Spatial lipidomics of carotid atherosclerosis

Mirjam Visscher

1 oktober 2021

ISBN: 978-94-6421-455-0

Layout & Cover: M. Visscher

Print: Ipskamp Printing

Spatial Lipidomics of Carotid Atherosclerosis

Lipide samenstelling van carotide atherosclerotische plaques

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam,
op gezag van de rector magnificus

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

vrijdag 1 oktober 2021 om 10:30 uur

door

Mirjam Visscher

geboren te Groningen.

Erasmus University Rotterdam

Promotiecommissie:

Promotoren:	Prof. dr. ir. A.F.W. van der Steen Prof. dr. G. van Soest
Overige leden:	Prof. dr. M.W. Vernooij Prof. dr. R.M.A. Heeren Prof. dr. S. Manohar

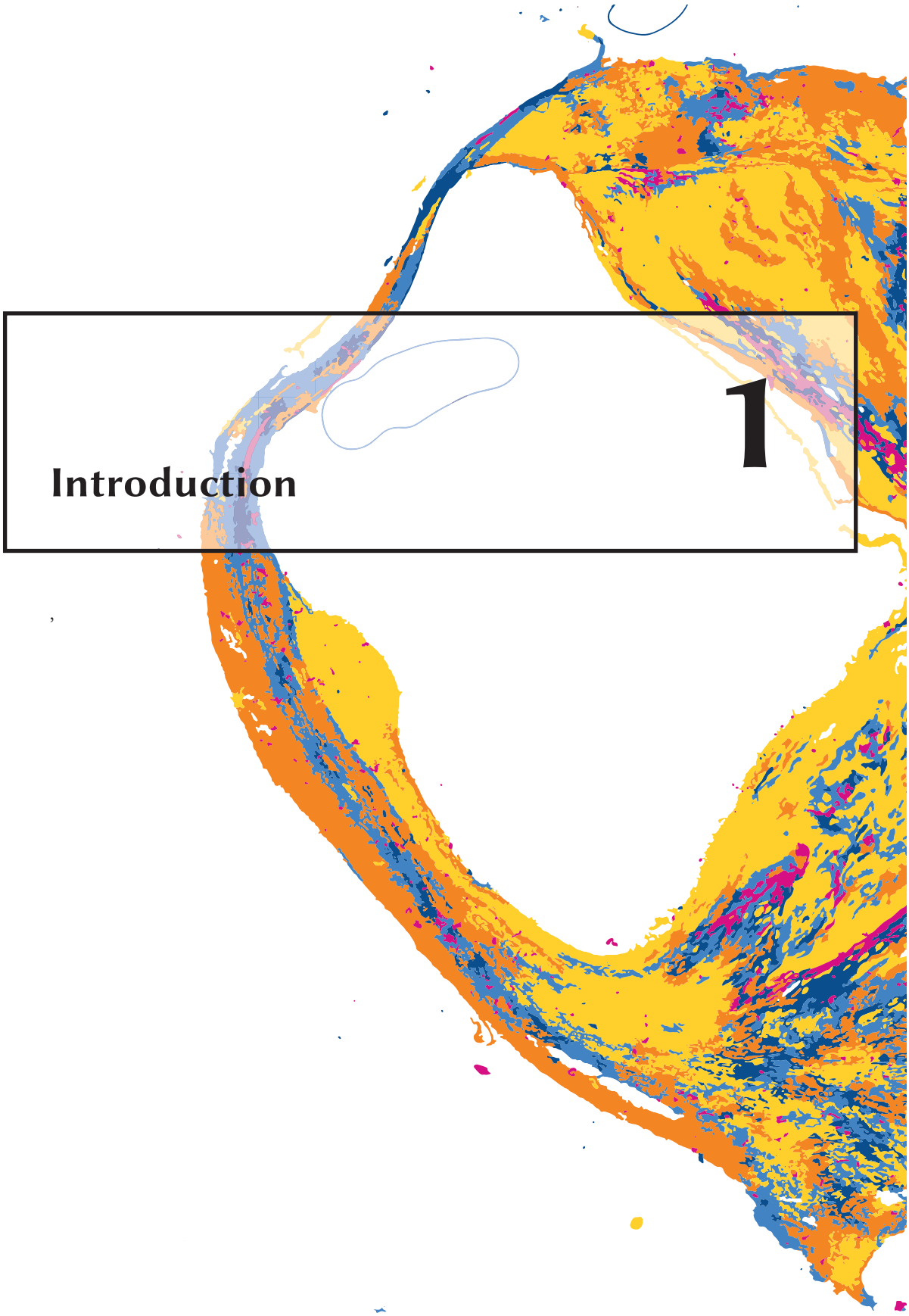
The research described in this thesis has been carried out at the Department of Cardiology, Thorax Center, Erasmus University Medical Center, Rotterdam, the Netherlands.

Financial support for the printing of this thesis was kindly provided by Waters Corporation.

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

Financial support for the printing of this thesis was kindly provided by the Erasmus University Medical Center.

Contents



1 Introduction

1.1 Atherosclerosis

Atherosclerosis is a disease of the arteries. Arteries deliver oxygen-rich blood from the heart to the body. Healthy arteries consist of three layers, the innermost layer is called the tunica intima and consists of a single cell layer of endothelial cells. It is separated by the second layer, the tunica media, through the internal elastic lamina. The media is the middle layer which consists of smooth muscle cells, elastin and collagen and is responsible for the mechanical compliance of the artery. The tunica media is separated from the tunica adventitia, the third and outermost layer, by the external elastic lamina. The adventitia is mainly composed of collagen and elastin, and is a supportive and protective layer of the arteries anchoring the arteries to nearby tissues.

1.1.1 Pathology of atherosclerosis

Atherosclerosis is a systemic disease of the arteries that is initiated by endothelial dysfunction[1–4]. The mechanisms for this are not fully known, but factors that have been proposed are reduced shear stress, turbulent blood flow and oxidative stress[1,5]. The incidence of atherosclerosis is higher at arterial bifurcations, where shear stress is locally reduced and turbulent flow is induced[1,2,5]. The most common sites for atherosclerotic plaques are the carotid bifurcation, often leading to stroke, and the coronary arteries, causing myocardial infractions.

In the initial phase of atherosclerosis, low-density lipoproteins (LDL) accumulate in the sub-endothelial space[5]. There the LDL are oxidized by the oxidative waste of vascular cells, which in turn stimulates the inflammatory process that drives atherosclerosis. The severity of this inflammation is influenced by several factors, including but not limited to, high blood pressure, smoking, presence of other chronic inflammatory diseases, obesity, genetics and diabetes type 2[2,5,6].

The smallest detectable lesions in the vessel wall are xanthomas (or fatty streaks), see Fig. 1.1a, where there is a local accumulation of foam cells in the vessel wall[5]. Foam cells are transformed macrophages after excessive accumulation of cholesteryl esters in the endoplasmic reticulum[5]. However, the majority of advanced atherosclerotic lesions do not originate from fatty streaks but from intimal thickening. In initimal thickening there is an accumulation of smooth muscle cells (SMCs) in the initima[5]. This accumulation becomes pathological thickening when there are some macrophages, small areas of lipid accumulation and proteoglycan-matrix filled with SMCs[5].

These two types of lesions can both develop into more advanced lesions, called fibrous cap atheromas or fibroatheromas, see Fig. 1.1a. These plaques consist of a necrotic core covered by a fibrous cap. The necrotic core consists of necrotic cells surrounded by macrophages, and when it develops further large amounts of extracellular lipids and cholesterol crystals accumulate. The content of this necrotic core is highly thrombogenic. The fibrous cap consists of SMCs and collagen in a proteoglycan-matrix[7]. The atheroma is defined as a thin-cap fibroatheroma (TCFA) if the thickness is $\leq$165 μm in carotid arteries and $\leq$65 μm in coronary arteries. If the cap is thicker, the plaque is defined as a thick-cap fibroatheroma (ThCFA)[5].

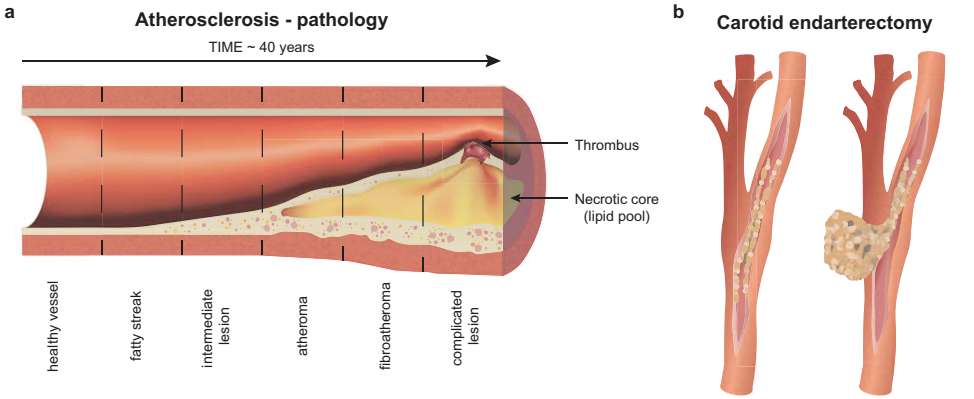


Figure 1.1: **Atherosclerosis pathology and treatment of carotid atherosclerosis. a)** schematic illustrating the development process of atherosclerosis and **b)** illustration of carotid endarterectomy. The procedure with which the atherosclerotic plaque is surgically removed from the arterial wall.

ThFCAs rarely rupture but can erode, causing an acute ischemic event. In absence of the endothelial cells, SMCs and extracellular matrix (ECM) are exposed to the bloodstream. These are thrombogenic, leading to clot formation. Erosion is responsible for about 35% of myocardial infarctions and 10% of strokes and transient ischemic attacks (TIAs)[5].

TCFAs on the other hand are assumed to be prone to rupture, upon which the thrombogenic content is released into the bloodstream[1,5]. Plaques of this type are hypothesized to be responsible for approximately 60% of acute ischemic events in both coronary and carotid atherosclerosis[5]. Over the last years the events initiated by plaque rupture have reduced, possibly because of the use of statins (lipid lowering medication) and a reduction in the tobacco consumption[5].

Coronary versus carotid atherosclerosis

Atherosclerosis is mainly a systemic disease, influenced by factors like blood lipoprotein levels, hypertension, inflammation along with others. Due to these systemic factors atherosclerosis in the carotid arteries is correlated with coronary atherosclerosis and vice versa[5,8]. However, there are some difference between the two which are mostly attributed to the difference in vascular beds[5]. The flow in the carotid artery is greatest during systole whereas the coronary arteries supply the heart during the diastolic phase[5]. This makes carotid atherosclerotic plaques more susceptible for changes in blood pressure[5,8]. Furthermore, the specific geometry of the carotid bifurcation slows down the blood flow, creating lower and non-laminar shear stress on the vessel wall. Additionally, intra-plaque hemorrhage and calcified nodules are more common and more severe in carotid plaque. All these factors explain the difference in critical cap thickness $\leq$165 µm for carotid plaques compared to $\leq$65 µm for coronary plaques[5].

Alternatively, plaques can also develop into fibrocalcific plaques. These plaque do not cause events, but, due to lumen intrusion, can still cause clinical symptoms[5]. Fibrocalcific plaques are characterized by large calcified areas, may contain a necrotic core, but contains only few inflammatory cells[5].

The development of the plaques however, is not a linear process. At first the artery tries to compensate plaque formation by positive remodeling; the vessel grows outward in an attempt to preserve the lumen[1]. However, at some point this is not possible anymore and the plaque will intrude into the lumen causing stenosis[1]. Additionally, over time the thickness of the fibrous cap can increase and decrease, also a TCFA and ThFCA can revert into a fibrocalcific plaque[2,5]. This means high-risk plaques can become stable and vice versa.

1.1.2 Current treatment and diagnostic methods of carotid atherosclerosis

Present-day treatment of atherosclerosis consists of life-style changes (e.g. quit smoking, more exercise and a change of diet)[2–4]. While effective for most, lifestyle changes are challenging for many to maintain. If patients have hyperlipidaemia or comorbidities such as hypertension, prior cardiovascular events or diabetes, life-style changes may not be sufficient. In these cases pharmaceutical treatment will be prescribed. The most common drugs for atherosclerosis patients are statins, these drugs increase the cholesterol uptake by the liver, which in turn increases the HDL production by the liver[2–4]. Patients who do not achieve sufficient cholesterol reduction may be treated with alternative lipid lowering drugs, such as fibrates, cholesterol absorption inhibitors and proprotein convertase subtilisin/kexin type 9 (PCSK-9)[3,4].

In patients with neurovascular symptoms and proven obstructive carotid atherosclerosis, surgical intervention is warranted by the present guidelines. There are two options: 1) carotid endarterectomy (CEA), where the plaque is surgically removed from the vessel wall (Fig. 1.1b), and 2) stenting, where a stent is placed to re-open the artery[9,10]. The latter is mostly performed in the US, whereas CEA is the preferred method of treatment in Europe. Both techniques are proven prevention methods of stroke, nevertheless, the number of procedures required to prevent a single stroke remains high, especially for asymptomatic patients, and the intervention comes with significant risk of complications[9,10].

Clinically, there is a need for, specifically, the identification of high-risk asymptomatic patients that may benefit from interventions. Biomarkers can be used to identify these asymptomatic patients, and classify their risk. A biomarker is defined as 'a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'[11].

Currently used biomarkers in cardiovascular disease (CVD) are; 1) functional assessments, such as blood pressure and stress testing, 2) imaging assessments, such as ultrasound imaging (intravascular, for coronaries, or external, for carotid arteries), CT and MRI-scans[10,12] and 3) biomarkers measured in bodily fluids, such as low-density lipoprotein (LDL) and C-reactive protein (CRP)[1,6,9]. However, these biomarkers are not

specific enough; they do not reflect treatable mechanisms, do not fully reflect disease severity and do not predict long-term prognosis well enough[1,2]. A combination of serum related biomarkers and imaging methods may identify asymptomatic patients to have a high risk of stroke, however, the clinical relevance hereof still has to be proven in large cohorts[9,10].

1.1.3 Lipids in atherosclerosis

Lipids play a major role in the development of atherosclerosis and therefore have potential as biomarkers. The disease is initiated by the infiltration of low density lipoproteins (LDL) into the endothelium. The main purpose of these lipoproteins is the transport of hydrophilic molecules in water, besides LDL there are many types of lipoproteins, e.g. HDL, IDL, VLDL. Methods of lipid assessment in current clinical assessment are mostly aimed at the evaluation of these lipoproteins in plasma. However, these lipoproteins only transport lipids to and from the plaque, but do not give us information on the lipid composition of the plaque itself.

Several lipodomics studies have evaluated the lipid content of plaques[13–16]. Free-cholesterol and cholesteryl esters (CE) species were more abundant in unstable plaque[16], whereas phospholipids and triglycerides were more abundant in stable plaques[15]. Stegemann *et al.* analyzed lipids in stable and unstable regions of the same plaque specimens and found that plaques had higher content of CEs, phosphatidylcholines (PC), lysoPCs and sphingomyelins (SM). Compared to normal vessels, atherosclerotic vessels have more CEs and lyso-phospholipids[14].

Regarding the spatial distribution of lipids in atherosclerosis the current gold-standard is Oil Red O staining, which only stains neutral lipids like triglycerides. Getting more information about the spatial distribution and differentiation of lipids may help to reveal their role in atherosclerosis.

1.2 Mass spectrometry imaging

Mass spectrometry imaging (MSI) is a family of label-free molecular imaging techniques which allows the visualization of molecules through-out a complex tissue. It allows for the imaging of the spatial distribution of lipids, proteins, metabolites and glycans[17]. Imaging experiments are performed on thin tissue sections, the molecules in these tissue sections are ionized and subsequently sent into a mass analyzer which then measures the mass-to-charge ratio (m/z) of the molecules in the sample. By raster scanning the tissue section, mass spectra are collected in every point and an image is acquired. This generates a large dataset, with x and y, defining the locations of the pixel in the image and z, defining the mass spectra in that particular location, see Fig. 1.2.

The various members of the MSI family differ primarily in their ionization mechanism, which impacts their chemical sensitivity, specificity, mass range and spatial resolution. In addition, several types of mass analyzers are employed for imaging experiments.

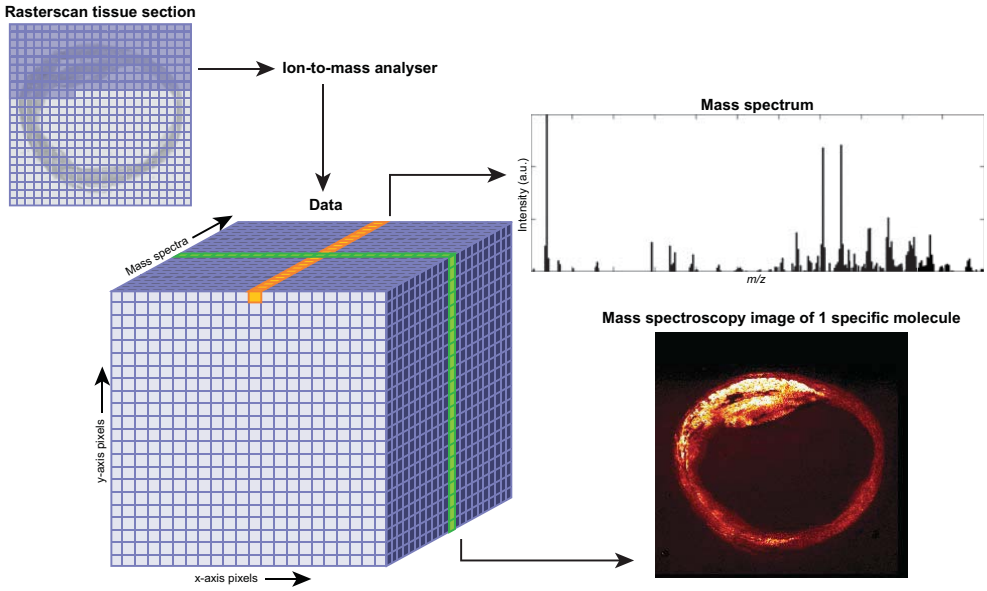


Figure 1.2: **Acquisition process mass spectrometry imaging** illustrating a raster-scan of a tissue section and the organization of the highly dimensional dataset acquired, for every individual pixel contains a spectrum, and by looking at all the pixels at a particular m/z value a mass intensity image can be generated.

1.2.1 Ionization methods

There are different ionization methods with which MSI can be performed. These methods are all 'soft ionization' sources, which minimize fragmentation of the analytes, providing the possibility to analyze complex mixtures of large intact biomolecules[18]. The ionization method used for the work in this thesis is matrix-assisted laser desorption/ionization (MALDI)[19].

MALDI

In matrix-assisted laser desorption/ionization (MALDI) a thin layer of crystals is applied on top of the tissue section, either by spraying or sublimation. This layer is referred to as the matrix and it enhances ionization of the analyte molecules by proton donation or metal ion adduction. Subsequently, a laser is focused on the sample, absorbed by the matrix compound, resulting in desorption of the matrix and analyte molecules. These molecules are then ionized by means of proton transfer to or from the matrix compound, or by ions present within the tissue, see Fig. 1.3. Requirements for a compound to be a matrix are the following; 1) it forms crystals, 2) desorbs under optical irradiation 3) does not evaporate under vacuum, and 4) the ability to aid ionization, for example by proton transfer, or by the matrix itself, in case of silver. Commonly used matrices are 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (CHCA), many other compounds have been investigated as matrices, in attempts to broaden the range of molecules to be desorbed and to improve the spatial resolution[17,18]. Many organic matrices require UV light, at for instance 226, 337 or 355 nm, to efficiently support

the MALDI process. As ions are formed by (de-)protonization or ion adduction, the overwhelming majority of MALDI ions is singly charged. Thus, *z=1* in *m/z* and the spectrum can be directly related to the compound mass in Da.

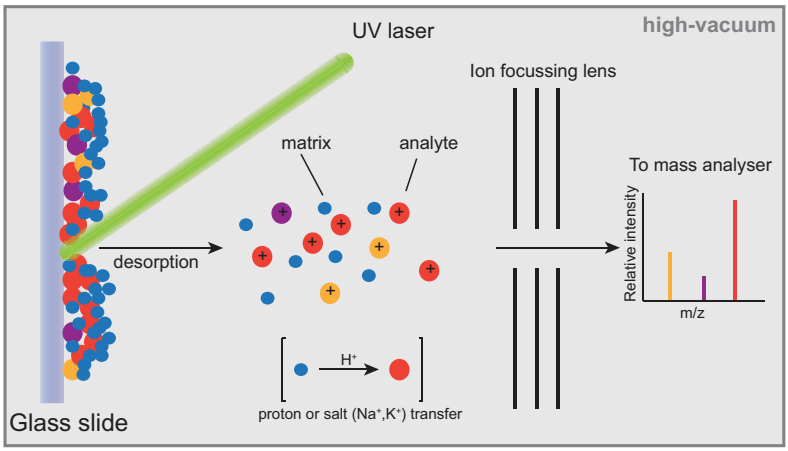


Figure 1.3: **MALDI ionization process** Schematic illustrating the ionization process of matrix-assisted laser desorption/ionization (MALDI), where a UV laser is absorbed by the matrix, causing desorption of both matrix and analytes. Hereafter, protons or salts are transferred to the analytes, resulting in ionization of the analytes. Analytes are typically singly charged, resulting in a 'simple' spectrum, where every analyte is represented by one peak in the spectrum.

Lipid imaging with MSI

Mass spectrometry imaging experiments can be performed in positive and negative mode. The ionization mode influences the type of molecules that can be detected with MSI[20,21]. The choice of matrix also needs to be tuned to different modes.

Positive mode	
Sphingolipids	Mainly sphingomyelins (SM) and ceramides (Cer)
Phospholipids	Mainly phosphatidylcholines (PC) and lysophosphatidylcholines (LPC)
Neutral lipids	Sterol lipids (e.g. cholesterol and cholesteryl esters (CE)) and glycerolipids (e.g. diacylglycerols (DAG) and triacylglycerols (TAG))
Negative mode	
Phospholipids	All other phospholipids (e.g. phosphatidic acids (PA), phosphatidylethanolamines (PE), phosphatidylglycerols (PG) and phosphatidylinositols (PI))
Sphingolipids	E.g. glycosphingolipids and cardiolipins

Application of this matrix compound to the tissue section can be done by spraying or by sublimation. Spraying allows for better extraction of the analytes, but generally has a larger crystal size, causes more diffusion of analyte, and is less homogeneous than sublimation. Sublimation, however, may result in a limited extraction of analyte, lowering the detection limit. This effect can be reduced by re-crystallization of the matrix by vapor of an organic solvent. The crystal size of the matrix is directly related to the spatial resolution of the MALDI imaging experiment, since one laser shot will ablate the entire crystal[21,22]. Typical spatial resolution achieved with MALDI-MSI is 5-50 μm.

Different matrices and application methods thereof may benefit or suppress the detection of different types of molecules[18]. The sample preparation of MALDI is a precise process that needs optimization when changing from one tissue to another or when targeting a different type of analyte on the same tissue[22].

1.2.2 Mass analyzers

The mass analyzer measures the weight to charge ratio (*m/z* value) of the analytes, the are many types, but the ones most used in MSI are Time-of-Flight (TOF), Fourier-transform ion cyclotron resonance (FT-ICR) and Orbitrap mass spectrometer (MS). A TOF MS measures the time it takes for the ions to travel through a field-free region. The ions are collected at the beginning of the time-of-flight tube and then pushed into the tube where they are accelerated by a high voltage at the detector plate. Ions with a higher mass are slower than smaller ions, the time between push and detection allows for mass separation of the analytes and is inversely proportional to the mass-to-charge ratio[18].

Identification of analytes

Mass analyzers measure mass-to-charge ratio of ions, however, the goal of the experiment is to measure molecules. Molecules can arise in mass spectra as different adducts. For lipids the most common ions are $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$ or $[M+H]^-$. However, in some cases the ion may also be a fragment ion of a large lipid for example cholesterol *m/z* 369.3 (loss of H_2O $[M-H_2O+H]^+$). Therefore, it is important to known the nature of the molecules observed. For example, triacylglycerol (TAG) in the form of $[M+H]^+$ is known to decompose rapidly and cannot be detected[20].

Identification of molecules can be performed based on its accurate mass, and matching this to databases of known molecules[20,21]. This method is acceptable for instruments with high mass accuracy, such as FT-ICR or Orbitrap measurements[20,22]. More certainty can be achieved by preforming MS^2 experiments, where a specific *m/z* value is isolated, fragmented, and analyzed by the MS. These fragments are molecule specific and allow reconstruction of molecular structure of the ion[17,20,22].

Both FT-ICR and Orbitrap mass analyzers are ion traps, where the ions are respectively magnetically and electrostatically trapped. In both analyzers the ions traverse an orbit within the ion trap and their oscillation frequency is measured, by the electric signal of ions passing near the detector. This frequency is dependent on its mass-to-charge ratio and can be retrieved by Fourier transformation. In an FT-ICR detector, the ions are trapped in a Penning trap, and are excited by an oscillating electric field orthogonal to the magnetic field[18]. In an Orbitrap the ions oscillate along the long-axis of the spindle-shaped electrode. This type of mass analyzers, both FT-ICR and Orbitrap, have a much higher mass-accuracy than a TOF MS, but are more expensive and slower, which can be a severe limitation for imaging which acquires data from many pixels[18].

1.2.3 Data analysis

Mass spectrometry generates large and complex datasets that are difficult to analyze due to the high degree of dimensionality, since thousands of mass images are generated in one experiment. Mass spectrometry imaging data requires pre-processing which includes; normalization of the data, baseline removal, smoothing, re-calibration and reduction of the data[17,23,24]. These steps are necessary to remove systematic artifacts originating from sample preparation, matrix application, ion suppression and different ionization efficiencies in complex samples. Despite the fact that the field agrees that these analysis steps are necessary for further processing, there is still a lot of discussion regarding which methods are best[23]. For example, MSI data can be normalized to; the total ion current (TIC), mean intensity, median intensity or by normalizing to the intensity of a specific peak in the spectrum, i.e. a peak which is known to be uniformly distributed in the sample or an internal standard[25].

Reduction of the data can be performed through a more targeted approach, e.g. histology driven, or by unsupervised algorithms, such as PCA or machine learning. The goal of all these approaches is to reduce the data to the most important, most discriminative, *m/z* values. Which approach is best is mostly dependent on the hypothesis of the experiment performed.

Ion suppression

In biological samples protein, lipids, metabolites and carbohydrates are present in the same tissue section, making these samples quite complex. These molecules can negatively influence each others ionization and desorption efficiency and therefore limit detection; this effect is called ion suppression[18]. This phenomenon occurs when one analyte is more abundantly present or ionizes more easily than others and results in a non-linear relation between the molecular concentration in the sample and the measured ion count. In lipid imaging phosphatiadylcholine (PC) and sphingomyelin (SM) have a permanently charged headgroup, making them very easy to detect, but often suppressing the detection of other phospholipids in positive mode[20].

Supervised methods

Supervised data analysis methods in mass spectrometry imaging are mostly used in experiments where there are two, or sometimes more, groups of tissue, that can be histologically separated. Another common application is in experiments aimed at detecting the presence and spatial distribution of a specific analyte, e.g. a pharmaceutical compound and metabolites.

Univariate analysis can be performed to compare the relative intensities of a specific *m/z* value between two regions of interest (ROI). Usually MSI data does not have a Gaussian distribution and therefore a non-parametric test like Wilcoxon Signed Rank test is best suited to test the data for statistically significant differences[17].

Nevertheless, in MSI often one *m/z* value does not have enough discriminatory value to explain the difference between two ROIs. Therefore often multivariate analysis (MVA) methods are necessary[17]. These type of algorithms can indicate how much each *m/z* contributes to the difference between groups. One of the methods used for MVA is partial least-squares regression (PLS). PLS is a supervised classification method where data is annotated with labels, for example tumor and normal tissue, based on histological segmentation. It uses linear regression to separate the components and can be combined with discriminant analysis (DA) to separate the two groups based on spectral changes[17].

Unsupervised methods

Unsupervised data analysis in mass spectrometry imaging aim to represent the high-dimensional MSI data using a reduced number of variables, meanwhile minimizing the loss of information. This reduced representation of the data allows for a simpler visualization of the data, giving insights towards, for example, co-located *m/z* values or pixels with similar mass spectra. This type of methods is particularity helpful in biomarker discovery type experiments. Unsupervised data analysis methods can be divided into three subgroups; (1) Factorization, (2) Clustering and (3) Manifold learning[23]. Examples of factorization techniques commonly used with MSI data are principal component analysis (PCA), independent component analysis (ICA), probabilistic latent semantic analysis (pLSA) and non-negative matrix factorization (NMF)[23]. These techniques typically perform a linear decomposition of the high-dimensional MSI datasets, and represent the data in lower-dimensional form. These methods assume a linear relation within the data, however MSI data is inherently non-linear due to ion suppression and other effects.

Clustering techniques group pixels with similar mass spectra, and attributes pixels with the same spectral content to an assigned cluster. This results in a segmentation image of the MSI-imaged target[23]. This approach is mostly used in pathology-directed and clinical applications of MSI. Clustering can also be preformed across the spectral domain where ion images with a similar spatial expression are grouped together, though this is less common than spatial domain clustering. Frequently used techniques are hierarchical clustering (HC) and K-means clustering, which assign one cluster to every pixel, so called 'hard segmentation'[17,23]. 'Soft segmentation' techniques allow for pixels to be in multiple clusters, examples are fuzzy c-means and spatial shrunken centroids[26]; the latter was especially designed for MSI data.

Manifold learning assumes a nonlinear subspace of the data, and is therefore better

at representing a larger portion of the data variation in a lower dimensional representation than linear methods like PCA. Examples of manifold learning techniques applied on MSI data are t-distributed stochastic neighborhood embedding (t-SNE)[27,28] and self-organizing maps (SOMs)[23]. However, these methods do not provide a simple decomposition of the new lower-dimensional representation, and therefore it may be more difficult to determine these. For example, Fonville *et al.*[28] maps the 3D t-SNE to RGB-values in order to give a visual representation of the data. However, t-SNE is a stochastic methods, and thus repeating the analysis will not result in the same color-coding of the MSI data. Efforts have been made to overcome these issues[29–32] however, this has not led to a consensus in the field on how to best deal with this issue.

In general, all data analysis approaches, both supervised and unsupervised, have their own advantages and disadvantages in dealing with MSI data. Which approach suits the data is dependent on the type of experiment, the target of the analysis and the size of the acquired data. A biomarker discovery experiment requires different analysis methods than a targeted pharmaceutical target metabolomics experiment. The field is still evolving and designing new analysis methods and tools.

1.2.4 Mass spectrometry imaging applied on atherosclerosis

Several efforts have been made to apply MSI to image lipids in atherosclerosis. Studies have been performed using different mass spectrometry ionization methods, next to MALDI[33–40], also DESI[41,42] and SIMS[39,43–45]. Cholesterol, CE, SM and LPC lipids are found to be related to atherosclerotic plaques by various researchers[33–43]. PC lipids are found to be less specific, homogeneously distributed throughout the tissue[33], or co-localized with smooth muscle cells[36,37].

Nevertheless MSI of atherosclerotic tissues is still a very unexplored field, which only a few studies which have mainly focused on the optimization of measurements and processing of the data. Therefore, most of these studies either focus on animal atherosclerosis models or have a limited amount of human atherosclerotic samples[21].

1.3 Photoacoustics

In this thesis we aim towards *in vivo* imaging of the lipid biomarkers. Mass spectrometry imaging in itself is not a technique that allows for non-invasive or minimally invasive procedures. Therefore, we aim to combine mass spectrometry imaging with photoacoustic imaging, a technique that allows for *in vivo* molecular imaging.

