapped onto the optical image used to define ROI's that can be used to extract spatially defined mass spectra out of large and complex data volumes after whole-tissue measurements.

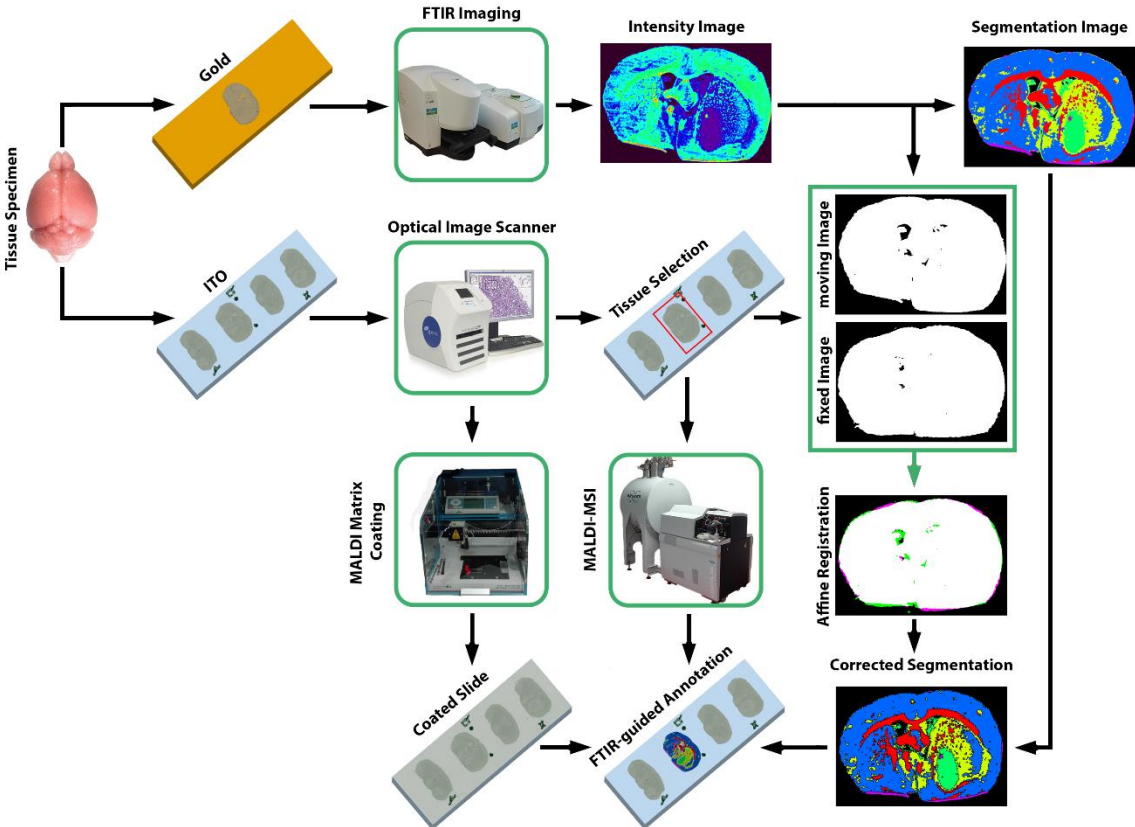


Figure 21 Schematic workflow diagram for FTIR-guided extraction of MSI data. FTIR image acquisition of subsequent tissue sections on gold can be utilized to denote measurement regions and/or ROIs. The spatial contours are defined by segmentation of the FTIR image and subsequent registration to the optical image used for later MSI measurement. Modification of the MSI acquisition file enables targeted mass feature extraction and unbiased annotation.

5.3 Analytic applications of multimodality FTIR-MS imaging

After developing a stable and versatile workflow for FTIR-MSI imaging, different case studies that exemplify the added benefit of the technique will be presented. In the following chapters, the workflow's potency to target cell populations of high vibrational homogeneity for improved mass feature retrieval will be presented. Moreover, an untargeted approach that enables automated, morphology-dependent biomarker identification was examined. The proposed workflow was also utilized to assess tumor heterogeneity in human-derived cancer tissues. Lastly, slight adjustments to enable same-slice multimodality imaging were made, thus enabling targeted MALDI-MSI acquisition of pre-defined tissue regions.

5.3.1 Automated FTIR-guided tumor-targeting and mass feature extraction

The strength of FTIR image segmentation lies in its ability to partition the examined tissue into regions of the same or a similar profile. It can assess the divergent biomolecular composition between distinct tissue regions with more detail than established histochemical methods. Moreover, it can generate images of superior spatial resolution at a faster rate than MALDI-MSI.

In that regard, the precise localization of tumor margins is a critical task in histopathology and, hence, serves as a relevant test case for the established multimodal workflow. As a suitable model, four female 6-12 weeks old CD1 *nu/nu* mice were inoculated with human U87-MG glioblastoma cells and organs were harvested 5-6 weeks after inoculation. Adjacent coronal brain cryosections were mounted on gold- and ITO-coated slides for FTIR and MALDI-MSI respectively. MALDI-MS images were acquired in reflector positive mode using DHB matrix. For reference purposes, the gold-coated section used for FTIR measurement was stained with hematoxylin and eosin and a microscopy image was taken (**Fig. 22A, left panel**). As FTIR images feature comparatively small volumes of data, it became possible to partition the entirety of acquired infrared spectra into ten groups by virtue of a single ($k = 10$) k-means++ cluster analysis. In contrast to single-section segmentation, this process guarantees the identification of similar tissue areas across all four examined brain tissue samples. In this context, the cluster number k was empirically found to partition the tissue into additional morphology-related segments. By linking the obtained spatial properties that

represent inter- and intra-section dependencies between a single FTIR and H&E image, it became possible to locate similar cells within the other images as they share the same index-value. This enabled identification of the segment corresponding to tumor (**Fig 22A, right panel**) without the need of additional histochemical stains. By means of intensity-based registration, the spatial properties derived from FTIR image segmentation were used to automatically define regions of interest within the subsequently acquired MALDI-MSI dataset. For further analysis of tumor-related mass shifts, the previously identified segment that resembled tumor contours (S1) was used to extract region specific MS profiles.

In addition, an adjoining segment (S2) which was not noticeable in the reference H&E image, was examined (**Fig. 22A**), as it may correspond to potential expansion of the tumor margin like an edema and/or an infiltration zone characteristic of glioma. The registration efficiency was quantified by calculating the DSC, which suggested an overlap of more than 97% ($DSC = 0.982 \pm 0.004$) between the registered FTIR and fixed optical MS image (**Fig. 22C**). Targeted extraction of mass spectra that lie within the identified segments borders (**Fig. 22D**) was achieved by transforming the respective FTIR coordinates into MSI coordinates. For the identification of divergent mass signals that may represent segment-specific biomolecular signatures, feature extraction using the ROC criterion was calculated. For each segment, the ten m/z -values that are most suitable to distinguish between mass spectra within and outside the respective segment were identified (**Fig. 22E**). For instance, m/z 797.1 was identified as a signal specifically elevated in S1. The m/z feature 204.2 was linked to the extended tumor margin S2 and m/z 867.2 was identified as an ion that was symmetrically distributed across the brain.

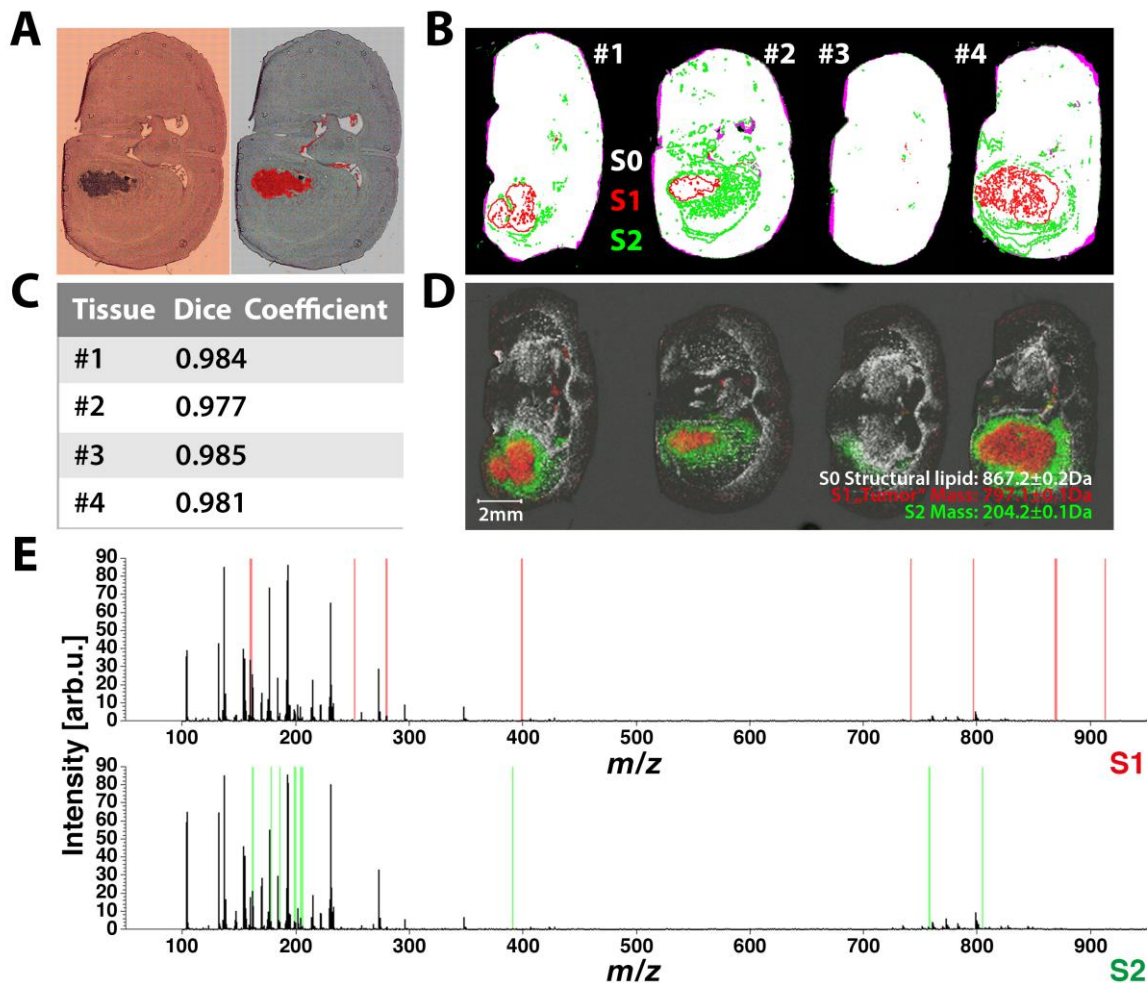


Figure 22 FTIR-based targeting of tumor lesions in xenografted CD1 *nu/nu* mice. Brains of four CD1 *nu/nu* mice (#1 to #4) were inoculated with 1×10^5 human U87-MG glioblastoma cells; organs were harvested 5-6 weeks later. Subsequently, adjacent 8 μm coronal brain sections for FTIR and MALDI MS image acquisition were prepared. FTIR data acquired from cryosections of all four mouse brain samples were simultaneously disaggregated by ($k = 10$) k-means++ segmentation to identify intra- and inter-specimen relationships. **(A)** By comparing the depicted spatial contours belonging to a given index in a single brain section (right panel) to its corresponding H&E stain on gold (left panel), it becomes possible to identify tumor-associated segments in all four specimen because they share the same index. **(B)** Binary images derived from FTIR data and microscopic images of subsequent sections used for MS acquisition are automatically registered. The resulting fusion images (white = matching areas, magenta = non-matching areas) are overlaid by contours of two segments (S1, red & S2, green) present in all observed specimen. **(C)** The quality of registration is further evaluated by means of DSC calculation, which reveals a mean overlap of 98.2% ($\text{DSC} = 0.982 \pm 0.004$) between all transformed binary FTIR images and the binary images used for MSI measurement **(D)**. Extracting only mass spectra that lie within the transformed spatial contours of S1 and S2 enables identification of m/z -values with matching distribution patterns (m/z 797.1 tumor-specific signal in S1, m/z 204.2 as signal specific for the expanded tumor margin S2) by **(E)** means of feature selection using the area between the empirical receiver operating characteristic (ROC) curve and the random classifier slope. For both S1 and S2, ten conspicuous features (highlighted in red for S1 and green for S2) were calculated.

In order to verify the FTIR-denoted tumor contours that were assigned based on comparison to a single H&E stain, the remaining sections were also designated to histochemical staining (**Fig. 23**). Without prior knowledge about tumor occurrence, the calculated tumor segments were automatically registered to the corresponding

H&E image ($DSC = 0.994 \pm 0.001$). In all sections, the predefined tumor contours precisely matched the tumor distribution of infiltrating glioblastoma cells. Moreover, the stain of section #2, for which nearly no (0.4%) infrared spectra were assigned to the tumor-associated allocation group, was proven to be cut off target and didn't contain glioblastoma cells. It is uncertain whether the positively determined pixels refer to false-positives or small tumor residues, i.e. intra-tumor vessels.

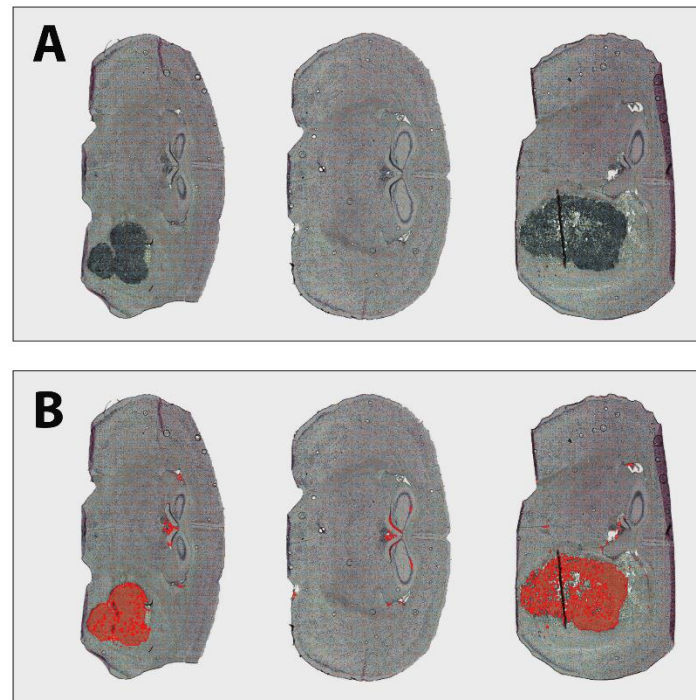


Figure 23 Correlation of tumor contours visible in H&E stained images and FTIR segments. Coronal brain sections of CD1 *nu/nu* mice engrafted with U87-MG glioblastoma cells were designated to subsequent FTIR image acquisition and histochemical staining. **(A)** The observed tumor cell infiltration could be linked to FTIR-derived image segments **(B)** that were automatically registered ($DSC = 0.994 \pm 0.001$) and visualized within the respective H&E stained image, thus demonstrating precise FTIR-based tumor localization.