Photoacoustic (PA) imaging, also referred to as opto-acoustic imaging, is a technique that is been researched for application in a wide variation of diseases. The technique uses a pulsed laser which is locally absorbed by chromophores, causing a thermal expansion of this chromophore, which in turn generates a pressure wave which can be detected using an ultrasound transducer, see Fig. 1.5a and b. PA imaging is predominantly used to image vascularization, e.g. detection of angiogenesis in tumors. However, different chromophores absorb light at different wavelengths, this fact can be exploited to perform spectroscopic PA (sPA), where spectral unmixing is performed to identify chro-

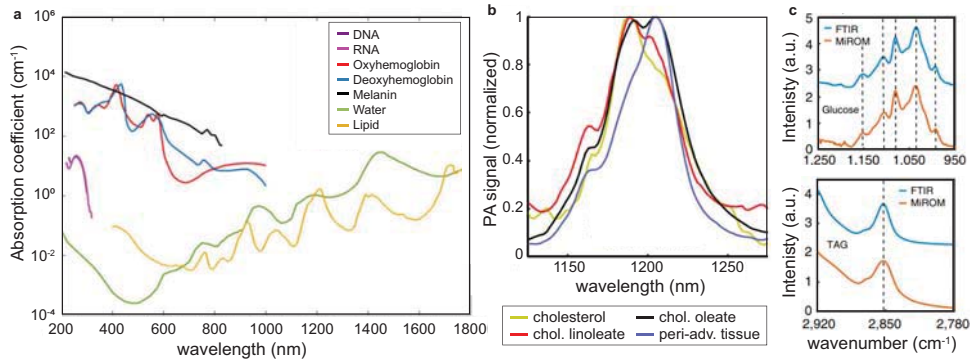


Figure 1.4: **Endogenous tissue chromophores in photoacoustic imaging. a)** Optical absorption spectra of the main tissue chromophores, including DNA, RNA, oxyhaemoglobin, deoxyhaemoglobin, melanin, water and lipid. (Adapted from Zhao *et al.*[46]) **b)** Normalized reference PA spectra of cholesterol, cholesteryl linoleate, cholesteryl oleate, and peri-adventitial tissue, measured on the pure compounds. (Adapted from Jansen *et al.*[47]) **c)** Comparison of mid-infrared photoacoustic and attenuated total reflection Fourier-transform infrared spectroscopy (ATR-FTIR) spectrum of glucose and triacylglycerol (TAG) *in vitro*. The vertical dashed lines indicate good spectral matching between both methods. (Adapted from M.A. Pleitez et al. (2020)[48])

mophores. This technique is used to, for example, differentiate oxygenated blood from non-oxygenated blood.

Traditionally, PA imaging has been performed mainly in the visible and in the near-infrared (NIR) wavelength range, aimed at absorption of chromophores such as hemoglobin (Hb and HbO_2) and melanin, or contrast agents, see Fig. 1.4a. There have been some efforts to expand this range, for example by using UV wavelengths to image cell

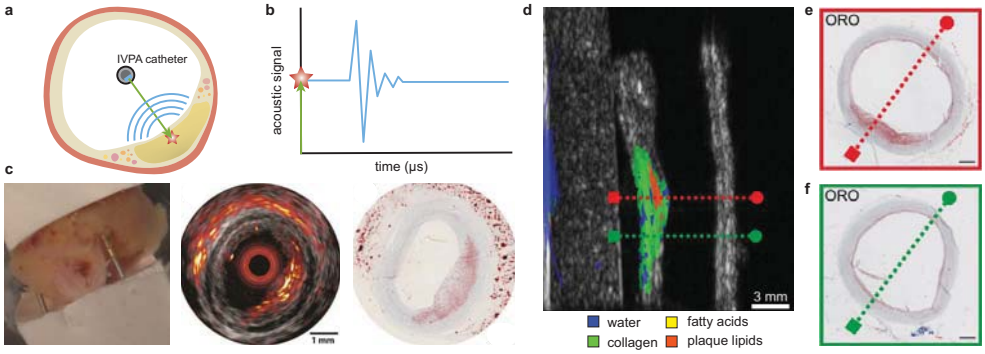


Figure 1.5: **Photoacoustic imaging of lipids in atherosclerosis. a)** IVPA principle, a catheter is inserted inside the vessel and a pulsed laser is fired at wavelengths ranging from 1130-1270 nm, the light is absorbed by the lipids in de vessel wall and consequently an acoustic wave is generated which can be detected by the catheter. **b)** Photoacoustic signal as received after laser excitation. (**a and b** Adapted from Jansen *et al.*[49]) **c)** intravascular photoacoustics (IVPA) image of an *ex vivo* human coronary artery and an ORO stain of the artery (Adapted from Wu *et al.*[50]) **d)** PA imaging of a lipid plaque in a carotid artery *ex vivo*, tissue composition (water, collagen, fatty acids, and plaque lipids) identified by PA spectroscopy are color coded and overlaid on the US image. Plaque lipid accumulation inside the collagen structure of the artery wall can be observed. **e)** Oil red O stained section of red dashed line, indicating presence of lipids. **f)** Oil red O stained section of green dashed line, indicating absence of lipids. (**d-f** Adapted from Kruizinga *et al.*[51])

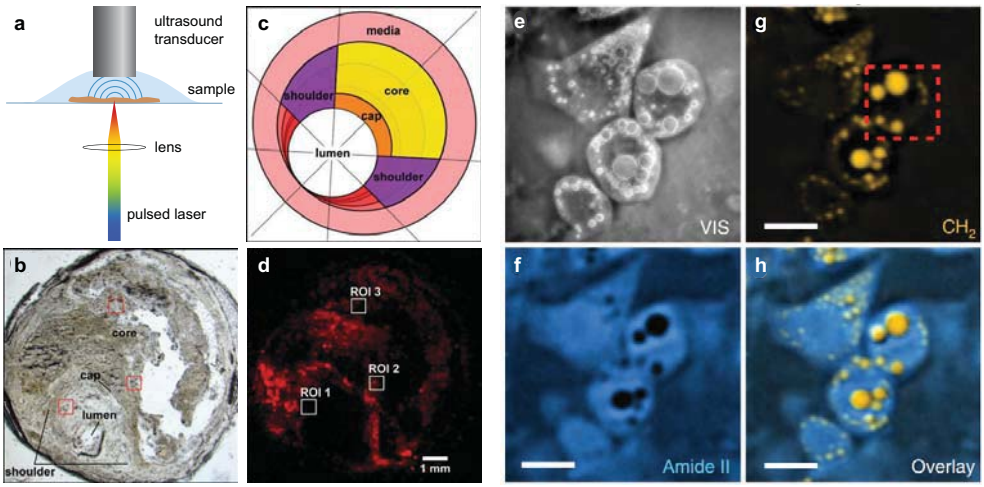


Figure 1.6: **Optical resolution photoacoustic microscopy (OR-PAM). a)** Transmission-mode OR-PAM system, where the ultrasound transducer and the focusing lens are on opposite sides of the object **b-d)** OR-PAM image of human carotid atheroma. **b)** Bright field observation of the unstained atheroma sample used for OR-PAM imaging and identification of lumen, cap, and shoulders. **c)** Schematic depiction of a typical atherosclerotic vascular cross-section. **d)** Coarse OR-PAM scan indicating a progressed intraplaque red blood cell embedding, images at 515 nm. (**b-d** Adapted from Seeger *et al.*[59]) **e)** Brightfield visible (VIS) image of adipocytes (fat cells). **f–h)** Mid-infrared optoacoustic microscope (MIROM) images of the cells with protein contrast **f)** at 1,550 cm^{-1} (amide II) and endogenous lipid contrast **g)** at 2,857 cm^{-1} (CH_2 vibration). **h)** Overlay of lipid and protein maps. (**e-h** Adapted from Pleitez *et al.*[48])

nuclei[52] (Fig. 1.4a), or by looking at vibrational modes of the atomic bonds using NIR[47,53,54] (Fig. 1.4b) and mid-infrared (mid-IR) wavelengths[48] (Fig. 1.4c).

The latter is particularly interesting when looking at atherosclerosis, since it has the potential to spectrally differentiate between specific lipids[47], see Fig. 1.4b. Most of the work in this topic so far has been towards intravascular photoacoustic imaging (IVPA) through catheters, focusing on visualizing the lipid content of atherosclerotic plaques[55], see Fig. 1.5a, b and c. With respect to carotid plaque imaging using PA, studies have been performed investigating the feasibility of imaging without the use of catheters[51,56–58]. Aiming at the visualization of blood-flow in the carotid artery[56,57], detection of hemorrhage[58] and lipid detection[51], see Fig. 1.5d, e and f.

1.3.1 Optical resolution photoacoustic microscopy

In the context of this thesis we aim to exploit the data and biomarkers found using mass spectrometry imaging to validate photoacoustic imaging of lipids. For this goal it is necessary to match the spatial resolution of the two imaging modalities, therefore we make use of optical resolution photoacoustic microscopy (OR-PAM).

OR-PAM makes use of a focused light beam illuminating the tissue, this ensures only local generation of photoacoustic signal. The light is raster-scanned over the region of interest and recorded by a (un)focused single element ultrasound transducer, see Fig. 1.6 a. The lateral resolution of OR-PAM is determined by the diffraction limited optical focus

and the axial resolution is determined by the frequency response of the US transducer[46].

OR-PAM was used to image a 10μm thick atherosclerotic carotid tissue section[59]. Seeger *et al.*[59] used a hybrid OR-PAM and multiphoton microscopy to look at collagen and elastin (using multiphoton) and red blood cells (using OR-PAM at 532nm), Fig. 1.6 b,c and d, showing good correlation with histological stainings. Lipids were also imaged by OR-PAM, this was performed by imaging in de mid-infrared range. Pleitez *et al.*[48] used the symmetric CH_2 vibrational mode at 2857 cm^{-1} to image the lipid content of live adipocytes.

1.4 Thesis aims and outline

In this thesis we aim to unravel the role of lipids in atherosclerosis related to the plaque, and the potential to evaluate stroke-risk based on the lipid distribution *in vivo*. We do this using two imaging techniques, MALDI mass spectrometry imaging and photoacoustic microscopy.

In **Chapter 2** we develope a mass spectrometry imaging pipeline with which we aim to set a framework for the analysis of carotid endarterectomy samples. To this end we optimize the tissue preparation and imaging methods by systematically analyzing data from three specimens: two human carotid endarterectomy samples (advanced plaque) and one autopsy sample (early stage plaque). We show a robust data reduction method and evaluate the variability of the endarterectomy samples.

Using this pipeline we evaluate spatial lipids distribution of twelve carotid endarterectomy samples in **Chapter 3**. We use unsupervised clustering methods and a histology driven multi-variate analysis. Additionally, we compare our MSI data to histological segmentation, and correlate lipid distribution of morphological plaque features that have been associated with enhanced risk of symptoms.

We further investigate the effect of several unsupervised clustering techniques of mass spectrometry imaging data in **Chapter 4**. We compare the clustering of (1) non-negative matrix factorization (NMF), (2) spatial shrunken centroids (SSC), developed for MSI data, (3) auto-encoder and t-SNE to histological segmentation. We also look at the combined effect of an auto-encoder and t-SNE, where first the dimensionality of the data is reduced by means of an auto-encoder and consequent t-SNE analysis.

In **Chapter 5** we use near-infrared photoacoustic microscopy to evaluate the presence of lipids carotid endarterectomies. We match spectral photoacoustic microscopy to MALDI-MSI and validation using pure lipids, correlating photoacoustic spectra to high-risk plaques. Additionally, we use photoacoustic microscopy to evaluate the feasibility of *in vivo* assessment of lipids in the carotid artery. In **Chapter 6** we move from the near-infrared to the mid-infrared, where the vibrational bands are more diverse enabling, imaging of lipids, proteins and carbohydrates. We validate the findings with several histological stainings.

Chapter 7 contains a general discussion of the work presented in this thesis and recommendations for future research.

2

Data processing pipeline for lipid profiling of carotid atherosclerotic plaque with mass spectrometry imaging

This chapter is based on:
Mirjam Visscher*, Astrid M. Moerman*, Peter C. Burgers, Heleen M.M. Van Beusekom, Theo M. Luider, Hence J. M. Verhagen, Antonius F. W. Van der Steen, Kim Van der Heiden and Gijs Van Soest: Data processing pipeline for lipid profiling of carotid atherosclerotic plaque with mass spectrometry imaging, *Journal of the American Society of Mass Spectrometry*, July 2019
*these authors contributed equally

Abstract

Atherosclerosis is a lipid and inflammation-driven disease of the arteries that is characterized by gradual buildup of plaques in the vascular wall. A so-called vulnerable plaque, consisting of a lipid-rich necrotic core contained by a thin fibrous cap, may rupture and trigger thrombus formation, which can lead to ischemia in the heart (heart attack) or in the brain (stroke). In this study, we present a protocol to investigate the lipid composition of advanced human carotid plaques using matrix-assisted laser desorption ionization (MALDI) mass spectrometry imaging (MSI), providing a framework that should enable the discrimination of vulnerable from stable plaques based on lipid composition. We optimized the tissue preparation and imaging methods by systematically analyzing data from three specimens: two human carotid endarterectomy samples (advanced plaque) and one autopsy sample (early stage plaque). We show a robust data reduction method and evaluate the variability of the endarterectomy samples. We found diacylglycerols to be more abundant in a thrombotic area compared to other plaque areas and could distinguish advanced plaque from early stage plaque based on cholesteryl ester composition. We plan to use this systematic approach to analyze a larger dataset of carotid atherosclerotic plaques.

2.1 Introduction

Atherosclerosis is the cause of more than 10 million deaths worldwide, and a major contributor to long-term morbidity and disability[60]. This disease of the artery walls is characterized by the gradual formation of depositions in the vascular wall that are called atherosclerotic plaques. Plaque growth is initiated by endothelial dysfunction and driven by impaired lipid transport mechanisms, which trigger an inflammatory response. Lipids thus play a key role in the pathogenesis of atherosclerosis. Besides a variety of lipid species, these plaques contain calcifications, fibrous tissue, smooth muscle cells and necrotic material[7,61]. Depending on the tissue composition and structure of the plaque, plaques can be either 'stable' or 'vulnerable', the latter means they are prone to rupture. Plaque rupture may trigger embolization of fragments of plaque tissue in the bloodstream, which, depending on the anatomical location of the plaque, can lead to stroke or myocardial infarction[7,62]. The mechanisms driving the evolution of a stable plaque to a vulnerable phenotype are not known in detail, although it is clear that defective or unbalanced lipid transport plays a key role[15,63–66]. The recent LRP study (presented at TCT2018[67]) demonstrated that coronary vessel segments with lipid-rich plaque have a >4 times higher clinical event rate than vessel segments without such plaque, highlighting the importance of plaque lipid content for triggering ischemia. We aim to study whether the lipid profile differs between stable and vulnerable plaque types. To determine whether a plaque's lipid profile can be correlated to plaque phenotype and, therefore, risk of rupture, variation between and within plaque samples needs to be assessed. In this study we present a mass spectrometry imaging (MSI) method for systematically studying plaque lipid distribution at various stages of development. MSI is an emerging experimental technique in the molecular characterization of biological samples. Using MSI, the molecular composition of thin tissue sections can be imaged at a micron-scale resolution. As an initial step, MSI requires ionization of the tissue section. Different methods of ionization exist, including matrix-assisted laser desorption ionization (MALDI), a commonly used ionization method for biological samples[68]. The main advantages of MALDI are its ability to generate singly charged ions, which greatly simplifies analysis of mass spectra, and its ability of so-called 'soft' ionization, preventing molecules from fragmenting during the ionization process[68]. MALDI-MSI is suitable for the detection of small molecules, such as lipids.

To date, few studies of (MALDI-)MSI on human arteries have been published[41,43,69,70] [34,36,37]. They explored the parameter space of atherosclerotic plaque composition as sampled by MSI along various dimensions. Prior work examined different arterial beds, i.e. aorta[34], carotid[37,41,43], coronary arteries[70], femoral or popliteal[36,69], and in the ionization method used, i.e. MALDI[34,36,37,69], secondary ion mass spectrometry (SIMS)[43,70] or desorption electrospray ionization (DESI)[41]. Clinically relevant plaques, those that are likely to precipitate ischemic events, contain a variable mix of lipids[15]. In studying the role of various lipids[64,71] in the atherosclerotic disease process, it is therefore important to optimize the analysis for accurately sampling the lipid spectrum, and to quantitatively understand the variance in measured ion abundances. We consequently chose to focus on lipid-rich plaques only, and moreover restricted ourselves to carotid plaques, as these are a major precursor of ischemic stroke[72].

Our aim is to develop a MALDI-MSI pipeline for the analysis of lipids in human carotid artery plaques, suitable for assessing the variation in lipid profile, both between different plaque specimens and within different locations in the same specimen. To that aim, we add the following insights to the existing MSI knowledge: (1) We describe how arterial tissue should be processed, such that the combination of MALDI-MSI and histological information enables contextualization of MSI data in terms of the known morphological features of plaque pathology. (2) We combine various MSI data analysis methods into a robust data processing pipeline that reliably reduces an MSI dataset into manageable lipid-specific results. We describe this pipeline in detail to encourage other researchers to follow similar pipelines in their own research. (3) We assess the reproducibility of measurements by calculating the coefficient of variation of MALDI-MSI measurement results of tissue sections that were located close together.

In addition, we discuss tentative observations in three human carotid plaques, that illustrate the potential for imaging lipid biochemistry in atherosclerosis. This method is readily applicable to larger collections, limited by scan time, of carotid plaque specimens, allowing to assess the relation between the carotid plaque's molecular lipid composition and plaque type using statistical methods.

2.2 Materials and experimental methods

2.2.1 Materials

DPBS (without Calcium and Magnesium), KP Haematoxyline mayer (VWRK4085-9002), Resorcin (1075930100), SuperFrostPlus glass slides (76 x 26 mm) were purchased from VWR (The Netherlands). Eosine (HT110232-1L), Oil Red O (00625-25G), Basic Fuchsin (857343-25G), Ferric(III) Chloride (157740-100G), 2,5-dihydroxybenzoic acid (98%, 149357-100G) and porcine type A gelatin were purchased from Sigma-Aldrich (The Netherlands). HPLC grade acetone was purchased from Fisher Scientific (The Netherlands). Mouse anti-human CD31 (JC/70A, ab9498) and, mouse anti-human CD68 (KP1, ab955) were purchased from Abcam (United Kingdom). EnVision+Kit (HRP, mouse, 1100 tests, 110 ml, K4001) was purchased from Agilent (the Netherlands). Lipid standards Cholest-5-en-3ß-yl heptadecanoate (CE 17:0) 1mg/mL (700186M-1mg) and 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine (PC 17:0) 10 mg/mL (850360C-25mg) were purchased from Avanti Polar Lipids (USA).

2.2.2 Tissue collection

Carotid endarterectomy (CEA) surgery was performed by carefully opening the outer wall of the carotid artery and excising the plaque by separating the intima and media. This protocol preserves an intact lumen and plaque morphology[73]. Two human carotid plaque specimens, harvested at carotid endarterectomy (CEA) surgery, were rinsed in phosphate-buffered saline (DPBS), snap frozen and stored at -80 °C until further processing. A third carotid artery specimen, without apparent atherosclerotic disease, was collected at autopsy within 24 hours postmortem, from a subject with a non-cardiovascular cause of death. This specimen was rinsed in DPBS and stored in the same manner.

Tissue collection was performed according to protocols sanctioned by the Ethics Board at Erasmus MC (MEC 2008-147 for carotid endarterectomy samples, MEC 2017-300 for the autopsy sample).

2.2.3 Sample preparation

One CEA specimen, referred to as P1, was transversally divided into 1 mm segments to study intraplaque variability. We hypothesize that the plaque's molecular lipid composition may change along the length of the vessel. Five 1 mm segments, originating ≈3mm apart, evenly spaced along the axial length of the CEA sample, were used for imaging. Three of the segments were located caudal to the carotid bifurcation and two segments originated from the internal carotid artery, see Fig. 2.1. The second CEA specimen, referred to as P2, and the autopsy specimen, referred to as A, were imaged in a single axial location. The imaged 1 mm segment of P2 was located cranial to the bifurcation, in the internal carotid artery, where atherosclerotic disease stage is generally maximal. The imaged 1 mm segment of A originated from the common carotid artery, where no or minimal plaque formation is expected.

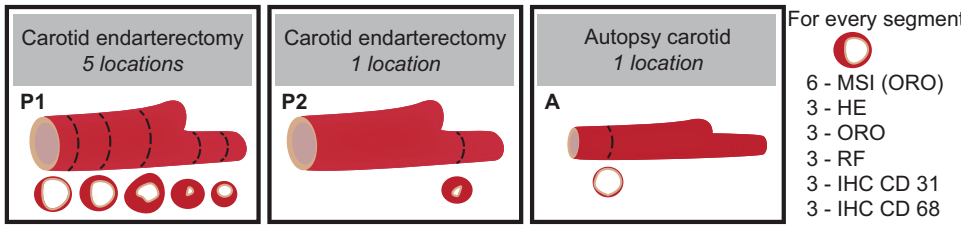


Figure 2.1: **Overview tissue samples.** Five segments from carotid endarterectomy specimen 1 (P1), one segment from carotid endarterectomy specimen 2 (P2) and, one segment from the autopsy carotid (A). Number of MALDI-MSI measurements and adjacent histological stains

All 1 mm segments were embedded in 10% porcine type A gelatin and cryo-sectioned (CM3050 S, Leica Biosystems (cutting temps: OT -21 °C; CT -19 °C)), into 10µm sections which were then thaw mounted onto non-conductive glass slides. After sectioning, the slides were stored at -80 °C until imaging or histological staining. Per segment, 6 sections were imaged; the distance between MALDI-MSI sections was 30-50 µm. All sections were imaged or stained within 2 months after sectioning.

Prior to imaging, the slides were desiccated under vacuum at room temperature for 10 minutes. The matrix, 2,5-dihydroxybenzoic acid (DHB), was applied using sublimation (home-built sublimation system as described in Dekker et al.[74]) at 125 °C for 10 minutes; 50 mg of DHB was dissolved in 5 ml acetone. No washing, re-crystallization, or addition of ionization agents was used.

2.2.4 Histological staining

Histological staining was performed on adjacent slides, which were histochemically stained by Hematoxylin-Eosin (HE), Oil Red O (ORO) and Resorcin-Fuchsin (RF) staining, to visualize general structure, lipids, and collagen and elastin, respectively. Immuno-

histochemical (IHC) staining was performed to visualize endothelial cells and macrophages. The slides imaged with MALDI-MSI were stained by ORO after MSI.

We assessed the plaque tissue according to the plaque classification scheme of Virmani et al.[7,61] and examined the tissue for features of plaque vulnerability (i.e. necrotic core, calcification, thrombus). We also assessed the tissue's structural variability over consecutive tissue sections that originate from the same 1mm segment. We used the histopathological data as a guidance to interpret the MALDI-MSI data.

2.2.5 Data acquisition

MALDI-MSI was performed on a MALDI-q-TOF mass spectrometer (Waters Synapt G2Si-TOF, Manchester, UK), operated in positive mode, at the instrument's "resolution mode" (single-pass reflectron TOF) with a spatial resolution of 50 μm. An Nd:YAG (355 nm) laser was operated at 2,000 Hz, at 50% of the maximum laser energy available (250/500 laser energy). At this energy level, the laser spot was ≈40 μm in size. The mass range was 300-1,200 *m/z* and the scan time per pixel was 0.5 seconds using a left-right scanning pattern. The data was acquired using MassLynx v4.1 software (Waters, Manchester, UK) and HDimaging v1.4 software was used to export the data to imzML format. Prior to imaging, the machine was calibrated with red phosphorus spotted onto a steel plate.

2.2.6 Mass identification

For the identification of the various lipid components, lipids were extracted[75] by homogenizing 5 mg of tissue in 250 μL of Milli-Q water. The tissue was taken between sections P1-4 and P1-5. Ten μL homogenized sample was combined with 10 μL 10% TEA solution (trimethylamine (10/90, v/v) in methanol/dichloromethane (DCM) (50/50, v/v)) and mixed thoroughly. Subsequently, 450 μL methanol/DCM (50/50, v/v) was added to this mixture. Samples were incubated on the roller bank for 30 minutes at 4 °C. After incubation, samples were centrifuged at 14,000 rpm for 20 minutes at 4 °C. Supernatant was transferred to a glass vial, and freeze dried until further use.

The extract was subjected to high resolution MS for accurate mass data and MS/MS experiments for structural information, as described in Angel et. al (2016)[76]. Ten μL of the extract was mixed with 190 μL of a methanol/chloroform (1/1, v/v) mixture and infused at a rate of 240 μL/hr. In a second set of experiments, ammonium acetate was added to efficiently detect the cholesteryl esters (CE), a 10 μL of the extract was mixed with 180 μL of a methanol/chloroform (1/1, v/v) and 10 μL of a 10 mM ammonium acetate solution was added; the cholesteryl esters appear as NH_4^+ adducts[77].

High resolution mass data were obtained with an LTQ Orbitrap XL mass spectrometer with an electrospray ion source operated in the positive ion mode using the following settings: source voltage 4 kV and capillary temperature 300 °C. In such experiments, the analyte molecules appear as $[M+H]^+$ or as $[M+NH_4]^+$ species, whereas in MALDI imaging experiments they are usually cationized by residual Na^+ and/or K^+. Peaks within 2 ppm of the expected *m/z* values were taken into consideration for identification by MS/MS experiments.

For lipids such as the phosphatidylcholines (PC) and sphingomyelins (SM), product ions in MS/MS spectra (for example *m/z* 184.1) may lie below the cut-off mass of the ion trap (27% of precursor mass). To circumvent such discrimination effects and to reliably identify such species, MS/MS spectra were recorded on a triple quadrupole mass spectrometer, in our case an AB Sciex API-3000 mass spectrometer operated in the positive ion mode, using the following settings: source temperature 200 °C, capillary voltage 5.5 kV and collision energy 20 V.

Many of the ionized lipids undergo a common dissociation reaction giving a common product ion, for example *m/z* 369.3 for the CE and *m/z* 184.1 for PC and SM. In such cases a precursor ion scan (i.e. a scan of MS1 at fixed MS2) allows the identification of the CE, PC and SM precursor ions. Thus, both PC and SM components are identified by setting MS2 at *m/z* 184.1, and then scanning MS1, whereas the CE molecules are identified by selecting MS2 at *m/z* 369.1, followed by an MS1 scan. Such experiments were performed to identify all CE components using cholesteryl heptadecanoate (CE 17:0) as internal standard as described by Liebisch et. al (2006)[77]. The identified peaks were matched with the internally calibrated MSI data (Fig. 2.3) and identification was deemed positive if the masses agreed to within 2 ppm. Identified PCs, SMs and CEs are listed in Tables 2.1 and 2.2.

Table 2.1: Identification of phosphatidylcholines (PC) and sphingomyelines (SM) by high resolution mass spectrometry (orbitrap) and MS/MS (triple quadrupole)

a. Experimental mass (Orbitrap) for protonated species
a. Calculated mass (http://www.lipidmaps.org/tools/structuredrawing/GP masscalc.php) for protonated species or for Na^+ species
a. Difference calculated and experimental mass (in ppm)
a. Identification by exact mass and MS/MS data (triple quadrupole). Numbers refer to chain lengths and number of unsaturations (e.g. SM 16 18 1 represents a sphingomyelin with one C16 and one C18 chain and one double bond; position double bond(s) not determined)

Experimental mass[a]	Calculated mass $[M+H]^{+}$[b]	Delta (ppm)[c]	Identification by MS^2 [d]
369.35175	369.35156 $[M-H_2O+H]^+$	-0.5	cholesterol
703.57471	703.57487	0.2	SM 16 18 1
725.55659	725.55922	3.6	Identified as $[M+Na]^+$
731.60597	731.60617	0.3	SM 18 18 1
734.56912	734.56938	0.4	PC 16 16 0
758.56916	758.5694	0.3	PC 16 18 2
760.58483	760.58504	0.3	PC 16 18 1
780.55191	780.55373	2.3	Identified as $[M+Na]^+$
782.56889	782.56938	0.6	PC 18 18 4
784.58494	784.58503	0.1	PC 18 18 3
786.60048	786.60068	0.3	PC 18 18 2
788.61586	788.61633	0.6	PC 18 18 1
810.60037	810.60068	0.4	PC 18 20 4
813.68395	813.68437	0.5	SM 18 24 2
Experimental mass[a]	Calculated mass $[M+Na]^{+}$[b]	Delta (ppm)[c]	Identification by MS^2 [d]
725.55659	725.55922	0.3	SM 16 18 1
780.55191	780.55373	-0.7	PC 16 18 2

Table 2.2: Identification of cholesteryl esters (CE) by triple quadrupole (precursor ion scan)
a. Experimental mass (triple quadrupole)
a. Calculated mass (http://www.lipidmaps.org/tools/structuredrawing/GP masscalc.php) for NH_4^+ adduct
a. Peak present at mz 369.3 in precursor ion scan

Experimental mass[a]	Calculated mass $[M+NH_4]^{+}$[b]	Identification by MS2 [c]
614.6	614.5870	14:0
628.6	628.6026	15:0
640.6	640.6027	16:1
642.6	642.6183	16:0
654.6	654.6183	17:1
656.6	656.6340	17:0
666.6	666.6183	18:2
668.7	668.6340	18:1
682.7	682.6497	19:1
684.7	684.6653	19:0
696.7	696.6653	20:1
698.7	698.6809	20:0
716.5	716.6340	22:5

2.3 Development of data processing pipeline

MSI data in the form of imzML files were loaded into MATLAB 2017a (Natwick, USA) unbinned, resulting in dataset with 20,000 to 40,000 pixels per section, depending on the imaged area, and 79,413 mass bins. Data processing steps that were used to reduce this complex data set and retain the lipid spectrum exclusively are described below.

2.3.1 Pre-processing

Spectral processing imaging data

For each image pixel, the spectrum was smoothed using a second order Savitzky-Golay filter with a frame length of nine data points. After smoothing, a post-measurement calibration was performed using DHB cluster peaks as lock-mass and a linear fit, see Fig. 2.3. In Fig. 2.2 a representative mean spectrum of calibrated MALDI-MSI data is shown.

Peak-picking

We perform peak-picking to make a pre-selection of the most important features of the data, for which we used the base peak (BP) spectrum because of the heterogeneous nature of our samples. The BP spectrum shows for each *m/z* value its maximum intensity measured over the entire MSI dataset[78]. Thus, the BP spectrum allows for detection of high intensity lipids that are only present in a small region. After calibration the BP spectrum was computed for the entire image and exported to mMass v5.5.0[79]. We performed peak-picking in mMass by: (1) smoothing (Savitzky-Golay filter: window size

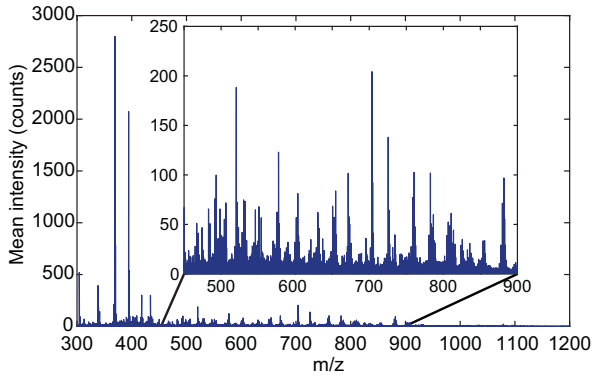


Figure 2.2: **Typical MALDI-MSI spectrum.** P1-4 calibrated, mean spectrum, non-normalized

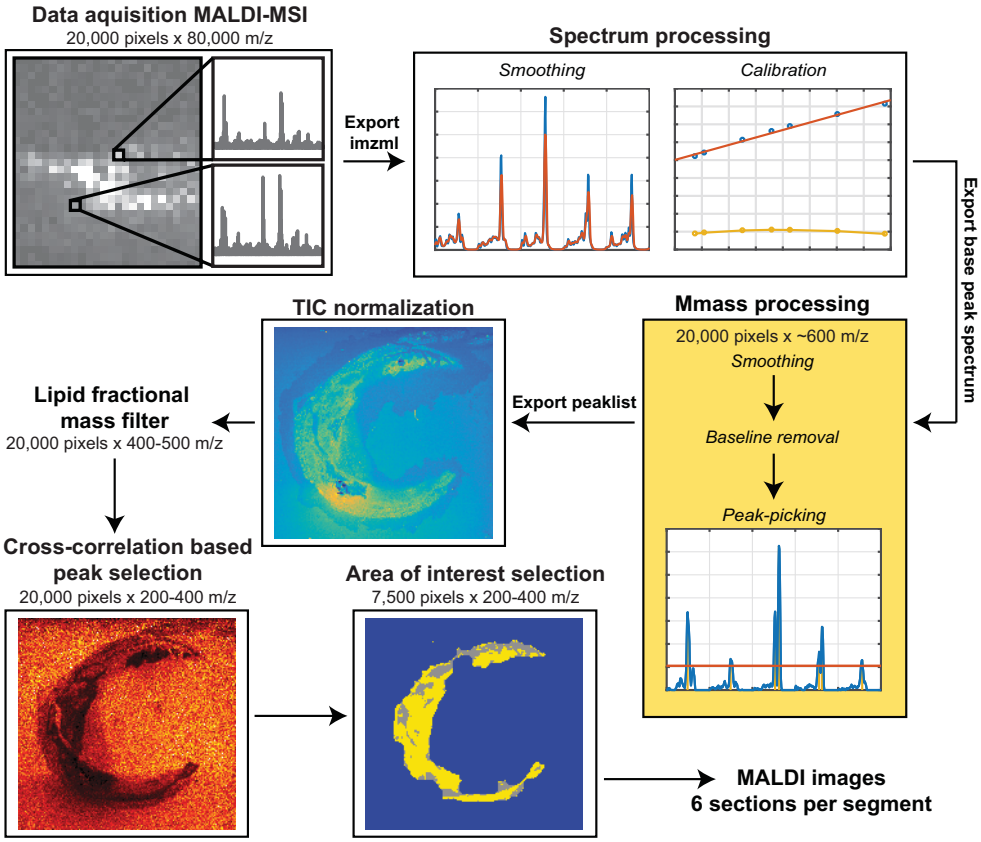


Figure 2.3: **Flowchart of MSI data processing**, showing the consecutive data processing steps of the MSI data. The number of m/z data points is reduced from 80,000 to 200-400 by peak picking, elimination of non-lipid peaks, and cross-correlation-based background filtering. Also the number of informative pixels is reduced from 20,000 to 7,500.