After successful H&E evaluation, the intensity distribution patterns of all 20 peaks whose presence or absence was identified to correlate with either S1 or S2 were visualized in order to verify the segment contours by multiple sensor types. A match between the spatial properties from FTIR and all identified masses in complex MSI datasets could be observed, which emphasizes the authenticity of the identified tumor localization (**Fig. 24**). Moreover, FTIR-denoted segmentation was proven to add substantial value by identifying the tumor-related segment S2 that was not observable

in conventional H&E staining, but supported by multiple masses identified in MALDI-MSI.

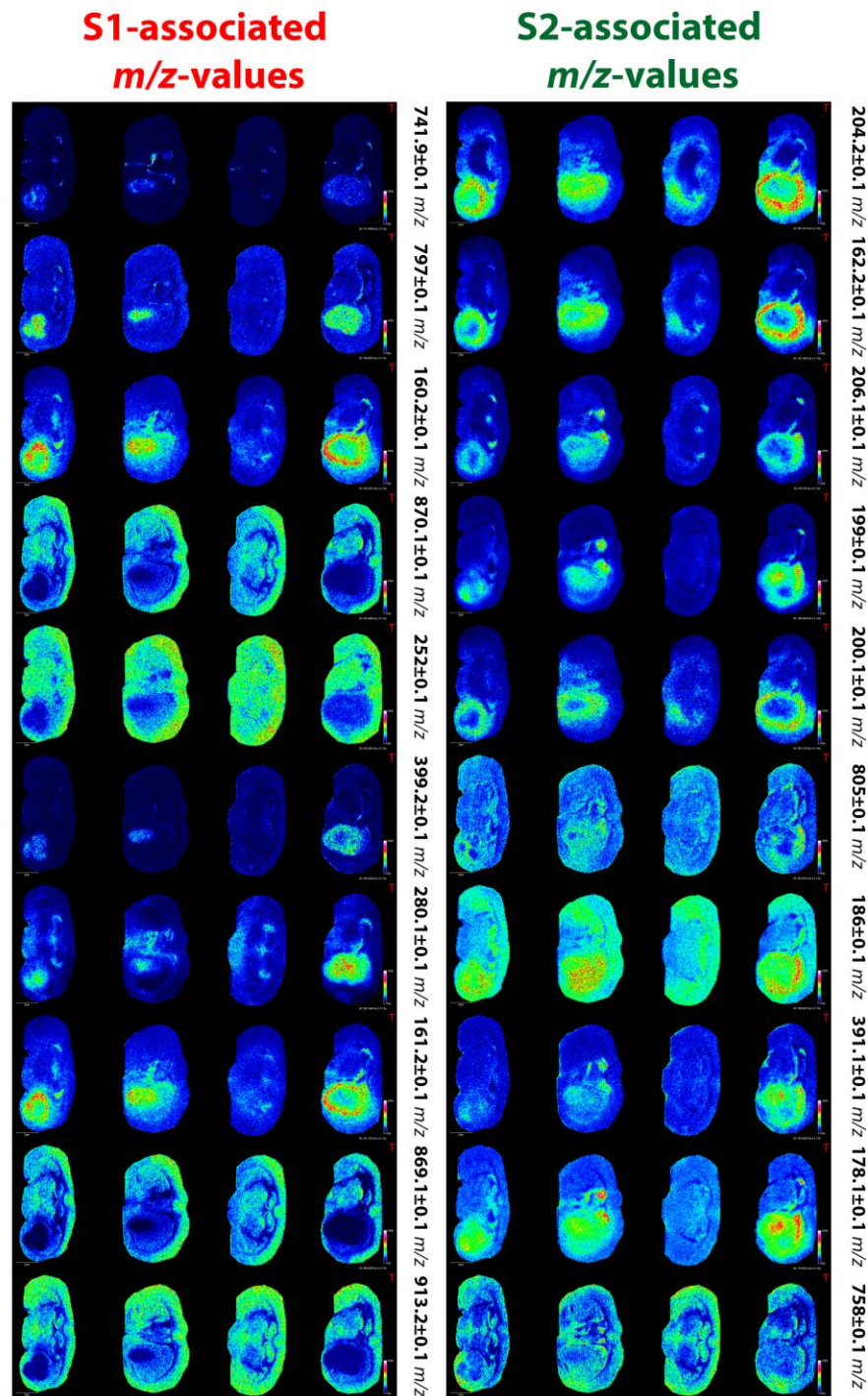


Figure 24 Distribution patterns of m/z features representing FTIR segment contours. Subsequent to MSI, FTIR images of four CD1 *nu/nu* mice brain sections inoculated with U87-MG glioblastoma cells were acquired and clustered for guided extraction of tumor-associated masses that resemble tumor/margin-contours. For two segments, S1 and S2, ten discriminant mass peaks were identified by means of feature extraction using the empirical ROC curve criterion. The identified peaks not only show importance for the respective segment, but also resemble their spatial pattern.

In summary, the automated pipeline of segmentation and subsequent co-registration enabled precise targeting of glioma tumor areas. In comparison to conventional histopathology which comprises of histochemical tissue staining, FTIR-guided MALDI-MSI provided additional spatial tissue characterization and in-depth molecular profiling. In addition, the possibility to segment multiple tissues at once and thus circumventing the necessity of additional stains was demonstrated as linkage to a single H&E stain proved to be sufficient to enable tumor localization in multiple biological replicates.

5.3.2 Automated MS marker signature retrieval for defined tissue morphologies

Comparison of FTIR cluster images to H&E stains enables targeted annotation of pathologically important tissue areas such as tumor margins. However, the needed a *priori* information (i.e. the histopathological examination) is a limiting factor in clinical mass spectrometry. A different application of FTIR-guided MSI is the dissection and segment-wise comparison of mass signals. This allows for a spatially focused retrieval of disease-related marker signatures without the need for prior knowledge about the examined specimen's structural makeup. As a suitable example, coronal cerebellum cryosections of NPC1 I1061T knock-in mice that resemble Niemann-Pick type C disease, a spatially restricted neuronal lipid disorder, were mounted on gold-coated object slides. In addition, sections of a C57BL/6 mouse brain were prepared at a similar region in order to serve as a control. FTIR images were recorded for both animals and subsequently portioned into five segments by means of k-means++ segmentation. In this context, the cluster number k was empirically found to result in comparable morphology-related segments in both brain tissue sections. Four segments were chosen in order to *in-silico* dissect subsequently acquired MSI datasets (**Fig. 25A**). The fifth segment was discarded as it represented each tissue's contours. Sections designated to MALDI-TOF-MSI acquisition were coated with PhCCA matrix in 90% acetone solution to ensure optimal acquisition of negatively charged lipids and metabolites. In comparison to the Allen Brain Reference Atlas, the defined segments were linked to their morphological counterparts (**Fig. 25B**). Hence, (k = 5) k-means++ segmentation of the recorded FTIR images revealed the spatial contours of an unspecified microstructure (I), cerebellar granule layer (II), cerebellar fiber tracts (III) and molecular layer (IV). After automated intensity-based registration

(DSC = 0.992 ± 0.003) to the optical image used for MALDI-MSI acquisition, the mass signatures provided by MALDI-MSI were extracted for each individual index to allow for a segment-wise and therefore cerebellar substructure-wise comparison. By applying t-test based feature extraction, the ten m/z -values that differ the most between NPC1 I1061T knock-in and control mice were identified for each morphological structure (Fig. 25C-F).

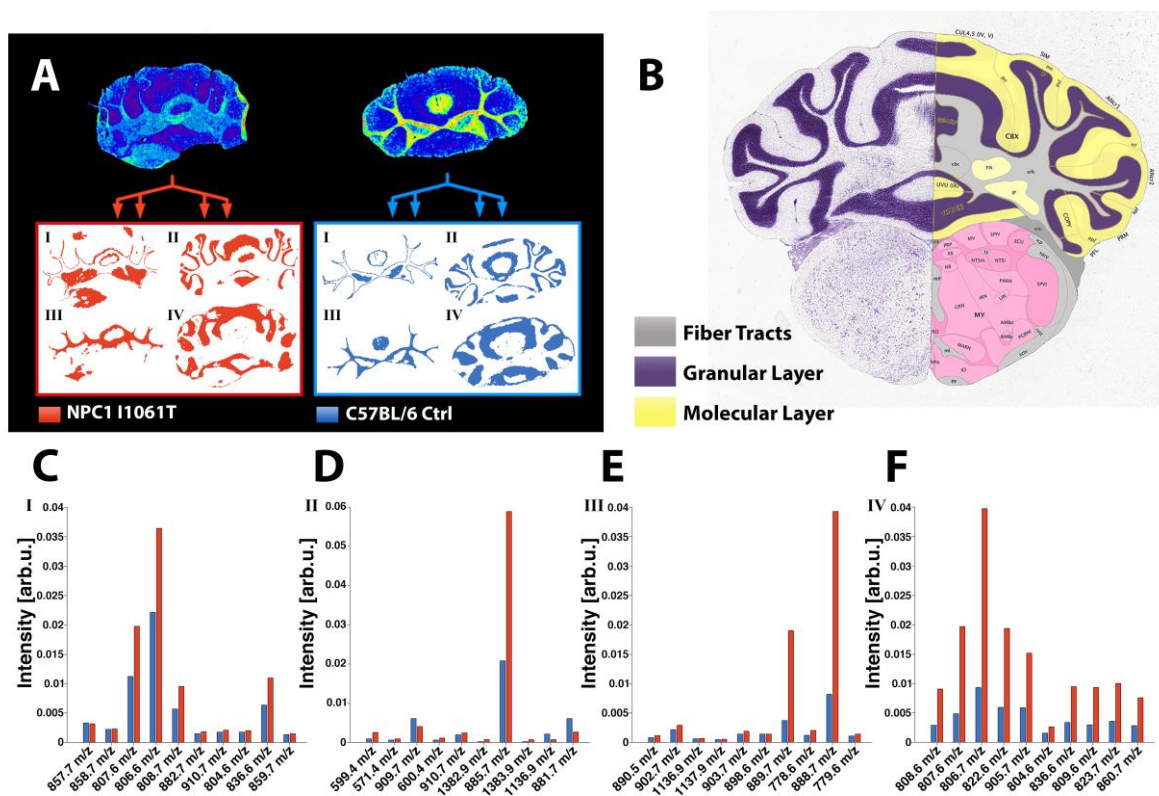


Figure 25 Comparison of MS lipid/metabolite signatures derived from FTIR-segmented NPC1 and Control mouse brains. (A) FTIR images of coronal mouse cerebellar sections (NPC1 I1061T knock-in mice versus C57BL/6 control mice) were divided into four segments of maximum data similarity by k-means++ clustering. (B) Comparison to the Allen Brain Reference Atlas (Image credit: Allen Institute Mouse Brain Atlas [149], P56, Coronal, Image 121 of 132) suggests the morphological identity of three FTIR-derived segments, namely granule cell layer (II) fiber tracts (III) and molecular layer (IV). The remaining segment (I) represents an intermediate region between the outer edge of the arbor vitae and the granule cell layer. The spatial contours reflecting distinct cell morphologies were automatically registered to the optical image used for reflector negative MSI acquisition, thus enabling extraction of the respective mass signatures. Comparison between the MS lipid/metabolite profiles of NPC1 and C57BL/6 mice guided to the obtained infrared segments was achieved by means of t-test based feature extraction (C-F) The mean intensity of ten distinct masses allowing for discrimination between healthy and diseased mice was visualized for each predefined cerebellar substructure.

Most of the received marker signals were found to differ between segments, which leads to the presumption of locally restricted lipid accumulation. The highest similarity was observable between the molecular layer and the observed unspecified microstructure that share five common features of which three are m/z 806.6 and its

potential isotopes m/z 807.6 and m/z 808.6. In NPC1 |1061T knock-in mice, the molecule of m/z feature 1136.9 was found to be up-regulated within the cerebellar fiber tract and concurrently down-regulated in the granule cell layer. In this context, the achieved morphological breakdown derived by FTIR segmentation enables the identification of local mass shifts that would otherwise balance out over the entire dataset. In this regard, by comparing the obtained results to the mouse model-describing literature [131], the granule cell layer-specific accumulation of m/z 1382.9 and its potential isotope m/z 1383.9 become particularly interesting (**Fig. 26A**), as it may refer to the previously reported disease-related ganglioside GM2 (d18:1/18:0). An argument against this hypothesis is the mass-difference of 0.2 Da when compared to previous mass spectrometry analyses of GM2 (d18:1/18:0) [158]. For further evaluation, the granule cell layer-specific increase of m/z 1382.9 was monitored in subsequent sagittal brain sections of NPC1 |1061T knock-in and C57BL/6 control mice. Fold change plots (**Fig. 26B**) of the mass features acquired within the sagittal cerebellar granule layer were created in Tableau (Tableau Software Inc., Seattle). The obtained data confirmed the identification of a locally-restrictive accumulation that is not noticeable without prior morphology-guided dissection.

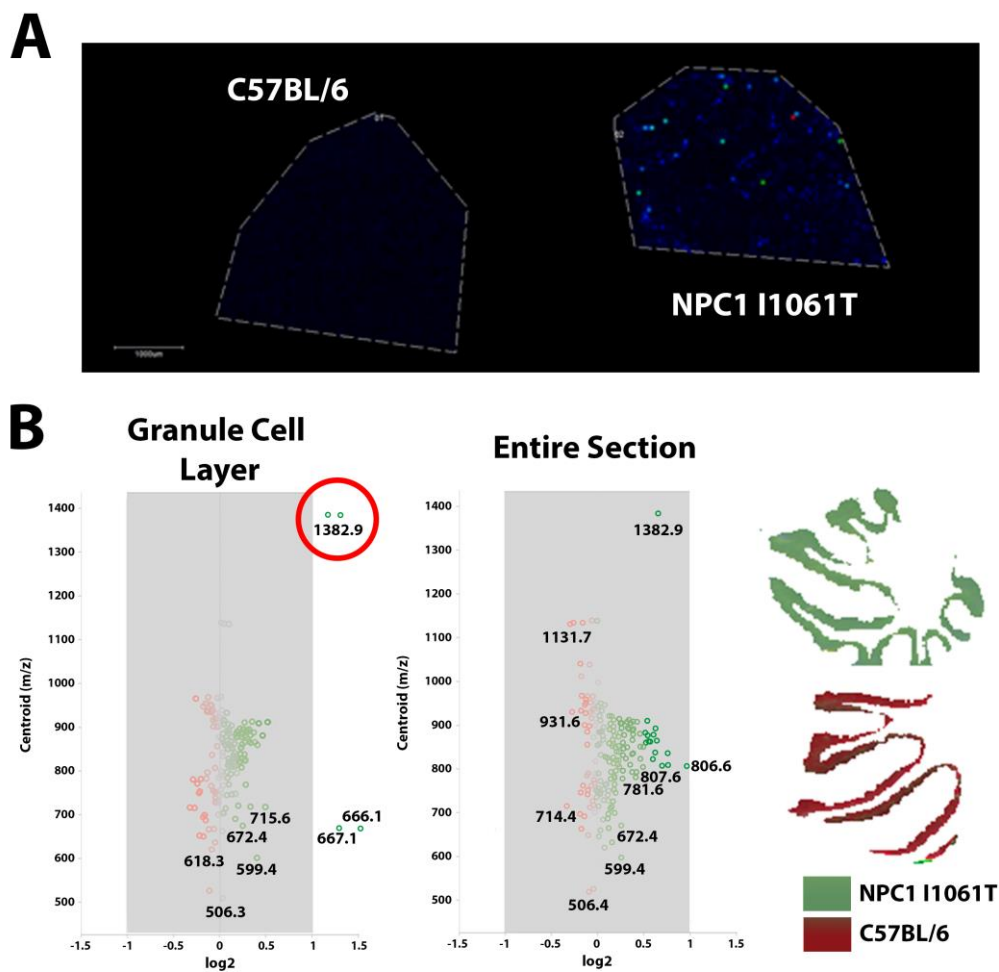


Figure 26 FTIR-guided identification of NPC1-related locally-restricted GM2 accumulation (A) The distribution and accumulation of m/z 1382.9 was monitored within cerebellar brain sections of NPC1 I1061T knock-in and C57BL/6 control mice. The mass was previously identified as a potential discriminant marker in granule cell segments denoted by FTIR imaging. (B) Comparison of the fold change between mass spectra originating from the entire section and mass spectra exclusively originating from the granule cell layer demonstrates the added value of prior FTIR dissection.