0.1 *m/z* and 2 cycles), (2) baseline correction (precision 15, relative offset 25), and (3) peak-picking (S/N threshold 10.0, peak-picking height 80%), see Fig. 2.3. The filter settings used in this process were chosen heuristically: the additional peaks that are selected with less restrictive parameter settings are found to be removed in later processing steps. This supports the internal consistency of the work flow we propose here. The peak list was exported as a text file.

2.3.2 Imaging data processing

TIC-normalization and peak matching

The calibrated imzML measurement and the peak list were imported into MATLAB™ (The MathWorks, Natick, MA, USA). The data was normalized using Total-Ion Current (TIC) normalization, see Fig. 2.3. The peak list exported from mMass was used to make a pre-selection and to that aim, the peaks of the mMass-derived peak list were matched to the MSI data. The peak list was generated in mMass because the mMass peak-picking algorithm is better suited to process mass spectrometry data than the MATLAB findpeaks algorithm. For instance, in mMass the noise level is calculated locally and adjusted for every *m/z* value, which results in fewer false-positive peaks compared to the findpeaks algorithm of MATLAB.

Because the data processing pipeline was developed in MATLAB, the peaks of the mMass-derived peak list were matched to the MSI data imported in MATLAB. Using the findpeaks algorithm of MATLAB™ the peaks in the mean TIC normalized spectrum of the image were detected. Subsequently, these peaks were matched to the peaks found in the mMass BP spectrum with a mass tolerance of ± 0.02 *m/z*; this resulted in ≈ 600 peaks. The data was subsequently binned around these peaks, with a bin width of ± 1 data point. The sample pitch of the TOF mass analyzer used varied with mass-to-charge ratio, increasing from ± 0.0075 *m/z* at 300 *m/z* to ± 0.015 *m/z* at 1,200 *m/z*.

Lipid fractional mass filter

At this point, the dataset contains only the 600 most intense *m/z* peaks, which includes the lipids of interest as well as background and other non-lipid ions. To eliminate these unwanted signals, we applied several data reduction steps, aiming to maintain only lipid-representing *m/z* values. First, we applied a lipid fractional mass filter[78,80] to remove non-lipid peaks from the dataset. Masses that have a fractional mass outside the lipid fractional mass boundaries cannot be lipids and are therefore discarded, see Fig. 2.4.

Cross-correlation-based peak selection

Subsequently, we performed peak selection based on background peaks to remove non-biological *m/z* values, similar to Fonville et al.[81]. In this method the correlation of images with a background reference peak was used to find matrix-related peaks and embedding medium-related peaks. Only masses that correlated negatively with this background reference mass were retained. The reference mass was chosen upon visual inspection; for all samples we used 419.2 *m/z* as the background reference mass, see Fig. 2.3 step 5.

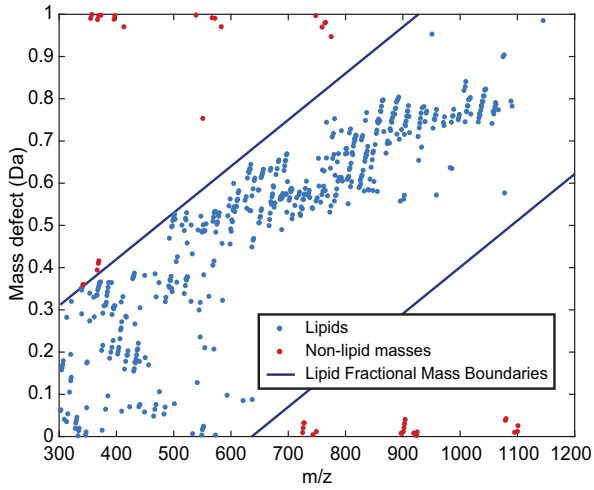


Figure 2.4: **Lipid fractional mass filter** masses outside the lipid fractional mass boundaries cannot be lipids

Features that are important for choosing a good reference mass are: (1) high intensity in background area, (2) low intensity in tissue area and, (3) smooth transitions between the two areas.

We use the background reference mass and compute the cross-correlation using the Pearson correlation coefficient;

$$R^2 = \frac{\sum((x_i - \overline{x})(y_i - \overline{y}))}{\sigma_x \sigma_y} \tag{2.1}$$

We then took the 9 highest correlations to this reference mass and used these to calculate the reference to all masses in the spectrum. We then calculated the mean correlation of all these 10 references. If this was negative a mass would be marked as being non-background and retained. This step reduced the number of *m/z* values to 200-400 peaks.

For sections with highly calcified regions, as were present in P1, an extra processing step was necessary due to noise allocated to the presence of these regions. This additional step is further explained in the supplementary information, available in the online version of this paper.

Area of interest selection

After elimination of the background and non-lipid peaks, we used a K-means clustering algorithm with N=2 clusters to identify lipid-rich tissue areas, see Fig. 2.5a. Small gaps in the data were filled by first dilating the positive structures in the image with a six pixel circular kernel (MATLAB™ function imdilate), see Fig. 2.5b, and subsequent removal with the same kernel (imerode), resulting in a continuous region of interest for lipid analysis, see Fig. 2.5c.

This selection of pixels containing lipid-rich tissue was additionally used to remove the last background peaks, by discarding masses that were more intense in the back-

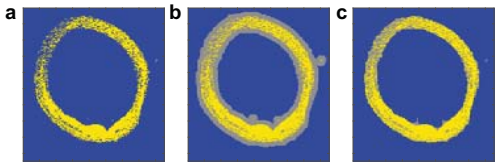


Figure 2.5: **Pixel classification. a)** Pixel selection based on K-means clustering (yellow) (N=2), **b)** Dilated pixel selection to ensure continuity (gray), and **c)** Final continuous pixel selection.

ground pixels than in the lipid-rich tissue pixels. The combination of data processing steps reduced the datasets to 3,000-14,000 pixels, depending on section size, and 200-400 *m/z* values per measurement.

2.3.3 Reproducibility

With the data processing pipeline presented here, we intend to compare lipid profiles along the length of the carotid artery within a patient and between different patients. Therefore, we need to estimate the variability between experiments and the biological variation within the samples.

The coefficient of variation (CV) is a standardized measure for the variability in relation to the mean of the population, and was previously applied in MSI by Tillner et al.[82,83]. It is defined as the ratio of the standard deviation of a population(σ) to the mean of a population (μ): CV=$\sigma/\mu\times$ 100, expressed as a percentage. CV is invariant to the size of the population which makes it applicable to our relatively small dataset of six sections per segment[84]. Also, CV is applicable to data with a wide variation in mean values, which is the case for our MSI dataset. The CV represents population variability: the lower the CV in a set of repeated experiments, the higher is the reproducibility of the MSI measurement.

The quantitative reproducibility of these MSI datasets depends on the repeatability of the measurement, which includes instrument parameters but also the stochastic ionization process and efficiency of transfer into the analyzer, the heterogeneity in the sample as prepared for imaging, and variability between the sections within one segment. We assessed the baseline reproducibility of the instrument by spotting two lipid standards on top of homogenized carotid artery tissue. 2 µL of ≈1000 pmol/µL PC 17:0/17:0 and 2 µL of ≈1560 pmol/µL CE 17:0 were spotted onto the homogenate. Both lipid standards were spotted and measured six times with the instrument settings reported in the methods section and from that data we calculated the CV value.

After data reduction, the tissue image dataset contains 200-400 *m/z* values, however, this data selection still contains isotopes. When calculating the CV, we used the de-isotoped *m/z* values (≈160) to avoid bias of similar isotope peaks. Every segment contains six tissue sections of which we calculate the mean intensity of the lipid-rich pixels for every *m/z* value. Resulting in 6 $\times$ 160 mean intensities, making up 160 populations for which we calculate the CV. We then calculated the mean CV value for every dataset ($\overline{CV}$) to allow comparison between the different segments.

We used the histopathology images to verify if all 6 sections within one segment were the same from a morphological perspective.

2.4 Results

2.4.1 Reproducibility

We found that the CV of the measurement of the lipid standard PC(17:0/17:0), is 10% for $[M+H]^+$, 8% for $[M+Na]^+$ and 16% for $[M+K]^+$. For cholesteryl esters we only looked at $[M+Na]^+$, since it is the only ion that is stable in time of flight[85], and we found a CV of 16%. The relative intensity of the PC $[M+H]^+$ peak is about 100 times higher than that of PC $[M+K]^+$ and CE $[M+Na]^+$, which have about the same relative intensity. This means that PC $[M+H]^+$ is less susceptible to the influence of noise on the signal, explaining the lower CV value.

In the imaging data, tissue structure and content was not apparently different in the series of histologically stained sections within 1 mm segments. We thus assume that the CV values per segment represent MALDI-MSI measurement variability and not biological variability.

Figure 2.6 shows the CV distribution of all measured segments. P2 shows the lowest $\overline{CV}$ of 12%, also P1-1, P1-2, P1-4 and P1-5 show relatively limited variability ranging from 14 - 25% for $\overline{CV}$. Furthermore, P1-1, P1-2 and P1-4 show some *m/z* values with a higher CV; most of these peaks could not be identified. P1-3 and A show significantly higher $\overline{CV}$ values of 44% and 40%, respectively. Computing the total variability across all peaks in all datasets without matching the segments, results in a $\overline{CV}$ of 54%.

Furthermore, we expected to find that $\overline{CV}$ would be lower when only comparing adjacent sections (30-50µm apart), since biological variation between these sections is relatively minimal. Also we hypothesized that two sections that were measured on the same glass slide, and thus have had the exact same sample preparation and storing conditions, would show a lower $\overline{CV}$. However, we did not find these conditions to influence $\overline{CV}$.

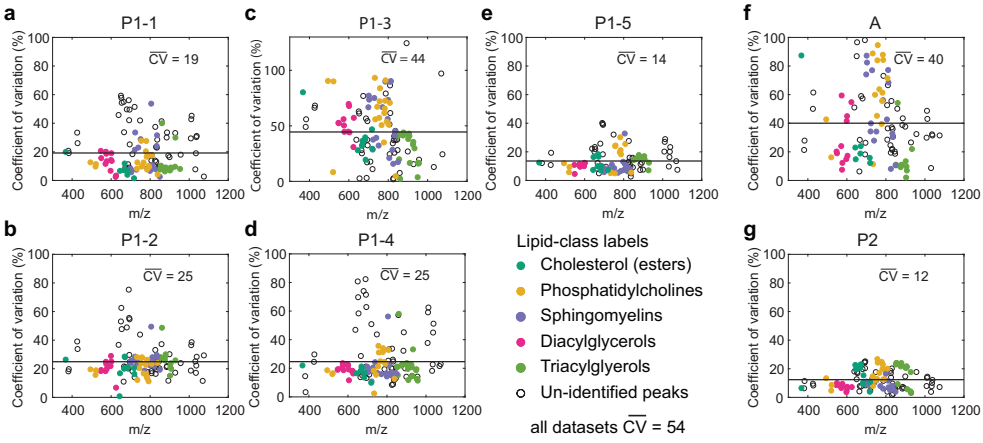


Figure 2.6: **Coefficient of variation calculated using the mean intensity of every *m/z* value and comparing these values over all six measurements per segment.** The mean CV value across the *m/z* values is given, and depicted as a horizontal line. The different assigned lipid classes have been clustered and are color coded. **a)** P1-1, **b)** P1-2, **c)** P1-3, **d)** P1-4, **e)** P1-5, **f)** A, and **g)** P2.

2.4.2 Relation between lipid content and plaque phenotype - preliminary observations

We hypothesize that the lipid composition of a plaque reflects plaque phenotype, and therefore expect to find specific lipid signatures in phenotypically different plaque structures. In our dataset we could examine thrombus, a key component in ischemic events, and changes in lipid profile occurring with development stage.

Thrombus

Segment P1-4 contained thrombus, see Fig. 2.7c, in which monoacylglycerols (MAG), diacylglycerols (DAG) and, triacylglycerols (TAG) were abundantly present. When we compared 300×300 μm regions of interest (ROIs) in tissue segments, we found the DAGs to be more abundant in the thrombus region of P1-4 compared to another region in the same segment and also compared to other segments and other tissue samples. As such, a high concentration of DAGs appears to be thrombus-specific. In Fig. 2.7 an example of the distribution and intensity of DAG 34:2 is shown, in comparison to the other tissue sections measured. In Table 2.3, a list of all peaks assigned as DAGs with distributions similar to DAG 34:2 are given.

Table 2.3: Diacylglycerols (DAG) associated with thrombus area

m/z $[M\text{-}H_2O+H]^+$	Peak assignment
549.49	DAG 32:1
551.50	DAG 32:0
573.49	DAG 34:3
575.50	DAG 34:2
577.52	DAG 34:1
579.53	DAG 34:0
599.50	DAG 36:4
601.52	DAG 36:3
603.53	DAG 36:2
605.55	DAG 36:1
607.57	DAG 36:0
623.50	DAG 38:6
625.52	DAG 38:5
627.53	DAG 38:4
629.55	DAG 38:3

Plaque development stage

The autopsy sample (sample A) contained a small fatty streak, which is an early manifestation of atherogenic changes in the vessel wall, as can be seen on the histological staining Fig. 2.8e, f and g. This sample was marked as an early stage plaque whereas the two CEA samples, P1 and P2, were advanced atherosclerotic plaques. We compared the differences in abundance of prominent lipid peaks between early- and late stage disease,

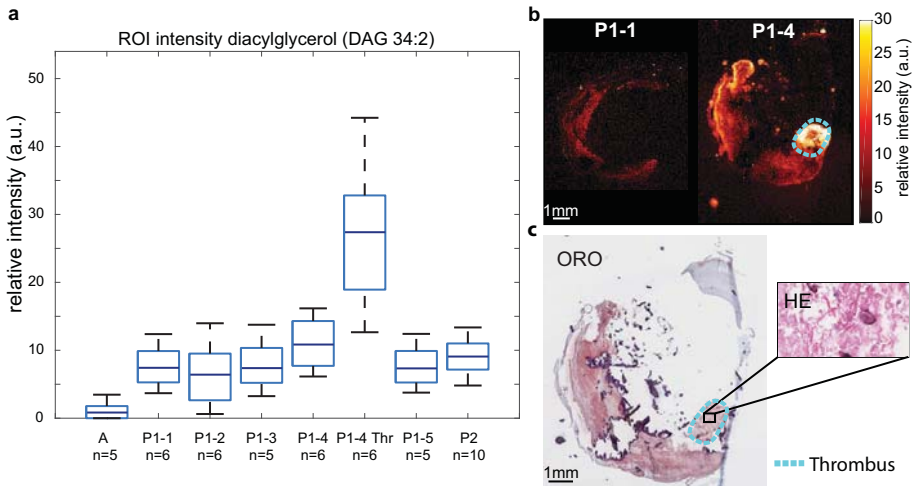


Figure 2.7: **Distribution of diacylglycerol (DAG) 34:2 (*m/z* 575.51). a)** Boxplot of average intensity of diacylglycerol (DAG 34:2, $[M+H\text{-}H_2O]^+$) in a 300×300 µm region of interest in tissue segments. Expression in the thrombus-containing ROI of segment P1-4 (P1-4 Thr) is highly elevated, **b)** MALDI MSI of P1 segment 1 and 4. The thrombus region is outlined in cyan **c)** ORO staining of MALDI section and HE staining of adjacent section indicating thrombus.

and found that the early lesion is significantly enhanced in cholesteryl oleate, whereas the advanced plaques exhibits higher concentrations of cholesteryl linoleate, see Fig. 2.8a and Fig. 2.8c.

2.5 Discussion

In this study, we composed a framework to investigate the lipid content of human carotid atherosclerotic plaques. Arterial tissue was systematically processed for both MSI and histology, ensuring that MSI findings could be contextualized in terms of histopathology. Also, the data-processing pipeline presented here, can effectively and reproducibly assess the complex MALDI-MSI data set in a way that focuses on the lipid distribution in the relevant tissue locations.

Our protocol differs from previous MALDI-MSI studies on human atherosclerotic tissue[34,36,37,69] in the following ways. To our knowledge, this is the first study in which multiple carotid artery plaque specimens have been measured by MALDI-MSI. Based on this collection of carotid artery measurements, we optimized our data processing protocol. We have observed that the settings for discriminating tissue signal from background signal by negative cross-correlation, are dependent on the composition of the tissue investigated, in particular the presence of calcifications. We expect that this data-processing step may need to be adjusted when investigating other arterial beds, the method allows these adjustments. Secondly, this is the first MSI study on arterial tissue that allows measuring both differences between- and within-specimen lipid content and determines intra and inter variations in the tissue. Finally, we quantified the reproducibility of the samples.

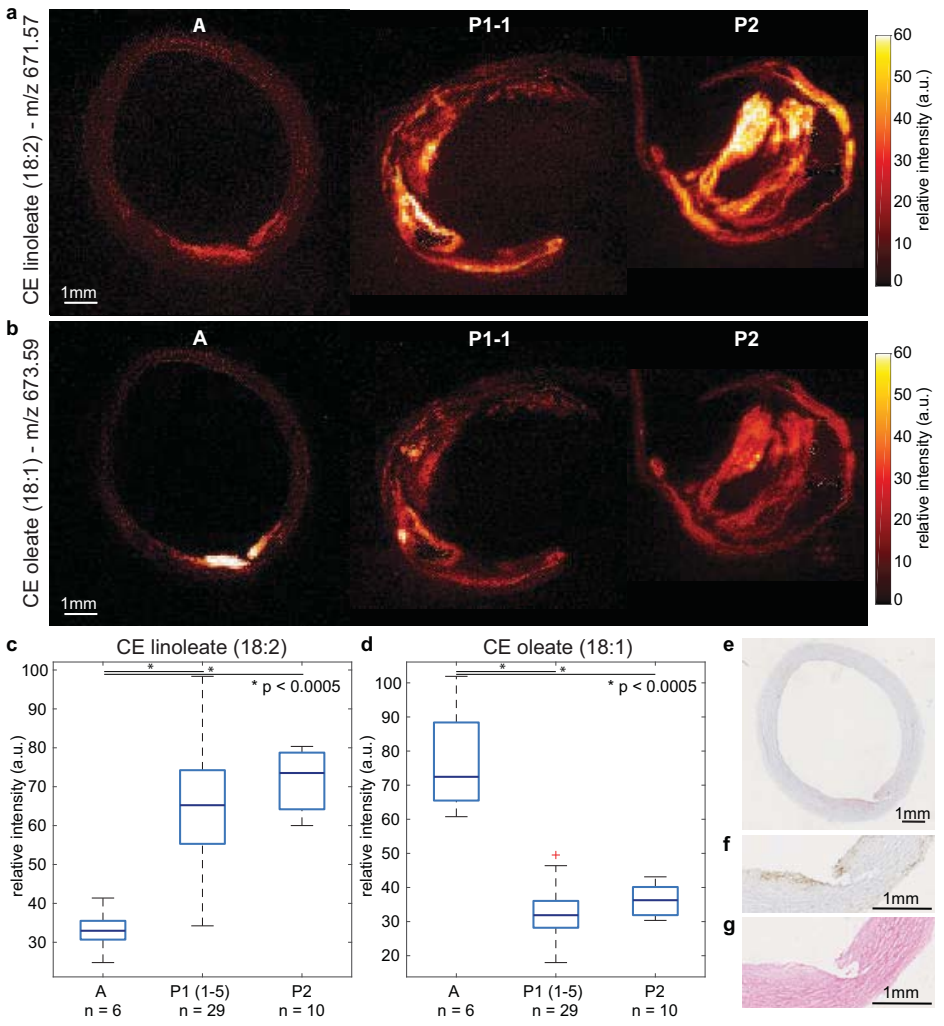


Figure 2.8: **Distribution of cholesteryl linoleate (*m/z* 671.57) and cholesteryl oleate (*m/z* 673.59). a)** MALDI MSI of CE 18:2 in autopsy sample, P1-1 and, P2, **b)** MALDI MSI of CE 18:1 in autopsy sample, P1-1 and, P2, **c)** Box plot of maximum intensity (99th percentile) cholesteryl linoleate (CE 18:2), with over-expression in advanced plaques (P1 and P2), *p-value<0.0005, **d)** Box plot of maximum intensity (99th percentile) cholesteryl oleate (CE 18:1), with over-expression in early plaque (Autopsy), *p-value<0.0005, **e)** Oil Red O staining of MALDI MSI section, showing fatty streak, **f)** Immunohistochemistry CD68 stain (zoom, adjacent section), showing macrophages in fatty streak **g)** HE stain (zoom, adjacent section), showing fatty streak.

We designed the data processing pipeline to be easily adjustable for other MALDI-MSI research purposes. We processed the data to the standard imzML format, which is supported by all MS instruments. Moreover, the processing pipeline is generated in mMass and MATLAB, which are both software packages in common use and easily available. The design of our processing pipeline makes the method applicable and adjustable for research in other labs.

We quantified the variability of the MALDI-MSI measurements by determining the coefficient of variation for the 6 tissue sections measured per segment. $\overline{CV}$ varied between 12-44%, with an average value of 25%. Two segments, P1-3 and A, showed relatively larger variability which may be attributed to different factors. In the case of segment P1-3, there were large calcifications present in the sample. This influenced the quality and amount of tissue per glass slide, which can be seen from the K-means pixel segmentation of these sections shown in the supplementary information, available in the online version of this paper. These large variations in sections most likely contribute to an elevated CV.

In the case of segment A, this tissue is biologically different from P1 and P2, because it also contains the outer vessel wall layers, i.e. the tunica media and tunica adventitia. These layers are typically not as lipid-rich as the intimal layer. Therefore, the MALDI-MSI signal is much lower in these areas, resulting in a lower mean intensity value. When the mean values in the population are smaller, the CV becomes more sensitive to relatively small changes. Also, differences in sample collection may have had an – unknown – impact. These results indicate that, in sections of lipid-rich atherosclerotic plaque, this protocol allows us to achieve a quantitative accuracy of approximately 12-20% for abundant lipids, which sets a lower limit for the detection of individual, local, or longitudinal variations in lipid profile.

Our analysis of three carotid plaque samples indicates that abundance of specific subtypes of cholesteryl esters may be related to atherosclerotic disease stage. Several studies have associated the presence of cholesteryl oleate (*m/z* 673.59) with early stage plaques[86–91]. Our MALDI-MSI data also showed significantly higher cholesteryl oleate intensity in the early plaque of sample A, compared to the advanced plaques in the CEA tissue specimens. In addition, cholesteryl linoleate (*m/z* 671.57) has been shown to be more abundant in advanced plaques, compared to early stage plaques[92]. While our limited number of samples does not allow us to draw quantitative conclusions, strikingly, we find cholesteryl oleate-to-linoleate ratios that are in close agreement with previous literature[93].

In our dataset, we also observed relatively high signal of DAG species in thrombus tissue. With the current dataset we cannot rule out the possibility that these DAG molecules represent fragmented TAGs. However, relative intensities of MALDI-MSI data of DAG molecules were different for thrombus compared to other segments, whereas their presumed TAG precursor molecules did not show this trend, indicating that we might be observing original DAG molecules, not TAG fragments. DAGs have at least two roles in biology, first, as an intermediate molecule for the synthesis of TAG and phospholipids, and second, as a regulator of protein kinase C (PKC) isozymes, which modulate several cellular processes[94]. Accumulation of DAGs has been associated with diabetes, cancer and cardiovascular disease[95]. PKC isoforms have also been associated with platelet activation and phosphatidylserine (PS) exposure[96], underlining our finding of a high abundance of DAGs in a thrombus.

In this study we imaged 6 sections per segment and quantified the reproducibility of these sections, which is indicative for the minimum fold change in intensity necessary to be significant. In future studies, in which we plan to investigate a larger collection of specimens, it is therefore possible to reduce the number of sections per site that we

measure. Together with the semi-automated analysis presented here, and technological advances such as continuous scanning modes, this makes it feasible to perform systematic analyses of molecular lipid composition in larger sets of atherosclerotic specimens.

In conclusion, we have developed a protocol for MALDI-MSI measurements of lipids in human carotid plaque tissue using a new analysis pipeline. This method can now be applied to larger collections of carotid plaque specimens, allowing to assess whether the carotid plaque's lipid signature is correlated to plaque phenotype and vulnerability. If so, this opens up new opportunities in terms of diagnosis and patient stratification, and could enable the investigation of the effect of lipid-lowering medication on plaque composition and rupture-risk.

Acknowledgments

The authors thank Kim van Gaalen for assistance with sectioning and histological staining. We thank Aad van der Lugt for helping us with obtaining tissue samples. We thank Bram Heijs for sharing his knowledge on MALDI-MSI with us. We thank Frits Mastik for helping us with importing imzML data into MATLAB and we thank Sandra den Hoedt for making the lipid extraction for us.

This project was funded by:
ZonMw - project number: 91113020
Netherlands Heart foundation - project number: NHS2014T096
Nederlandse Organisatie voor Wetenschappelijk Onderzoek - project number: 16131

3

Lipid signature of advanced human carotid atherosclerosis assessed by mass spectrometry imaging

This chapter is based on:
Astrid M. Moerman*, **Mirjam Visscher***, Nuria Slijkhuis, Kim Van Gaalen, Bram Heijs, Theo Klein, Peter C. Burgers, Yolanda B. De Rijke, Heleen M.M. Van Beusekom, Theo M. Luider, Hence J.M. Verhagen, Antonius F.W. Van Der Steen, Frank J.H. Gijsen, Kim Van der Heiden and Gijs Van Soest: Lipid signature of advanced human carotid atherosclerosis assessed by mass spectrometry imaging, *Journal of Lipid Research*, January 2021
*these authors contributed equally

Abstract

Carotid atherosclerosis is a risk factor for ischemic stroke, one of the main causes of mortality and disability worldwide. The disease is characterized by plaques, heterogeneous deposits of lipids and necrotic debris in the vascular wall, which grow gradually and may remain asymptomatic for decades. However, at some point a plaque can evolve to a high-risk plaque phenotype, which may trigger a cerebrovascular event. Lipids play a key role in the development and progression of atherosclerosis, but the nature of their involvement is not fully understood. Using matrix-assisted laser desorption/ionization mass spectrometry imaging, we visualized the distribution of approximately 200 different lipid signals, originating of > 90 uniquely assigned species, in 106 tissue sections of 12 human carotid atherosclerotic plaques. We performed unsupervised classification of the mass spectrometry dataset, as well as a histology-directed multivariate analysis. These data allowed us to extract the spatial lipid patterns associated with morphological plaque features in advanced plaques from a symptomatic population, revealing spatial lipid patterns in atherosclerosis and their relation to histological tissue type. The abundances of sphingomyelin and oxidized cholesteryl ester species were elevated specifically in necrotic intima areas, while diacylglycerols and triacylglycerols were spatially correlated to areas containing the coagulation protein fibrin. These results demonstrate a clear co-localization between plaque features and specific lipid classes, as well as individual lipid species in high-risk atherosclerotic plaques.

3.1 Introduction

Atherosclerosis is a disease causing plaque formation in the inner arterial walls and remains the largest cause of death worldwide[97]. The main driving mechanism for atherosclerotic plaque initiation is the gradual accumulation of lipids, originating from circulating lipoproteins, at sites of endothelial dysfunction in the vessel wall[98]. The influx of lipids and their subsequent modification in the vessel wall triggers inflammatory reactions that exacerbate the atherogenic process[63,99]. With disease progression, lipids start to play a dual role: they are deposited as metabolites of inflammatory processes but also form lipoproteins that act as signaling molecules, binding to a variety of macrophage-borne receptors[64]. Advanced plaques are complex heterogeneous structures that eventually can destabilize or erode to form a nidus for thrombogenesis, triggering ischemic vascular syndrome[65]. Previous studies have found that the lipid content of atherosclerotic plaques is different at various stages of disease progression[66,93], but the detailed interaction between the local vessel wall pathology and lipid biology has not been resolved. Insight into this interaction, and its relation with disease stage may be gained by visualizing the molecular lipid composition of plaques[15].

Matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) is a label-free molecular imaging technique and is suitable for detection and visualization of lipids in tissue sections[68,100]. MSI studies of lipid distribution in human atherosclerotic plaques that have been performed to date, reported only small numbers of samples or a limited spectrum of lipid species[21,37,41,43,69,70,101]. We use a previously established MALDI-MSI pipeline for systematic imaging of lipids[102] to visualize the spatial distribution of 194 lipids in 106 tissue sections of 12 carotid atherosclerotic plaques, harvested at carotid endarterectomy surgeries in 12 patients. Analysis of this dataset identifies spatial lipid patterns in advanced atherosclerosis that transcend individual variability.

We processed tissue sections adjacent to those studied by MSI for a series of histological staining procedures to identify compositional features of plaque vulnerability, i.e. necrotic core, the thrombus-associated protein fibrin, erythrocytes and foam cells[7,61,103]. In this way we were able to compare spatial lipid patterns with gold standard histological assessment of plaque composition. In this paper, we describe the systematic analysis of the high-dimensional MALDI-MSI dataset. We assessed spatial correlations between lipids and performed unsupervised cluster analysis. Secondly, we considered the histological results and investigated correlations between lipids and compositional features of plaque vulnerability.

In this study, we systematically visualize the molecular lipid content of 12 human atherosclerotic plaques with high resolution, and contextualize it in terms of gold standard histological plaque assessments, to provide insight towards the lipid signature of specific plaque features related to plaque vulnerability, such as necrotic core and fibrin.

3.2 Methods

3.2.1 Tissue collection and processing

Twelve human carotid endarterectomy (CEA) plaque specimens were surgically excised and were snap frozen and stored at -80 °C until further processing. The surgery was performed using a protocol that preserves an intact lumen and plaque morphology[73]. Upon processing, CEA specimens were divided in 2 mm thick cross-sections. Each cross-section was embedded in 10% porcine type A gelatin (Sigma-Aldrich, The Netherlands) and cryosectioned (CM3050 S, Leica Biosystems (cutting temps: OT -21 °C; CT -19 °C)) into 10 μm thick sections. Tissue sections were thaw-mounted on glass slides and stored at -80 °C. One slide was processed for MALDI-MSI, 6 other slides were histochemically stained. This study was performed according to the ethical guidelines sanctioned by the Ethics Board of Erasmus MC.

3.2.2 MALDI-MSI sample preparation, measurement and data reduction

MALDI-MSI experiments and data processing were performed according to the methods described in Chapter 2. In short, we desiccated the tissue sections and applied 2,5-dihydroxybenzoic acid (DHB) matrix by sublimation (home-built sublimation system as described in Dekker et al.[74]). MALDI-MSI experiments were performed on a Synapt G2Si-TOF system (Waters, Manchester, UK), operated in positive ion mode at the instrument's resolution mode (single-pass reflectron TOF), using a 2000 Hz Nd:YAG (355 nm) laser with a pixel-size of 45×45 μm^2 and using Waters Research Enabled Software suite (WREnS). The mass range was 300-1,200 *m/z* and the laser fired 100 shots per pixel. The data was acquired using MassLynx v4.2 software (Waters, Manchester, UK), and HDI v1.4 was used to export the MSI data in imzML format and processed using an in-house developed data processing pipeline in MATLAB™ 2017a (Natick, USA) and mMass software[79] was used to select tissue-specific lipid *m/z* values. We selected the lipid *m/z* values that were measured in more than 30% of all tissue sections and removed isotopes from the dataset, different adducts of the same molecule were retained. The data was log transformed for statistical analysis to comply with the assumptions of these methods[104].

3.2.3 Spatial cross-correlation

Per tissue section, the spatial correlation of lipid images was calculated using the Pearson correlation coefficient:

$$R^2 = \frac{\sum((x_i - \overline{x})(y_i - \overline{y}))}{\sigma_x \sigma_y} \tag{3.1}$$

The results of all tissue sections were averaged to obtain the average Pearson correlation coefficients for our dataset.

3.2.4 Unsupervised clustering algorithm

We calculated the optimal number of spectral components for this dataset and we applied non-negative matrix factorization (NMF) dimension reduction to the combined spectral data of all tissue sections, using an NMF toolbox for biological data-mining[23,105,106]. We determined the optimum number of components based on dispersion coefficients[107].