In conclusion, the obtained results suggest that, without the need for H&E staining, FTIR image-based tissue segmentation and segment-wise comparison of MS data allow for unbiased extraction of distinguishing features. The presented workflow is fully automatable.

5.3.3 Identification of GIST by means of FTIR cluster analysis

An essential and challenging task for the identification of diagnostic markers in mass spectrometry imaging is the occurrence of histological heterogeneities. , as the cellular origin of acquired mass signatures is uncertain without annotation. For this reason, it has become common practice to utilize subsequent H&E stained sections and pathological guidance in order to define striking tissue regions of similar morphological makeup. However, the manual assignment in current praxis makes it impractical for the analysis of large tissue cohorts. In this context, recent publications have demonstrated direct cluster analysis of MSI data in order to *de novo* identify heterogeneities and tumor-subpopulations in breast and gastric cancer [159, 160]. Moreover, linkage between intra-tumor segments that are *a priori* not observable by conventional histopathology and disease outcome has been presented [118, 161]. The depicted studies reflect the high potential of unsupervised discovery of tissue heterogeneities. However, size and complexity of ultrahigh-resolution MALDI-MSI data (especially when using a FTICR detector) makes clustering of large sample cohorts impractical and demands for excessive pre-processing and dimensionality reduction. In contrast, FTIR images are of comparably small size which makes it possible to virtually micro-dissect a multitude of tissue sections in a faster and computationally less expensive fashion when compared to MALDI-MSI (chapter 4.1.4.). Moreover, the derived information about intra- and inter-sample dependencies can automatically be transferred to MSI in contrast to manual histopathological annotation. In order to examine histological heterogeneity in a suitable test case of clinical relevance, three series of 89 fresh-frozen tissue samples derived from 27 patients suffering from GIST were cut to 8 μm thickness and mounted on gold-coated object carrier slides. FTIR images of all specimens were acquired at 25 μm spatial and 8 cm^{-1} spectral resolution. Sectioning and infrared measurements were performed over the course of multiple days to balance out practical and technical variability. For later histopathological reference analytic, subsequent tissue sections were placed on Starfrost adhesive slides. H&E staining was performed by research assistants at the Pathology Institute of University Medical Center Mannheim, and optical reference images were acquired using an Aperio CS2 Scanner. In order to identify tissue heterogeneity in an unsupervised fashion, processed FTIR images were grouped by means of k-means++ cluster analysis. It was assumed that related morphologies would display similar vibrational properties across all patients. For this reason, segmentation was

simultaneously performed across sections on a total number of 1231560 MIR-spectra. As the true number of occurring tissue morphologies was unknown, segmentation was performed with different k-values in the range of 3-10 (k = 5 exemplified in **Fig. 27A**). Thereby, calculation and spectral allocation was rapidly performed in ~90 seconds resulting in a color-coded pattern of inter- and intra-tumor dependencies. In this context, heterogeneities could be observed in almost all examined tissue sections. However grouping was possible, as the prevalent color remained mostly consistent within section series of the same patient recorded on different days. Hence, sections of patient A, AA, AB, C, G, J, M, N, O, U, W and Z were assigned to belong to the same allocation group “red” and patient E, L, Q, R, S and Y were assigned to “purple”. Similarly, patient D, G, X and T who underwent surgery an additional time after tumor relapse were assigned to “blue”, patient I, K, P and V to “green” and patient B and L to “yellow”. Differences between sections derived from the same patient were observed in AA, E, J, W and Y for which allocation changed between the red and purple segment and V for which two section was assigned to “green” and one section was assigned to “yellow”. In order to assess the corresponding morphological characteristics, the respective samples were linked to clinical metadata (**Supplement Table**) and annotated subsequent H&E stains (exemplified in **Fig. 27B**). It turned out that the red- and purple-colored segments represent tumor areas, whereas the blue-colored segment corresponds to the occurrence of fibrosis and edema. The green allocation group could be correlated to adipose tissue in patient I and necrosis in patient P and K. In patient A, the green-allocated residue represented normal liver tissue. The remaining yellow-colored segment was found to correspond to sectioning artifacts like folds and freezing damage.

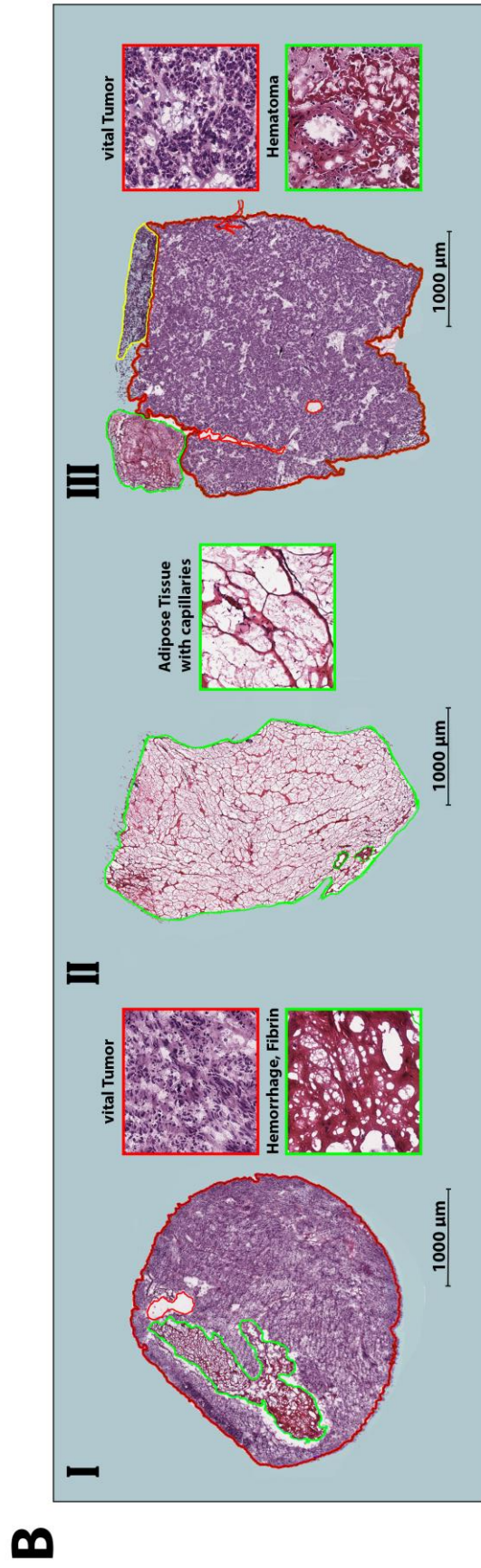
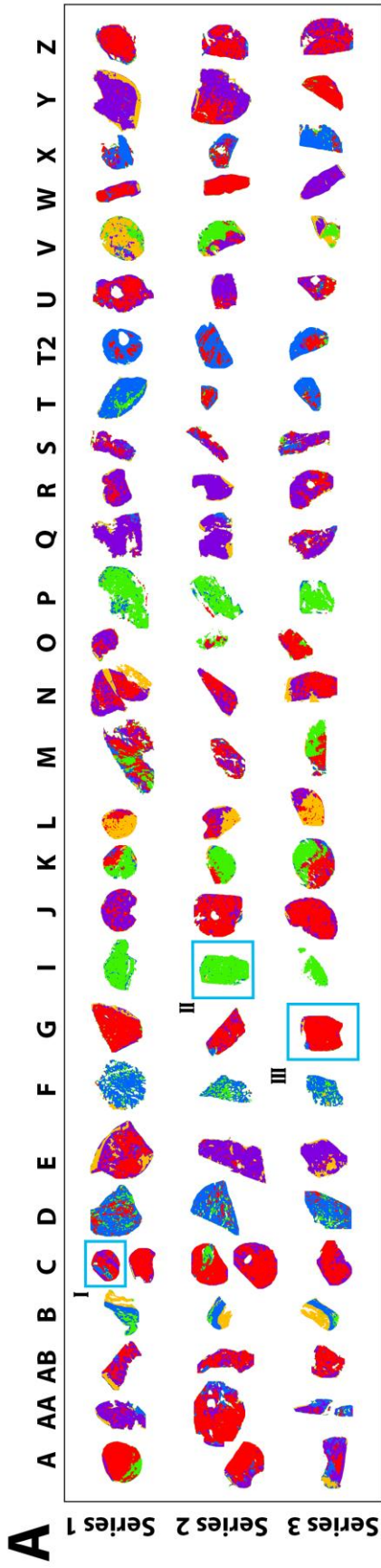


Figure 27 Heterogeneity and Grouping in GIST. (A) The recorded three FTIR image series of 89 tissue sections derived from 27 patients (patient T undergoing surgery twice) suffering from GIST were simultaneously partitioned by means of (k=5) k-means++ cluster analysis. Segmentation among all recorded tumor tissues revealed substantial tumor heterogeneity. (B) Magnified subsequent H&E Images of patient C, patient I and G exemplify FTIR-based distinction of tumorous areas and cancer-free tissue in accordance with histologically observed differences.

Moreover, it turned out that the tumor origin might influence segmentation, as the tissue sections of the blue-coded patients B, D, F, T and X were all resected out of the colon. However, no such correlation was observable for sections derived from other organs. Furthermore, a hematoma observed in the liver section of patient G also perfectly matched the blue-allocated pixels in that section. (**Fig. 27B**). Although the red- and purple-colored segment resemble the annotated tumor regions in all samples, a relation to different mutation subtypes of GIST was not observable, as all red and purple sections were genetically distinct. For example the spectra derived from patient C and Z were clustered similarly, although the tumor of patient C features a mutation of c-KIT in exon 9, whereas the tumor in patient Z features mutations in exon 11 and 13 as well as a mutation in exon 18 of the PDGFR α gene. It can therefore be concluded that FTIR image segmentation is suitable to identify dense tumors in large sample cohorts in a fast and computationally inexpensive fashion. The presented subgrouping of patient-derived tissues could serve as a prescreening in order to guide subsequent MALDI-MSI examination. As the prior selection of “pure” cell groups of similar molecular makeup is a necessary condition for the identification of biomarkers, the method can be used to boost classification efficacy [162]. Moreover, differentiation and segmentation can be performed without the need of any *a priori* information, except the number of expected tumor subpopulations k . In that regard, the depicted example of $k = 5$ segmentation appears inappropriate, as the resulting red and purple segment do not correspond to divergent histological features. In order to identify diagnostic features that reflect tumor subpopulations and patient outcome, an increased chemical specificity as provided by mass spectrometry would be necessary.

5.4 Same-slice targeted FTIR-guided FTICR-MS acquisition

Extraction of mass features targeted to pre-defined tissue morphologies and/or divergent cell populations by means of FTIR segmentation improves the interpretation and characterization of MALDI-MSI data in an unbiased and histopathology-independent manner. As a negative side effect, the integration of FTIR images prolongs the experimental acquisition time and induces additional data load. This is especially unfavorable in a clinical setting in which speed and ease of a technique are necessary for its applicability. For this, the analysis of large tissue series typically demands a compromise between resolution and throughput in order to keep data load, acquisition time and computational cost manageable. In addition, only a small, manually-defined fraction of the recorded MSI data is actually used for later characterization [21]. To this end, it becomes obvious that the pre-definition of spatially restricted morphological structures for target-exclusive MALDI-MSI measurements has the potential to drastically increase throughput. FTIR images are beneficial as their acquisition only takes a fraction of time when compared to MSI, especially ultra-high resolution MSI. Moreover, segmentation of FTIR images provides guidance to predefined tissue morphologies without the introduction of user bias. In order to overcome the computational and practical challenges related to ultrahigh-resolution MS, the developed workflow was modified to allow for automated acquisition of FTIR and ultrahigh-resolution MALDI-MSI images using a FTICR-MS (**Fig. 28**). Moreover, the use of gold-coated object slides enables multimodality imaging on the same tissue section and thus guarantees the needed precision for high-resolute registration and morphology definition.