3.2.5 Histology, histology segmentation and image registration

The tissue section used in the MALDI-MSI experiment was stained for lipids by Oil Red O. Adjacent tissue sections were histochemically stained by: Miller's elastic stain, Martius scarlet blue trichrome and hematoxylin-eosin. Based on the histological information, a segmentation image of the tissue section was drawn, in which we identified the following plaque components: necrotic core (NC), fibrin, foam cells (FC), erythrocytes and calcium. The segmentation images were registered to the MALDI tissue section by translation and scaling using an in-house developed point-based rigid image registration framework in MeVisLab (MeVis Medical Solutions AG, Germany), to enable correlation of histology and MALDI-MSI data. After registration, the mean spectrum for each plaque component was calculated per tissue section.

3.2.6 Multivariate analysis

To investigate the presence of lipids characteristic for a plaque component, we performed multivariate analysis. For all plaque components, we fitted an Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) model[108] in SIMCA 15 (Umetrics, Sweden) comparing the mean spectra of the plaque component to the mean spectra of the rest of the tissue. In these models the intensities of the *m/z* values in the mean spectrum were defined as variables and the histological components were defined as the observations. Quality of fit and predictability of the model were reported by R^2 and Q^2 values respectively. Significance of OPLS-DA models was checked by sevenfold cross-validation analysis of variance (CV-ANOVA)[109] and the models were validated by permutation tests. If the model was significant, we extracted the Variable Influence on Projection (VIP) values and defined *m/z* values with a VIP > 1.0 to have a significantly different expression in the two histological components[110].

3.2.7 Comparison unsupervised NMF clustering to histology-based multivariate analysis

Per NMF component spectrum we calculated a cut-off value of 0.4 times the normalized weight of that NMF component. We subsequently identified the *m/z* values with intensities above this cut-off value, resulting in a list of *m/z* values that dominated the NMF component. We compared these *m/z* values to the *m/z* values with VIP > 1.0 resulting from the multivariate analyses to check the correspondence between the NMF unsupervised clustering and the histology-based multivariate analysis.

3.2.8 Statistics

The VIP values > 1.0 of the significant OPLS-DA models were separately tested on all data, including the patients for which the MVA model was not significant. Wilcoxon Signed Rank test was used to statistically test the non-log fold changes in intensity between the segmented and all other tissue areas. *M/z* values that had a p-value ≤ 0.05 were considered statistically significant. Receiver operating characteristic (ROC) was used to determine the optimal cut-off intensity (maximum Youden's index) for the VIP *m/z* values.

3.2.9 Lipid identification

We performed lipid identification experiments in two ways. Firstly we performed a lipidomics analysis on homogenized carotid endarterectomy tissue using the Lipidyzer platform (Sciex, Framingham, MA). Nine pieces of 50 mg CEA tissue was homogenized in 2 ml ice-cold methanol/MilliQ water (50/50 % (v/v)) with three 10 s bursts using an ultraturrax homogenizer. Homogenates were transferred to 2 ml Eppendorf tubes and stored on ice until centrifugation at 20,000 x g. The supernatant was collected and submitted to lipidomics analysis. 50 µL of tissue homogenate was mixed with the software pre-calculated amount of isotopically-labelled internal standard compounds in 13 lipid classes before liquid-liquid extraction with dichloromethane (DCM) /methanol. After 30 minutes, the lower phase was collected and the sample re-extracted with DCM. The combined extracts were dried under flowing N2 and re-suspended in running buffer (10 mM ammonium acetate in DCM/methanol 50/50 (% v/v)). The extracted samples were analyzed using the man-ufacturer-provided Lipidyzer SRM methods on a Shimadzu Nexera 20 UPLC system hyphenated to a Sciex 5500 Q-TRAP mass spectrometer equipped with a Selexion differential ion mobility module run with 1-propanol as a carrier. After analysis, lipid species quantification was performed using the Lipidyzer software.

Additionally, we performed MALDI-FTICR-MSI on a 12 T Bruker Daltonics solariX xR mass spectrometer equipped with dynamically harmonized ParaCell™, and a CombiSource™(Bruker Daltonics, Bremen, Germany). The instrument was operated using ftmsControl (v2.1.0 Build 98; Bruker Daltonics), and data were collected using a transient length of 0.4194 s (512k data point time domain), resulting in an estimated resolution of 97,000 at *m/z* 400. MALDI-measurements were performed using the SmartBeam™-II laser (λ = 355 nm) operating at 200 Hz, at 21% power, and using the "Small" focusing setting (ablation area approximately 70×70 µm^2). A total of 50 shots were acquired per pixel, with a mass range covering 294.86-1,500 *m/z*. MSI analyses were performed at 70×70 µm^2 and 100×100 µm^2 pixel-size.

MSI analysis was performed on one section using a transient length of 3.3554s (4M data point time domain) resulting in an estimated resolution of 776,000 at *m/z* 400, in order to resolve isobaric species. MALDI-measurements were performed using the SmartBeam^{TM}-II laser (λ = 355 nm) operating at 200 Hz, at 15% power, and using the "Small" focusing setting (ablation area approximately 70×70 µm^2). A total of 50 shots were acquired per pixel, with a mass range covering 300-1,500 *m/z*. MSI analyses were performed at 100×100 µm^2 pixel-size.

Acquired data was loaded in SCiLS Lab (v2016b; Bruker Daltonics), and converted to imzML before further processing. We uploaded the FTICR results to the METASPACE annotation platform[111] and we searched the available annotation databases with a false-discovery rate (FDR) < 10%.

3.3 Results and discussion

3.3.1 MALDI-MSI of plaque lipids shows lipid class-specific spatial patterns

Twelve human carotid endarterectomy samples were processed into a series of 106 axial tissue sections to be imaged by MALDI-MSI and histology. Processing of MALDI-MSI data resulted in lipid images of 194 unique lipid-related signals of which 93 were uniquely identified or assigned lipid species, see Table 3.3, at the end of this chapter. Demographic and histological information of the twelve human carotid plaque specimens can be found in Table 3.1 and Fig. 3.1, this demonstrates the heterogeneity in histological tissue composition between patients and within patients. Fig. 3.2 illustrates the variety in spatial distributions over 6 tissue sections, these tissue sections embody a representative subset of the 106 tissue sections, in terms of histological plaque composition. That is, all representative sections contained NC, four out of six sections contained fibrin areas of variable sizes (section 3, 4, 5, 6), section 2 and 4 contained a large FC area and in section 3 a large erythrocyte area was observed, see Fig. 3.2. We describe the observations in lipid patterns using these as examples, although the quantitative results we report were obtained on the complete data set. For the masses measured with FTICR we refer to the uploaded data in METASPACE, which is publicly available, and can be found as Human CEA Patient I - section 4 (https://tinyurl.com/y4p8prtq), Human CEA Patient H - section 4 (https://tinyurl.com/y2f4jbuv) and Human CEA Patient K – section 8 (https://tinyurl.com/yyqyatod).

In our dataset we observed lipids belonging to different lipid classes: cholesterol and cholesteryl esters (CEs), lysophosphatidylcholines (LPCs), phosphatidylcholines (PCs), sphingomyelin (SMs), diacylglycerol (DAGs) and triacylglycerol (TAGs). When stratifying the data by lipid class, we found that lipids belonging to the same lipid classes were distributed similarly over the plaque cross-sections, as quantified by their positive correlation coefficients. The average correlation heatmap of all tissue sections (Fig. 3.3a) and the heatmaps of the 6 representative tissue sections (Fig. 3.3b) are shown.

The cholesteryl ion (*m/z* 369.350) was abundantly present in the mass spectra of all tissue sections. This ion may originate from cholesterol ($[M\text{-}H_2O+H]^+$) or as a result of fragmentation of CEs. Its presence was detected over the whole intimal area, though with local variations in intensity. No clear co-localization with histological components was seen. In general, cholesteryl patterns showed moderate cross-correlation with intact CEs (Fig. 3.3). The spatial patterns of two prominent CEs, CE(18:2) and CE(18:1), show striking differences (Fig. 3.2). While CE(18:2) is often present throughout the cross-section, CE(18:1) is observed in high intensity spots around the lumen, while not abundant throughout the rest of the tissue sections (Fig. 3.2). CE lipids in general showed mild spatial cross-correlation with triacylglycerols (Fig. 3.3).

Table 3.1: **Average histological tissue compositions (percentages of intima area) per patient.**
% area of histological component relative to total intima area. μ = mean, SD = standard deviation.
μ ± SD intima area is given in mm2, areas were averaged over all tissue sections of a patient. Intima area of largest tissue section is reported between brackets
μ ± SD: percentages were averaged over all tissue sections of a patient
* Representative tissue sections are referred to in Fig. 3.2, Fig. 3.3 and Fig. 3.5.

Patient	**Age** [yrs]	**Sex**	**Intima area** μ ± SD (max) [mm^2]	**% Necrotic core** μ ± SD	**% Foam cells** μ ± SD	**% Fibrin** μ ± SD	**% Erythrocytes** μ ± SD	**Number of sections**	**Representative tissue section***
A	56	F	19.0 ± 10.4 (34.4)	10.1 ± 9.0	1.1 ± 1.7	0.9 ± 1.9	0.0 ± 0.1	9	
B	69	M	22.2 ± 16.2 (54.5)	14.4 ± 18.9	6.1 ± 7.8	10.3 ± 17.4	0.6 ± 2.1	12	
C	65	M	22.6 ± 6.9 (30.1)	5.2 ± 4.8	5.1 ± 6.7	1.7 ± 3.5	0.0 ± 0.0	7	1
D	75	M	23.0 ± 13.9 (54.5)	11.9 ± 17.6	4.6 ± 4.8	0.6 ± 1.1	3.7 ± 6.5	11	
E	75	M	25.4 ± 11.4 (47.2)	8.5 ± 9.4	2.9 ± 1.8	14.4 ± 13.6	0.0 ± 0.1	8	
F	63	M	26.6 ± 11.6 (43.8)	16.1 ± 11.0	3.1 ± 6.3	1.5 ± 2.3	0.3 ± 0.7	8	2
G	81	M	28.9 ± 17.2 (59.2)	24.8 ± 17.1	0.6 ± 1.3	5.0 ± 5.9	0.0 ± 0.1	10	
H	62	F	35.7 ± 11.1 (51.0)	8.7 ± 3.1	0.3 ± 0.4	5.6 ± 4.4	10.2 ± 7.7	5	3
I	69	M	49.6 ± 16.4 (70.1)	48.6 ± 23.8	3.0 ± 5.1	15.7 ± 14.2	0.0 ± 0.1	7	4
J	79	M	52.1 ± 28.5 (101.0)	35.3 ± 17.5	0.6 ± 1.1	23.6 ± 22.9	0.0 ± 0.0	8	5
K	69	M	53.3 ± 18.3 (77.4)	47.5 ± 21.3	0.9 ± 1.4	13.8 ± 14.5	0.3 ± 1.1	12	6
L	82	M	57.3 ± 37.9 (120.0)	25.6 ± 21.4	0.4 ± 0.7	32.0 ± 31.3	0.8 ± 1.4	9	

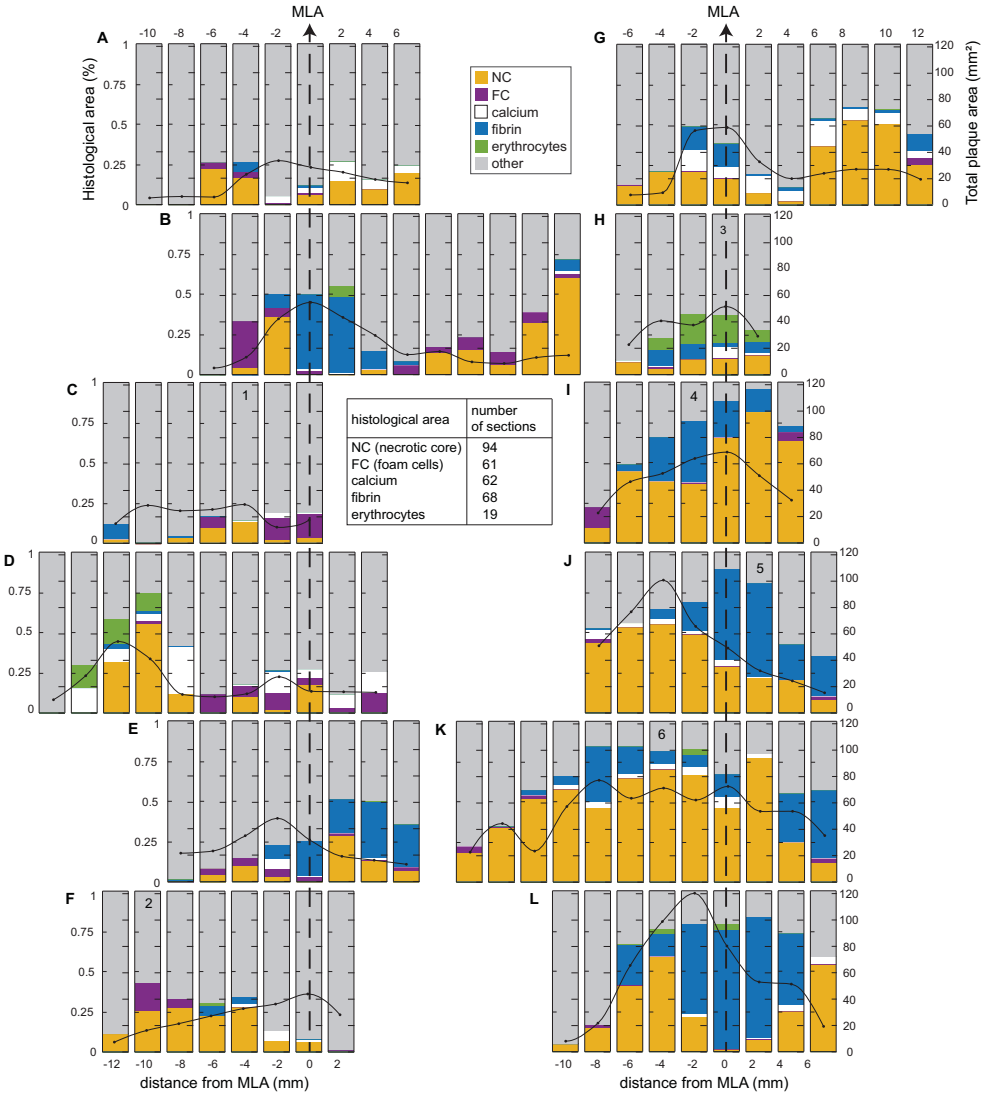


Figure 3.1: **Overview of plaque components present in each carotid endarterectomy (CEA) sample. A-L**: Bar graphs show the relative proportions (%) of necrotic core (NC), foam cells (FC), calcifications, fibrin and erythrocytes compared to the total intima area, see left axes.Tissue labeled as other was not segmented, this part of the tissue was histologically heterogeneous, but did not fit into any accepted classifications of atherosclerotic tissues. Total intima area represents the area of the intima layer of the vessel. The black line superimposed on the bargraph shows the total intima area in mm^2, see right axis. The numbers in the graphs correspond to the exemplatory sections shown in Fig. 3.2,Fig. 3.3 and Fig. 3.5. The table in the middle shows the frequency of occurance of each histological component. MLA = minimal lumen area. Per CEA sample, the tissue section containing the minimal lumen area is denoted.

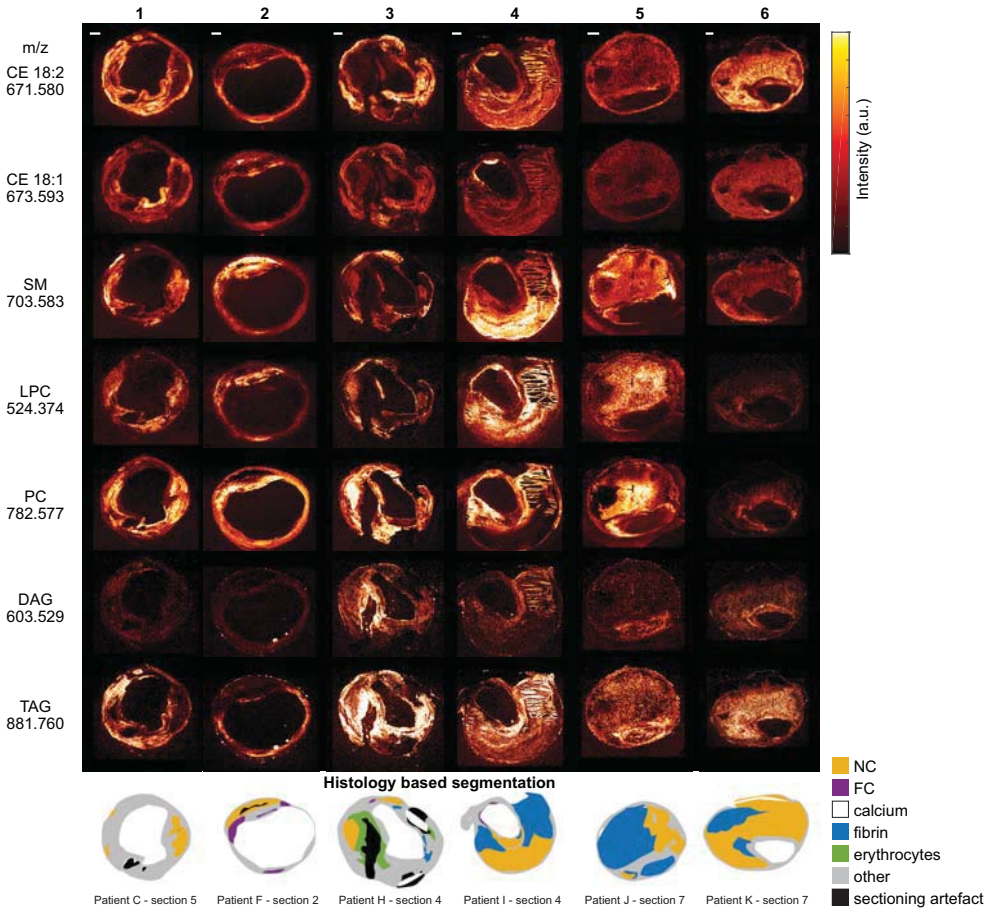


Figure 3.2: **Overview of MALDI-MSI data for a selection of 7 lipids.** Data is shown for 6 tissue sections representative of the 106 section dataset, in terms of histological plaque composition. For comparison, histology segmentation images are also shown. The lipid images were normalized per *m/z* value, to enable reliable comparison of the distribution of a specific lipid in different tissue sections. Also, CE(18:2) and CE(18:1) are shown on the same intensity scale, as are DAG and TAG. Scalebars are 1 mm.

Although the spatial distribution of most CEs was similar to the distribution of cholesterol itself, one oxidized cholesterol lipid with *m/z* 401.343, was distinct from that of other cholesterol-related molecules (Fig. 3.3a-oxoChol/CE first row). The spatial location of this species and cholesteryl ion, identified as cholesterol or fragmented cholesterol ester, is depicted in Fig. 3.4a for two tissue sections. In addition, this molecule showed a unique spatial distribution as it did not show any cross-correlation with other lipids in our dataset and was also not correlated with histological components.

We found that the co-localization of SM, LPC and PC lipids varied strongly depending on the tissue section. In general, PCs were found to surround the lumen (Fig. 3.2 – section 2, 6) and sometimes to exude into thrombotic or inflamed areas (Fig. 3.2 – section 1, 3, 4, 5). The distribution of SMs did not deviate much of that of PCs in some tissue sections (Fig. 3.2 – section 1) but in others SMs were clearly more abundant in the necrotic parts

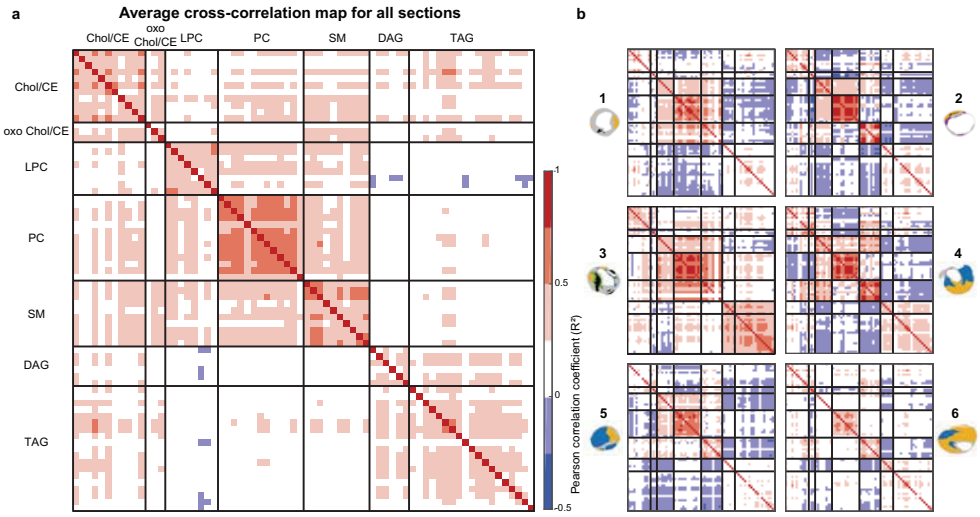


Figure 3.3: **Spatial cross-correlation heatmaps.** X- and Y-axes are the same and represent a selection of 70 *m/z* values (Table 3.3), classified by lipid class. Per lipid class, the lipids are sorted from low to high *m/z* value. **a)** Average heatmap of all 106 tissue sections, showing a correlation within lipid classes and showing a moderate correlation between LPCs, PCs and SMs and also between DAGs and TAGs. **b)** Individual heatmaps of the 6 representative sections shown in Fig. 3.2. Heatmaps of section 1 and 3 show moderate positive correlation between PC and SM lipids, whereas section 2 shows a negative between these lipids.

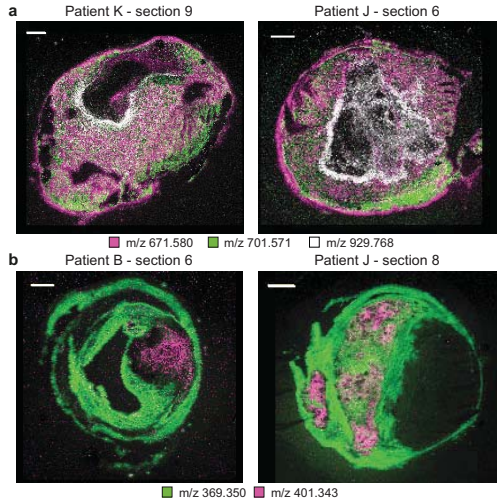


Figure 3.4: **a)** Combined MALDI-MSI image of spatial distributions of a CE in purple (*m/z* 671.580), SM in green (*m/z* 701.571) and a TAG in white (*m/z* 929.767) for three tissue sections. CE is found throughout the tissue section, while SM and TAG show a more localized spatial pattern. **b)** Spatial distribution of *m/z* 401.343, annotated as 7-ketocholesterol, in purple and *m/z* 369.350, identified as cholesterol, in green. 7-ketocholesterol shows a spatial distribution distinct from that of other lipids in our dataset. Scalebars are 1 mm.

of the tissue section (Fig. 3.2 – section 2, 4). These observations were quantified in the cross-correlation analysis: the degree of spatial correlation between SMs and PCs ranged from strong positive correlation, i.e. R_2>0.75, to strong negative correlation, i.e. R_2<0.25 (Fig. 3.3b – tissue section 2). In most tissue sections, the highest mutual spatial correlation was found between PC lipids (Fig. 3.3). The distribution of PCs and LPCs was found to be weakly positively correlated in most tissue sections (Fig. 3.3). LPCs are structurally similar to PCs but lack one fatty acid branch. When assessing the tissue sections that showed marked variations in PC and SM distributions (i.e. Fig. 3.2 – section 2, 4, 5, 6) for the LPC distribution pattern, LPCs were highly expressed at the intersection of the locations that showed high SM and PC intensity.

In contrast, in most tissue sections, the spatial patterns of the DAG and TAG lipids showed moderate to high correlation and were distinctive from the distributions of other lipids except for some CEs (Fig. 3.3b - section 2, 4, 5, 6). However the intensities of measured DAG were much lower than those of TAG. Figure 3.4a shows overlay images of three lipids identified as a CE, a SM and a TAG for two tissue sections. The variation in spatial distributions between these lipids, which belong to different lipid classes, can clearly be seen.

3.3.2 The spectral variety in the dataset can be captured by six spectral components

We reduced the dimensionality of the spectral data by unsupervised NMF, thereby extracting the major spectral patterns, using 6 components. Figure 3.5a and b, shows the spectra of the NMF components and the spatial distributions of the 6 spectral components, respectively. Component I consisted mostly of oxidized and non-oxidized cholesterol and CE lipids. Component II was a combination of cholesterol, both oxidized and non-oxidized, DAGs and a proportion of LPCs. Component III was dominated by DAGs and TAGs. Component IV represented unidentified *m/z* values. Component V consisted of LPCs, SMs and cholesterol, both oxidized and non-oxidized, while component VI represented the PC moiety of the data.

3.3.3 Correlation between spatial lipid patterns and histological plaque components

We compared the average spectra of different histological components by fitting a multivariate model. Four multivariate models, discriminating NC, fibrin, FCs and erythrocytes from all other tissue, were significant, i.e. R^2 > 0.5 and Q^2 > 0.5, for the combined data of at least 6 out of 12 patients. Model parameters of the significant models are listed in Table 3.2. The multivariate models determine the *m/z* values that have the highest influence on separation between two histological components, reported as Variable Influence for Projection (VIP) values. VIP values > 1.0 are considered significant[110] and these are reported in the supplementary information, available in the online version of this paper. We found SMs and oxidized CEs to be more abundant in NC than in other intima tissue, while DAGs and TAGs were found to be distinctive for areas containing fibrin. The VIP

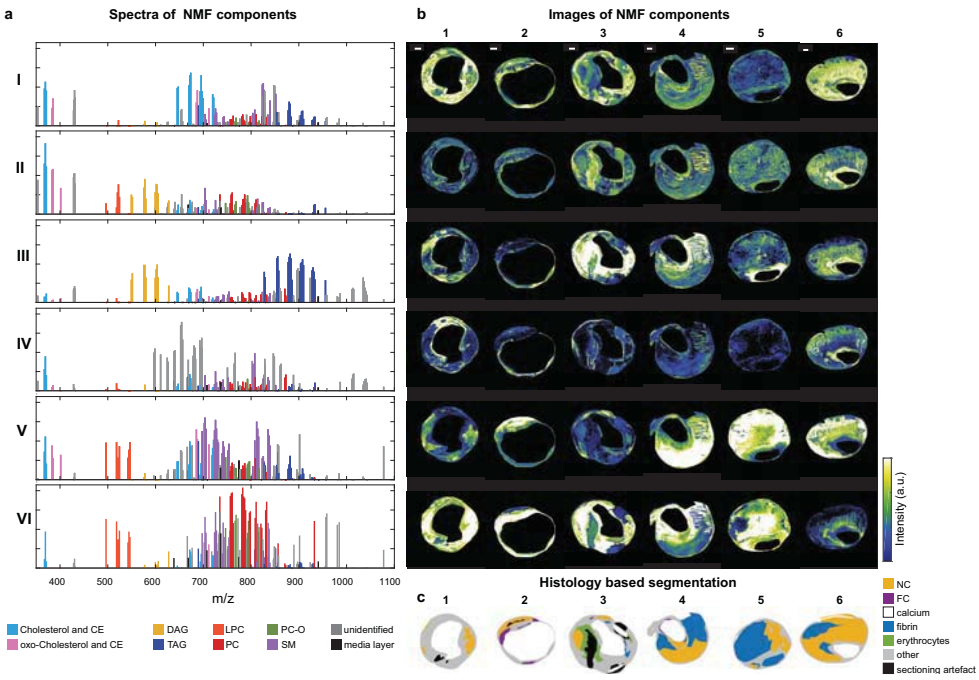


Figure 3.5: **Unsupervised non-negative matrix factorization (NMF) of the 194 *m/z*-containing MALDI-MSI data. a)** NMF spectra of the components showing the weight of each *m/z* value to that NMF component. *m/z* values in the NMF component spectra are labelled according to their assigned lipid class, showing a clear separation of lipid classes by means of NMF. **b)** Corresponding NMF-weighted images of the representative sections, showing the spatial distribution and relative intensities of the NMF components. **c)** Corresponding histology-based segmentation suggesting co-localisation between certain NMF components and histologically-relevant features. Scalebars are 1 mm.

values resulting from the FC model were mostly SMs, while the erythrocyte-related VIP values were mostly PCs.

In order to quantitatively establish the contrast in lipid com-position between various tissues, we performed two univariate analyses of the intensities of *m/z* values with VIP > 1.0 appearing in the multivariate models: one contrasting NC and not-NC, and another contrasting fibrin and not-fibrin. Figure 3.6 illustrates the results for two *m/z* values. All *m/z* values that had significant VIP values in the multivariate model were significantly (p-value < 0.005) different between NC and not-NC in the univariate model, and the difference became stronger with increasing NC sizes. In the univariate analysis of fibrin, 58 out of 62 VIPs > 1.0 were significantly different between fibrin and not-fibrin. The four non-significant *m/z* values had VIP values lower than 1.14.

3.3.4 Comparison NMF – multivariate analysis - histology

Since certain NMF components appeared to co-localize with areas identified in histology (Fig. 3.5b and c), we investigated the similarities between the NMF spectra and the VIP values that were distinctive for histological components. The majority of the NC *m/z*

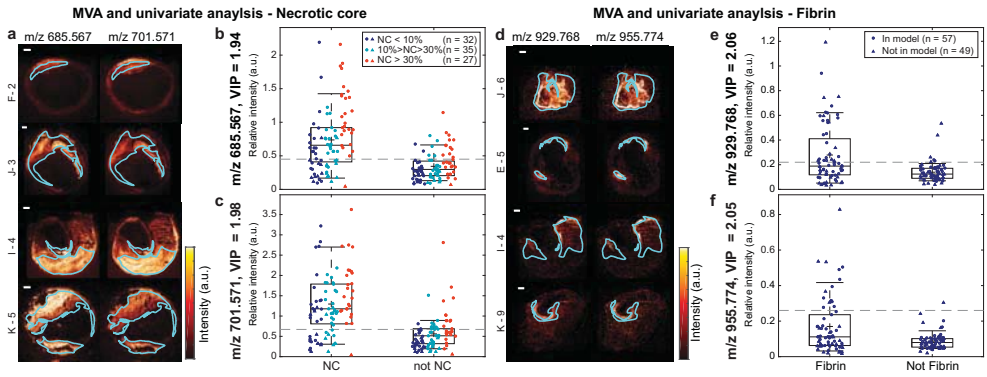


Figure 3.6: **Univariate analysis comparing intensities in necrotic core (NC) versus not-NC areas and fibrin versus not-fibrin areas for two lipids. a)** MALDI-MSI images of *m/z* 685.567 (oxoODE-CE $[M+Na]^+$) and *m/z* 701.571 (SM(34:2) $[M+H]^+$) of 4 different tissue sections with superimposed the outlines of the NC segmentation **b)** Boxplots of *m/z* 685.567 and **c)** *m/z* 701.571 showing the mean intensity of the pixels in the NC area and in the not-NC area. Datapoints are colored according to the size of NC present in the corresponding tissue section. **d)** MALDI-MSI images of *m/z* 929.768 (TAG(58:9) $[M+H]^+$ or TAG(56:6) $[M+Na]^+$) and *m/z* 955.774 (TAG(60:10) $[M+H]^+$ or TAG(58:7) $[M+Na]^+$) of 4 different tissue sections with the outlines of the fibrin segmentation **e)** Boxplots of *m/z* 929.768 and **f)** *m/z* 955.774 showing the mean intensity of the pixels in the fibrin area and in the not-fibrin area. Only tissue sections containing NC or fibrin were included in these analyses. The dotted lines depict the maximum Youden's index, representing the optimal cut-off intensity value above which the *m/z* value is more likely to be associated with NC or fibrin. For *m/z* 685.567 the threshold is 0.45 and for *m/z* 701.571 it is 0.67, for *m/z* 929.768 it is 0.22 and for *m/z* 955.774 it is 0.26. All *m/z* values showed significantly different intensities between NC and not-NC or between fibrin and not-fibrin (p-value < 0.001). ○: data of the patients included in the multivariate model, △: data of patients that did not fit the model. Scalebars are 1 mm.

values with VIP > 1.0, was found in NMF component V, i.e. mostly SMs and oxidized CEs (Fig. 3.7a).

For fibrin, most VIP *m/z* values were found in NMF component III (Fig. 3.7c). These VIP values are a set of DAG and TAG lipids that discriminated areas containing the thrombus-associated protein fibrin. Figure 3.7b and d illustrate the correspondence between the histology-based multivariate analysis and the NMF analysis for NC and fibrin areas. The majority of the significant VIP values for erythrocyte areas were most abundant in components VI and III and the VIP values associated with FCs corresponded to NMF component VI and these are reported in the supplementary information, available in the online version of this paper.