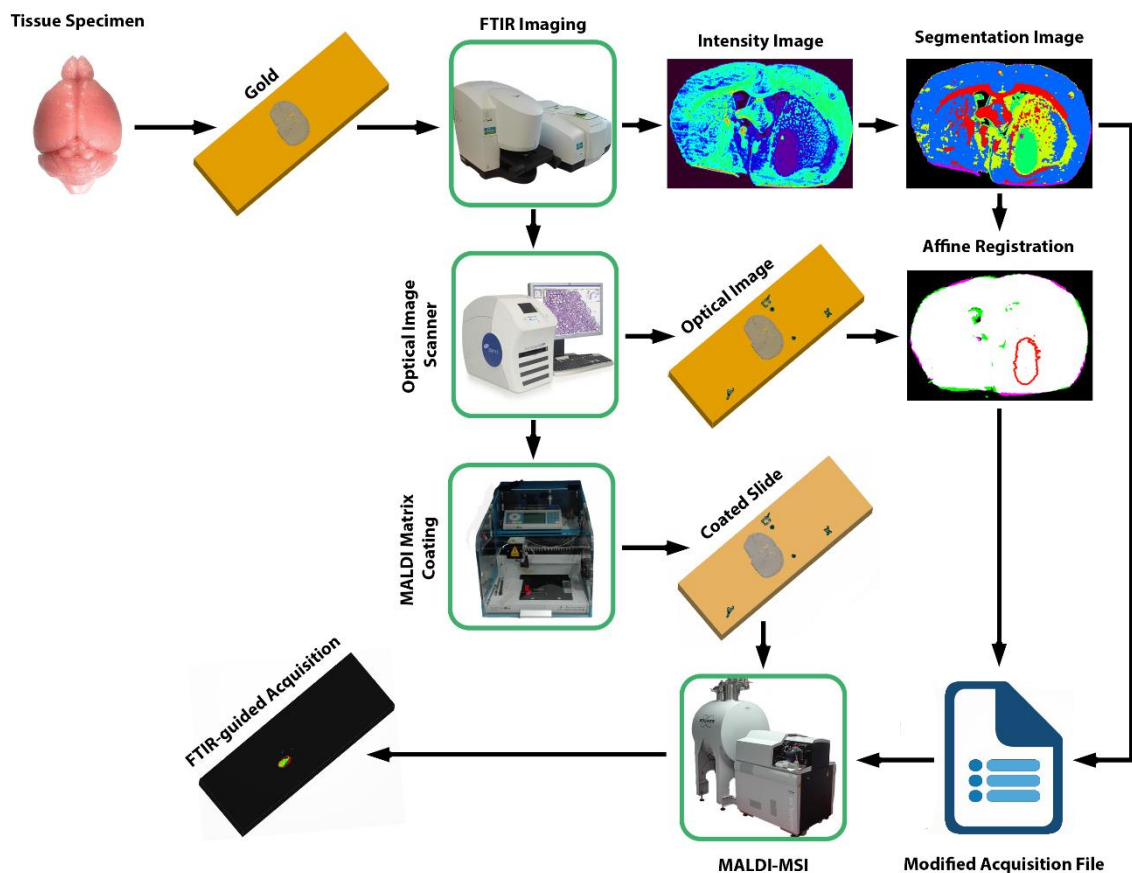


Figure 28 Schematic workflow diagram for FTIR-guided single-slice acquisition of MSI data. FTIR image acquisition on gold can be utilized to denote measurement regions and/or ROIs prior to MALDI-MSI acquisition from the same slide. The spatial contours are defined by segmentation of the FTIR image and subsequent registration to the optical image used for later MSI measurement. By choosing a particular segment, its respective MSI contour points can be calculated in order to modify the MSI acquisition file. The presented workflow enables unbiased targeted mass feature extraction and multimodal acquisition of the same tissue specimen, thus guaranteeing optimal registration.

Most importantly, transfer of FTIR-derived spatial information to the FTICR mass spectrometer before MSI acquisition enables the exclusive measurement of a pre-defined number of segments in subsequent MALDI-FTICR-MSI. In order to examine the capabilities of this new workflow, coronal and sagittal brain sections of C57BL/6 wildtype mice were prepared for FTIR- and subsequent MALDI-FTICR-MSI measurement. PhCCA in 90% acetone solution was used as a matrix to enable optimal crystallization and recording of lipid and metabolite signals in reflector negative mode. Utilizing FTIR image-based k-means++ segmentation, MALDI-FTICR-MSI measurements were directed at morphological structures of interest, namely the dentate gyrus of the hippocampus in the sagittal mouse brain section and the cerebellar fiber tracts in the coronal section. Targeted MALDI-FTICR measurement of the dentate gyrus was achieved by partitioning the respective FTIR image into eight segments (**Fig. 29A**). After successful registration ($DSC = 0.99$), masses were

recorded using a common 512kB data point size (**Fig. 29B**). The resulting image was acquired in 24 min, which indicates a reduction of 97.8% compared to an acquisition time of 18.2 h for the whole tissue section (**Fig. 29C**). Accordingly, the data volume of the whole tissue section was reduced from 35.3 GB to 0.7 GB. MALDI-FTICR measurement of the fiber tracts in coronal brain sections was achieved by subdividing the respective FTIR image into five segments (**Fig. 29D**). After registration to the optical image used for MSI (DSC = 0.992), cerebellar fiber tracts were exclusively measured using a more excessive 4M data point size (**Fig. 29E**) that resulted in a total data load of 189.5 GB. This implies a reduction of 79.3% when compared to whole-tissue image acquisition that would have resulted in a total data load of 915.3 GB (**Fig. 29F**). Similarly, the data acquisition time was reduced to 15.4% from 25.4 hours down to 3.9 hours. In both test cases, the cluster number k was gradually increased until the desired morphological sub-region became observable.

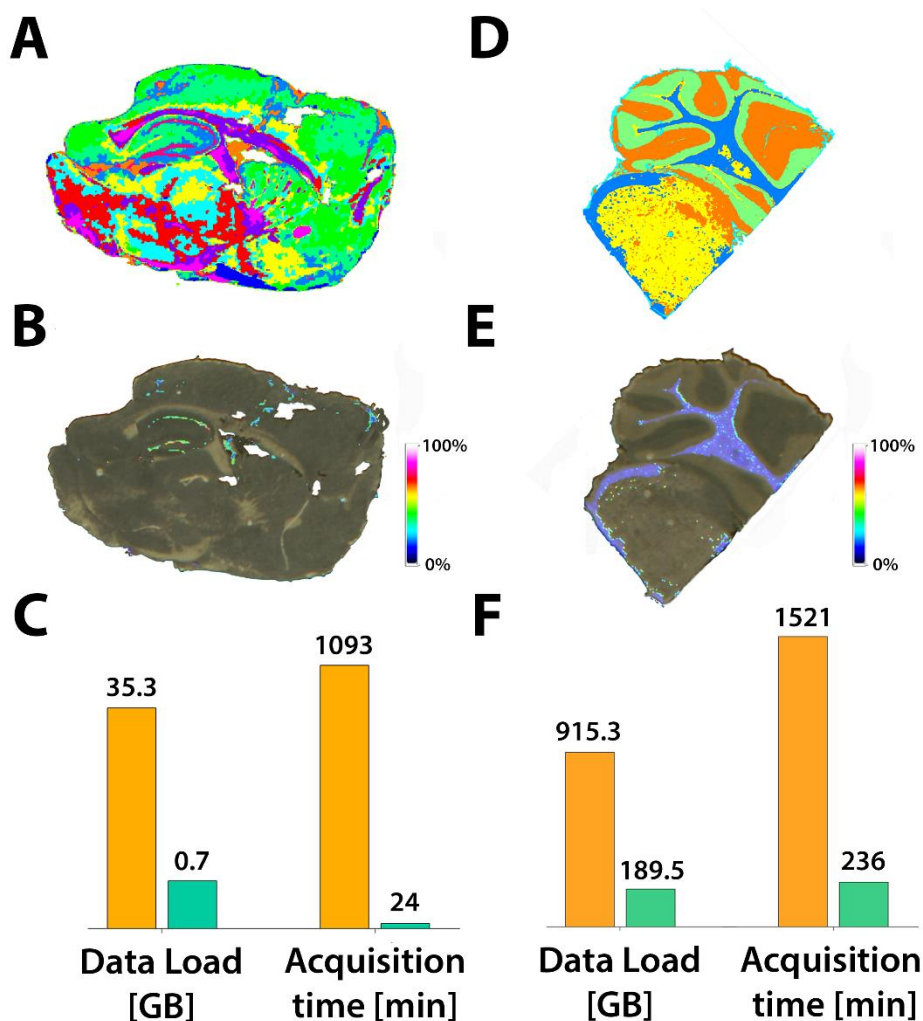


Figure 29 FTIR-guided, spatially restricted data acquisition by ultrahigh-resolution MALDI-FTICR-MSI (A) Based on a FTIR-image of a sagittal CB57BL/6 brain tissue section, segments were calculated by means of ($k = 8$) k-means++ cluster analysis. (B) The segment that represents the spatial contours of hippocampal dentate gyrus was picked and registered to the optical image used for subsequent ultrahigh-resolution MALDI-FTICR-MSI, thus enabling targeted segment-exclusive recording of mass features. The depicted example image of m/z 1133.64 (± 0.05) was acquired in reflector negative mode after matrix coating with PhCCA in 90% acetone solution. (C) Targeting of the dentate gyrus region using a common 512k data point size drastically decreases (97.8%) data load and admission time. (D) In a similar fashion, the coronal brain tissue section of a CB57BL/6 mouse was dissected into five segments to enable targeted acquisition of the cerebellar fiber tracts. (E) The depicted intensity distribution of m/z 906.61 (± 0.05) was acquired in reflector negative mode using a more excessive 4M data point size, thus (F) reducing data load and acquisition time by 79.3%.

Ultimately, time saving and data load decrease are determined by the size of the respective segment used for acquisition. In conclusion, the potency of FTIR-guidance to drastically improve throughput and data handling of ultrahigh-resolution MSI was demonstrated using a Bruker Daltonics MALDI-FTICR-MSI system.

6 DISCUSSION

In this thesis, a robust workflow that combines the speed and ease of vibrational spectromicroscopy with the in-depth chemo-analytical capabilities of mass spectrometry imaging was developed. Instead of combining the obtained spectral features in order to potentially discriminate between different cell populations with higher accuracy, a multimodal guidance approach was investigated in which FTIR imaging guides the interpretation and acquisition of subsequent MSI experiments. In this context, a variety of (semi-)automated methods for the targeted extraction and acquisition of mass signatures was evaluated.

6.1 Technical evaluation of FTIR and MS imaging

A major part of this work focused on the development and evaluation of an automated workflow that enables combinatory FTIR-MS multimodality imaging. With respect to the work of Wehbe *et al.* [145], the first challenge during development was the evaluation of different coating-materials and their influence on MS and FTIR imaging. In case of FTIR, imaging of homogeneous porcine liver sections using ITO-coated slides that represent state-of-the-art material for MS image acquisition was proven to significantly impair spectral quality and dispersion when compared to slides coated with gold or Ag/SNO₂ [146, 147]. For this reason, ITO was considered impractical for vibrational imaging, although FTIR-based segmentation results were comparable. In case of MSI, no critical differences between gold- and ITO-coated object carriers were observable. In comparison to ITO, signal intensity of mass spectra acquired on porcine liver sections was slightly increased when obtained on gold. However, as the improved signal was accompanied by increased scattering, the influence of different conductive slide-coating materials on MS imaging remains uncertain. MS imaging on MirrIR-slides was not evaluated, as the slides didn't fit into the target plate holder of the mass spectrometer and no compatible replacement was found. Taken together, these findings don't impede the usage of ITO-coated object slides for parallel acquisition of FTIR and MS images, but same-slide measurements appear unfavorable. In comparison to gold-coated carriers, ITO- and MirrIR-slides hold the advantage of being transparent and thus allow for histopathological staining following MALDI-MSI and FTIR imaging measurements. By contrast, gold can be utilized for both imaging

modalities and thus enables subsequent FTIR- and MS-image acquisition on the same specimen. This represents a significant advantage for multimodality imaging purposes by further improving registration efficiency [19, 21].

Besides practical considerations necessary for the realization of combinatory MS-FTIR imaging, a conceptual strategy had to be thoroughly investigated. In this sense, two major strategies for multimodal imaging exist. The former combines the orthogonal information obtained from different sensor types or acquisition modes to create hyperspectral data that allows for joint analysis and a deeper insight into the chemistry of complex samples [13, 163]. The latter utilizes the sensor-dependent capabilities of one modality for the purpose of improving or guiding another downstream modality [14, 20, 164]. In this work, the segmentation and classification capabilities of MS and FTIR imaging were investigated. In this context, FTIR-based segmentation of glioblastoma xenograft tissues and human-derived GIST was proven superior when compared to MALDI-MSI, as the respective tumor contours were depicted with higher accuracy. Advanced state-of-the-art clustering algorithms improved the dissection of MSI data sets when compared to the k-means++ method and thus indicate that proper clustering of MSI is possible [144]. However, the FTIR-derived structural information connote a significant improvement. In order to extend the capabilities of MSI-segmentation, a variety of segmentation pipelines have been presented in recent publications (reviewed in [165]). In a comparative study of Sarkari *et al.*, enhanced MSI segmentation was presented by applying prior dimensionality reduction and alternating distance metrics [166]. Widlak *et al.* demonstrated unsupervised iterative k-means clustering of squamous cell carcinoma by including a Gaussian mixture model in order to detect spectral components [167] and Bemis *et al.* examined the potency of probabilistic segmentation by applying spatial shrunken centroids [168]. Moreover, the applicability of dimensionality reduction techniques such as PCA and t-SNE to visualize structures in MSI data was examined [169]. Despite these findings, FTIR-based tissue segmentation holds major advantages. The technique is unaffected by ion suppression, analyte-diffusion or other effects that may suppress the identification of morphology-dependent differences. Moreover, the technique operates comparatively fast at higher spatial resolution and demands for no excessive processing. The feasibility to record FFPE tissue materials was also demonstrated, although segmentation failed to identify tumor areas in human-derived prostate cancer

sections. It is uncertain, if this is due to the noise increase caused by embedding medium residues or if FTIR segmentation is in general not able to visualize cancer in this particular tumor type, as more comprehensive studies of different tissue types would be needed. The presented long-term experiments demonstrated high stability and technical reproducibility of FTIR-denoted tissue segmentation. In that sense, the establishment of a tissue databases for quality control purposes appears to be possible. The non-destructive nature of FTIR microscopy makes it also applicable prior to MSI, thus providing structural information prior to MSI acquisition. However, FTIR lacks the ability to directly visualize individual molecules in a multiplex manner which makes its combination with mass spectrometry expedient. Taken together, these considerations indicate for the potency of guided multimodality imaging that utilizes the rapid virtual dissection of tissues obtained from FTIR imaging to define homogeneous cell populations in subsequent MSI experiments.

6.2 Development of FTIR-guided MSI

According to the depicted single modality results, a workflow for FTIR-guided, ultrahigh-resolution MSI was developed that offers interpretation and data acquisition focused on defined tissue sub-compartments independent of histopathological staining. This represents a substantial advance when compared to previous image fusion approaches that rely on H&E images acquired on adjacent sections for interpretation [14, 18, 21]. The ability of infrared spectroscopy to visualize different classes of biomolecules provides a particularly powerful tool for the analysis of tissue samples as traditional histology is only able to visualize a few molecules simultaneously [170]. The presented concept for multimodal imaging can be easily transferred to any other non-destructive optical upstream modality (i.e. Raman imaging), thus providing guidance in cases where FTIR imaging is incapable of identifying tissue regions of interest. Spatially guided imaging is applicable to MSI data obtained using any ionization method and mass spectrometer (i.e. DESI-MSI). The only requirement is the alignment of both imaging planes onto the same coordinate system for which a registration routine was developed in MATLAB. Thereby, the acquired FTIR image and the optical image used for MSI acquisition are transferred to binary and subsequently registered.