3.4 Discussion

With this study, we aimed to visualize the spatial distribution of lipids over the human carotid atherosclerotic plaque and to investigate the relation between lipids and compositional features of plaque vulnerability. We imaged the spatial distribution of 194 lipids related signals, originating from >90 unique lipid species, in 106 tissue sections of carotid atherosclerotic plaques. This study collects the largest MSI atherosclerosis dataset to date, spatially mapping the molecular lipid content of a series of human plaques, ob-

Table 3.2: **OPLS-DA model parameters comparing the mean spectrum of a histological area (NC, fibrin, FCs or erythrocytes) to the mean spectrum of tissue outside this histological area.** The patients included in the model are reported in the last column of this table, all tissue sections of included patients were added to the model. R^2 and Q^2 values represent the quality of fit and predictability of the model respectively. Significance of OPLS-DA models was checked by sevenfold cross-validation analysis of variance (CV-ANOVA). * The number of patients for which R^2 and Q^2 values were highest. Minimal 6 out of 12 patients were included in the model.

Histological area	R^2	Q^2	CV-ANOVA	Patients included in model (number of patients in model / total number of patients) *
NC	0.57	0.51	5.34E-19	All except D, H, L (9/12)
Fibrin	0.68	0.52	2.53E-07	B, C, F, G, J, K (6/12)
Foam cells	0.63	0.55	7.33E-15	All except D (11/12)
Erythrocytes	0.67	0.61	5.57E-05	B, F, H, I, K, L (6/12)

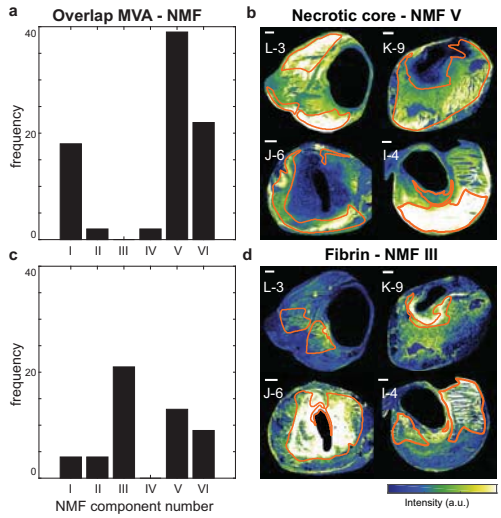


Figure 3.7: **Overview of the correspondence between the histology-based multivariate analysis and the NMF analysis for NC- and fibrin-containing areas. a)** Bar graphs showing the number of NC VIP values above 1.0 that were allocated to the different NMF component spectra. Most NC VIP values were allocated to NMF component V. **b)** The results in a are illustrated by depicting the NMF weight images of component V for four tissue sections with superimposed NC segmentation. **c)** The number of fibrin VIP values above 1.0 allocated to the different NMF component spectra. Most fibrin VIP values were present in NMF component III. **d)** The NMF weight images of component III with fibrin segmentation for four tissue sections. Scalebars are 1 mm.

taining information on both the cross-sectional and the longitudinal composition of these plaques. Moreover, since we simultaneously processed adjacent tissue sections for histology, this study enabled the comparison of MALDI-MSI results with gold standard histological assessment of plaque stage.

In our dataset we observed lipids belonging to different lipid classes: cholesterol and CEs, LPCs, PCs, SMs, DAGs and TAGs. When stratifying the data by lipid class, we found that lipids belonging to the same lipid classes were distributed similarly over

the plaque cross-sections, as quantified by their correlation coefficients. PC lipids in particular showed high mutual spatial correlation. PCs also showed spatial overlap with SMs and LPCs, while CEs were spatially correlated to DAGs and TAGs. However, the degree of spatial correlation was found to differ per tissue section. SMs and oxidized CEs co-localize in NC areas, and this co-localization became more apparent with greater NC size and thus with bigger intima area. The presence of oxidized CEs in NC is in line with theories on pathogenesis of atherosclerosis: during plaque progression, LDL proteins are thought be become oxidized and trapped inside the vessel wall[98,112,113]. The presence of oxidized CE species in plaques, and their increase in abundance with lesion complexity, has been reported previously[69,114–116]. Increased SM levels in plaques, compared to healthy artery have also been described before[41,117], as well as the higher abundance of SMs, relative to PCs, with increasing plaque severity[70,118]. Yet, no previous study was able to visualize a clear co-localization of SM and oxidized CE lipid patterns with NC as identified on histology. We showed that the increased SM and oxidized CE levels concentrate in necrotic regions of the plaque, a finding that forms an important addition to the current understanding of the pathophysiology of atherosclerosis, which also provides us with a potential target for detecting NC presence or size in advanced plaques by lipid-sensitive diagnostic imaging technologies[51,119,120].

High levels of ceramides in serum, lipids that are structurally-related to SMs, have previously been associated with high-risk plaque[121,122]. We detected low levels of ceramides in CEA tissue using the Lipidyzer platform, but ceramide levels were probably below the detection limit of the MALDI-MSI experiments in this study or were suppressed by other ions.

Most cholesterol esters showed a similar spatial distribution. Interestingly, however, the spatial distribution of an oxidized cholesterol lipid with *m/z* 401.343 was distinct from that of other cholesterol-related molecules. In addition, this molecule showed a unique spatial distribution as it did not show any cross-correlation with other lipids in our dataset, and was also uncorrelated with histological components. According to literature, *m/z* 401.343 can be annotated as 7-ketocholesterol[43], however, since identification based on MS/MS or chromatographic data is missing other possible isomers or isobars cannot be excluded. 7-Ketocholesterol has been found in atherosclerotic tissue before[114,115,123] and shown to play a pro-atherogenic role, e.g. by mediating cell death[124,125].

DAGs and TAGs showed a spatial distribution that was distinct from that of PCs, SMs and LPCs. DAGs are present in atherosclerotic tissue[39] but may also form by in-source fragmentation of TAGs and PCs. In the spatial NMF analysis, abundant DAGs but not TAGs and only few PCs appear in NMF component II. Likewise, no DAGs are seen in NMF component VI, which features abundant PCs. The observation of these distinct spatial patterns supports the hypothesis that we detect intact species, although partial fragmentation of TAGs and PCs cannot be excluded. When comparing the MALDI-MSI images of DAG and TAG molecules, similarities in the spatial patterns of DAG molecules and their presumed TAG precursor molecules appear. However, some specific patterns are only visible in the DAG images, which confirms the detection of endogenous DAG species, see Fig. 3.8. Similarly, this occurs for PCs, a figure hereof is reported in the supplementary information, available in the online version of this paper. Our

multivariate analysis showed that a set of DAG and TAG lipids discriminated areas containing the thrombus-associated protein fibrin. A subset of TAG lipids was also associated with erythrocyte areas in the multivariate model. Previously[102], we described a co-localization between DAGs and thrombus and this observation is thus confirmed in our current dataset. We hypothesize that the co-localization of DAGs and thrombotic plaque elements can be explained by the role that DAGs play in regulating protein kinase C (PKC), an enzyme known to induce an array of pro-atherogenic responses[126], including impairment of fibrinolysis[127] and platelet activation[128,129].

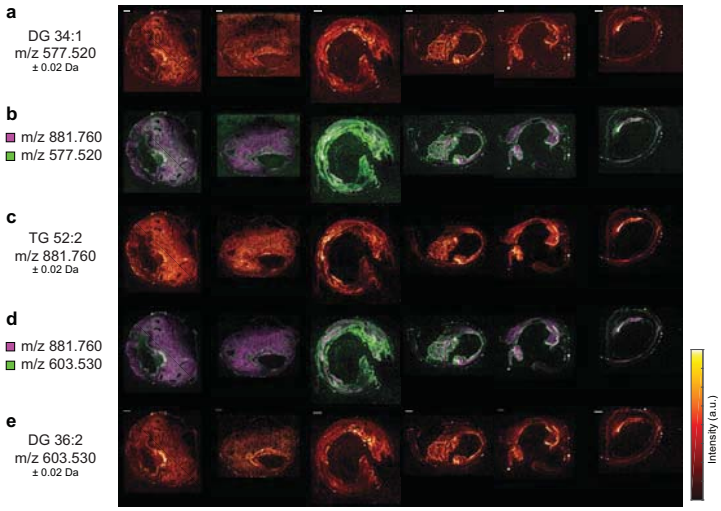


Figure 3.8: **MALDI-MS images of a)** DG(34:1) with m/z 577.520, **b)** an overlay of lipid images with DAG(34:1) shown in green and TG(52:2) shown in purple **c)** TG(52:2) with m/z 881.760, **d)** an overlay of lipid images with DAG(36:2) shown in green and TG(52:2) shown in purple and **e)** DG(36:2) with m/z 603.530 Scalebars are 1 mm.

The tissue areas that contained FCs and erythrocytes were generally smaller than those of NC and fibrin. The lipids that were found to be distinctive for FC areas in the multivariate model were mostly PCs. When comparing the FC VIP values and the NMF component spectra, the majority of FC VIPs were present in the PC-dominated NMF component VI. Thus, there was large spatial overlap between the outcomes of the histology-based multivariate model and the NMF classification (Fig. 3.7). A subset of erythrocyte-associated *m/z* values – as found by the multivariate analysis – were also PC lipids. PC lipids are major constituents of healthy cell membranes[130], which probably explains why the spatial distribution of PC lipids was found to correlate with areas of high cellularity, for instance FC- and erythrocyte-rich areas.

3.4.1 Limitations

In order to facilitate processing of a large number of sections, we opted to perform experiments in positive mode only. Lipids that have been associated with atherosclerosis previously, e.g. SMs[13,34,41,43,70,71], CEs[13,15,33,36,37,41,69,89], PCs[13,15,33,36,37,41,43,69,70] have been shown to be detectable in positive mode.

Lipid identities could not be assigned to a subset of the 194 *m/z* values that were measured in the MALDI-MSI experiments. Most of the unknown masses in the MALDI-MSI data were not detected in the data captured by the Lipidyzer and FTICR measurements that were used for mass identification purposes. In addition, these unknown *m/z* values were not reported in the databases we consulted (Human Metabolome Database, SwissLipids and LipidMaps). The unidentified *m/z* values were dominating the spectrum of NMF component IV, while barely contributing to the spectra of the other NMF components. The distribution images of NMF component IV indicate that the unidentified *m/z* values were mostly present at the edges of the MALDI-MSI cross-sections. We assume that these images represent artefacts of the MALDI-MSI measurement method.

In the annotation by exact mass in FTICR, a few isobaric species are candidates for assignment. In addition to mass, we took into account the relative sensitivity of MALDI-MSI and detection likelihood in confidently annotating these peaks. For the Synapt data, we found a mass error of on average 7.02 ppm with a standard deviation of 4.40 when calculating the mass error based on the exact mass of the identified lipids. To strengthen lipid identification, we have tried to perform MS/MS analysis on tissue for lipid identification, however, our system was lacking sensitivity for the lipids that are ambiguous. We therefore chose to use the Lipidyzer platform and the FTICR system to get the best possible annotation and identification. The mass error of lipids annotated from the FTICR measurement is on average 0.11 ppm with a standard deviation of 0.015.

3.5 Conclusion

In our endeavor to identify lipids characteristic of high-risk plaque phenotype, we conclude that certain SMs and oxidized CEs are significantly higher expressed in NC and that the abundance of these molecules is correlated to the size of the NC. Interesting observations to build further studies on are the correlation between DAGs and TAGs and thrombus fragments, i.e. erythrocytes and fibrin, providing a possible marker for intra-plaque bleeding. Additionally, erythrocytes and FCs are correlated with PCs. These findings add to the current understanding of atherosclerosis pathogenesis. In the future, these marker lipids related to "vulnerable" plaque features may serve as targets for *in vivo* imaging and therapeutic applications, aiding the lipidomic identification of plaques, and thus patients, in need of focal or systemic preventive treatment.

Acknowledgments

We would like to thank Aad van der Lugt for helping us with clinical sample collection. This project was funded by:

ZonMw - project number: 91113020
Netherlands Heart foundation - project number: NHS2014T096
Nederlandse Organisatie voor Wetenschappelijk Onderzoek - project number: 16131

Table 3.3: **List of 194 lipids measured in human carotid atherosclerotic plaques**
m/z = mass measured in MALDI-MSI experiment using Synapt G2Si TOF system. For exact mass measured with FTICR and ppm values, we refer to the online available METASPACE at https://metaspace2020.eu/ data named Human CEA Patient H - section 4 and Human CEA Patient I - section 4
lipid group = assigned lipid group based on database search, i.e. LipidMaps, HMDB.
ID = proven lipid identity, superscript denotes identification method: [a] Lipidyzer MRM analysis, [b] FTICR measurement combined with METASPACE database (FDR of maximum 10%), [c] Identified in Chapter 2, [d] Identified from literature[43,69,131,132], [LM] assigned from search in LipidMaps, [HMDB] assigned from search in HMDB.
adduct = positive ion adduct in identification experiment.
* Asterisks denote the lipids that were included in the cross-correlation analysis

m/z ± 0.02	lipid class	ID	adduct
353.331	unknown		
367.343	Chol	Cholesterol derivative[LM,d]	$[M\text{-}H_2O+H]^+$
369.350*	Chol	Cholesterol[c]	$[M\text{-}H_2O+H]^+$
371.358	Chol	Cholesterol derivative[LM,d]	$[M\text{-}H_2O+H]^+$
383.333	oxChol	Cholesterol derivative[LM,d]	$[M+H]^+$
385.348	oxChol	Dehydrocholesterol[LM,d]	$[M+H]^+$
401.343*	oxChol	7-ketocholestol[d]	$[M+H]^+$
428.370	unknown		
429.375	unknown		
430.379	unknown		
431.385	unknown		
496.346*	LPC	LPC(16:0)[d]	$[M+H]^+$
518.335*	LPC	LPC(18:3)[d]	$[M+H]^+$
520.351*	LPC	LPC(18:2)[d]	$[M+H]^+$
522.371*	LPC	LPC(18:1)[d]	$[M+H]^+$
524.374*	LPC	LPC(18:0)[d]	$[M+H]^+$
542.322*	LPC	LPC(20:5)[d]	$[M+H]^+$
544.340*	LPC	LPC(20:4)[d]	$[M+H]^+$
546.351*	LPC	LPC(20:3)[d]	$[M+H]^+$
		LPC(18:0)[b]	$[M+Na]^+$
549.488*	DAG	DAG(32:1)[a]	$[M\text{-}H_2O+H]^+$
551.509	DAG	DAG(32:0)[a]	$[M\text{-}H_2O+H]^+$
575.510*	DAG	DAG(34:2)[a]	$[M\text{-}H_2O+H]^+$
		DAG(O-32:1)[b]	$[M+Na]^+$
577.520*	DAG	DAG(34:1)[a]	$[M\text{-}H_2O+H]^+$
		DAG(O-32:0)[b]	$[M+Na]^+$
579.534	DAG	DAG(34:0)[a]	$[M\text{-}H_2O+H]^+$
596.540	unknown		

Table 3.3 continued from previous page

m/z ± 0.02	lipid class	ID	adduct
598.558	unknown		
599.497*	DAG	DAG(36:4)[a]	$[M\text{-}H_2O+H]^+$
601.512*	DAG	DAG(36:3)[a]	$[M\text{-}H_2O+H]^+$
		DAG(O-34:2)[b]	$[M+Na]^+$
603.530*	DAG	DAG(36:2)[a]	$[M\text{-}H_2O+H]^+$
		DAG(O-34:1)[b]	$[M+Na]^+$
605.555	DAG	DAG(36:1)[a]	$[M\text{-}H_2O+H]^+$
610.572	unknown		
624.567	unknown		
626.591	unknown		
627.544	DAG	DAG(38:4)[a]	$[M\text{-}H_2O+H]^+$
		DAG(O-36:3)[b]	$[M+Na]^+$
638.472	unknown		
638.587	unknown		
640.603	unknown		
645.561*	CE	CE(16:1)[a,b,c]	$[M+Na]^+$
647.577*	CE	CE(16:0)[a,b,c]	$[M+Na]^+$
652.610	unknown		
654.628	unknown		
656.630	unknown		
666.495	unknown		
668.525	unknown		
668.622	unknown		
669.563	CE	CE(18:3)[a]	$[M+Na]^+$
671.580*	CE	CE(18:2)[a,b,c]	$[M+Na]^+$
673.593*	CE	CE(18:1)[a,b,c]	$[M+Na]^+$
675.573	SM	SM(d32:1)[a,b]	$[M+H]^+$
676.560	unknown		
678.623	unknown		
680.636	unknown		
681.643	unknown		
682.653	unknown		
683.658	unknown		
685.567*	oxCE	Oxo-ODE-CE[d]	$[M+Na]^+$
687.566*	oxCE*	HODE-CE[b]	$[M+Na]^+$
	CE	CE(18:2)[a]	$[M+K]^+$

Table 3.3 continued from previous page

m/z ± 0.02	lipid class	ID	adduct
689.570	CE	CE(18:1)[b]	$[M+K]^+$
	TAG	TAG(38:1)[b]	$[M+Na]^+$
692.638	unknown		
693.567*	CE	CE(20:5)[a,b]	$[M+Na]^+$
695.581*	CE	CE(20:4)[a,b]	$[M+Na]^+$
695.659	unknown		
696.673	unknown		
697.535*	SM	SM(32:1)[a,b]	$[M+Na]^+$
697.591*	CE	CE(20:3)[a,b]	$[M+Na]^+$
697.672	unknown		
699.606	CE	CE(20:2)[a,b]	$[M+Na]^+$
701.571	SM	SM(34:2)[b]	$[M+H]^+$
703.583*	SM	SM(34:1)[a,b,c]	$[M+H]^+$
705.589	SM	SM(34:0)[b]	$[M+H]^+$
707.542	unknown		
711.550*	oxCE*	CE(11:1D3)[HMDB]	$[M+Na]^+$
	CE	CE(20:4)[a]	$[M+K]^+$
717.578	unknown		
719.571*	CE	CE(22:6)[a,b]	$[M+Na]^+$
721.586*	CE	CE(22:5)[a,b,c]	$[M+Na]^+$
723.544	SM	SM(34:2)[b]	$[M+Na]^+$
725.565*	SM	SM(34:1)[a,b,c]	$[M+Na]^+$
727.566	SM	SM(34:0)[b]	$[M+Na]^+$
729.590*	SM	SM(36:2)[a]	$[M+H]^+$
731.614*	SM	SM(36:1)[a,b,c]	$[M+H]^+$
732.584	unknown		
733.561	unknown		
734.580*	PC	PC(32:0)[a,b,c]	$[M+H]^+$
735.572	unknown		
736.577	unknown		
739.550	SM	SM(34:2)[LM,HMDB]	$[M+K]^+$
741.545*	SM	SM(34:1)[a]	$[M+K]^+$
743.555	SM	SM(34:0)[LM,HMDB]	$[M+K]^+$
746.606	PC-O	PC(O-34:1)[LM,HMDB]	$[M+H]^+$
750.565	unknown		

Table 3.3 continued from previous page

m/z ± 0.02	lipid class	ID	adduct
751.574	SM	SM(36:2)[a]	$[M+Na]^+$
753.585	SM	SM(36:1)[a,b]	$[M+Na]^+$
755.552	unknown		
756.558*	PC	PC(32:0)[a,b]	$[M+Na]^+$
758.574*	PC	PC(34:2)[a,c]	$[M+H]^+$
760.587*	PC	PC(34:1)[a,b,c]	$[M+H]^+$
762.618	PC	PC(34:0)[HMDB]	$[M+H]^+$
764.652	unknown		
766.578*	PC-O	PC(O-36:5)[LM,HMDB]	$[M+H]^+$
766.754	unknown		
768.581*	PC-O	PC(O-36:4)[LM,HMDB]	$[M+H]^+$
772.542	PC	PC(32:0)[a]	$[M+K]^+$
774.580	unknown		
776.584	unknown		
780.564*	PC	PC(34:2)[a,b,c]	$[M+Na]^+$
782.577*	PC	PC(36:4)[a,c]	$[M+H]^+$
		PC(34:1)[a,b]	$[M+Na]^+$
784.591*	PC	PC(36:3)[a,b,c]	$[M+H]^+$
786.610*	PC	PC(36:2)[a,b,c]	$[M+H]^+$
788.620*	PC	PC(36:1)[c]	$[M+H]^+$
790.585	unknown		
792.600	PC-O	PC(O-38:6)[LM,HMDB]	$[M+H]^+$
794.614	PC-O	PC(O-38:5)[LM,HMDB]	$[M+H]^+$
796.541	PC	PC(34:2)[a]	$[M+K]^+$
796.615	PC-O	PC(O-38:4)[LM,HMDB]	$[M+H]^+$
798.548	PC	PC(34:1)[a,b]	$[M+K]^+$
802.610	unknown		
803.617	unknown		
807.641*	SM	SM(40:2)[a,b]	$[M+Na]^+$
808.609	PC	PC(36:2)[b]	$[M+Na]^+$
810.610*	PC	PC(38:4)[c]	$[M+H]^+$
811.676*	SM	SM(40:0)[b]	$[M+Na]^+$
813.695*	SM	SM(42:2)[a,c]	$[M+H]^+$
815.710	SM	SM(42:1)[a,b]	$[M+H]^+$
816.597	PC-O	PC(O-38:5)[LM,HMDB]	$[M+Na]^+$

Table 3.3 continued from previous page

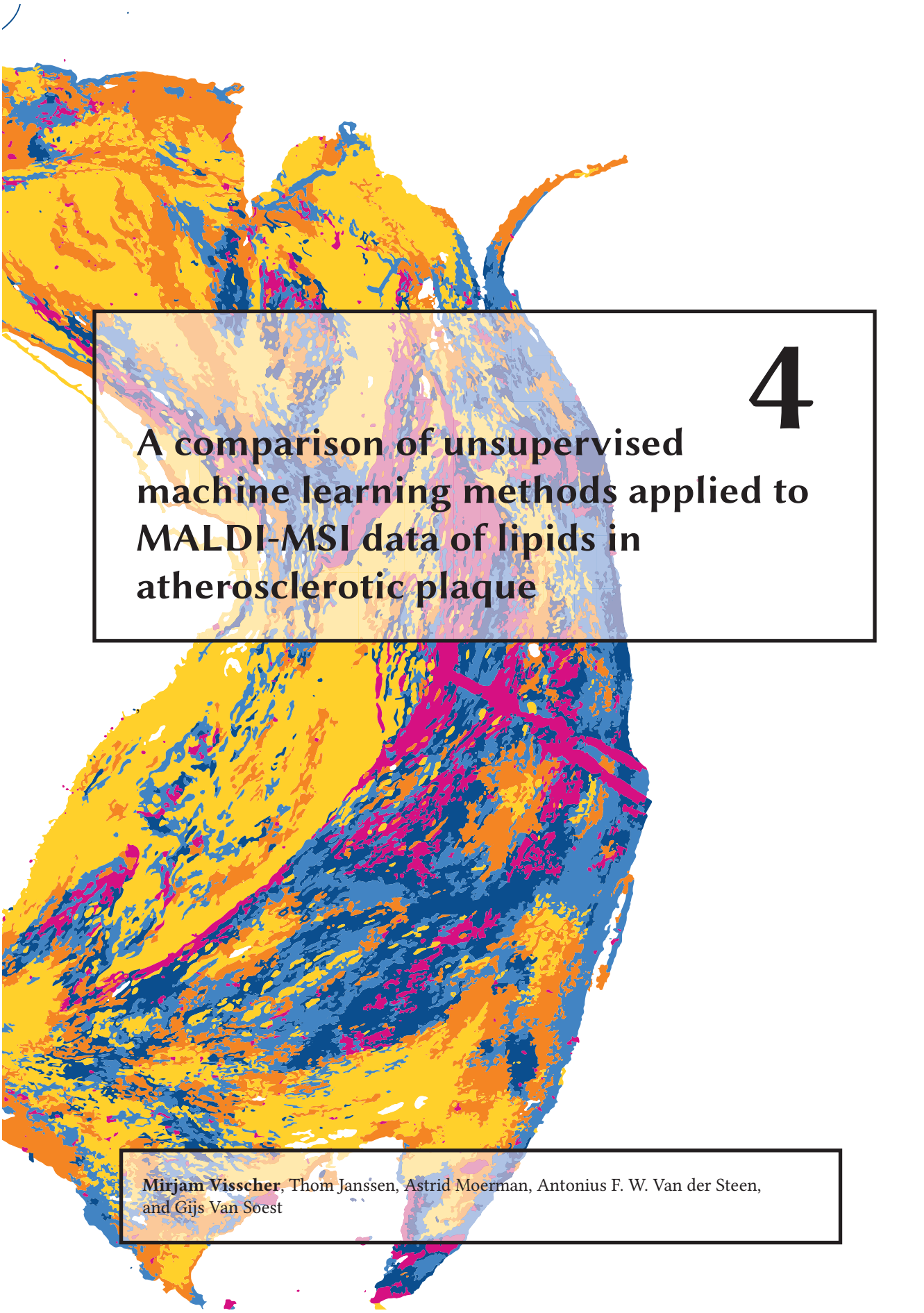
m/z ± 0.02	lipid class	ID	adduct
820.539*	PC	PC(36:4)[a]	$[M+K]^+$
822.555	PC	PC(36:3)[a]	$[M+K]^+$
823.592	SM	SM(40:2)[a]	$[M+K]^+$
824.584	unknown		
825.610	SM	SM(40:1)[a]	$[M+K]^+$
827.622	unknown		
827.709*	TAG	TAG(50:4)[a]	$[M+H]^+$
		TAG(48:1)[a]	$[M+Na]^+$
828.620	unknown		
829.619	unknown		
829.717*	TAG	TAG(50:3)[a]	$[M+H]^+$
		TAG(48:0)[a]	$[M+Na]^+$
830.574	PC	PC(40:8)[b]	$[M+H]^+$
		PC(38:5)[b]	$[M+Na]^+$
832.600	PC	PC(38:4)[b]	$[M+Na]^+$
834.637	unknown		
835.673*	SM	SM(44:5)[b]	$[M+H]^+$
		SM(42:2)[a,b]	$[M+Na]^+$
837.690	SM	SM(42:1)[a]	$[M+Na]^+$
839.590	unknown		
846.661	unknown		
847.593	unknown		
848.588	unknown		
848.676	unknown		
849.611	unknown		
850.605	unknown		
851.629	SM	SM(42:2)[a]	$[M+K]^+$
853.721*	TAG	TAG(52:5)[a,b]	$[M+H]^+$
		TAG(50:2)[a]	$[M+Na]^+$
855.592	unknown		
855.737	TAG	TAG(52:4)[LM,HMDB]	$[M+H]^+$
		TAG(50:1)[LM,HMDB]	$[M+Na]^+$
856.596	PC	PC(42:9)[HMDB]	$[M+H]^+$
		PC(40:6)[HMDB]	$[M+Na]^+$
857.596	unknown		

Table 3.3 continued from previous page

m/z ± 0.02	lipid class	ID	adduct
857.750*	TAG	TAG(52:3)[a]	$[M+H]^+$
		TAG(50:0)[a]	$[M+Na]^+$
860.774	unknown		
862.790	unknown		
870.684	PC	PC(42:2)[HMDB]	$[M+H]^+$
872.693	PC	PC(42:1)[HMDB]	$[M+H]^+$
877.738*	TAG	TAG(54:7)[a]	$[M+H]^+$
		TAG(52:4)[a]	$[M+Na]^+$
879.738*	TAG	TAG(54:6)[a]	$[M+H]^+$
		TAG(52:3)[a]	$[M+Na]^+$
881.760*	TAG	TAG(54:5)[a]	$[M+H]^+$
		TAG(52:2)[a]	$[M+Na]^+$
883.768*	TAG	TAG(54:4)[a]	$[M+H]^+$
		TAG(52:1)[a]	$[M+Na]^+$
885.781*	TAG	TAG(54:3)[a]	$[M+H]^+$
		TAG(52:0)[a]	$[M+Na]^+$
888.804	unknown		
889.726	unknown		
895.725	unknown		
897.736	PE-Cer	PE-Cer(t38:0)[b]	$[M+Na]^+$
901.569	unknown		
901.732*	TAG	TAG(56:9)[a]	$[M+H]^+$
		TAG(54:6)[a]	$[M+Na]^+$
903.751*	TAG	TAG(56:8)[a]	$[M+H]^+$
		TAG(54:5)[a]	$[M+Na]^+$
905.751*	TAG	TAG(56:7)[a]	$[M+H]^+$
		TAG(54:4)[a]	$[M+Na]^+$
907.773*	TAG	TAG(56:6)[a]	$[M+H]^+$
		TAG(54:3)[a]	$[M+Na]^+$
909.794*	TAG	TAG(56:5)[a]	$[M+H]^+$
		TAG(54:2)[a]	$[M+Na]^+$
923.750	unknown		
927.744*	TAG	TAG(58:10)[a]	$[M+H]^+$
		TAG(56:7)[a]	$[M+Na]^+$
929.768*	TAG	TAG(58:9)[a]	$[M+H]^+$
		TAG(56:6)[a]	$[M+Na]^+$

Table 3.3 continued from previous page

m/z ± 0.02	lipid class	ID	adduct
931.778*	TAG	TAG(58:8)[a]	$[M+H]^+$
		TAG(56:5)[a]	$[M+Na]^+$
932.570	unknown		
933.790*	TAG	TAG(58:7)[a]	$[M+H]^+$
		TAG(56:4)[a]	$[M+Na]^+$
940.770	unknown		
955.774*	TAG	TAG(60:10)[a]	$[M+H]^+$
		TAG(58:7)[a]	$[M+Na]^+$
956.570	unknown		
958.583	unknown		
980.570	unknown		
984.602	unknown		
986.769	unknown		
1,010.774	unknown		
1,012.783	unknown		
1,014.798	unknown		
1,034.772	unknown		
1,036.783	unknown		
1,038.805	unknown		
1,040.817	unknown		
1,042.824	unknown		
1,077.583	unknown		



4

A comparison of unsupervised machine learning methods applied to MALDI-MSI data of lipids in atherosclerotic plaque

Mirjam Visscher, Thom Janssen, Astrid Moerman, Antonius F. W. Van der Steen, and Gijs Van Soest

Abstract

Traditional processing of mass spectrometry imaging (MSI) data relies on a form of peak-picking in the physically continuous spectrum, followed by analysis using supervised or unsupervised processing methods, which were reviewed in Chapter 1. Data acquired by MSI are extremely high-dimensional and intrinsically nonlinear. These properties make these data sets challenging to analyze: most algorithms either work well only on low-dimensional or linear data, or require a large amount of memory to compute. In this chapter we investigate the use of an auto-encoder to reduce the dimensionality of the data to allow the use of a sophisticated manifold learning algorithm (t-distributed stochastic neighbor embedding), and compare the results with two traditional linear methods, relying on matrix factorization (non-negative matrix factorization) and clustering (spatial shrunken centroids). We demonstrate that, compared to histological analysis, analysis methods with higher complexity yield results that are qualitatively similar to traditional clustering methods.

4.1 Introduction

Mass spectrometry imaging (MSI) is a label-free molecular imaging technique that allows the visualization of the distribution of molecules throughout a sectioned sample with high chemical specificity. The technique is particularly suitable for exploratory analyses because it is capable of measuring a large variety of molecules in a single experiment. Mass spectrometry imaging relies on the generation of ions from a small area by means of: (1) a laser, in matrix-assisted laser desorption/ionization (MALDI), (2) a spray of charged droplets, in desorption electrospray ionization (DESI) or (3) by a focused ion beam, in secondary ion mass spectrometry (SIMS). Scanning the ionization region across a tissue slide, and analyzing the generated ion masses using a suitable spectrometer, results in a complex data set which can be represented as abundance images for all sampled ions in the section. All ionization and mass analysis techniques have their advantages and disadvantages, but they all produce very large, high-dimensional data sets, which are challenging to process.

MSI is often used to map molecules throughout a tissue sample with the goal of biomarker discovery. Some studies allow for the use of supervised data analysis methods, when either the masses to analyze are known, or the spatial distribution of target analytes. Examples of such research questions are discovering the difference in molecular content between tumorous and healthy tissues, or the detection of drugs in tissue. In such cases, data analysis can be relatively straightforward. However, for MSI experiments aimed at seeking underlying structures or patterns in tissue, an unsupervised method is more appropriate.

Most unsupervised machine learning algorithms attempt to represent a dataset using a reduced number of dimensions or variables while minimizing the loss of information. This serves the purpose of revealing underlying structures, patterns or trends in the data, which can be discovered by identifying high covariance or correlation in the data. Factorization techniques, like principal component analysis (PCA) and non-negative matrix factorization (NMF), are linear by nature and are less suitable if the data has a non-linear structure[23,133].

Manifold learning, such as t-Distributed Stochastic Neighbor Embedding (t-SNE) or self-organizing maps (SOM), take into account non-linear structures in data. These techniques assume that the dimensionality of the data is high in the measurement space, but underlying structures can be projected, through a nonlinear mapping, on a lower-dimensional manifold. The goal of the technique is to discover this non-linear subspace[23].