6.3 Tumor targeting and spatially restrictive mass feature identification

Integration of the molecular information provided by MALDI-MSI and tissue anatomical structures that can be resolved by microscopy is mandatory for its clinical implementation. Utilizing spatial features derived from FTIR imaging, the precise identification of tumor-associated contours was presented in mouse brain sections inoculated with human U87-MG glioblastoma cells. This was achieved by joint clustering of all acquired mid-infrared images at once, as it reveals cellular intra- and inter-section dependencies. For this reason, the linkage between a single segmentation image and its corresponding H&E stain sufficed to enable tumor localization in all biological replicates. In contrast to MSI, clustering of FTIR data acquired for multiple sections was easily performed without the need of excessive preprocessing. In order to extract tumor-specific mass signatures with increased precision, regions of interest were defined by registering the retrieved tumor-segments to MSI space. The registration quality was quantified by calculating the DSC that suggested a precise alignment between FTIR image and MSI data set ($DSC = 0.982 \pm 0.004$). Whereas the capability to visualize sparse tumor populations might represent a challenging task, automated segmentation and registration of tumor-lesions was readily performed. In addition to tumor-contours, molecular differences were observed in an expansion of the tumor margin (referred to as the 'S2 segment'). Both the retrieved distribution pattern of glioblastoma and altered surrounding tissue were backed up with chemical mass features. These observations emphasize the potency of FTIR-imaging to reveal molecularly substantially altered cell populations that are invisible to conventional H&E staining. In order to further characterize the segment, additional histopathological and/or in-depth mass spectrometry techniques such as tandem mass spectrometry may be considered. Without further knowledge, the hypothetical origin of the S2 contour might be a brain edema caused by the inoculation and/or a glioma-characteristic infiltration zone. It is also possible that S1 and S2 refer to the glioblastoma-characteristic necrotic tumor center and surrounding anaplastic cells.

It is noteworthy that the identification of U87-MG xenograft patterns in mice represents a simplified test case in comparison to clinical human-derived glioblastoma biopsy analysis. On top of this, the origin of U87-MG was critically scrutinized by Allen *et al.* in a recently presented study [171]. Based on genetic profiling and transcriptome

analysis, they discovered that the DNA profile differed from the cells originally isolated by J. Ponten and associates from 1966 to 1969. However, the automatic alignment between FTIR and MSI holds great promise as the mapping of data from different sensors to the same coordinate space allows for an improved biomolecular profiling of the examined specimen. Moreover, the structural annotation provided by FTIR segmentation enables the extraction of mass features in context of their anatomical localization and mitigates the lack of *a priori* information that mass spectrometry laboratories often deal with.

6.4 Automated screening for local mass biomarker signatures

In the course of this work, a method that utilizes segmentation of FTIR images in order to explore tissue morphologies in MSI data was developed. Serving as a suitable example, lipid disorders were examined in mouse brain, arguably one of the most-used models to investigate human neurological diseases [172]. The technique was able to identify spatially restrictive mass signature shifts between NPC1 and wild-type mice by using the FTIR-denoted anatomical structures as a template for pairwise feature extraction. This way, disease-related effects could be associated to distinct brain regions without any *a priori* information about their spatial occurrence. In comparison to the ABA the origin of three segments could be identified, namely cerebellar fiber tracts as well as granule and molecular cell layer. In addition, an intermediate region between white and grey matter was identified. Although the morphological origin of this structure is unknown, it was repeatedly observed in FTIR images acquired throughout the course of this thesis. The most likely reason for its occurrence is the partly acquisition of granule cells and myelinated axons in a single pixel. As the infrared signatures of the granule cell layer and the cerebellar fiber tract differ considerably, mixed signals have the potential to form a distinct cluster during segmentation. This hypothesis is supported by two observations. First of all, the structure's calculated thickness mostly amounts to a single pixel (25x25 μm) which could account for the fuzzy intersection between grey and white matter. Secondly, a similar border region was obtained at the outer edge of the tissue. The respective segment was discarded prior to the analysis, as it is conceivable that it represents the intersection of spectra acquired on- and off-tissue.

Notwithstanding the above, the presented technique substantially improves MSI analysis by predefining cerebellar sub-sections. This way, the local accumulation of distinct masses, such as m/z 1382.9, which was insignificant over the entire dataset became detectable. This particular mass might refer to the previously reported accumulation of GM2(d18:1/18:0) although the reported mass differs [131, 158]. This allows for two possible explanations: Either the depicted m/z 1382.9 refers to a different molecule that has not been reported in the mouse model-describing literature or the mass axis was misaligned during the presented MS experiment. Future ultrahigh-resolution MSI experiments will further investigate the origin chemical identity of m/z 1382.9.

Biomedical MSI is typically not feasible for intraoperative applications and therefore carried out in analytical laboratories for which tissue annotation is a major bottleneck. The automatic alignment and dissection of MS data into subsections facilitates comparative studies of structured organs (e.g. brain, kidney) or infiltrating tumors of different origin. On the negative side, it doesn't improve comparative analysis of homogeneous specimen (e.g. liver sections). It is noteworthy that the presented spatial decomposition occurs solely at data level and can therefore be adjusted should the initial experiment fail to detect sub-structures of interest. For this reason, FTIR-guided dissection holds decisive advantages when compared to disruptive procedural steps such as laser-capture microdissection that are irreversible [20, 164]. In addition, joint segmentation of multiple images would guarantee the comparative analysis of intersection compartments with a similar morphological composition, thus circumventing the need for atlas linkage [173].

6.5 Assessment of histological heterogeneity

A major challenge in (pre)clinical mass spectrometry imaging is the occurrence of histological heterogeneities. The occurrence of morphological differences impedes true molecular marker identification by introducing variability and mixed class data sets. Without precise information regarding the identity of each acquired signature, it is highly possible that tumor-related signals are overlooked. For mass spectrometry, the conventional approach of integrating histopathological tissue stains can prove to be insufficient [21, 97, 174, 175], as they demand for manual annotation and are thus

not automatable. Moreover, the unbiased exploration of alterations demands for an untargeted analysis technique that accounts for the spatial tumor environment. In order to solve the issue, direct segmentation as well as t-SNE dimensionality reduction and color-coding of MSI data have been presented [118, 159, 176]. Thereby, heterogeneities and subpopulations were identified *de novo* in breast and gastric cancer and tumor-types that were not distinguishable by conventional histopathology could be linked to disease outcome prognoses [118, 159-161]. Despite the reported potential of direct MSI dissection, this method has major disadvantages. In a practical sense, size and complexity of MALDI-MSI data (especially ultrahigh-resolution MSI using a FTICR detector) makes unsupervised clustering of large sample cohorts impractical and time consuming as it demands for excessive pre-processing. Moreover, the high information content of MSI data is not necessarily beneficial for segmentation purposes as the high dimensionality introduces multiple challenges. Although some dimensions (mass features) are characteristic of a given subgroup, a larger number of attributes will usually not be meaningful and can mask existing clusters in high-dimensional data sets [177, 178]. Furthermore, it is likely that the disease-related effects are depicted in multiple mass features. This presents a challenge as samples should only be grouped together if they exhibit a common correlation among the respective set of m/z -values. The phenomenon that different features or correlations may be relevant for varying clusters is called local feature relevance. Lastly, pattern recognition from a finite number of data samples in high-dimensional feature space suffers from what is called “Curse of Dimensionality”, where the concept of distance substantially loses precision and consequently degrades the predictive power of the classification models [112].

In comparison to MSI data, FTIR images are of comparably small size which allows for the joint segmentation of a multitude of tissue sections in a faster and computationally less extensive fashion. For example, FTIR-based segmentation of 89 GIST-patient derived tissue samples was carried out in ~90 seconds on a standard desktop computer. In that regard, the technique may hold great potential as a supporting tool in clinical MSI rather than a contender, as the identified group’s origin was mostly linked to the occurrence of tumor or fibrosis. In order to identify diagnostic features that reflect the tumor’s genetic variance and patient outcome, the chemical specificity of mass spectrometry would be necessary. Nevertheless, joint FTIR-segmentation of tissues

offers fast initial subgrouping of large tissue cohorts and can pre-select tumorous areas prior to MSI acquisition and mostly independent of H&E staining. The obtained allocation information can contribute to in-depth mass feature analysis in two ways. First and foremost, it becomes possible to pre-select samples of similar morphological composition, thus reducing the locale feature relevance and boosting subtype identification in MSI [162]. Secondly, FTIR-based dissection can be utilized to identify tissue inhomogeneity prior to MSI, thus minimizing subgroup-irrelevant spectra that reduce classification efficacy.

6.6 Targeted spatially-restrictive high-resolution MSI acquisition

Lastly, FTIR-denoted, segment-exclusive MALDI-FTICR-MSI was showcased as a way to overcome the compromise between spatial and spectral resolution needed for high-throughput measurements. The technique also enables the exclusive processing of similar cell populations, thus reducing data variance when compared to whole-organ images. These aspects are beneficial for very specific analyses of tissue structures, as time and data savings are achieved by discarding the pixels of all remaining sections before acquisition. Although the selection of an appropriate segment for further in-depth mass spectrometry experiments requires knowledge regarding its location, shape or spectral composition, the presented workflow is independent of conventional histopathological staining. This represents a substantial advance when compared to previously reported solutions [19, 21], as the acquisition of spatially-restrictive regions solely relies on unbiased FTIR image segmentation results. In addition, the results of this thesis have emphasized the advantages of FTIR-based image dissection when compared to conventional histopathological stains. First and foremost, MIR-spectra contain additional biomolecular information that enables a more detailed differentiation of cells. Secondly, the use of gold-coated object slides enables FTIR image acquisition of the same tissue specimen prior to the MSI experiment. Same-slice acquisition represents a significant upgrade by further improving registration efficiency when compared to consecutive section registration. Moreover, same-slice measurements minimize practical deficiencies that can be caused by sectioning artifacts such as tears, strains and folds. This is essential for the exclusive acquisition of smaller tissue compartments, i.e. the granule cell layer of the dentate gyrus, as small misalignments can cause the subsequent MSI acquisition to miss the desired region. However,

preparation and acquisition on gold-coated slides also produces disadvantages, as the opaque surface limits the applicability of subsequent same-slice stains and microscopy. Moreover, the use of gold introduces additional costs when compared to ITO. For these reasons, MirrIR-slides of MS-compatible size might be better suited for multimodal FTIR-guided MSI depending on their conductive property. Although Ag/SNO₂ has been reported to possess excellent electrical and mechanical qualities [179, 180], the exact coating-composition of MirrIR-slides is unknown and its effect on MSI has not been evaluated in the course of this thesis.

The capability to automatically identify and exclusively acquire small-sized tissue compartments is ultimately denoted by the guiding optical modality and cluster algorithm. In this work, FTIR-guided high-resolution MSI was demonstrated using a Bruker Daltonics MALDI-FTICR-MSI system. However, the presented workflow is vendor-neutral and applicable to any subsequent imaging technique that allows to define measurement regions prior to recoding – be it by courtesy of an optical image or point definition within a vendor-specific acquisition file. The presented concept for multimodal imaging can be easily transferred to any other non-destructive optical modality (i.e. Raman imaging), thus providing guidance in cases where FTIR imaging is incapable of identifying the structure of interest. In addition, the workflow is applicable to MSI data obtained using different ionization methods and/or mass analyzer (i.e. DESI-MSI). The only requirement is the alignment of both imaging planes onto the same coordinate system. In summary, the presented technique for multimodality imaging bypasses computational and practical challenges in cases where either the region of interest only represents a comparatively small tissue fraction or the acquisition of whole-tissue sections is not feasible. In a similar fashion, the technique can be utilized to better focus high spatial and mass resolution analyses by limiting acquisition to morphological structures of high spectral homogeneity. Multimodality imaging of the same tissue slide is especially advantageous as it circumvents structural dissimilarities between subsequent sections and artifacts caused by tissue handling, such as stretches, tears or folds. In addition, it renders the need for additional tissue preparation steps unnecessary, which speeds up the workflow. On the negative side, the region-specific recording of mass data limits retrospective analyses and requires the region of interest to be known in advance.

6.7 Different cluster approaches

The current work has focused on k-means++ clustering for its potential to rapidly segment complex data and therefore increase the throughput of subsequent ultrahigh-resolution mass spectrometry analysis. Although k-means++ represents an advantageous technique that can partition big data sets at a low cost of central processing unit (CPU) load, it also features several disadvantages. The specified number of computed segments k needed for clustering represents a remaining source of user-bias. In most experiments, the true number of identifiable tissue sub-structures is unknown and it is therefore up to the experimenter to determine a fitting value. However, this process creates two particular risks, as k-means++ represents a hard cluster technique that strictly allocates each pixel to one group only. If k is chosen too low, structures of importance might be overlooked. If k is chosen too high, the image is forcefully partitioned into sub-compartments that do not represent real structures, thus creating noise. This issue is reinforced by the fact that tissues of identical origin do not necessarily share a common number of mid-infrared sub-structures, which limits the applicability of organ-specific k -values. There are two specific reasons for this problem. First, tissues of the same organ can still vary in their composition, depending on the cutting side and disease-related effects. Second, folds and other sectioning artifacts typically represent separate segments, as the increased tissue thickness substantially alters the acquired mid-infrared spectrum. In this work, a possible solution that might mitigate the k -bias was proposed by applying cluster evaluation methods, i.e. the presented Calinski-Harabasz criterion [125] or the Davies-Bouldin index [181]. However, the CHc was found to be sometimes inappropriate to define a k -value that represents the true number of within-tissue morphologies. In multiple cases throughout this study, the most stable number of allocation groups was too low and cell populations that could be verified in H&E were only visible for manually increased k -values. For this reason, multiple k -values were tested in many of the depicted studies and empirically evaluated. It is up to additional experimental testing to verify the accuracy of cluster evaluation techniques in order to pre-define a suitable k -value for tissue segmentation.

An additional limitation of k-means++ is the strict partitioning of the acquired image. Each obtained spectrum is assigned to exactly one cluster without overlapping. In biology however, it is conceivable that a given cell or cell population is representing

multiple characteristics simultaneously. In that sense, alternative segmentation tools, i.e. agglomerative hierarchical cluster analysis [182, 183] that is independent of the value k or fuzzy c -means [184, 185] that allows each data point to belong to multiple clusters, could enable a more detailed and independent coverage of tissue areas. However, hard clustering is essential for segment-exclusive MSI acquisition. For the technical implementation, it is of importance to denote the exact coordinates of all acquired pixels, as partial recording is not possible. Moreover, the depicted alternate algorithms demand for additional CPU load which minimizes throughput.

7 SUMMARY AND OUTLOOK

Mass spectrometry imaging provides ultrahigh-resolution mass information needed to uncover molecular features and heterogeneities related to the disease-state of tissues. The technique holds great potential in cancer research and patient personalized medicine, as it has demonstrated its ability to uncover tumorous cell populations and disease-related effects. Modern ultrahigh-resolution mass analyzers (i.e. FTICR) have greatly increased cell-profiling accuracy at the cost of ever increasing data load. For this reason, a major challenge that impedes routine clinical implementation to date has been the identification of regions of interests within the highly complex datasets. Analysis of large tissue cohorts typically demands prior histopathological annotation, in order to keep data load and acquisition time manageable. In contrast, the speed, simplicity and non-perturbing nature of vibrational microscopy has made it a valuable tool to complement current histopathological practice. The technique provides a rough overview of biomolecular distribution patterns under native conditions, but lacks MSI's potential to identify specific molecules. Therefore, the development of a multimodal imaging workflow that combines both technique's capabilities while overcoming modality-dependent restraints would have substantial impact on future research and perhaps clinical practice.

For these reasons, this thesis focused on the following four objectives:

- 1.) Evaluation of suitability of different slide-coating materials for FTIR and MS single modality imaging.
- 2.) Development and technical evaluation of an automated workflow that enables combined FTIR and MALDI-MSI tissue analysis.
- 3.) Application of this technique for the discrimination of healthy and diseased tissues in various disease models.
- 4.) Use of FTIR-guidance in order to pre-define acquisition regions in subsequent MALDI-MSI experiments, thus reducing data load and acquisition time.

In the first part of this study, different coating materials were tested for their applicability in FTIR and MS imaging. Thereby, it could be shown that the respective composition had a substantial influence on spectral quality in both modalities. For FTIR imaging,

gold- and Ag/SNO₂ (MirrIR)-coating provided superior spectral quality and less dispersion on homogeneous liver tissue samples when compared to ITO. For MSI, the use of gold- and ITO-coated slides resulted in comparable mass signatures and dispersion. MirrIR-slides could not be used for mass spectrometry as they didn't fit into the Bruker instrument. For these reasons, multimodal experiments were carried out on gold-coated slides. Moreover, the feasibility of long-term FTIR studies and measurements of FFPE tissues was briefly evaluated.

In the second part of this study, a concepts for guided multimodality imaging was developed. In this context, FTIR image segmentation was able to identify tissue structures and cancerous lesions in human primary GIST and murine glioblastoma xenograft tissues with higher precision when compared to MSI. These findings emphasized the benefits of a guidance approach and a multistep workflow with the purpose of defining homogeneous tissue structures based on FTIR images in order to partition complex MSI data was developed. By registering the FTIR-derived cluster contours to the optical image used for MSI acquisition and interpretation, it became possible to identify and define molecularly distinct tissue morphologies for MALDI MSI guidance.

In the third part of this study, the utility of the developed multimodality imaging workflow was exemplified in multiple disease models. Based on FTIR imaging, tumor-related contours that were partially invisible to conventional histopathological staining could be visualized in murine glioblastoma xenograft models. By registration to MSI that was acquired in parallel, it became possible to screen for tumor-associated mass signatures. For each of two tumor-segments, ten masses were identified by means of feature extraction using the empirical ROC curve criterion. The distribution pattern of all 20 masses matched the respective FTIR-segment, thus emphasizing its authenticity and the advantages of the developed workflow. In cases for which no *a priori* information about the spatial occurrence of disease-specific effects is available, automated segment-wise mass signature exploration was presented. Without user-bias or the need of stains, specific mass shifts could be detected and linked to cerebellar subsections in a murine Niemann-Pick Disease, type C1 model that were not striking in whole-tissue fold change plots. In another approach, human derived GIST tissue sections of multiple patients were simultaneously clustered in order to

identify tumor subpopulations. Although the pathological relevance of the depicted groups could not be assessed conclusively, allocation results of technical replicates were proven consistent. Future comparison to the respective MSI imaging results may clarify the observed subgroup-identities.

In the last part of this thesis, FTIR-guided acquisition was presented as a suitable tool to overcome computational challenges related to ultrahigh-resolution FTICR-MSI. The adjusted workflow for single-slice multimodality imaging ensures the exclusive analysis of pre-defined tissue structures of similar morphological makeup. In coronal and sagittal murine brain tissue sections, a registration efficacy of ~99% and a reduction of acquisition time and data load of up to 97.8% were demonstrated, thus enabling high resolution FTICR-MSI in cases for which the slow scan speed impedes the analysis of complete tissue sections or large patient cohorts.

In conclusion, an automated registration pipeline that enables FTIR guidance for targeted acquisition or mass feature extraction in subsequent MSI experiments was demonstrated. Future work on the subject will likely be extended to alternate vibrational microscopy techniques, i.e. Raman spectroscopy that potentially discover yet unseen tissue morphologies, i.e. cerebellar Purkinje cells. In addition, the applicability of cluster evaluation techniques in order to automatically define a suitable k-value will be examined. Moreover, t-SNE dimensionality reduction [176] and varying cluster algorithms will be tested for their potential to provide better segmentation results. Furthermore, ultrahigh-resolution experiments of I1061T NPC1 mice will be carried out in order to verify the chemical identity of m/z 1382.9. Finally, it is noteworthy that the depicted registration pipeline would also enable FTIR-MSI image fusion as reported by van de Plas *et al.* [14]. It is conceivable that the additional biomolecular information of FTIR images greatly improves the workflow and enables forecasting distribution patterns of additional mass features.

8 REFERENCES

1. Behar, J.M., et al., *The role of multi modality imaging in selecting patients and guiding lead placement for the delivery of cardiac resynchronization therapy*. *Expert Rev Cardiovasc Ther*, 2017. **15**(2): p. 93-107.
2. Lin, E. and A. Alessio, *What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT?* *Journal of cardiovascular computed tomography*, 2009. **3**(6): p. 403-408.
3. Cormack, A.M., *Representation of a Function by Its Line Integrals, with Some Radiological Applications*. *Journal of Applied Physics*, 1963. **34**(9): p. 2722-2727.
4. Kuhl, D.E., J. Hale, and W.L. Eaton, *Transmission Scanning: A Useful Adjunct to Conventional Emission Scanning for Accurately Keying Isotope Deposition to Radiographic Anatomy*. *Radiology*, 1966. **87**(2): p. 278-284.
5. Hasegawa, B.H., et al. *Description of a simultaneous emission-transmission CT system*. 1990.
6. Blankespoor, S.C., et al. *Attenuation correction of SPECT using X-ray CT on an emission-transmission CT system: Myocardial perfusion assessment*. in *1995 IEEE Nuclear Science Symposium and Medical Imaging Conference Record*. 1995.
7. Beyer, T., et al., *A combined PET/CT scanner for clinical oncology*. *J Nucl Med*, 2000. **41**(8): p. 1369-79.
8. Townsend, D.W., *Dual-modality imaging: combining anatomy and function*. *J Nucl Med*, 2008. **49**(6): p. 938-55.
9. Chughtai, S., et al., *A multimodal mass spectrometry imaging approach for the study of musculoskeletal tissues*. *International Journal of Mass Spectrometry*, 2012. **325–327**: p. 150-160.
10. Bocklitz, T.W., et al., *Deeper understanding of biological tissue: quantitative correlation of MALDI-TOF and Raman imaging*. *Anal Chem*, 2013. **85**(22): p. 10829-34.
11. Lotz, J.M., et al., *Integration of 3D multimodal imaging data of a head and neck cancer and advanced feature recognition*. *Biochim Biophys Acta*, 2016.
12. Bocklitz, T., et al., *Novel workflow for combining Raman spectroscopy and MALDI-MSI for tissue based studies*. *Anal Bioanal Chem*, 2015. **407**(26): p. 7865-73.
13. Lasch, P. and I. Noda, *Two-Dimensional Correlation Spectroscopy for Multimodal Analysis of FT-IR, Raman, and MALDI-TOF MS Hyperspectral Images with Hamster Brain Tissue*. *Analytical Chemistry*, 2017. **89**(9): p. 5008-5016.
14. Van de Plas, R., et al., *Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping*. *Nat Methods*, 2015. **12**(4): p. 366-72.
15. Abdelmoula, W.M., et al., *Automatic generic registration of mass spectrometry imaging data to histology using nonlinear stochastic embedding*. *Anal Chem*, 2014. **86**(18): p. 9204-11.
16. Van Malderen, S.J.M., et al., *Three-Dimensional Reconstruction of the Tissue-Specific Multielemental Distribution within *Ceriodaphnia dubia* via Multimodal Registration Using Laser Ablation ICP-Mass Spectrometry and X-ray Spectroscopic Techniques*. *Analytical Chemistry*, 2017. **89**(7): p. 4161-4168.

17. Seeley, E.H., et al., *Co-registration of multi-modality imaging allows for comprehensive analysis of tumor-induced bone disease*. *Bone*, 2014. **61**: p. 208-216.
18. Tata, A., et al., *Rapid Detection of Necrosis in Breast Cancer with Desorption Electrospray Ionization Mass Spectrometry*. *Sci Rep*, 2016. **6**: p. 35374.
19. Woolman, M., et al., *Optimized Mass Spectrometry Analysis Workflow with Polarimetric Guidance for ex vivo and in situ Sampling of Biological Tissues*. *Scientific Reports*, 2017. **7**: p. 468.
20. Großerueschkamp, F., et al., *Spatial and molecular resolution of diffuse malignant mesothelioma heterogeneity by integrating label-free FTIR imaging, laser capture microdissection and proteomics*. *Scientific Reports*, 2017. **7**: p. 44829.
21. Heijs, B., et al., *Histology-Guided High-Resolution Matrix-Assisted Laser Desorption Ionization Mass Spectrometry Imaging*. *Anal Chem*, 2015. **87**(24): p. 11978-83.
22. Diem, M., et al., *A decade of vibrational micro-spectroscopy of human cells and tissue (1994-2004)*. *Analyst*, 2004. **129**(10): p. 880-5.
23. Cozzolino, D., *Near infrared spectroscopy in natural products analysis*. *Planta Med*, 2009. **75**(7): p. 746-56.
24. Raman, C.V., *A new radiation [Reproduced from Indian J. Phys., 1928, 2, 387–398]*. *Current Science*, 1998. **74**(4): p. 382-386.
25. Blout, E.R. and R.C. Mellors, *Infrared Spectra of Tissues*. *Science*, 1949. **110**(2849): p. 137-8.
26. Woernley, D.L., *Infrared absorption curves for normal and neoplastic tissues and related biological substances*. *Cancer Res*, 1952. **12**(7): p. 516-23.
27. Hester, R.E., *Biological applications of Raman spectroscopy, Vols 1 & 2*. *FEBS Letters*, 1988. **230**(1-2): p. 222-222.
28. Clark, R.J.H.H., R.E. , *Advances in infrared and Raman spectroscopy*. 1975.
29. Helm, D., et al., *Classification and identification of bacteria by Fourier-transform infrared spectroscopy*. *J Gen Microbiol*, 1991. **137**(1): p. 69-79.
30. Salman, A., et al., *Early detection of colorectal cancer relapse by infrared spectroscopy in "normal" anastomosis tissue*. *J Biomed Opt*, 2015. **20**(7): p. 75007.
31. Wald, N., et al., *Identification of melanoma cells and lymphocyte subpopulations in lymph node metastases by FTIR imaging histopathology*. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 2016. **1862**(2): p. 202-212.
32. Bellisola, G. and C. Sorio, *Infrared spectroscopy and microscopy in cancer research and diagnosis*. *American Journal of Cancer Research*, 2012. **2**(1): p. 1-21.
33. Ellis, D.I., et al., *Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool*. *Analyst*, 2013. **138**(14): p. 3871-3884.
34. Downes, A. and A. Elfick, *Raman Spectroscopy and Related Techniques in Biomedicine*. *Sensors (Basel, Switzerland)*, 2010. **10**(3): p. 1871-1889.
35. Michelson, A.A. and E.W. Morley, *On the relative motion of the Earth and the luminiferous ether*. *American Journal of Science*, 1887. **Series 3 Vol. 34**(203): p. 333-345.
36. Griffiths, P.R., J.A. De Haseth, and J.D. Winefordner, *Fourier Transform Infrared Spectrometry*. 2007: Wiley.
37. Silva, J.C.D., et al., *Advances in the Application of Spectroscopic Techniques in the Biofuel Area over the Last Few Decades*, in *Frontiers in Bioenergy and*

- Biofuels*, E. Jacob-Lopes and L.Q. Zepka, Editors. 2017, InTech: Rijeka. p. Ch. 03.
38. Walsh, M.J., et al., *FTIR Microspectroscopy Coupled with Two-Class Discrimination Segregates Markers Responsible for Inter- and Intra-Category Variance in Exfoliative Cervical Cytology*. *Biomarker Insights*, 2008. **3**: p. 179-189.
 39. Baker, M.J., et al., *Using Fourier transform IR spectroscopy to analyze biological materials*. *Nat Protoc*, 2014. **9**(8): p. 1771-91.
 40. Butler, H.J., et al., *Using Raman spectroscopy to characterize biological materials*. *Nat Protoc*, 2016. **11**(4): p. 664-87.
 41. Movasaghi, Z., S. Rehman, and I.U. Rehman, *Raman Spectroscopy of Biological Tissues*. *Applied Spectroscopy Reviews*, 2007. **42**(5): p. 493-541.
 42. Ellis, D.I. and R. Goodacre, *Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy*. *Analyst*, 2006. **131**(8): p. 875-85.
 43. Varma, V.K., et al., *A label-free approach by infrared spectroscopic imaging for interrogating the biochemistry of diabetic nephropathy progression*. *Kidney international*, 2016. **89**(5): p. 1153-1159.
 44. Martinez-Marin, D., et al., *Infrared spectroscopic imaging: a label free approach for the detection of amyloidosis in human tissue biopsies*. *Amyloid*, 2017. **24**(sup1): p. 163-164.
 45. Sreedhar, H., et al., *Infrared spectroscopic imaging detects chemical modifications in liver fibrosis due to diabetes and disease*. *Biomedical Optics Express*, 2016. **7**(6): p. 2419-2424.
 46. Bird, B., et al., *Detection of breast micro-metastases in axillary lymph nodes by infrared micro-spectral imaging*. *Analyst*, 2009. **134**(6): p. 1067-1076.
 47. Walsh, M.J., et al., *Attenuated total reflectance Fourier-transform infrared spectroscopic imaging for breast histopathology*. *Vibrational Spectroscopy*, 2012. **60**: p. 23-28.
 48. Verdonck, M., et al., *Characterization of human breast cancer tissues by infrared imaging*. *Analyst*, 2016. **141**(2): p. 606-19.
 49. Lasch, P., et al., *Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis*. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 2004. **1688**(2): p. 176-186.
 50. Nallala, J., et al., *Infrared imaging as a cancer diagnostic tool: Introducing a new concept of spectral barcodes for identifying molecular changes in colon tumors*. *Cytometry Part A*, 2013. **83A**(3): p. 294-300.
 51. Kallenbach-Thieltges, A., et al., *Immunohistochemistry, histopathology and infrared spectral histopathology of colon cancer tissue sections*. *Journal of Biophotonics*, 2013. **6**(1): p. 88-100.
 52. Akalin, A., et al., *Classification of malignant and benign tumors of the lung by infrared spectral histopathology (SHP)*. *Lab Invest*, 2015. **95**(4): p. 406-21.
 53. Bird, B., et al., *Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer*. *Lab Invest*, 2012. **92**(9): p. 1358-73.
 54. Grosserueschkamp, F., et al., *Marker-free automated histopathological annotation of lung tumour subtypes by FTIR imaging*. *Analyst*, 2015. **140**(7): p. 2114-20.
 55. Gazi, E., et al., *A Correlation of FTIR Spectra Derived from Prostate Cancer Biopsies with Gleason Grade and Tumour Stage*. *European Urology*, 2006. **50**(4): p. 750-761.