In MSI data, manifold learning techniques are able to pack more data variation into the lower dimensional representation than factorization techniques, meaning linear techniques often require more dimensions to represent the information compared to manifold learning techniques. Additionally, most linear techniques optimize their analysis by separating dissimilar data points, as opposed to grouping similar data which manifold learning techniques aim to[23,27].

A downside of manifold learning, however, is that the nonlinear mapping to the manifold can be relatively complex. This results in an indirect relationship between the new low dimensional representation and the underlying mass spectra, complicating the chemical interpretation of processed data[23]. In factorization methods, on the other

hand, variables contributing to each component emerge naturally when applying the algorithm.

In this chapter we investigate the use of both factorization and manifold learning algorithms on three advanced atherosclerotic carotid tissue sections. We cluster the MSI data using non-negative matrix factorization (NMF), spatial shrunken centroids (SSC), auto-encoder (AE), t-Distributed stochastic neighbor embedding (t-SNE), and a combination of the last two methods. We compare the results to histology based segmentation of the tissue sections and evaluate the algorithms in terms of performance as well as the usability and computational costs.

4.2 Theory

4.2.1 Non-negative matrix factorization

Non-negative matrix factorization (NMF) is a matrix factorization technique used in unsupervised data analysis, which attempts to compute a lower dimensional representation of a high dimensional dataset. All factorization methods are based on linear unmixing and can be mathematically written as:

$$X \approxeq AY + E \tag{4.1}$$

where X is an $n \times m$ matrix and contains the non-negative values of m *m/z* spectra in n pixels. Y is the lower-dimensional basis matrix of size $k \times n$, $k << m$, A is the coefficient matrix of size $m \times k$, mapping the low-dimensional to high-dimensional space, and E is the residual or the error[23,105].

A well-known matrix factorization technique is principal components analysis (PCA). PCA reduces the number of dimensions while retaining the variation: it maximizes the orthogonal variation between different principal components. PCA allows the signs of entries A and Y to be both positive and negative, which makes the results less intuitive to interpret. NMF does not permit entry of negative values, and is limited to only additive combinations, but makes no assumptions about the statistical dependencies. This causes the results of NMF to be a more intuitive sparse representation of parts of the original image[105]. The process is iterative and tries to minimize the error E.

4.2.2 Spatial shrunken centroids

Spatial shrunken centroids (SSC) is a statistical model-based method developed especially for MSI data[26]. Most data analysis methods used in MSI are "spatially unaware", i.e. do not take the relative location of the pixels into account. SSC takes into consideration spectral similarities of spatially proximate pixels. The method characterizes the statistical probability of a pixel belonging to a specific cluster taking into account spectral shape as well as local and global spatial structures in the data.

N is the number of spectra and pixels in the data set, with (i, j) the location of the pixel in space and $x_{ij} = \{x_{ijm}, m = 1, ..., M\}$ the spectrum at that location containing M *m/z* values and K is the number of clusters in the data set. First, the algorithm calculates the average spectrum of all clusters k: $\bar{x}_k$. These clusters are initialized

randomly and their location changes iteratively. Subsequently the average spectrum of a cluster, $\bar{x}_k$, is compared to the average spectrum of the entire image, $\bar{x}$. Clusters that show deviation from the average spectrum possible contain interesting biomarkers[134,135]

To determine the difference between $\bar{x}$ en $\bar{x}_k$, t-statistics is calculated for every *m/z* value, m, for every $\bar{x}_k$, as follow:

$$t_{km} = \frac{\bar{x}_{km} - \bar{x}_m}{\hat{\tau}_m \sqrt{\frac{1}{N_k} - \frac{1}{\sum_{k=1}^{K} N_k}}} \tag{4.2}$$

where $\hat{\tau}_m$ is the pooled estimate of the within-class standard deviation for *m/z* value m. Consequently, this t-statistic is shrunken to 0 using a soft thresholding operator[136]:

$$t'_{km} = sign(t_{km})(|t_{km}| - s)_+ \tag{4.3}$$

The value of the shrunken t-statistic t'_{km} determines which *m/z* value is specific for that cluster and thus which features are different from the average spectrum. s is the shrinkage parameter, higher values of s causes more t'_{km} to be set to 0[26].

$$t'_{km} = \begin{cases} > 0 & \text{specific for cluster k} \\ < 0 & \text{not-specific for cluster k} \\ 0 & \text{not informative.} \end{cases}$$

The average intensity of the spectrum of cluster k is determined as follows[136]:

$$\bar{x}'_{km} = \bar{x}_m + t'_{km} \hat{\tau}_m \sqrt{\frac{1}{N_k} - \frac{1}{\sum_{k=1}^{K} N_k}} \tag{4.4}$$

where $\bar{x}'_k$ is the shrunken centroid for cluster k. Depending on these values, pixels can be assigned to a cluster, which is determined by the equality between the spectrum of a pixel and $\bar{x}'_k$. This is then entered into the spectral distance function[134]:

$$d(x_{ij}, \bar{x}'_k) = \sum \alpha_{\delta_i \delta_j} \|x_{(i+\delta_i)(j+\delta_j)} - \bar{x}'_k\|^2 \tag{4.5}$$

where d is the difference between the spectrum at pixel (i, j) and $\bar{x}'_k$, (δ_i, δ_j) is the geometric distance vector from pixel (i, j), and α denotes the gaussian weights.

$$\alpha_{\delta_i \delta_j}(x_{ij}) = exp\left\{ -\frac{\delta_i^2 + \delta_j^2}{\sigma^2} \right\} \beta_{\delta_i \delta_j}(x_{ij}). \tag{4.6}$$

with:

$$\sigma = (2r + 1)/4 \tag{4.7}$$

and

$$\beta_{\delta_i \delta_j}(x_{ij}) = exp\left\{ -\frac{1}{2\lambda^2} \|x_{i+\delta_i, j+\delta_j} - \bar{x}'_k\|^2 \right\}. \tag{4.8}$$

The Gaussian weights α ensure that pixels that are far apart spatially are assigned lower weights compared to pixels in close proximity, such that local structures are maintained. The weighting factor β accounts for the similarity between the local spectrum at $(i + \delta_i, j + \delta_j)$ and the spectrum of the shrunken centroid $\bar{x}'_k$ [134,135]. These weights are iteratively determined updating the probability of pixels being assigned to a segments.

4.2.3 Auto-encoder

An auto-encoder (AE) is a type of machine learning algorithm which attempts to learn itself; mathematically it is trying to learn its identity matrix. Modern machine learning frequently employs neural networks, which consist of one or more connected layers containing nodes. In its simplest form, every node has a weight, $\mathbf{W}$, and bias $\mathbf{b}$, the minimum weight a node needs to have to contribute to the next layer. An activation function σ ensures that $\mathbf{W}$ and $\mathbf{b}$ are contained within a chosen domain.

An auto-encoder consists of 2 parts:

- Encoder: $g(\phi) : \mathbf{X} \Rightarrow \mathbf{Y}$
- Decoder: $f(\theta) : \mathbf{Y} \Rightarrow \mathbf{X}'$

Where $g(\phi), f(\theta)$, represent the transformations through the different hidden layers in the neural network in the encoder and decoder. $\mathbf{X}$ is the input, $\mathbf{Y}$ is the encoded state and $\mathbf{X}'$ is the output. In the ideal situation $\mathbf{X}$ and $\mathbf{X}'$ are equal. Figure 4.1 shows a schematic representation of an auto-encoder network.

The auto-encoder takes input $\mathbf{x} \in \mathbb{R}^{\mathbf{d}} = \mathbf{X}$ and encodes it to $\mathbf{y} \in \mathbb{R}^{\mathbf{k}} = \mathbf{Y}$ through:

$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}), \tag{4.9}$$

with $\mathbf{x}$ the input and the sigmoid function as activation function:

$$\sigma = \frac{1}{1 + e^{-\mathbf{x}}}. \tag{4.10}$$

with $\mathbf{d}$ the dimension of the input and $\mathbf{k}$ the dimension in the encoded state. Consequently, the decoder transforms $\mathbf{y}$ to the output $\mathbf{X}'$ through:

$$\mathbf{x}' = \sigma'(\mathbf{W}'\mathbf{y} + \mathbf{b}'), \tag{4.11}$$

where $\sigma', \mathbf{W}', \mathbf{b}'$ are independent from $\sigma, \mathbf{W}, \mathbf{b}$.

The auto-encoder is trained by minimizing the difference between $\mathbf{x}$ and $\mathbf{x}'$, the loss function have various forms but in this chapter we use mean square error:

$$\mathcal{L}(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|^{\mathbf{2}} = \|\mathbf{x} - \sigma'(\mathbf{W}'(\sigma(\mathbf{W}\mathbf{x} + \mathbf{b})) + \mathbf{b}'\|^{\mathbf{2}}. \tag{4.12}$$

This function is minimized using the gradient descent technique.

Dimensionality of the data is reduced when the encoder, $g(\phi)$, is chosen such that $\mathbf{d} > \mathbf{k}$, the encoded state $\mathbf{h}$ then has a lower number of dimension then the input $\mathbf{x}$. This allows auto-encoders to be used as a clustering method, where $\mathbf{k}$ defines the number of clusters.

4.2.4 t-Distributed stochastic neighbor embedding

t-Distributed stochastic neighbor embedding (t-SNE) is a manifold learning technique, developed by Van der Maaten[27], that is capable of projecting high dimensional data in a 2D or 3D space.

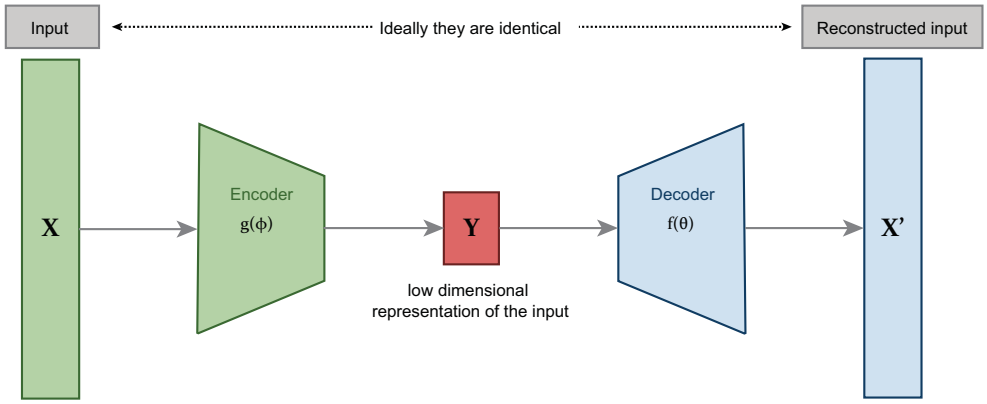


Figure 4.1: **Schematic representation of an auto-encoder.** On the left, the input $\mathbf{X}$ with dimension $\mathbf{d}$. The input goes through the encoder $g(\phi)$ and is transformed to $\mathbf{Y}$ with dimension $\mathbf{p}$. In this case $\mathbf{d} > \mathbf{k}$. In the ideal case $\mathbf{X}'$, a transformation of $\mathbf{Y}$ through $f(\theta)$, is equal to $\mathbf{X}$

The original high-dimensional data set is defined as $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\}$, which is converted to a lower dimensional representation $\mathbf{Y} = \{\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_n}\}$. The algorithm works in two steps. First, the Euclidean distances between pairs of data points, x_i and x_j, are converted into conditional probability based on their similarity. As a result, similar points have a higher probability of being neighbors than dissimilar pairs. The similarity, p_{ij}, between point x_i and x_j is written as:

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2/2\sigma^2)}, \tag{4.13}$$

see Fig. 4.2. Points that are similar will be closer together on the normal distribution curve, therefore the value of p_{ij} will be higher. p_{ii} does not have any meaning and therefore it is defined as 0. When all similarities between points are determined, the average between p_{ij} and p_{ji} is defined t o avert anti-symmetric effects, points that are far apart or that are located in less dense areas would otherwise have less contribution to the cost-function. This results in a matrix with p_{ij} similarity scores:

$$p_{ij} = \begin{array}{c} \\ x_1 \\ \vdots \\ \vdots \\ x_n \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & \cdots & \cdots & x_n \end{array} \\ \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \end{array} \tag{4.14}$$

For the second step in the t-SNE algorithm the data points of $\mathbf{Y}$ are randomly plotted in the low-dimensional representation space of the data. At this point the similarity between points is determined again in the same fashion as in the high-dimensional space. The difference is that this time a t-distribution is used to define the similarity, which ensures that not all clusters will be centered around the middle of the space, see Fig. 4.2.

This results in a second matrix of similarity scores, q_{ji}:

$$q_{ij} = \begin{array}{c} \\ y_1 \\ \vdots \\ \vdots \\ y_n \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & \cdots & \cdots & y_n \end{array} \\ \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{pmatrix} \end{array} \tag{4.15}$$

with q_{ij} defined as:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{4.16}$$

To ensure that matrix q_{ij} approaches p_{ij} as much as possible, the Kullback-Leibler divergence is used:

$$C = KL(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{4.17}$$

with gradient:

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j). \tag{4.18}$$

The gradient, C, is then minimized using the gradient-descent method. The result is a map reflecting the similarities between low dimensional representation $\mathbf{Y}$ and the high dimensional input $\mathbf{X}$. As a result of the randomization in the second step, the results from repeated t-SNE analyses cannot be directly compared, although the clustering structure is similar.

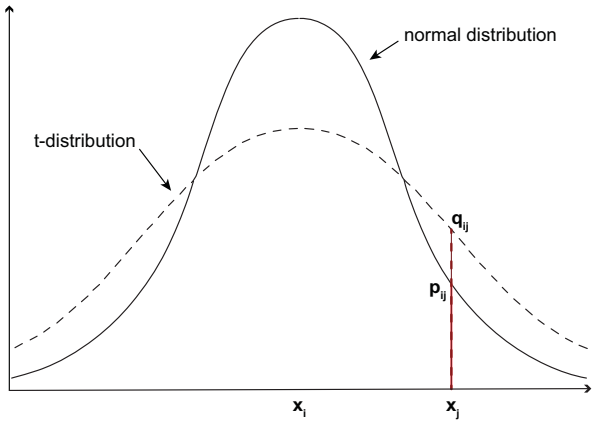


Figure 4.2: **Determination of similarity between pairs** p_{ij} is determined by the distance between x_j and the normal curve with point x_i as its center, and q_{ij} is determined by the distance between x_j and the t-statistics curve with point x_i as its center

One of the drawbacks of t-SNE is that it works best when the input dimensionality of the data is smaller than 50[27]. Therefore in this chapter we combine this technique with an auto-encoder which first reduces the initial dimensions to 50 before t-SNE is applied.

4.3 Methods

4.3.1 Tissue collection and processing

Three human carotid endarterectomy (CEA) were surgically removed using a protocol preserving an intact lumen and plaque morphology[73], according to the ethical guidelines sanctioned by the Ethics Board of Erasmus MC. The specimens were snap frozen and stored at -80 °C until further use. The CEA specimens were embedded into 10% porcine gelatin (type A, Sigma-Aldrich, The Netherlands) and cryo-cut into 10 μm cross-sections (CM3050 S, Leica Biosystems (cutting temps: OT -21 °C; CT -19 °C)). Sections were thaw-mounted onto glass slides and stored at -80 °C until the MSI measurement and/or histology staining.

Samples were prepared for MALDI-MSI experiments by 10 minutes desiccation at RT and sublimation of 2,5-dihydroxybenzoic acid (DHB) matrix (home-built sublimation system as described in Dekker et al.[74]). Sections were measured using a Synapt G2Si-TOF system (Waters, Manchester, UK) operated with Waters Research Enabled Software suite (WREnS) in positive resolution mode, using a 2000 Hz Nd:YAG (355 nm) laser and a pixel size of 45×45 μm^2. 100 laser shots were fired per pixel measuring a mass range of 300-1200 *m/z*. Data was captured using MassLynx v4.2 software (Waters, Manchester, UK), and data was exported to an imzML format using HDI v1.4.

4.3.2 Histology based segmentation

Tissue sections adjacent to the MALDI-MSI section were histochemically stained by Miller's elastic stain, Martius scarlet blue trichrome and hematoxylin-eosin. Using these stainings, the tissue was delineated into segments of necrotic core (NC), foam cells (FC), calcium, fibrin, erythrocytes and sectioning artefacts.

4.3.3 Unsupervised machine learning

Pre-processing of the data was performed using the Cardinal toolbox in R[137]. Data was pre-processed by means of total-ion current (TIC) normalization, filtering with a Savitsky-Golay filter, peak calibration using the mean spectrum as a reference and peak-picking using a signal-to-noise-ratio (SNR) of 10. The background pixels were removed from the data by means of SSC with two cluster (s = 30 and r = 1), except with NMF.

Non negative matrix factorization

Non-negative matrix factorization (NMF) dimension reduction was performed in Matlab using an NMF toolbox for biological data-mining[106]. The background pixels were removed from the data by means of NMF with two clusters. Thereafter, the optimum

number of components present in the tissue pixels was determined based on dispersion coefficients[107]. Clustering was achieved by assigning a pixel to the cluster with the highest NMF coefficient value.

Linear factorization techniques work well under the assumptions that the measurement noise is consistent across the intensity range. However, in MSI data, the variability of higher intensity peaks in MSI data was larger than low intensity peaks. Using a log-based variance-stabilizing normalization (VSN)[104] ensured that that the data is consistent with the statistical assumption of the algorithm.

Spatial Shrunken Centroids

SSC was performed using the Cardinal toolbox in R[137], using a gaussian method. Values of K, the initial number of segments, s, the shrinkage parameter and r, the spatial smoothing radius, are shown in Table 4.1. SSC determines if clusters are statistically significant, if not, the number of clusters is reduced.

Table 4.1: Parameters for spatial shrunken centroids

Parameter	Value
Method	Gaussian
r	5
s	20
K	10

Auto-encoder

The auto-encoder neural network was trained in R using the 'Keras' library. The architecture of the network is schematically depicted in Fig. 4.3. The algorithm used several parameters shown in Table 4.2. The number of dimensions of the lower dimensional representation $\mathbf{F}$ was chosen such that it is the same as the number of clusters determined by SSC, plus one extra to account for the background, when the AE was used before t-SNE the number of dimensions is 50.

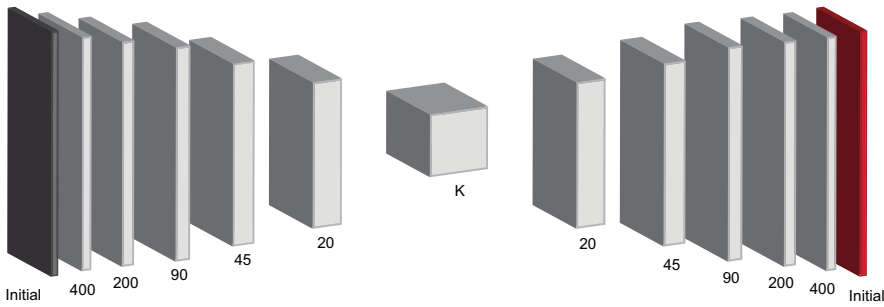


Figure 4.3: **Architecture of the auto-encoder.** On the left, the input with the initial number of dimension (usually ≈ 600 *m/z* values). The input is encoded using five layers between in input and the low dimensional representation, F, with K dimensions. The input is the reconstructed again using the decoder, containing the same five layers.

Table 4.2: Parameters auto-encoder

Parameter	Value
Activation	Sigmoid
Optimizer	Adam
Epochs	150
Batch size	5
Loss function	Mean squared error

t-SNE

t-SNE was determined using the package Rtsne[27]. t-SNE was used twice, once on the pre-processed data, and a second time after first using AE to reduce the initial dimensions to 50. The algorithm used the parameters shown in Table 4.3.

Table 4.3: Parameters t-SNE

Parameter	Value
Initial dimensions	50 or $\approx$600
Reduced dimensions	3
Perplexity (σ)	15

The result of of t-SNE algorithm was a 3D scatterplot, which was transformed into a color image, such that all three dimensions represent a color channel, R (red), G (green) and B (blue), as also shown in Fonville *et al.*[28].

4.4 Results

Five different unsupervised machine learning methods were used to cluster three carotid atherosclerotic tissue sections. The first section, shown in Fig. 4.4a, is a more intermediate type of lesion, containing a relatively small necrotic core (NC), multiple areas of foam cells (FC), an area with erythrocytes contained within an area of fibrin. There was an area with the media layer around the plaque, which is not depicted in Fig. 4.4a, but is present in Fig. 4.4b-f.

Figure 4.4b showed the result of NMF, containing 5 well defined clusters, with some mixing of cluster 1 and 3. In Fig. 4.4b the SSC showed similarly shaped clusters as the NMF. There were some differences though, the first is SSC contains 7 clusters, opposed to the NMF which contains only 5. One of the additional clusters, number 2, was located around the small fibrin spot. Secondly, there was a difference around the FC areas where the clusters are distributed differently and the SSC has an additional cluster, number 3. Finally, the clusters in SSC were smoother compared to the more noisy NMF clusters. Neither method showed a spatial pattern which can be clearly associated with particular histological segments.

In Fig. 4.4e and f the result of t-SNE without the auto-encoder and with the auto-encoder as additional pre-processing step are shown. It was clear that the auto-encoder causes a blurring effect in the results of the t-SNE algorithm. This reduced the amount of noise, which is clearly visible in Fig. 4.4e, but also generated edge effects, on the edges

of the tissue and the image itself. Moreover, it caused less sharp delineation between the structures in the data. Overall, the structures found in the tissue section are similar to those found by NMF and SSC.

Section 2, shown in Fig. 4.5a, was a more advanced lesion almost completely consisting of NC, with some fibrin areas, calcifications, and a sectioning artifacts at the top. Both NMF and SSC recognized these artifacts as being background and removed these pixels from the data set in the first round of clustering, with only two clusters.

Figure 4.5b and c showed the results of the NMF and SSC algorithm, respectively. Contrary to section 1 now the NMF algorithm assigned more clusters than the SSC algorithm. Cluster 2 and 5 of the NMF, and cluster 1 and 6, were combined together in SSC and form cluster 3 and cluster 2, respectively. Apart from those two differences, the results of both methods were similar. NMF cluster 7 and SSC cluster 4 seem to be related to fibrin, and NMF cluster 2 and 5 and SSC cluster 3 were co-localized with necrotic regions. Nevertheless, there was no clear resemblance of the histological segmentation to be recognized in the clustering algorithm, which showed more variation than the histology does.

In Fig. 4.5e and f and in Fig. 4.6e and f, the result of t-SNE with and without the auto-encoder were shown for CEA sections 2 and 3, showing the same blurring effects

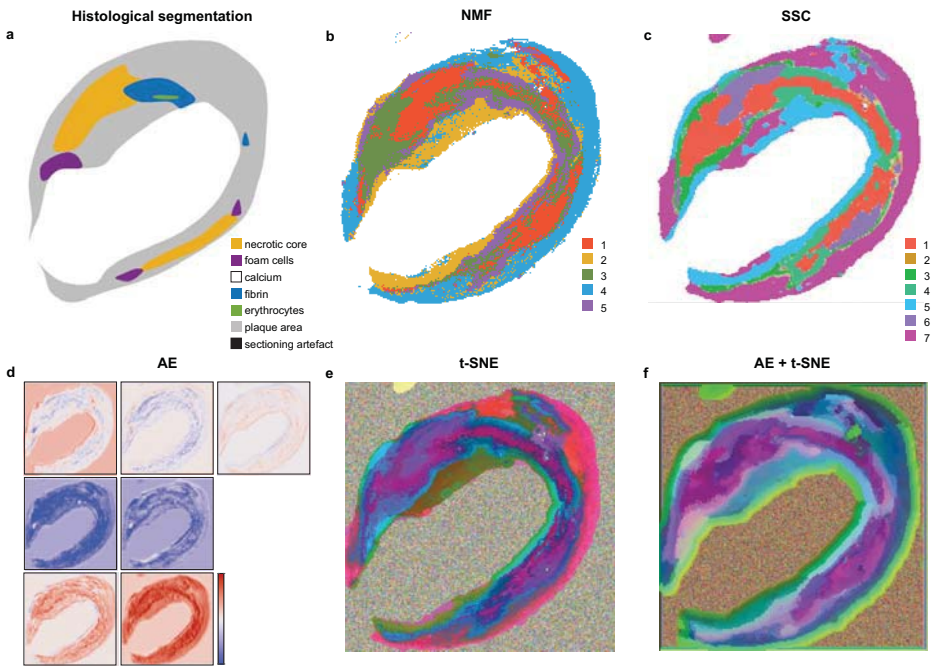


Figure 4.4: **Histology based segmentation and 5 clustering methods on CEA section 1. a)** Histology based segmentation on staining HE, Miller and MSB. Delineating; necrotic core, foam cells, calcification, fibrin, erythrocytes, plaque, and sectioning artifacts. **b)** Segmentation using non-negative matrix factorization (NMF) (N = 5) **c)** Segmentation using spatially shrunken centroids (SSC) (N = 7) **d)** Dimension reduction using auto-encoder (AE) (N = 8) **e)** Segmentation using t-distributed stochastic neighbor embedding (t-SNE) (3D) **f)** Segmentation using a combination of AE (N = 50) and subsequently t-SNE (3D).

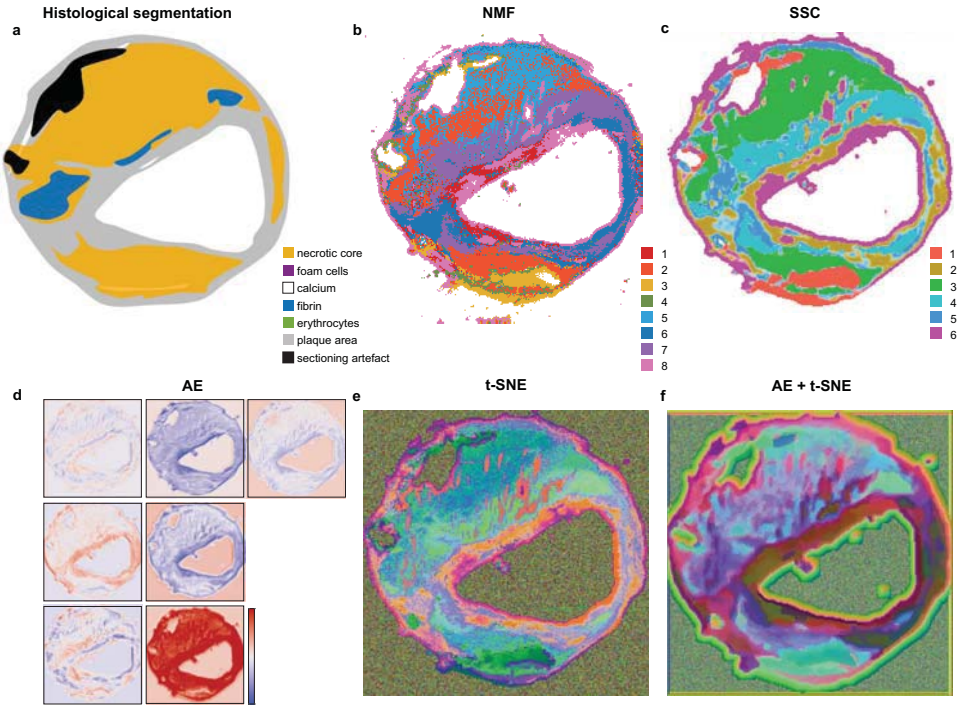


Figure 4.5: **Histology based segmentation and 5 clustering methods on CEA section 2. a)** Histology based segmentation on staining HE, Miller and MSB. Delineating; necrotic core, foam cells, calcification, fibrin, erythrocytes, plaque, and sectioning artifacts. **b)** Segmentation using non-negative matrix factorization (NMF) (N = 7) **c)** Segmentation using spatially shrunken centroids (SSC) (N = 6) **d)** Dimension reduction using auto-encoder (AE) (N = 7) **e)** Segmentation using t-distributed stochastic neighbor embedding (t-SNE) (3D) **f)** Segmentation using a combination of AE (N = 50) and subsequently t-SNE (3D).

as were present in CEA section 1, Fig. 4.4e and f.

Section 3, depicted in Fig. 4.6, was a section which has possibly experienced a recent plaque rupture, as it contained a large area of fresh erythrocytes close to the lumen. Furthermore, it contains two necrotic areas, which were medium size, a small area with foam cells and multiple small areas of fibrin. The structure of the fresh erythrocytes was clearly delineated by almost all unsupervised algorithms indicating that its molecular content was distinctly different to the rest of the plaque.

Figure 4.4d, Fig. 4.5d,Fig. 4.6d showed the low dimensional representations of the auto-encoder. The images were scaled to the same colormap, maximized at the 95th percentile of absolute values, with 0 in the middle. The low dimensional representation showed similar structures in all images, with different scaling in intensity.

4.4.1 Computation time

Pre-processing was performed on an Amazon EC3 cloud computer with 256 GB RAM. Clustering methods were performed on a computer with 16 GB RAM. The computation times for all methods are shown in Table 4.4.

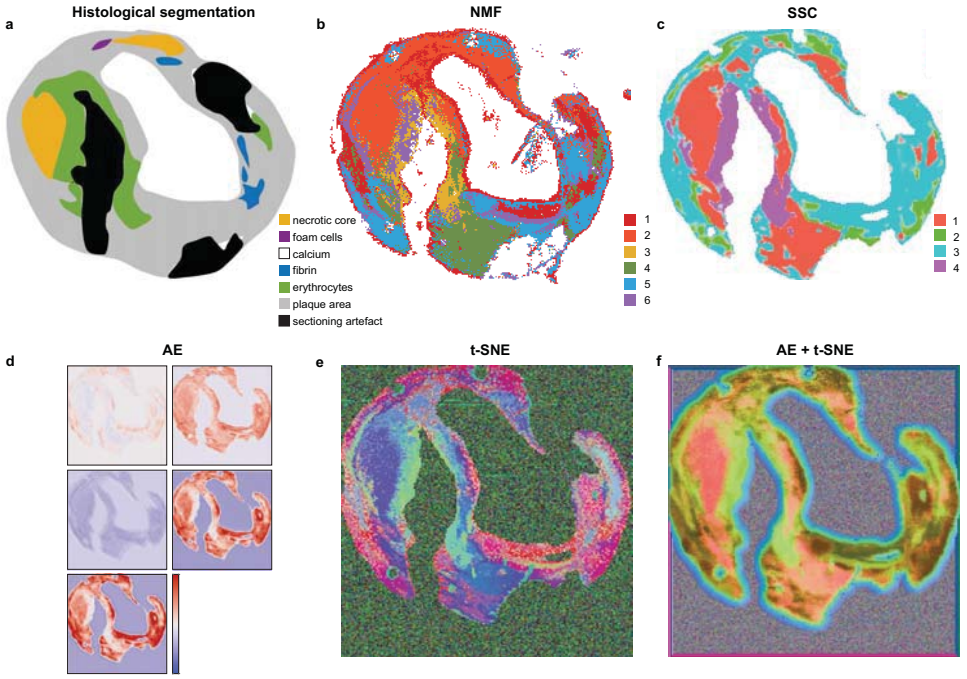


Figure 4.6: **Histology based segmentation and 5 clustering methods on CEA section 3. a)** Histology based segmentation on staining HE, Miller and MSB. Delineating; necrotic core, foam cells, calcification, fibrin, erythrocytes, plaque, and sectioning artifacts. **b)** Segmentation using non-negative matrix factorization (NMF) (N = 5) **c)** Segmentation using spatially shrunken centroids (SSC) (N = 4) **d)** Dimension reduction using auto-encoder (AE) (N = 5) **e)** Segmentation using t-distributed stochastic neighbor embedding (t-SNE) (3D) **f)** Segmentation using a combination of AE (N = 50) and subsequently t-SNE (3D).

4.5 Discussion

In this chapter we evaluated five different unsupervised machine learning methods, applied to three carotid atherosclerotic cross-sections. Although the methods are very different mathematically, the results were very similar, showing corresponding underlying structures regardless of the method used. Different unsupervised data analysis methods consistently identified structures within homogeneous areas in the histology classification, exhibiting the molecular heterogeneity of these sections and the potential wealth of information present in these data sets.

In comparing these methods, we used histological classification as a standard. We observed that all methods identified more structure than was visible in histological stainings annotated by an experienced pathology image reader. These structures were consistent between methods. We did not have access to an independent gold standard for clustering tissue type as a function of its lipid composition, and therefore could not perform a quantitative comparison between the various methods.

One of the main advantages of traditional clustering methods like SSC and NMF is the direct relation of the ouput of these methods to the mass spectra. This makes the results easy to interpret in terms of molecular abundance and makes translations

Table 4.4: Computational times of different algorithms

Algorithm	Computational time (min)
Pre-processing	40
NMF	6
Spatial shrunken centroids	15
Auto-encoder	30
t-SNE	40

towards pathological and clinical relevance of results relatively straightforward. In deep learning methods, and also with t-SNE, this translation of clustering image to *m/z* value contribution is not trivial.

Secondly, computational cost of SSC and NMF was less than half compared to AE and t-SNE, while the results in the cases presented here are qualitatively comparable. The use of linear or non-linear methods did not seem to affect the result of the algorithm significantly, if the data was pre-processed accordingly. The global structures found in the data were very similar.

In MSI, the variability of the noise scales with the intensity of the spectrum. The use of VSN ensures that the noise level is consistent throughout the measurement, which allows the NMF algorithm to perform well. Figure 4.7 shows an example comparing the performance of NMF with and without the log transformation. This illustrated the power of performing this correction to the data which is a very simple and inexpensive way to allow the use of linear methods on this non-linear data set.

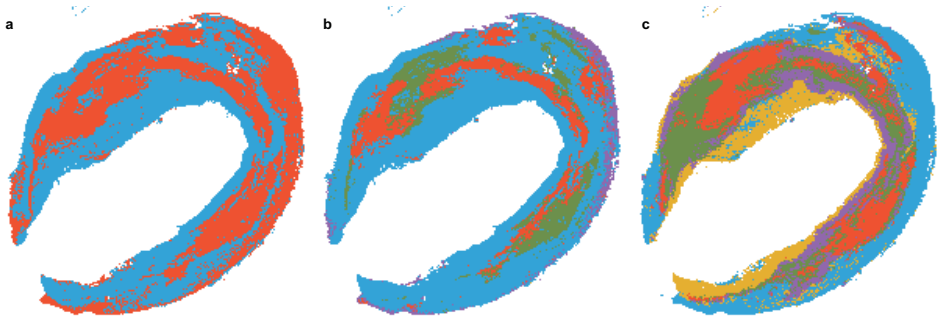


Figure 4.7: **Non-negative matrix factorization (NMF) in CEA section 1. a)** Segmentation using NMF without VSN (N = 2, determined by dispersion coefficients) **b)** Segmentation using NMF without VSN (N = 5) **c)** Segmentation using NMF with VSN (N = 5, determined by dispersion coefficients)

SSC is a 'soft' segmentation method, meaning that it allows pixels to be assigned to multiple clusters simultaneously, and takes this into account while assigning colors. This contributes to less sharp edges between the clusters. Additionally, because the algorithm takes into account the neighborhood of the pixels while assigning clusters, the clustering is less noisy compared to NMF, which only looks at individual pixels.

An advantage of SSC was that it gives statistical values for every *m/z* value, high values, for peaks that are distinctive for that specific cluster, but also negative values for peaks that are negatively associated with that cluster. This was different to NMF in the

sense that NMF only assigns positive values, therefore it only gives information about *m/z* values that contribute to that cluster.

With auto-encoders, the reconstructed image was compared to the input, which is a measure for the quality of the neural network. The agreement between the original and reconstructed images in our data set was 99,95%, meaning the lower dimension representation is a successful presentation of the input.

The results of the auto-encoder showed similar patterns in several images of the low dimensional representation. This was a sign that the algorithm is possibly separating dimensions based on spectral intensity opposed to spectral content, which could be tested by using a different normalization. Alternatively, it could be a sign of overgeneralization of the algorithm. Because of the relatively simple structure of the neural network, the latter is most likely in this case. This could be further investigated by extending the architecture of the network.

t-SNE was used with and without an auto-encoder as pre-processing step. Both methods showed similar patterns of clusters, however when pre-processed with the auto-encoder, the clusters were less delineated and show edge effects at the edges of the tissue and the image itself. The clustering without the auto-encoder showed more isolated pixels assigned as a different cluster within a cluster.

Furthermore, the auto-encoder in its current implementation requires the input images to be square. After removing the background, we therefore set the background pixels to zero and pad to equal x and y dimensions. This has an effect on the t-SNE algorithm, which tries to cluster the background together in one big cluster, however, there were so many background pixels that it created noise in the background when transformed to an RGB image. t-SNE itself however did not require a square input. Removing these background pixels may improve the results of the t-SNE algorithm, also increasing the space allowed to distribute the remaining tissue-related pixels.

Recently, researchers have tried to apply manifold learning algorithms on MSI data while trying to maintain the spectral information in the process or by calculating the spectra of assigned areas after clustering[138,139]. An advantage of the proposal of Abdelmoula *et al.* is that the method does not require pre-processing, which is time consuming and may introduce biases into the dataset. Dexter *et al.* proposed a neural network trained t-SNE algorithm which allows uniform color coding of large data sets. The advantage of this method is that new data can be added without re-running the complete algorithm in case of combined data sets, significantly reducing the computational costs and facilitating consistent interpretation.

In this chapter we evaluated the performance of unsupervised methods on single data sets. The results of these methods is similar but when extending to larger data sets, containing multiple tissue sections, results may be different. Computing power certainly will present a challenge when significantly increasing the data size, and our results suggests that gains of increased complexity in the analysis may not be worth the extra cost.

4.6 Conclusion

In our efforts to compare the performance of different unsupervised machine learning methods, we conclude that more simple algorithms do not under perform compared to more complex methods. Additionally, these simpler methods are more intuitive to interpret, especially regarding the spectral contributions towards different structures in the data. Also the computational costs are significantly lower.

5

Micro spectroscopic photoacoustic (μsPA) imaging of advanced carotid atherosclerosis

This chapter is based on:
Sophinese Iskander-Rizk*, **Mirjam Visscher***, Astrid Moerman, Suze-Anne Korteland, Kim van der Heiden, Antonius F.W. van der Steen and Gijs van Soest: Micro spectroscopic photoacoustic (μsPA) imaging of advanced carotid atherosclerosis, *Photoacoustics*, June 2021
*these authors contributed equally

Abstract

Atherosclerosis is a lipid-driven and an inflammatory disease of the artery walls. The composition of atherosclerotic plaque stratifies the risk of a specific plaque to cause a cardiovascular event. In an optical resolution photoacoustic microscopy setup, of 45 μm resolution, we extracted plaque lipid photoacoustic (PA) spectral signatures of human endarterectomy samples in the range of 1150 to 1240 nm, using matrix assisted laser desorption ionization mass spectrometry imaging as a reference. We found plaque PA signals to correlate best with sphingomyelins and cholesteryl esters. PA signal spectral variations within the plaque area were compared to reference molecular patterns and absorption spectra of lipid laboratory standards. Variability in the lipid spectroscopic features extracted by principal component analysis of all samples revealed three distinct components with peaks at: 1164, 1188, 1196 and 1210 nm. This result will guide the development of PA-based atherosclerosis disease staging capitalizing on lipidomics of atherosclerotic tissue.

5.1 Introduction

Cardiovascular diseases, predominantly associated with advanced atherosclerosis, account for about 50% of deaths worldwide and are a major cause of population morbidity [140,141]. Atherosclerosis gradually progresses over time; substances in the blood, such as lipids and inflammatory cells, are deposited at sites of dysfunctional endothelium, accumulating into the arterial wall and forming so-called plaques. These plaques may narrow the vessel, hindering blood flow. Unstable plaques can rupture; contact between plaque contents and blood can cause a thrombus to form, interrupting the blood supply to organs causing lethal events such as a Stroke or Myocardial Infarction.

The development of methods for monitoring the progress of atherosclerosis to detect and predict unstable plaque formation is an active research area which aims to prevent acute ischemic events, thereby reducing the disease burden of atherosclerosis[142–145]. Hypotheses regarding the development of the disease to unstable lesion types and triggers for rupture have been formed and reformed over time, describing atherosclerosis as a multi-factorial inflammatory disease[146–152]. Atherosclerotic plaques are classified into different stages based on plaque composition and size with histology as a golden standard for assessment[7]. In autopsy studies of sudden cardiac death cases[7], the plaque phenotype which caused most thrombotic events has been described as a large lipid-rich necrotic core contained by a thin fibrous cap, named a thin-cap fibroatheroma.

Lipids are involved in all stages of the disease. Lesions first appear as a fatty streak phenotype, which consists of accumulated smooth muscle cells and macrophages having internalized oxidized lipoproteins. The advanced stage includes a lipid-rich necrotic core, with a very heterogeneous lipid composition[13,66,102]. Previous research has studied the relation between plaque phenotype and evolution on one hand, and lipid composition on the other[13,153,154]. This opens the perspective of lipid typing of plaques *in vivo*, which could potentially be applied for diagnosis or prognosis, and used to guide treatment decisions in patients with coronary and carotid atherosclerosis. Current imaging modalities enabling atherosclerosis-related lipid identification *in vivo* are low resolution MRI[119,142], and catheter-based near-infrared spectroscopy (NIRS)[155], which is clinically available, and catheter-based photoacoustic (PA) imaging[55]. NIRS, in its commercial realization, derives the probability of the presence of lipid-core plaque[156–158] based on detection of scattered light, and does not provide depth information. PA imaging, on the other hand, probes tissue molecular composition in depth and thus can create images of tissue type. In addition, it can be straightforwardly integrated with ultrasound imaging to deliver morphological information and plaque sizing. Spectroscopic photoacoustic imaging was previously shown to distinguish between peri-adventitial and plaque lipids, not only owing to its capability of depth resolution, but also through sensitivity to differences in absorption spectra of lipid types[159]. Spectral differentiation of different cholesteryl compounds, namely cholesterol, cholesterol linoleate and cholesterol oleate could be achieved with PA imaging[120,160], showing the potential of the modality to identify lipid profiles classifying plaques as stable and unstable.

In this study we extract the lipid signature from human carotid endarterectomy samples, in the range of 1150 to 1240 nm using an in-house developed Optical Resolution Photoacoustic Microscopy (OR-PAM) system capable of spectral imaging, that we

refer to as micro-spectroscopic PA (μsPA), with images of lipid composition acquired by matrix-assisted laser desorption ionization mass spectrometry imaging (MALDI-MSI) as a reference. This is the first work, to our knowledge, which explores lipid PA spectral signatures and their distribution within advanced plaques at a microscopic level with corresponding molecular validation. Findings from this study can possibly be extrapolated for generation of optimal contrast in non-invasive carotid PA imaging or catheter-based coronary intravascular PA (IVPA) imaging.

5.2 Material and methods

5.2.1 Tissue collection and processing

Six human carotid endarterectomy (CEA) samples were surgically removed and snap frozen and stored at -80 °C until further processing. Tissue collection was performed according to Erasmus MC Ethics Board protocols (MEC 2008-147). The surgical protocol used preserves plaque morphology and intact lumen[73]. CEAs were transversally divided into 2 mm thick cross sections, which were subsequently embedded into 10% porcine type A gelatin (Sigma-Aldrich, The Netherlands). CEA cross sections were cryosectioned into 10 μm thick sections and thaw-mounted onto glass slides, and stored at -80 °C. One slide was used for MALDI-MSI and an adjacent slide was used for μsPA, after the measurement the sections were histochemically stained for co-registration purposes.

5.2.2 Micro-spectroscopic Photoacoustic Imaging

Experimental Setup

A micro-spectroscopic photoacoustic imaging (μsPA) setup was built by focusing the beam from a tunable diode pumped laser OPO (Spitlight EVO-OPO, 200 Hz PRF, 5 ns pulse width, Innolas GmbH, Germany) achromatically to match the resolution of the MALDI images at a 45 μm spot. Pulse-to-pulse laser intensity variations were monitored using a photodiode. On average, the fluence varied between 300 μJ/cm^2 to 500 μJ/cm^2 depending on the wavelength. An unfocused 50 MHz transducer (V358-SU, Panametrics-NDT, USA) was spatially aligned with the optical beam to acquire transmission mode photoacoustic signals. A standard microscope slide (thickness 1 mm) with the tissue section was spatially raster scanned by two motor stages (Newport Corporation, USA). One motor was scanned, meanwhile allowing for a 4-fold averaging within the 45 μm step size, while the other motor was used to step 45 μm between scan lines. Photoacoustic signals were filtered by a low pass filter with cutoff 70 MHz (BLP-70, Mini-Circuits,) and high pass filtered with cutoff 25 MHz (BHP+25, Mini-Circuits) and amplified 55 dB (AU-3A-0110, Miteq, Long Island, NY,USA) then digitized at 400 MHz (DP310, Aquiris).

Tissue measurements

Prior to imaging, tissue cryosections were fixated by submersion into 4% paraformaldehyde (PFA) for 10 minutes[161,162]. A photoacoustic image of the complete tissue section

was acquired at 1210 nm after which a region of interest (ROI) was chosen for spectroscopic imaging. During spectroscopic imaging the wavelength was swept from 1150 to 1240 nm in steps of 2 nm. Depending on the size of the endarterectomy sample cross-section the 1210 nm scan could take up to 25 minutes, while the spectroscopic scan of the sub-region could take up to 6 hours.

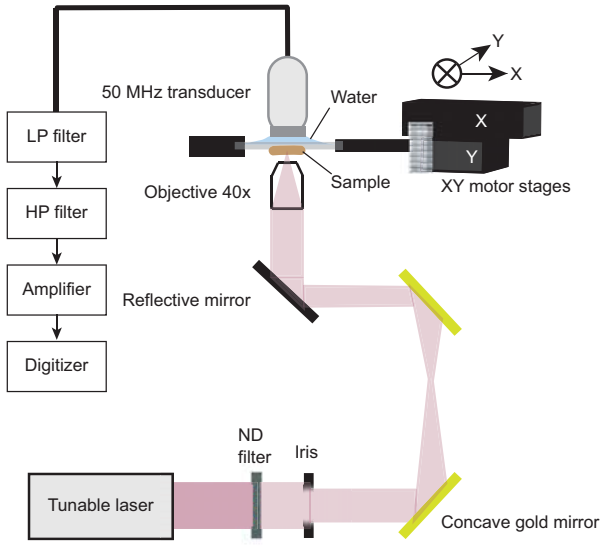


Figure 5.1: **Experimental setup of micro spectroscopic photoacoustic setup**

Lipid extract measurements

Six lipid extracts: a phosphatidylcholine PC(16:0/18:1) (850457C-25 mg, Avanti Polar Lipids, USA), two sphingomyelin species SM(d18:1/16:0) (1 mg, 6254-89-3, Cayman Chemical, USA), SM(d18:1/24:1) (860593P-5 mg, Avanti Polar Lipids, USA), a triacylglycerol TAG(18:1/18:1/16:0) (2190-30-9, 10 mg, Sigma-Aldrich, The Netherlands), cholesteryl ester (CE) linoleate, CE(18:2) (700269P-100 mg, Avanti Polar Lipids, USA), and its oxidized form, CE ox-18:2 (700192P-500 µg, Avanti Polar Lipids, USA) were purchased. Lipids purchased in powder form were dissolved in 1 ml HPLC-grade chloroform. Lipids were deposited onto a glass slide and the chloroform was left to evaporate, leaving a lipid deposit on the slide. Spectroscopic imaging was performed using the same protocol as used for the tissue measurements. In order to assess the effect of fixation and gelatin embedding on the results, we additionally measured multiple CE 18:2 prepared slides under all possible combinations of the following two conditions: (1) fix the slide in 4% paraformaldehyde (PFA) for 10 minutes, (2) deposit lipid on gelatin cryo-sections.

Data processing

µsPA data was processed by taking the maximum intensity projection in two time windows of the analytical signal per pixel (absolute Hilbert) to generate two photoacoustic

images. The tissue optical absorption image was generated using a square window corresponding to the time of flight of the acoustic signal through the glass slide. A tissue optical transmission image was generated using a square window in the PA signal trace, with a timing corresponding to the transducer response to light absorption in its surface[163]. Maximum intensity projection images were median filtered and convolved with a circular filter to account for the beam circular shape remapped to square pixels spatially. Signal intensity was corrected for laser pulse-to-pulse variations and filtered by a Savitsky-Golay filter of order 1 and frame length 5 all using MATLAB 2019b (Natick, MA, USA). Photoacoustic spectra of lipid extracts were processed similarly, averaging all spectra over the drop area to yield a standard spectrum per lipid standard. Selected images and spectra displayed in figures are normalized to the 95th percentile of the selected set.

Spectral variation in the data was extracted using principal component analysis (PCA). To ensure high quality spectral information, data was normalized to the noise level of every individual tissue section and masked using a threshold at 15% above the mean noise level. The number of components shown was determined using the elbow-method, retaining only the statistically significant components. Spatial distribution of the PCA components is depicted using diverging color scales, limits of the color scale are set to the smallest absolute median value of a certain PCA component, and the scaling is symmetrical. The overlay image is a sum of all three components scaled from -0.5 to 0.5 combined.

5.2.3 MALDI-MSI

Measurements

Samples were prepared for MALDI-MSI by sublimation of 50 mg 2,5-dihydroxybenzoic acid (DHB) onto the tissue using a home-built sublimation system[74]. The protocol uses 50 mg of DHB dissolved in 5 ml of aceton, sublimated at 125 °C for 10 minutes. MALDI-MSI was performed using a Synapt G2Si-TOF system (Waters, Manchester, UK) operated in the systems resolution mode. The spatial resolution of the experiment was 45×45 μm^2, using a 2000 Hz Nd:YAG (355 nm) laser, 100 laser shots per pixels and a mass range of 300-1200 *m/z*. Data was acquired using Waters Research Enabled Software suite (WREnS) and MassLynx v4.2 and exported to imzML format using HDI v1.4.

M/z values were identified as lipid species by means of FTICR MALDI-MSI (Bruker Daltonics, Bremen, Germany) uploading the data to METASPACE annotation platform[111] with a false-discovery rate (FDR) of < 10%[154]. Additionally, we homogenized CEA samples and used the Lipidyzer platform (Sciex, Framingham, MA)[154] and an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) with an ESI source and MS/MS analysis to further confirm the lipid identification[102].

Data processing

The exported imzML was processed using an in-house developed pipeline[102], executing the following data reduction procedures (1) pre-possessing, smoothing and recalibration of the data using DHB cluster peaks, (2) peak picking using the base peak spectrum, (3)

total-ion-current normalization and (4) a fractional mass filter and cross-correlations to remove background *m/z* values and selection of lipid *m/z* values.

5.2.4 Image registration and combined data processing

The tissue section used for μsPA was stained with hematoxylin-eosin and the MALDI-MSI section was stained using Oil Red O. The MALDI-MSI sections were registered to the μsPA image by means of a point-based rigid image registration framework in MeVislab (MeVis Medical Solutions AG, Germany). Per tissue section the Pearson correlation coefficient was calculated between 70 highly abundant *m/z* values and the 1210 nm full slide image taken with μsPA. This was performed on data masked by the plaque area outline, from histological segmentation, to ensure the correlation coefficient to be independent from background correlation.

5.3 Results

5.3.1 Micro-spectroscopic PA lipid images compared to MSI

Thirteen tissue sections were imaged using the μsPA system, and an example of such a measurement is shown in Fig. 5.2. In Fig. 5.2a, the 1210 nm absorption PA image (in semi-transparent red-orange-yellow scale) of the entire section is shown, overlaid on top of the tissue optical transmission image in grayscale. The tissue optical transmission image shows the structure of the tissue section, mapping the optical attenuation of the tissue in the wavelength range under study, whereas the absorption image shows the presence of lipids in the sample. From this image a region of interest was selected, depicted by the black box, and spectral PA imaging was performed within this region, shown in Fig. 5.2b. To illustrate spectral variations within a sample, spectra from five tissue locations are depicted in Fig. 5.2c. This shows a variety of spectral shapes present throughout the sample, with local maxima at 1164, 1188, 1196, 1210 and 1230 nm (Fig. 5.2c, indicated with vertical lines) in locations 5, 2, 1, 4 and 2 (Fig. 5.2a), respectively.

The Pearson correlation coefficients between 70 lipids, detected by MALDI-MSI, and the 1210 nm μsPA full slide image was calculated (n=70), see Fig. 5.3a. The list of the annotations of all MALDI-MSI detected lipids can be found in Table 3.3. The correlation with the PA image is generally lowest for diacylglycerol (DAG) and triacylglycerol (TAG) lipids, and highest for sphingomyelins (SM) and (oxidized) cholesterol and cholesteryl esters (CE).

In Fig. 5.3b and c, examples of four μsPA images and corresponding MALDI-MSI images are shown. In the analysis of the MALDI-MSI data we see that different lipid classes show the same spatial distribution, i.e. SM show the same spatial pattern independent of the tail length[154]. We also found that phosphatidylcholines (PC) are sometimes co-localized with SM, like in P5, see Fig. 5.3c. However, there are also cases in which there is limited overlap in spatial distribution, as in P3-1 and P4, see Fig. 5.3c. In cases where co-localization between PC and SM is low, the correlation coefficient of PC between the 1210 nm μsPA image and the MALDI-MSI drops significantly to that of the level of DAG and TAG, indicating that the signal may not be originating from PC lipids but rather

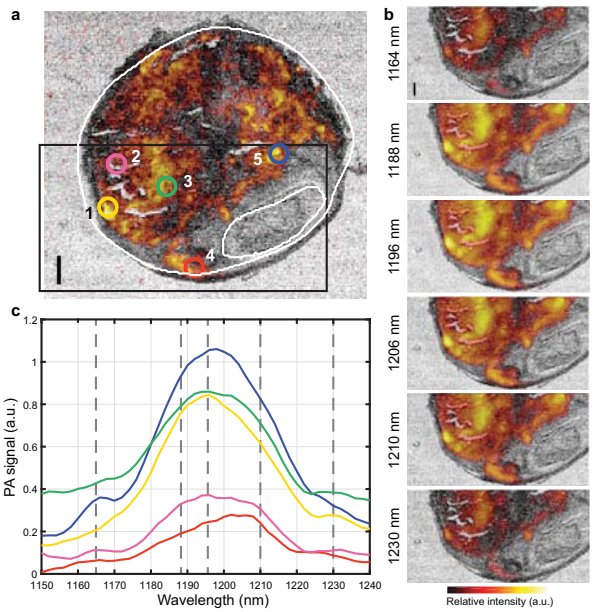


Figure 5.2: **µsPA measurement a)** µsPA full field of view image at 1210 nm, the black box indicates the ROI, white outlines indicate plaque and lumen **b)** ROI images showing six wavelengths, 1164, 1188, 1196, 1206, 1210 and 1230 nm, with subtle difference between the images and **c)** the normalized spectra at five different locations (indicated in figure a) showing differences in the local absorption spectra. Scalebars are 1 mm

from SM lipids, see the supplementary information, available in the online version of this paper.

5.3.2 Analysis of spectral variations

Principal component analysis of all samples combined reveals 3 spectral components to explain the variation in the data, see Fig. 5.4. The first component has the same shape as the average spectrum, showing a broad peak centered at 1196 nm. The second component shows a peak at 1210 nm and the third component shows two additional peaks at 1164 and 1188 nm.

These components have different weights per pixel. Fig. 5.5a, c, e and g show the relative weights of the spectral components in each pixel for selected tissue sections. Magenta, cyan and yellow reflect the positive coefficients per pixel of the PCA components whereas red, green and blue depict the spatial distribution of the negative PCA coefficient value. Spectral variations in PA amplitude reflect the relative abundance of the different lipids throughout the sample (Fig. 5.5), relative to the mean composition.

MALDI-MSI is a molecular imaging method enabling visualization of the spatial distribution of individual lipid molecules separately. Whereas in µsPA, spectral images correspond to a summation of the absorption of all tissue chromophores at the excitation wavelength, which shall reflect a mixture of the lipids identified by MALDI-MSI present at different concentrations. Therefore, we investigated the spectra of individual lipids to explore possible contributions from different individual lipids towards the complex

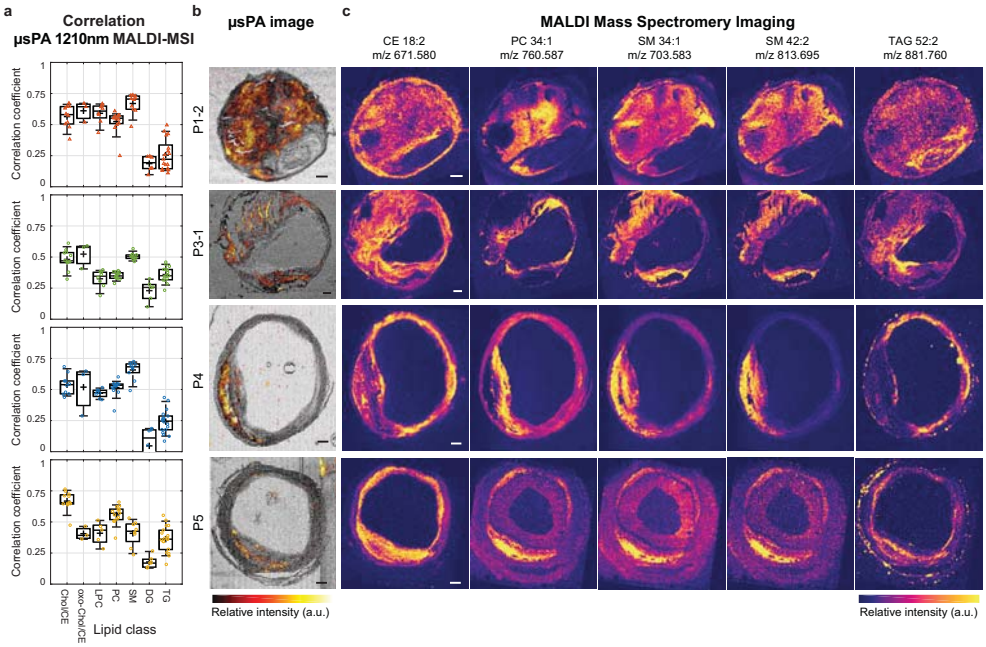


Figure 5.3: **Correlation of MALDI-MSI to µsPA. a)** Pearson correlation of 70 *m/z* MALDI-MSI images to the fullFOV 1210 nm, **b)** Overlay of maximum intensity projection (MIP) of transmission (gray scale) and MIP of absorption (red scale) µsPA image at 1210 nm and **c)** MALDI images showing the spatial distribution of 5 different *m/z* values spanning 4 lipid classes. Two column 3 and 4 are both sphingomyelins, showing the same spatial distribution. Scalebars are 1 mm.

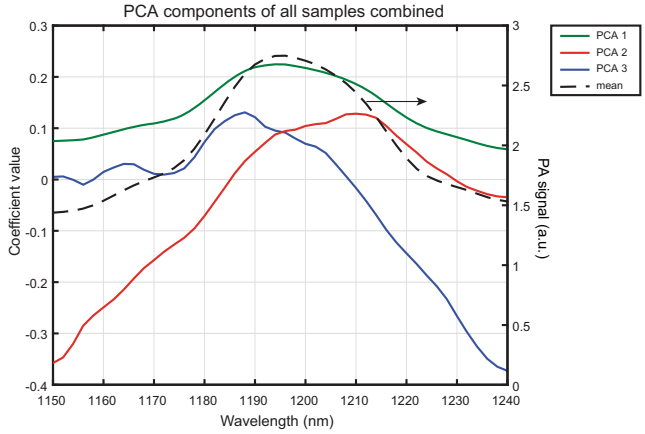


Figure 5.4: **Spectral information and decomposition of µsPA data.** PCA analysis of all CEA samples simultaneously showing the three spectral components (axis on the left) and the mean PA signal (axis on the right)

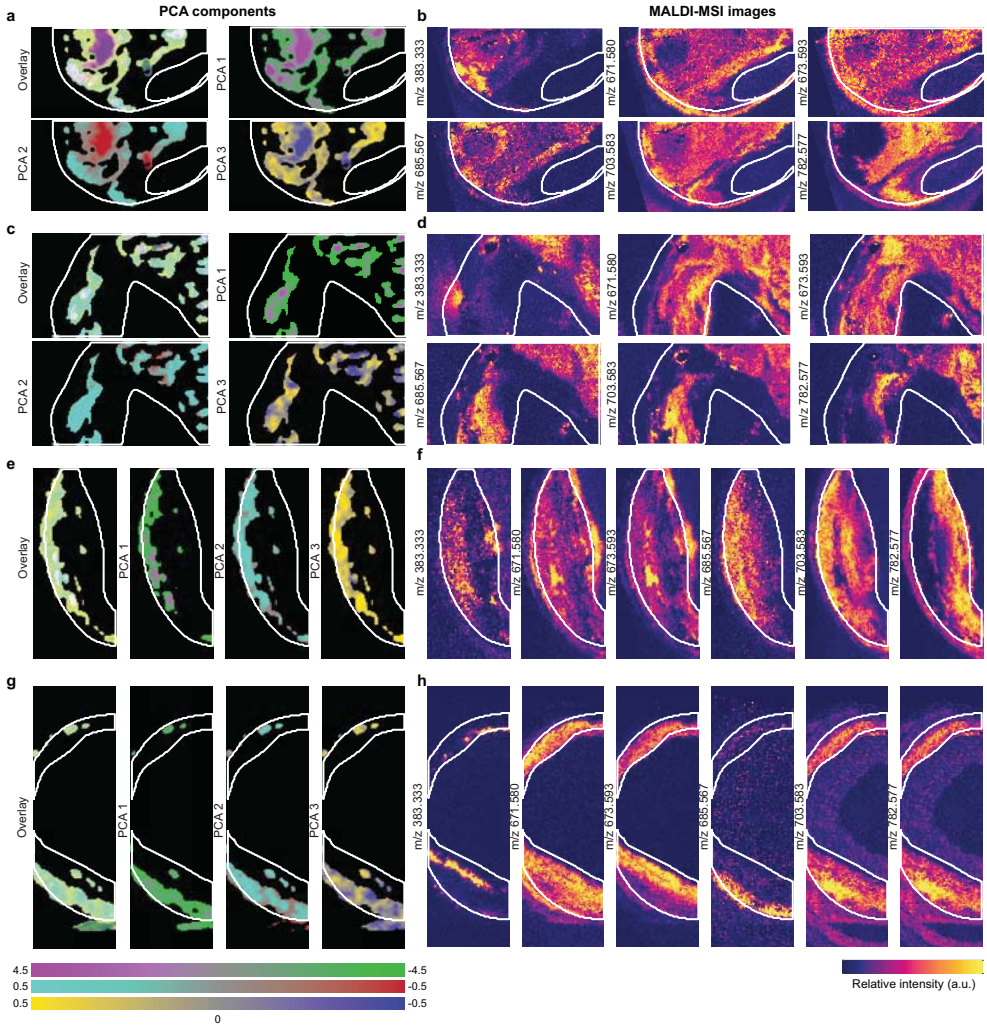


Figure 5.5: **Spatial distribution of all 3 principal components in and an overlay image, in magenta-green the first principal component, in cyan-red the second principal component and in yellow-blue the third component. Single PCA component images are a maximum intensity projection of the distributions, scaled between -0.5 and 0.5, the overlay images are summed PCA components. a-b)** sample P1-2, showing **a)** the spatial distribution of the PCA components and **b)** 6 lipids that imaged by MALDI-MSI, m/z 383.333, m/z 671.580 (CE 18:2 $[M+Na]^+$), m/z 673.593 (CE 18:1 $[M+Na]^+$), m/z 685.567 (ox-CE18:2 $[M+Na]^+$), m/z 703.583 (SM 34:1 $[M+H]^+$) and m/z 782.577 (PC 34:1 $[M+Na]^+$) **c-d)** sample P3-1 , **e-f)** sample P4 and **g-h)** sample P5

spectra from the tissue sections. Six lipids from five lipid classes were selected based on highest abundance in the tissue.

Figure 5.6 shows the spectra of the pure cholesteryl linoleate (CE 18:2) in normal form and in oxidized form (Fig. 5.6a), phosphatidylcholine (PC 34:1) (Fig. 5.6b), sphingomyelins with two different chain lengths (SM 34:1 and SM 42:2) (Fig. 5.6c) and triacylglycerol (TAG 52:2) (Fig. 5.6d). As expected CE 18:2 and oxidized CE 18:2 have

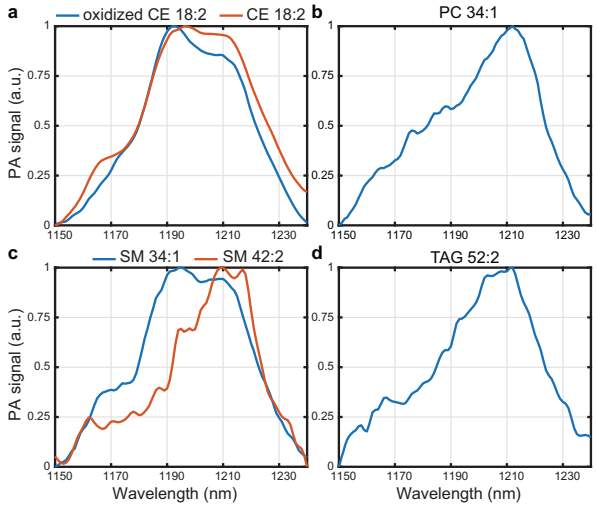


Figure 5.6: **µsPA recorded spectrum of lipid extracts. a)** Cholesteryl linoleate (CE 18:2) in red and oxidized cholersterol linoelate in blue, **b)** Phosphadiatylcholine 16:0/18:1 (PC 34:1), **c)** Sphingomyelin d18:1/16:0 (SM 34:1) in blue and sphingomyelin d18:1/24:1 (SM 42:2) in red and **d)** Triacylglycerol 18:1/18:1/16:0 (TAG 52 :2).

a similar shape, however for the oxidized form the shoulder peak at 1164 nm is missing and the peak at 1212 nm is not as strong. The spectrum of SM 34:1 is similar to that of CE 18:2. Also the spectra of SM 42:2, PC 34:1 and TAG 52:2 are comparable in shape, with a main peak at 1210 nm. Interestingly the spectra of both SM lipids differ substantially.