56. Baker, M.J., et al., *Investigating FTIR based histopathology for the diagnosis of prostate cancer*. Journal of Biophotonics, 2009. **2**(1-2): p. 104-113.
57. Zawlik, I., et al., *FPA-FTIR Microspectroscopy for Monitoring Chemotherapy Efficacy in Triple-Negative Breast Cancer*. Scientific Reports, 2016. **6**: p. 37333.
58. Schleusener, J., et al., *In vivo study for the discrimination of cancerous and normal skin using fibre probe-based Raman spectroscopy*. Exp Dermatol, 2015. **24**(10): p. 767-72.
59. Desroches, J., et al., *Characterization of a Raman spectroscopy probe system for intraoperative brain tissue classification*. Biomedical Optics Express, 2015. **6**(7): p. 2380-2397.
60. Jermyn, M., et al., *A review of Raman spectroscopy advances with an emphasis on clinical translation challenges in oncology*. Phys Med Biol, 2016. **61**(23): p. R370-r400.
61. Kallaway, C., et al., *Advances in the clinical application of Raman spectroscopy for cancer diagnostics*. Photodiagnosis and Photodynamic Therapy, 2013. **10**(3): p. 207-219.
62. Jermyn, M., et al., *Highly Accurate Detection of Cancer In Situ with Intraoperative, Label-Free, Multimodal Optical Spectroscopy*. Cancer Research, 2017. **77**(14): p. 3942-3950.
63. Heuke, S., et al., *Multimodal nonlinear microscopy of head and neck carcinoma — toward surgery assisting frozen section analysis*. Head & Neck, 2016. **38**(10): p. 1545-1552.
64. Le Naour, F., et al., *In situ chemical composition analysis of cirrhosis by combining synchrotron fourier transform infrared and synchrotron X-ray fluorescence microspectroscopies on the same tissue section*. Anal Chem, 2012. **84**(23): p. 10260-6.
65. Passot, S., et al., *Understanding the cryotolerance of lactic acid bacteria using combined synchrotron infrared and fluorescence microscopies*. Analyst, 2015. **140**(17): p. 5920-8.
66. Davis, E.A., *J.J. Thomson and the discovery of the electron*, J.J. Thomson and I.J. Falconer, Editors. 1997, Taylor & Francis: London ;.
67. Squires, G., *Francis Aston and the mass spectrograph*. Journal of the Chemical Society, Dalton Transactions, 1998(23): p. 3893-3900.
68. Aston, F.W., *LXXIV. A positive ray spectrograph*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1919. **38**(228): p. 707-714.
69. Wolff, M.M. and W.E. Stephens, *A Pulsed Mass Spectrometer with Time Dispersion*. Review of Scientific Instruments, 1953. **24**(8): p. 616-617.
70. Cameron, A.E. and D.F.E. Jr., *An Ion "Velocitron"*. Review of Scientific Instruments, 1948. **19**(9): p. 605-607.
71. Comisarow, M.B. and A.G. Marshall, *Fourier transform ion cyclotron resonance spectroscopy*. Chemical Physics Letters, 1974. **25**(2): p. 282-283.
72. Munson, M.S.B. and F.H. Field, *Chemical Ionization Mass Spectrometry. I. General Introduction*. Journal of the American Chemical Society, 1966. **88**(12): p. 2621-2630.
73. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons*. Anal Chem, 1988. **60**(20): p. 2299-301.
74. Karas, M., et al., *Matrix-assisted ultraviolet laser desorption of non-volatile compounds*. International Journal of Mass Spectrometry and Ion Processes, 1987. **78**: p. 53-68.

75. Tanaka, K., et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry*. Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.
76. Goodwin, R.J., *Sample preparation for mass spectrometry imaging: small mistakes can lead to big consequences*. J Proteomics, 2012. **75**(16): p. 4893-911.
77. Shariatgorji, M., P. Svenningsson, and P.E. Andrén, *Mass Spectrometry Imaging, an Emerging Technology in Neuropsychopharmacology*. Neuropsychopharmacology, 2014. **39**(1): p. 34-49.
78. Fülöp, A., et al., *4-Phenyl- α -cyanocinnamic Acid Amide: Screening for a Negative Ion Matrix for MALDI-MS Imaging of Multiple Lipid Classes*. Analytical Chemistry, 2013. **85**(19): p. 9156-9163.
79. Baluya, D.L., T.J. Garrett, and R.A. Yost, *Automated MALDI matrix deposition method with inkjet printing for imaging mass spectrometry*. Anal Chem, 2007. **79**(17): p. 6862-7.
80. Eberlin, L.S., et al., *Classifying Human Brain Tumors by Lipid Imaging with Mass Spectrometry*. Cancer Research, 2012. **72**(3): p. 645-654.
81. Balog, J., et al., *Intraoperative tissue identification using rapid evaporative ionization mass spectrometry*. Sci Transl Med, 2013. **5**(194): p. 194ra93.
82. McDonnell, L.A. and R.M. Heeren, *Imaging mass spectrometry*. Mass Spectrom Rev, 2007. **26**(4): p. 606-43.
83. Mamyrin, B.A., *Time-of-flight mass spectrometry (concepts, achievements, and prospects)*. International Journal of Mass Spectrometry, 2001. **206**(3): p. 251-266.
84. Liebl, H., *Ion Microprobe Mass Analyzer*. Journal of Applied Physics, 1967. **38**(13): p. 5277-5283.
85. Caprioli, R.M., T.B. Farmer, and J. Gile, *Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS*. Anal Chem, 1997. **69**(23): p. 4751-60.
86. Castellino, S.G., M. R.; Wagner, D., *MALDI imaging mass spectrometry: bridging biology and chemistry in drug development*. Bioanalysis, 2011. **3**(21): p. 2427-2441.
87. Addie, R.D., et al., *Current State and Future Challenges of Mass Spectrometry Imaging for Clinical Research*. Analytical Chemistry, 2015. **87**(13): p. 6426-6433.
88. Longuespee, R., et al., *MALDI mass spectrometry imaging: A cutting-edge tool for fundamental and clinical histopathology*. Proteomics Clin Appl, 2016. **10**(7): p. 701-19.
89. Belu, A.M., et al., *TOF-SIMS characterization and imaging of controlled-release drug delivery systems*. Anal Chem, 2000. **72**(22): p. 5625-38.
90. Giordano, S., et al., *3D Mass Spectrometry Imaging Reveals a Very Heterogeneous Drug Distribution in Tumors*. Scientific Reports, 2016. **6**: p. 37027.
91. Pietrowska, M., et al., *Molecular profiles of thyroid cancer subtypes: Classification based on features of tissue revealed by mass spectrometry imaging*. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2017. **1865**(7): p. 837-845.
92. Everest-Dass, A.V., et al., *N-glycan MALDI Imaging Mass Spectrometry on Formalin-Fixed Paraffin-Embedded Tissue Enables the Delineation of Ovarian Cancer Tissues*. Mol Cell Proteomics, 2016. **15**(9): p. 3003-16.

93. Balluff, B., et al., *MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer*. *Am J Pathol*, 2011. **179**(6): p. 2720-9.
94. Rauser, S., et al., *Classification of HER2 Receptor Status in Breast Cancer Tissues by MALDI Imaging Mass Spectrometry*. *Journal of Proteome Research*, 2010. **9**(4): p. 1854-1863.
95. Aichler, M., et al., *Clinical response to chemotherapy in oesophageal adenocarcinoma patients is linked to defects in mitochondria*. *The Journal of Pathology*, 2013. **230**(4): p. 410-419.
96. Lou, S., et al., *High-grade sarcoma diagnosis and prognosis: Biomarker discovery by mass spectrometry imaging*. *Proteomics*, 2016. **16**(11-12): p. 1802-13.
97. Carreira, R.J., et al., *Large-Scale Mass Spectrometry Imaging Investigation of Consequences of Cortical Spreading Depression in a Transgenic Mouse Model of Migraine*. *Journal of the American Society for Mass Spectrometry*, 2015. **26**(6): p. 853-861.
98. Crecelius, A.C., et al., *Three-dimensional visualization of protein expression in mouse brain structures using imaging mass spectrometry*. *J Am Soc Mass Spectrom*, 2005. **16**(7): p. 1093-9.
99. Lou, S., et al., *Prognostic Metabolite Biomarkers for Soft Tissue Sarcomas Discovered by Mass Spectrometry Imaging*. *Journal of the American Society for Mass Spectrometry*, 2017. **28**(2): p. 376-383.
100. Ermini, L., et al., *Imaging mass spectrometry identifies prognostic ganglioside species in rodent intracranial transplants of glioma and medulloblastoma*. *PLoS ONE*, 2017. **12**(5): p. e0176254.
101. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. *Int J Cancer*, 2015. **136**(5): p. E359-86.
102. Burningham, Z., et al., *The Epidemiology of Sarcoma*. *Clinical Sarcoma Research*, 2012. **2**: p. 14-14.
103. Chetty, R. and S. Serra, *Molecular and morphological correlation in gastrointestinal stromal tumours (GISTs): an update and primer*. *Journal of Clinical Pathology*, 2016. **69**(9): p. 754-760.
104. Casali, P.G. and J.Y. Blay, *Gastrointestinal stromal tumours: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. *Annals of Oncology*, 2010. **21**(suppl_5): p. v98-v102.
105. Bauer, S., et al., *Phase I study of panobinostat and imatinib in patients with treatment-refractory metastatic gastrointestinal stromal tumors*. *Br J Cancer*, 2014. **110**(5): p. 1155-62.
106. Stupp, R., et al., *Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma*. *N Engl J Med*, 2005. **352**(10): p. 987-96.
107. Krex, D., et al., *Long-term survival with glioblastoma multiforme*. *Brain*, 2007. **130**(Pt 10): p. 2596-606.
108. Zong, H., L.F. Parada, and S.J. Baker, *Cell of Origin for Malignant Gliomas and Its Implication in Therapeutic Development*. *Cold Spring Harbor Perspectives in Biology*, 2015. **7**(5): p. a020610.
109. Bleeker, F.E., R.J. Molenaar, and S. Leenstra, *Recent advances in the molecular understanding of glioblastoma*. *Journal of Neuro-Oncology*, 2012. **108**(1): p. 11-27.
110. Vanier, M.T. and G. Millat, *Niemann-Pick disease type C*. *Clin Genet*, 2003. **64**(4): p. 269-81.

111. Vanier, M.T., *Niemann-Pick disease type C*. Orphanet Journal of Rare Diseases, 2010. **5**: p. 16-16.
112. Hughes, G., *On the mean accuracy of statistical pattern recognizers*. IEEE Trans. Inf. Theor., 2006. **14**(1): p. 55-63.
113. Alexandrov, T., *MALDI imaging mass spectrometry: statistical data analysis and current computational challenges*. BMC Bioinformatics, 2012. **13 Suppl 16**: p. S11.
114. Pearson, K., *{On lines and planes of closest fit to systems of points in space}*. Philosophical Magazine, 1901. **2**(6): p. 559-572.
115. Maaten, L.J.P.V.D. and G.E. Hinton, *Visualizing High-Dimensional Data using t-SNE*. Vol. 9. 2008. 2579-2605.
116. Jolliffe, I.T., *Principal Component Analysis*. 2002: Springer.
117. Ravi, D., et al., *Manifold Embedding and Semantic Segmentation for Intraoperative Guidance with Hyperspectral Brain Imaging*. IEEE Transactions on Medical Imaging, 2017. **PP**(99): p. 1-1.
118. Abdelmoula, W.M., et al., *Data-driven identification of prognostic tumor subpopulations using spatially mapped t-SNE of mass spectrometry imaging data*. Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(43): p. 12244-12249.
119. Kullback, S. and R.A. Leibler, *On Information and Sufficiency*. Ann. Math. Statist., 1951. **22**(1): p. 79-86.
120. Bonnel, D., et al., *Multivariate analyses for biomarkers hunting and validation through on-tissue bottom-up or in-source decay in MALDI-MSI: application to prostate cancer*. Anal Bioanal Chem, 2011. **401**(1): p. 149-65.
121. Arthur, D. and S. Vassilvitskii, *k-means++: the advantages of careful seeding*, in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. 2007, Society for Industrial and Applied Mathematics: New Orleans, Louisiana. p. 1027-1035.
122. Jain, A.K., *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters, 2010. **31**(8): p. 651-666.
123. McDonnell, L.A., et al., *Imaging mass spectrometry data reduction: Automated feature identification and extraction*. Journal of the American Society for Mass Spectrometry, 2010. **21**(12): p. 1969-1978.
124. Fawcett, T., *ROC graphs: Notes and practical considerations for researchers*. Machine learning, 2004. **31**(1): p. 1-38.
125. Caliński, T. and J. Harabasz, *A dendrite method for cluster analysis*. Communications in Statistics, 1974. **3**(1): p. 1-27.
126. Erich, K., et al., *Scores for standardization of on-tissue digestion of formalin-fixed paraffin-embedded tissue in MALDI-MS imaging*. Biochim Biophys Acta, 2017. **1865**(7): p. 907-915.
127. Pearson, K., *Note on regression and inheritance in the case of two parents*. Proceedings of the Royal Society of London, 1895. **58**(347-352): p. 240-242.
128. Foskey, M., et al., *Large deformation three-dimensional image registration in image-guided radiation therapy*. Phys Med Biol, 2005. **50**(24): p. 5869-92.
129. Sundaram, T.A. and J.C. Gee, *Towards a model of lung biomechanics: pulmonary kinematics via registration of serial lung images*. Med Image Anal, 2005. **9**(6): p. 524-37.
130. Dice, L.R., *Measures of the Amount of Ecologic Association Between Species*. Ecology, 1945. **26**(3): p. 297-302.

131. Praggastis, M., et al., *A Murine Niemann-Pick C1 I1061T Knock-In Model Recapitulates the Pathological Features of the Most Prevalent Human Disease Allele*. The Journal of Neuroscience, 2015. **35**(21): p. 8091-8106.
132. Yang, J. and R.M. Caprioli, *Matrix sublimation/recrystallization for imaging proteins by mass spectrometry at high spatial resolution*. Anal Chem, 2011. **83**(14): p. 5728-34.
133. Puchtler, H., et al., *Carnoy fixation: practical and theoretical considerations*. Histochemie, 1968. **16**(4): p. 361-71.
134. Pereira, M.A., et al., *Carnoy's solution is an adequate tissue fixative for routine surgical pathology, preserving cell morphology and molecular integrity*. Histopathology, 2015. **66**(3): p. 388-397.
135. Lasch, P., *Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging*. Chemometrics and Intelligent Laboratory Systems, 2012. **117**: p. 100-114.
136. Alexandrov, T. and P. Lasch, *Segmentation of confocal Raman microspectroscopic imaging data using edge-preserving denoising and clustering*. Anal Chem, 2013. **85**(12): p. 5676-83.
137. Grasmair, M., *Locally Adaptive Total Variation Regularization*, in *Scale Space and Variational Methods in Computer Vision: Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings*, X.-C. Tai, et al., Editors. 2009, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 331-342.
138. Peng, J., et al., *Asymmetric least squares for multiple spectra baseline correction*. Anal Chim Acta, 2010. **683**(1): p. 63-8.
139. H C Eilers, P. and H. F M Boelens, *Baseline Correction with Asymmetric Least Squares Smoothing*. 2005.
140. Barnes, R.J., M.S. Dhanoa, and S.J. Lister, *Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra*. Applied Spectroscopy, 1989. **43**(5): p. 772-777.
141. Gibb, S. and K. Strimmer, *MALDIquant: a versatile R package for the analysis of mass spectrometry data*. Bioinformatics, 2012. **28**(17): p. 2270-1.
142. van Herk, M., *A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels*. Pattern Recognition Letters, 1992. **13**(7): p. 517-521.
143. Friedman, J.H., *A variable span smoother*. J.Am.Stat.Assoc., 1984.
144. Alexandrov, T. and J.H. Kobarg, *Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering*. Bioinformatics, 2011. **27**(13): p. i230-i238.
145. Wehbe, K., et al., *The effect of optical substrates on micro-FTIR analysis of single mammalian cells*. Analytical and Bioanalytical Chemistry, 2013. **405**(4): p. 1311-1324.
146. Schwartz, S.A., M.L. Reyzer, and R.M. Caprioli, *Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation*. Journal of Mass Spectrometry, 2003. **38**(7): p. 699-708.
147. Chughtai, K. and R.M.A. Heeren, *Mass Spectrometric Imaging for biomedical tissue analysis*. Chemical reviews, 2010. **110**(5): p. 3237-3277.
148. Chaurand, P., et al., *Integrating Histology and Imaging Mass Spectrometry*. Analytical Chemistry, 2004. **76**(4): p. 1145-1155.
149. Lein, E.S., et al., *Genome-wide atlas of gene expression in the adult mouse brain*. Nature, 2007. **445**(7124): p. 168-76.

150. Ho Kim, J., et al., *Proteome-wide characterization of signalling interactions in the hippocampal CA4/DG subfield of patients with Alzheimer's disease*. *Sci Rep*, 2015. **5**: p. 11138.
151. Araujo, B., et al., *Decreased expression of proteins involved in energy metabolism in the hippocampal granular layer of rats submitted to the pilocarpine epilepsy model*. *Neuroscience Letters*, 2014. **561**(Supplement C): p. 46-51.
152. Ploski, J.E., S.S. Newton, and R.S. Duman, *Electroconvulsive seizure-induced gene expression profile of the hippocampus dentate gyrus granule cell layer*. *Journal of Neurochemistry*, 2006. **99**(4): p. 1122-1132.
153. McCombie, G. and R. Knochenmuss, *Enhanced MALDI Ionization Efficiency at the Metal-Matrix Interface: Practical and Mechanistic Consequences of Sample Thickness and Preparation Method*. *Journal of the American Society for Mass Spectrometry*, 2006. **17**(5): p. 737-745.
154. Srimany, A., et al., *Identification of effective substrates for the direct analysis of lipids from cell lines using desorption electrospray ionization mass spectrometry*. *Rapid Commun Mass Spectrom*, 2015. **29**(4): p. 349-56.
155. Guinan, T.M., et al., *Silver Coating for High-Mass-Accuracy Imaging Mass Spectrometry of Fingerprints on Nanostructured Silicon*. *Analytical Chemistry*, 2015. **87**(22): p. 11195-11202.
156. Thevenaz, P., U.E. Ruttimann, and M. Unser, *A pyramid approach to subpixel registration based on intensity*. *IEEE Trans Image Process*, 1998. **7**(1): p. 27-41.
157. Klein, S., et al., *Preconditioned stochastic gradient descent optimisation for monomodal image registration*. *Med Image Comput Comput Assist Interv*, 2011. **14**(Pt 2): p. 549-56.
158. Woods, A.S., et al., *Gangliosides and Ceramides Change in a Mouse Model of Blast Induced Traumatic Brain Injury*. *ACS Chemical Neuroscience*, 2013. **4**(4): p. 594-600.
159. Deininger, S.O., et al., *MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers*. *J Proteome Res*, 2008. **7**(12): p. 5230-6.
160. Jones, E.A., et al., *Imaging mass spectrometry-based molecular histology differentiates microscopically identical and heterogeneous tumors*. *J Proteome Res*, 2013. **12**(4): p. 1847-55.
161. Balluff, B., et al., *De novo discovery of phenotypic intratumour heterogeneity using imaging mass spectrometry*. *The Journal of Pathology*, 2015. **235**(1): p. 3-13.
162. Freund, Y., R. Schapire, and N. Abe, *A short introduction to boosting*. *Journal-Japanese Society For Artificial Intelligence*, 1999. **14**(771-780): p. 1612.
163. Kaya, I., et al., *Novel Trimodal MALDI Imaging Mass Spectrometry (IMS3) at 10 μ m Reveals Spatial Lipid and Peptide Correlates Implicated in A β Plaque Pathology in Alzheimer's Disease*. *ACS Chemical Neuroscience*, 2017.
164. Dilillo, M., et al., *Mass Spectrometry Imaging, Laser Capture Microdissection, and LC-MS/MS of the Same Tissue Section*. *Journal of Proteome Research*, 2017. **16**(8): p. 2993-3001.
165. Rafols, P., et al., *Signal preprocessing, multivariate analysis and software tools for MA(LDI)-TOF mass spectrometry imaging for biological applications*. *Mass Spectrom Rev*, 2016.

166. Sarkari, S., et al., *Comparison of clustering pipelines for the analysis of mass spectrometry imaging data*. Conf Proc IEEE Eng Med Biol Soc, 2014. **2014**: p. 4771-4.
167. Widlak, P., et al., *Detection of molecular signatures of oral squamous cell carcinoma and normal epithelium – application of a novel methodology for unsupervised segmentation of imaging mass spectrometry data*. Proteomics, 2016. **16**(11-12): p. 1613-1621.
168. Bemis, K.D., et al., *Probabilistic Segmentation of Mass Spectrometry (MS) Images Helps Select Important Ions and Characterize Confidence in the Resulting Segments*. Molecular & Cellular Proteomics : MCP, 2016. **15**(5): p. 1761-1772.
169. Fonville, J.M., et al., *Hyperspectral Visualization of Mass Spectrometry Imaging Data*. Analytical Chemistry, 2013. **85**(3): p. 1415-1423.
170. Walch, A., et al., *MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology*. Histochemistry and Cell Biology, 2008. **130**(3): p. 421-434.
171. Allen, M., et al., *Origin of the U87MG glioma cell line: Good news and bad news*. Science Translational Medicine, 2016. **8**(354): p. 354re3-354re3.
172. Waerzeggers, Y., et al., *Mouse models in neurological disorders: applications of non-invasive imaging*. Biochim Biophys Acta, 2010. **1802**(10): p. 819-39.
173. Abdelmoula, W.M., et al., *Automatic registration of mass spectrometry imaging data sets to the Allen brain atlas*. Anal Chem, 2014. **86**(8): p. 3947-54.
174. Balluff, B., M. Hanselmann, and R.M.A. Heeren, *Chapter Eight - Mass Spectrometry Imaging for the Investigation of Intratumor Heterogeneity*, in *Advances in Cancer Research*, R.R. Drake and L.A. McDonnell, Editors. 2017, Academic Press. p. 201-230.
175. Kunzke, T., et al., *Native glycan fragments detected by MALDI-FT-ICR mass spectrometry imaging impact gastric cancer biology and patient outcome*. Oncotarget, 2017. **8**(40): p. 68012-68025.
176. Inglese, P., et al., *Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer* †Electronic supplementary information (ESI) available. See DOI: 10.1039/c6sc03738k Click here for additional data file. Click here for additional data file. Click here for additional data file. Chemical Science, 2017. **8**(5): p. 3500-3511.
177. Kriegel, H.-P., et al., *Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering*. ACM Trans. Knowl. Discov. Data, 2009. **3**(1): p. 1-58.
178. Parsons, L., E. Haque, and H. Liu, *Subspace clustering for high dimensional data: a review*. SIGKDD Explor. Newsl., 2004. **6**(1): p. 90-105.
179. Krätzschar, A., et al., *Basic Investigations on the Behavior of Advanced Ag/SnO₂ Materials for Contactor Applications*. 2010. 1-7.
180. Xiuqing, Q., et al., *A Novel Method for the Preparation of Ag/SnO₂ Electrical Contact Materials*. Rare Metal Materials and Engineering, 2014. **43**(11): p. 2614-2618.
181. Davies, D.L. and D.W. Bouldin, *A Cluster Separation Measure*. IEEE Trans. Pattern Anal. Mach. Intell., 1979. **1**(2): p. 224-227.
182. Zhong, Q., et al., *Similarity maps and hierarchical clustering for annotating FT-IR spectral images*. BMC Bioinformatics, 2013. **14**: p. 333.
183. Bellisola, G., et al., *Rapid recognition of drug-resistance/sensitivity in leukemic cells by Fourier transform infrared microspectroscopy and unsupervised hierarchical cluster analysis*. Analyst, 2013. **138**(14): p. 3934-45.

184. Dunn, J.C., *A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters*. Journal of Cybernetics, 1973. **3**(3): p. 32-57.
185. Sebiskveradze, D., et al., *Automation of an algorithm based on fuzzy clustering for analyzing tumoral heterogeneity in human skin carcinoma tissue sections*. Lab Invest, 2011. **91**(5): p. 799-811.

9 SUPPLEMENT

Table: GIST Patient metadata including tissue type, necrosis rate, sex, risk factor and mutations

ID	tissue type	Nekrose (%)	Sex	Risk assessment	Mutation
A	Liver	0	F	High	-
AA	GIST DD	0	M	-	cKIT exon9
AB	GIST	50	M	High	cKIT exon 9,11,17 PDGFRA exon 12,14,18
B	Small intestine jejunum	70	M		cKIT exon 13
C	Liver	0	M		cKIT exon 9
D	Stomach / Colon	99	F		cKIT exon 11
E	Thoraxwand	5	M		cKIT exon 11,17 PDGFRA exon 12,14,18
F	Stomach	40	M		cKIT exon 11
G	Liver	-	F		cKIT exon 9,11,13,17 PDGFRA exon 18
H	Liver	30	M		cKIT exon 11
I	GIST	99	F		cKIT exon 11
J	Liver	30	M		cKIT exon 11,17 PDGFRA exon 18
K	Liver	20	M		cKIT exon 11,13
L	Liver	40	F	High	cKIT exon 11
M	Liver	40	M	High	-
N	GIST	5	M		cKIT exon 13
O	Colon	30	F	High	cKIT exon 9,11,13,17
P	GIST	90	M		PDGFRA exon 12
Q	Metastasis	5	M		cKIT exon 11
R	Pelvis peritoneum	0	M		cKIT exon 9,11,13,17
S	Lung	-	M		cKIT exon 11
T	Stomach	99	M		cKIT wt PDGFRA wt
U	Liver	0	M		-
V	Peritoneum	10	F	High	cKIT exon 9,11,13,17
W	Liver	0	M		cKIT exon 9,13,17 PDGFRA exon 18
X	Stomach	20	F	High	cKIT exon 9,13,17,18
Y	Abdomen	30	M	High	cKIT exon 15
Z	Peritoneum	5	M	High	cKIT exon 11,13 PDGFRA exon 18

10 CURRICULUM VITAE

PERSONALIEN

Name und Vorname: Rabe, Jan-Hinrich
Geburtsdatum: 07.08.1988
Geburtsort: Holzminden
Familienstand: ledig
Vater: Dr. Holger Rabe
Mutter: Carola Rabe

AKADEMISCHER WERDEGANG

1998 - 2007 Paul-Gerhardt Gymnasium, Abitur
2008 – 2010 Universität Bayreuth, Biologie BSc
2010 – 2013 Hochschule Mannheim, Biotechnology MSc
2013 – 2017 Intitute for Medical Technology of the University Heidelberg
and the University of Applied Sciences Mannheim,
Dr.sc.hum.

11 ACKNOWLEDGEMENT

Eine wissenschaftliche Arbeit resultiert nie aus dem Schaffen eines Einzelnen. Daher möchte ich mich an dieser Stelle bei all den Menschen, die mich auf meinem langen Weg begleitet haben, ganz herzlich bedanken. Es war eine lehrreiche und spannende Erfahrung, die ich um nichts in der Welt missen möchte.

Der allergrößte Dank gebührt zunächst meinem Doktorvater Prof. Dr. Carsten Hopf, nicht nur für die Möglichkeit, dieses spannende Projekt bearbeiten zu dürfen, sondern auch für seinen Rat und seine übermenschliche Unterstützung in jeder Lebenslage.

Weiterhin möchte ich mich bei meinen Kollegen am Institut für Instrumentelle Analytik und Bioanalytik der Hochschule Mannheim bedanken, denn in den vergangenen knapp fünf Jahren fühlte ich mich stets als Teil einer großen Familie.

Besonderer Dank in dieser Hinsicht gilt Denis Abu-Sammour, der stets gewillt war, konstruktive Gespräche gegen unnützes Deutsch-Wissen einzutauschen. Bei Dr. Sandra Schulz, Dr. Annabelle Fülöp und Dr. Christian Marsching möchte ich mich für ihre Freundschaft und die vielen guten Ratschläge bedanken. Dr. Bogdan Munteanu danke ich für seine Hingabe zur Massenspektrometrie. Ihm gebührt mein gesamtes heutiges Wissen in diesem Bereich. Darüber hinaus möchte ich mich herzlichst bei Dr. Carolina von Reitzenstein für ihre geradezu mütterliche Fürsorge bedanken. Alexander Geisel und Malgosia Muhm danke ich für ihre tatkräftige Unterstützung.

Auch meinen anderen Kollegen, Freunden und Weggefährten, Katrin Erich, Johanna von Gerichten, Julia Neitsch, Nathalie Gerhardt, Karina Ramallo-Guevara, Quin Zhou, Alexander Roth, Matthias Schwartz, David Weight, Peter Reis und Kevin Reinle, bin ich ganz besonders dankbar für eine schöne Zeit.

Prof. Dr. Alexander Marx, Prof. Dr. Peter Hohenberger und Maria Deligiani möchte ich für die Bereitstellung von humanen Geweben und ihre Annotation danken. Außerdem möchte ich mich bei den Projekt-Teams von Dr. Elke Burgermeister und Prof. Dr. Michael Platten für die Vergabe von Mausxenograft-Geweben bedanken, sowie bei Prof. Dr. Daniel S. Ory und seinem Team für ihre Beteiligung in unserem gemeinsamen NPC1-Projekt.

Diese Arbeit wurde finanziert durch das Zentrum für Angewandte Forschung ABIMAS des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg und dem European Community's Seventh Framework Programme MITIGATE, wofür ich mich ebenfalls bedanken möchte.

Zu guter Letzt gebührt mein persönlicher Dank meinen Eltern, die mich immer unterstützt und motiviert haben. Ich bin sicher, dass ich sie mit dieser Arbeit sehr stolz mache. Weiterer großer Dank gebührt meiner Paten-Oma, ohne die ich es wohl niemals geschafft hätte, und natürlich meinem Bruder; dafür, dass er immer für mich da ist. Außerdem möchte ich meiner Freundin Michaela für ihre Liebe und Zuneigung sowie die Korrektur meiner Arbeit danken.