5.4 Discussion

In this study, we explored the imaging lipid vibrations by means of photoacoustic spectroscopy, using a novel PA slide microscope. Carotid endarterectomy cryosections were spectrally imaged using this system revealing vibrational modes generating peaks at or near 1164, 1188, 1196, 1210 and 1230 nm. The µsPA intensity images acquired at 1210 nm showed the strongest correlation with (oxidized) Cholesterol/CEs and SMs in the MALDI-MSI data, lipids that have been found to be abundant in necrotic core plaque. We characterized the variability in the spectral absorption by PCA, which showed three independent components with maxima at 1196 nm (first), 1210 nm (second), and 1164/1188 nm (third component), reflecting the contrast generated by the peak structures in the acquired spectra. As a reference for PA lipid spectra, we imaged pure lipids and found that the relation between molecular structure and the spectral shape of this absorption band is complex. Lipids of composed of differing molecular groups show spectra with high similarity (such as CE 18:2 and SM 34:1) while lipids from the same class, differing only in carbon chain length and saturation may exhibit considerable spectral differences (SM 34:1 and SM 42:2). Since lipid classes tend to spatially cluster in the tissue[154], a direct mapping of atherosclerotic spectral features to molecular composition could not be established.

Based on near-infrared spectroscopic studies, the absorption peaks observed in the spectral range under study mainly result from the second C-H overtone asymmetric stretch (noted as 3 ν_a) molecular vibration[164]; the absorption peaks arising from phosphate and amide groups occur at higher wavelengths[164]. Hydrocarbon chain length and saturation affect the absorption spectra obtained with near-infrared spectroscopy. In particular, the absorption spectra of alkanes, 1-alkenes and aldehydes all exhibit a decrease in the absorption peak ratio of 1190 nm/1210 nm as the hydrocarbon chain length increases; moreover, unsaturated 1-alkenes display a decrease in the absorption peak at 1163 nm in comparison to alkanes[164]. In fact, for longer hydrocarbon chains, the absorption peak at 1210 nm associated with the methylene (CH_2) group increases and that at 1192 nm, associated with the methyl group (CH_3), decreases. In addition to the ratio of absorption peaks, slight shifts in the absorption peaks are also observed as the chain length increases. Since lipids that belong to the same lipid class differ only in fatty acid chain length and saturation, we also observed the same spectral shifts phenomena when examining the photoacoustic spectra from the sphingomyelins, as well as those of cholesterol, and cholesteryl esters[120]. Moreover, the observed peaks in the data correspond well with the documented absorption peak ranges in the NIRS studies. The variations associated with the strengths of the methyl, and methylene second overtone absorption peaks are also reflected in the PCA analysis derived from our data set.

Our results may slightly vary under different temperature and pH conditions of the specimens; such changes have been demonstrated to induce shifts ± 2 nm in peak position[164–166]. The applied embedding (vs. no embedding) and/or fixation (vs. no fixation) did not influence the spectral results.

In our previous study using MALDI-MSI only, we discovered that lipids from the same lipid classes show similar spatial distributions[154]. In this study we revealed that, photoacoustically, the spectra of lipids within the same lipid classes, such as SM 34:1 and SM 42:2, are not necessarily the same. Additionally, we found that spectra of different lipid classes, with different spatial distribution within the plaque, can have the same PA spectrum, SM 34:1 and CE 18:2. In fact, the spectra of SM 34:1 and CE 18:2 are so similar that within a tissue sample these would be indistinguishable from each other. Both of these lipids are major contributors to atherosclerosis and have previously been shown to be highly abundant in CEA samples[13,66,154].

Previous publications looking at spectral information to reveal plaque composition mainly focused on the detection of CE 18:1 and CE 18:2[51,120,160]. The spectra acquired by μsPA are comparable to previously published spectra of cholesteryl esters[120] (Figure 5.7). The only difference is in the relative strength of the shoulder peak at 1164 nm.

Considering applicability to *in vivo* imaging, we selected the 1150-1240 nm second C—H stretch overtone band for the spectroscopic microscopy of atherosclerosis in a trade-off between optical attenuation and absorption strength. Other absorption bands have also been described and used for PA characterization of plaque[47,54]. The longer-wavelength infrared absorption bands exhibit a larger absorption coefficient and complementary information about molecular structure[164,165,167]. The increased absorption at for instance 1650-1800 nm may generate a stronger PA signal, but will also limit the penetration depth because of attenuation by other substances such as water. Even at the

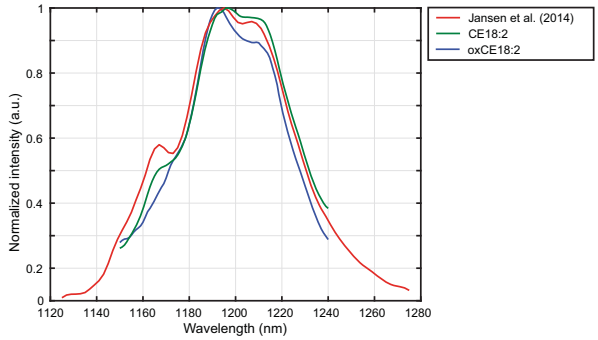


Figure 5.7: **Comparison of CE18:2 PA signal** measured by K. Jansen et.al (2014) to the measurement of CE 18:2 and oxCE18:2 by µsPA.

chosen wavelengths, delivery of sufficient optical power *in vivo* will remain challenging. While a unique mapping between spectral shapes and molecular composition could not be established, the observed variation in spectral absorption is expected to persist in intact specimens and may be related to plaque stage. The relative PA signal strength and peak positions should be investigated in intact specimens and *in vivo* in future studies. Since macroscopic imagers used in those settings have lower spatial resolution, it is possible that the subtle spectral variations we report here may mix and be difficult to discern.

Our setup is, to our best knowledge, the first to allow for spectral imaging of a region of interest, rather than just collecting spectra at pixelwise locations. Indeed, we have shown feasibility of a spectroscopic imaging system allowing simultaneously high spectral and spatial resolution within reasonable scanning times. This system allows for more detailed examination of spectral differences within different regions, and may be applied to other tissue types.

5.5 Conclusions

In conclusion, we have characterized lipid absorption of carotid atherosclerotic plaque at a resolution of 45 µm in the 1150-1240 nm band. We built a micro-spectroscopic photoacoustic imaging setup with which we spectrally imaged thirteen tissue sections originating from six carotid endarterectomies. Acquired spectra, supplemented by PCA decomposition, revealed four spectral peaks at wavelengths 1164, 1188, 1196, and 1210 nm which reflect most of the spectral variation within and between images. Spectral differences observed were primarily explained by differences in molecular structure such as chain length and number of unsaturated bonds. PA signal intensity was most strongly correlated with sphingomyelin and cholesteryl ester concentration as determined by MALDI-MSI, while diacylglycerols and phosphatidylcholines were associated with lipid-positive but low PA signal areas. The species SM 34:1 and CE 18:2 are abundant in necrotic core and were found to exhibit similar PA spectra, suggesting that this spectral feature can be used for *in vivo* detection of advanced atherosclerotic plaque with features of instability.

Acknowledgments

This project was funded by:
Netherlands Heart foundation - project number: NHS2014T096
Nederlandse Organisatie voor Wetenschappelijk Onderzoek - project number: 16131

6

Label-free analytic histology of carotid atherosclerosis by mid-infrared photoacoustic microscopy

This chapter is based on:
Mirjam Visscher*, Miguel Angel Pleitez*, Kim Van Gaalen, Ingeborg M. Nieuwenhuizen-Bakker, Antonius van der Steen, Vasilis Ntziachristos and Gijs Van Soest: Label-free analytic histology of carotid atherosclerosis by mid-infrared photoacoustic microscopy, *In preparation*
*these authors contributed equally

Abstract

Atherosclerotic plaque molecular composition reflects the pathological metabolic processes that contribute to plaque growth. In this work we visualize the constitution of human carotid plaques by mid-infrared opto-acoustic microscopy (MiROM), a form of label-free analytic histology. We imaged endarterectomy specimens (n=3 from three patients) at specific vibrational modes, targeting carbohydrates, lipids and proteins. Additionally, we recorded spectra in the CH region (2,930-2780 cm^{-1}) and fingerprint region (1,735-909 cm^{-1}) at selected locations. Variability in the spectral information was characterized using non-negative matrix factorization (NMF) of all locations (n=37), which revealed 3 components, relating to different molecular classes. Also, we observed a co-localization of cholesterol crystals and carbohydrates. These results will support further understanding of the molecular pathways involved in atherosclerosis and demonstrate the value of MiROM for imaging of complex human pathology.

6.1 Introduction

Atherosclerosis is the single largest cause of morbidity and mortality worldwide[60]. It is an inflammatory disease of the arteries that is characterized by the deposition of lipids in the arterial wall, called plaque. Lesions develop over decades from a superficial fatty streak to a lumen-obstructing plaque. Initially the vessel reacts to plaque formation by outward remodeling, ensuring conservation of the free lumen area. However, as the plaque progresses this is not possible and the lumen gets intruded by the plaque. Some plaques are high in fibrous material and are highly calcified, this type of plaques are usually stable. However, there are also plaques which have a complex composition, including a large necrotic core, these types of plaque are more prone to rupture. Proteoglycan-rich plaques frequently have a compromised endothelial layer[168]. These latter two types may initiating a thrombotic response, resulting in stroke or myocardial infarction depending on the location of the plaque, carotid artery or coronary circulation respectively[7].

Lipids play a major role in the development of atherosclerosis[13–16]. The disease is initiated by the infiltration of low density lipoproteins (LDL) into the sub-endothelial layer[169,170]. In histochemical staining the current gold-standard to investigate the spatial distribution of lipid is the Oil Red O stain, however the stain is unspecific and only stains neutral lipid species. Lipidomics studies have shown elevated levels of cholesterol and cholesteryl ester (CE) species in plaques that exhibit an unstable phenotype[16], whereas apparently stable plaques exhibit a higher abundance of phospholipid (PC) and triacylglycerol (TG)[15]. The mechanisms that lead to this variance in lipidomic appearance, as well as the interactions with local metabolic and enzymatic activity, remain largely unknown, however.

The severity of atherosclerosis has been correlated with lipid content, specifically with cholesterol and cholesteryl esters[171]. It has been shown that cholesterol crystals (CC), often found in the necrotic core of plaques, are associated with an increased risk of plaque rupture and occurrence of ischemic events[172–174]. Solid cholesterol in atherosclerosis has been described in different forms and shapes, needle-shaped, plate-shaped filamentous or helicals[172]. Furthermore the presence of CC has been shown to; (1) activate the NLRP3 inflammasome of macrophages[63,99,175], (2) induce apoptosis of macrophages[176] and, (3) impact the cell membrane of macrophages[177]. Despite extensive research, knowledge about underlying processes and mechanisms of cholesterol crystallization remains scarce.

Mass spectrometry imaging (MSI), a label-free molecular imaging technique, has been used to study the spatial distribution of lipids in atherosclerosis, and localized CE, sphingomyelin (SM) and lyso-PC lipids to in plaques, showing associations with high-risk plaque features[37,39,41,43,154]. In Chapter 3[154] we demonstrated that tissues that are homogeneous by histological classification have a large variety in lipid content. Challenges in MSI are a strong dependency of the chemical sensitivity on experimental parameters, expensive equipment, and long scan times, which inhibit routine high-resolution imaging examinations of excised tissue. In the case of atherosclerotic plaque specifically, cholesterol has been difficult to examine using MSI, precisely the lipid species to which a central role in pathogenesis has been attributed. This demonstrates the need for addi-

tional and more accessible characterization with chemically specific imaging to further illuminate the etiology of atherosclerosis.

The chemical specificity of vibrational spectroscopy in the mid infrared (mid-IR) allows the imaging of endogenous biomolecules[121,178,179] by techniques such as mid-IR absorption and Raman microscopy, which have also been exploited in imaging of atherosclerosis[171,180–185]. These investigations yielded a clear distinction between cholesterol crystals and condensed cholesteryl ester structures in intact plaques[171], characterization of vulnerable plaque features such as the fibrous cap, calcification and lipid changes in the necrotic core[183], and lipid distribution in the sub-endothelial space of the intima[185]. These studies show the ability of mid-IR imaging to add chemical specificity to histology.

Mid infrared spectroscopy is label-free, has high spatial and spectral resolution, and allows for imaging of several molecular species in a single experiment, but lacks depth resolution and has limited sensitivity. In this manuscript we used the mid-infrared optoacoustic microscopy (MiROM)[48], to image the molecular composition of advanced carotid atherosclerotic plaques. MiROM has the benefits of mid-IR microspectroscopy, but adds inherent three-dimensionality (two scan directions and acoustic time-of-flight) and high sensitivity. MiROM has been applied to live cells and animal tissues[186], but not for the systematic biochemical characterization of human pathology.

In this study we image human carotid endarterectomy samples (n=12) at selected vibrational modes with multi-stain histochemical analysis as a reference. Additionally, we collected spectral information from several locations in the imaged plaques in the CH and fingerprint spectral regions at high spectral resolution. With these data, we illustrate the distribution of a variety of biological molecules, lipids, but also proteins and glycans, in advanced human atherosclerotic plaque.

6.2 Materials and Methods

6.2.1 Tissue collection and processing

Three human carotid endarterectomy (CEA) plaque samples were surgically harvested and were washed in PBS, snap frozen and stored at -80 °C until further processing. A special surgical protocol was performed to preserve an intact lumen and morphology of the specimen[73]. When processing the CEA samples were divided into 2 mm thick cross-sections which were embedded into 10% porcine gelatin type A (Sigma-Aldrich, The Netherlands). These cross-sections were cryo-sectioned (CM3050 S, Leica Biosystems (cutting temps: OT -21 °C; CT -19 °C)) into 10 μm thick sections and thaw mounted onto glass slides and stored at -80 °C. Sections were used for histological staining and the remaining part, 1-1.5 mm in thickness, was used for MiROM microscopy. This study was performed according to the ethical guidelines sanctioned by the Ethics Board of Erasmus MC (MEC 2008-147).

6.2.2 Histology, histology segmentation and image registration

The tissue sections were histochemically stained by: Hematoxylin and eosin stain (HE), Miller's elastic stain, Martius scarlet blue trichrome (MSB), Oil Red O (ORO) and Periodic acid–Schiff (PAS) combined with Alcian blue (AB), to stain for general structures, collagen and elastin, fibrin and erythrocytes, lipids, and polysaccharides, respectively.

Based on the histological information the tissue section was segmented into the following plaque components; necrotic core (NC), fibrin, foam cells (FC), erythrocytes, calcium and cholesterol crystals (CC).

6.2.3 MiROM microscopy

A mid-infrared optoacoustic microscopy (MiROM)[48] was used to perform label-free, bond-selective imaging of carbohydrates, proteins and lipids. The microscope comprises a tunable pulsed quantum cascade laser (QCL) (MIRcat, Daylight Solutions) to generate a photoacoustic signal, which is detected in transmission-mode using a focused ultrasound transducer with a central frequency of 25 MHz (Sonaxis). The spectral range of the QCL is 3.4–11.0 μm (2,941-909 cm^{-1}) with a spectral line width of $\leq$1 cm^{-1} (full width at half maximum, FWHM). It has a pulse duration of 20 ns and a repetition rate of 100 kHz. The laser is focused onto the sample using a 0.5 NA reflective objective (36×, Newport Corporation). The laser and the ultrasound transducer are co-axially aligned so that the focii overlap. The image is obtained by moving the sample along of focal plane using motorized stages (Prior Scientific and Physik Instrumente). The raw photoacoustic signal was amplified by 63 dB (MITEQ) and filtered using a low pass filter 50 MHz

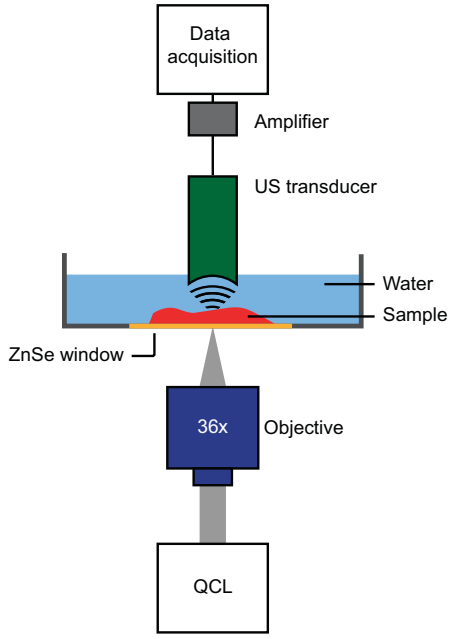


Figure 6.1: **Schematic representation of an auto-encoder.**

(Mini-Circuits). Data was recorded using 12-bit DAQ card (Gage Applied) at a sampling rate of 250 MSs^{-1}. The images show the peak-to-peak amplitude of 50 averages of the photoacoustic A-lines recorded, which corresponds to a pixel dwell time of 1 ms.

The sample was thawed and placed onto a custom-made mid-IR transparent ZnSe window (Edmund Optics) in a 3D printed acrylic–glass dish. The sample was then desiccated under vacuum for 10 minutes to remove bubbles between the sample and the window, and subsequently fixed using a transparent film and covered with deionized water as a coupling medium. A schematic of the microscope setup is depicted in Fig. 6.1.

Images of the full field of view of the CEA samples were recorded, at a number of wavelengths specifically aimed at known vibrational features, at a spatial resolution of 25 μm. Per cross-section, one or more regions of interest were additionally selected and imaged at higher spatial resolution, 2.5 and 5 μm. At specific locations within the region of interest photoacoustic spectra were recorded with a spectral resolution of 2 cm^{-1} or 4 cm^{-1} and 10,000 averages per wavelength, for two spectral regions, CH region (2920 - 2780 cm^{-1}) and the finger-print (FP) region (1737 - 901 cm^{-1}). Prior to every measurement a reference spectrum was recorded using carbon tape.

To enhance visibility and to improve the spatial resolution, the images contrast enhanced to 0.3% saturation, histogram normalization, were bi-cubic interpolated and post processed using a 2-pixel Gaussian filter.

6.2.4 Data processing

We performed two unsupervised clustering analyses by means of non-negative matrix factorization (NMF), using an NMF toolbox for biological datamining[105,106] implemented in MATLAB 2019b (Natick, MA, USA). We applied this analysis in the spectral domain, identifying co-occurring patterns in all high-resolution spectra acquired at sites of interest. A similar cluster analysis was performed in the image domain, using images at six salient wavelengths to identify features in all acquired images. In both, the optimum number of components was determined based on the dispersion coefficient[107].

6.2.5 Image registration

Using an in-house developed point-based rigid image registration framework in MeVisLab (MeVis Medical Solutions AG, Germany) the segmentation images were registered to the MiROM lipid images by translation and scaling, to enable correlation of the imaging modalities.

6.3 Results

6.3.1 MiROM imaging

Twelve carotid endarterectomy (CEA) cross sections, originating from three plaques of three different patients, were imaged using the MiROM system. Fig. 6.2a and b depict the full field of view MiROM recorded maximum intensity projections of an example CEA cross-section at 2850 cm^{-1} and 1550 cm^{-1}. The signal at 2850 cm^{-1} originates from the

symmetric stretching of CH_2, this vibration is not unique for lipids, but is significantly stronger for lipids than other molecules[184]. The signal at 1550 cm^{-1} is introduced by amide II bonds, mainly present in proteins.

High-resolution spectroscopy in Fig. 6.2c and d was recorded at three different locations, depicted in the ROI in Fig. 6.2e, corresponding to the red box in Fig. 6.2a. This area was scanned at high spatial resolution. Spectral peaks were assigned to vibrational modes with tentatively assigned molecular basis, based on literature review of Raman and FTIR measurements of atherosclerotic tissues, see Table 6.1. Where possible, we used histological staining to validate these assignments, e.g. the patterns in the ORO staining are the same as in the 2850 cm^{-1} images.

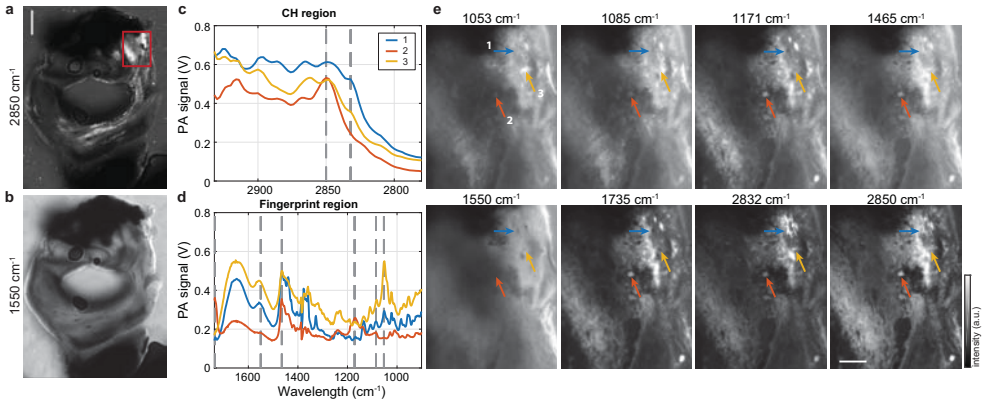


Figure 6.2: **Example of MiROM measurement of CEA cross-section (P1-B06). a)** MiROM image of cross-section at 2850 cm^{-1}, depicting the spatial distribution of lipids **b)** MiROM image of cross-section at 1550 cm^{-1}, depicting the spatial distribution of proteins **c)** MiROM spectrum of CH region with vibrational bands at 2850 cm^{-1} (CH_2 symmetric stretching) and 2832 cm^{-1} (shoulder peak of cholesterol) **d)** MiROM spectrum of FP region with vibrational bands at 1735 cm^{-1} (CH_2 scissoring), 1550 cm-1 (Amide II), 1465 cm^{-1}(CH_2 scissoring), 1171 cm^{-1} (C-O-C asymmetric stretching), 1085 cm^{-1} (C-O deformation), 1053 cm^{-1} (C-C stretching) **e)** MiROM images of CEA sample with marked the locations where spectra were recorded, scale bar is 250 µm. Scale bar a is 1 mm

Table 6.1: Overview of vibrational modes and wavelength at which they occur in MiROM with tentative molecular assignments based on literature[185,187,188]

Wavelength (cm^{-1}; ± 2)	**Vibrational mode**	**Tentative molecular assignment**
2850 cm^{-1}	CH_2 symmetric stretching	Lipids
2832 cm^{-1}		Cholesterol
1735 cm^{-1}	C=O stretching	Cholesteryl esters and triacylglycerols
1550 cm^{-1}	Amide II (C=N and N-H stretch)	Protein
1465 cm^{-1}	CH_2 scissoring	Lipids
1171 cm^{-1}	C-O-C asymmetric stretching	Cholesteryl esters
1085 cm^{-1}	PO_2^- symmetric stretching	Phospholipids and nucleic acids
	C-O stretching	Glycogen, oligosaccharides and glycolipids
1053 cm^{-1}	C-C stretching	Cholesterol and carbohydrates

6.3.2 Spectral cluster analysis

Spectra were recorded at 37 locations, these spectra were grouped and clustered using non-negative matrix factorization; this resulted in 3 NMF components, depicted in Fig. 6.3. Component 1 shows a correlation between the peak at 2832 cm^{-1} and 1053 cm^{-1}, however the peak at 1053 cm^{-1} is also present in components 3, combined with peaks of 1085 cm^{-1}, 1550 cm^{-1}. Component 3 is strong in the fingerprint region but low in the CH region, suggesting that these vibrations are not related to lipids. Component 2 has strong peaks at 2850 cm^{-1}, 1735 cm^{-1}, 1465 cm^{-1} and 1171 cm^{-1}, all peaks that are tentatively assigned to vibrational modes dominated by lipids.

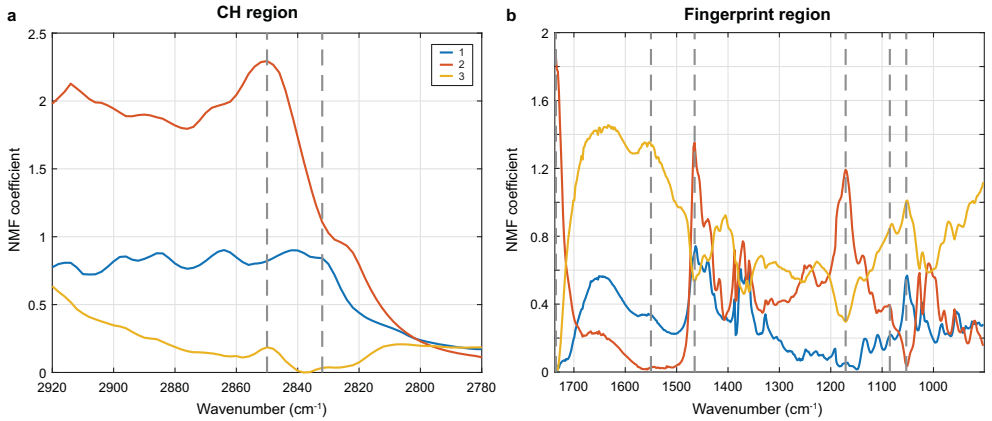


Figure 6.3: **Non-negative matrix factorization of MiROM spectra, resulting in 3 NMF components, depicting how spectral features are correlated to each other. a)** NMF spectral decomposition of CH region and **b)** NMF spectral decomposition of fingerprint region. Grey vertical lines depict the tentative molecular vibrational assignments, see Table 1.

Component 1 in the spectral NMF analysis showed a correlation between 2832 cm^{-1} and 1053 cm^{-1}, this is visible in the imaging data as co-localization of the two signals, showing needle shaped structures at both 1053 and 2832 cm^{-1}, see Fig. 6.4. There is a consistent spatial shift between the two signals, see Fig. 6.4e. Adjacent histological sections revealed that 2832 cm^{-1} can most likely be attributed to cholesterol, and needle structures correspond to the presence of clefts. Furthermore, cholesterol sheets, typically found in atherosclerotic plaques[189], are visible in sample P2-1 zoom 2. Additionally, the PAS-AB stain shows more intense staining directly around these clefts, revealing the presence of carbohydrates adjacent to cholesterol crystals, see Fig. 6.4.

6.3.3 Spatial cluster analysis

Furthermore, NMF was also used to decompose high resolution MiROM images into a lower dimensional space, see Fig. 6.5. This resulted into 3 components, of which one was dominated by the protein band at 1550 cm^{-1}; component II which shows little spatial structure, except general lower intensity in the necrotic area. Component I reveals contributions from 1171, 1550, 1735 and 2850 cm^{-1}, and shows high resemblance with the ORO stain, meaning it is most likely dominated by lipid signals. Components III

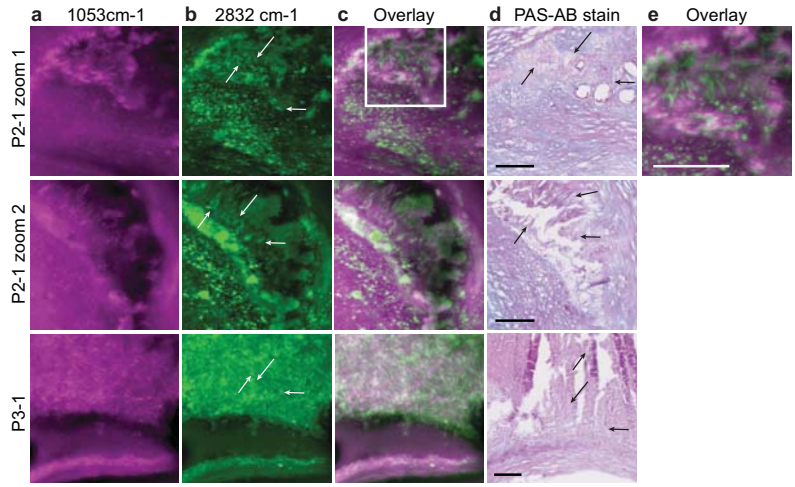


Figure 6.4: **MiROM images showing co-localization of cholesterol and carbohydrates. a)** MiROM image at 1053 cm^{-1}, C-C stretching of cholesterol and carbohydrates **b)** image at 2832 cm^{-1}, shoulder peak of cholesterol with arrows indicating cholesterol crystals, **c)** Overlay of 1053 cm^{-1} (in magenta) and 2832 cm^{-1} (in green) and **d)** Combined Periodic acid-Schiff (PAS) and Alcian Blue (AB) staining of tissue, showing the presence of carbohydrate macromolecules like glycogen and glycolipids (PAS, stained purple) and acidic mucosubstances (AB, stained blue). Arrows indicating cleft indicating presence of cholesterol crystals. **e)** Magnification of figure in c showing the slight shift in location between the signal at 1053 cm^{-1} (in magenta) and 2832 cm^{-1} (in green). Scalebars are 250 μm.

displays contributions from 1053, 1171, 2832 and 2850 cm^{-1} and is co-localized with the necrotic area shown in histological staining. Compared to component I, component III has a significantly higher contribution from 1053 and 2832 cm^{-1}, consistent with the collocation of cholesterol and glycans presented in Fig. 6.4.

6.4 Discussion

With this study, we aim to show the spatial distribution of different types of molecules in carotid atherosclerotic plaques. We imaged twelve cross sections from three carotid endartectomy samples, at several different wavelengths. This revealed the spatial distribution of different vibrational modes, illuminating the presence of cholesterol crystals and sheets thereof, the distribution of lipids, protein and carbohydrates in plaques. Additionally, we imaged smaller ROIs at high spatial resolution, 2.5 - 5 μm, disclosing the co-localization of cholesterol crystals and polysaccharides, and validated our findings using histochemical analysis of adjacent sections. Furthermore, we decomposed spectral and imaging information by means of NMF, each resulting in three distinct components.

Cholesterol crystals are commonly found in atherosclerotic plaques. They originate in early stage plaque development and trigger local inflammation[190]. During the formation of the crystalline structure the cholesterol expands in volume. Sometimes cholesterol crystals are found that perforate the fibrous cap of the plaque. In histology stains however, the presence of these cholesterol crystals is only detectable by absence, the cholesterol dissolves leaving behind a cleft in the tissue section[7,63,99,172]. With this

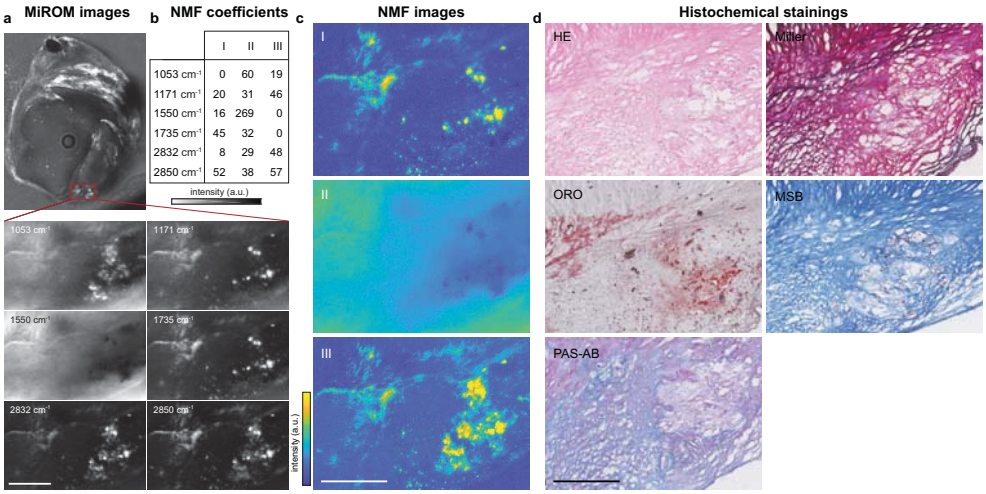


Figure 6.5: **MiROM data analyzed using NMF compared to histochemical stain (example shown P1-3). a)** MiROM image of cross-section at 2850 cm^{-1}, depicting the spatial distribution of lipids, and showing the ROI. ROI images, at 1053, 1171, 1550, 1735 2832 and 2850 cm^{-1}, showing the differences in spatial distribution between vibrational modes. **b)** Table showing the NMF coefficients of the 6 vibrational modes, illustrating the contribution different modes have to a NMF component. **c)** Images of the 3 NMF components, where I is dominated by lipid modes, mainly 1735 and 2850 cm^{-1}, II by protein and III by a mix of lipid related vibrational modes. **d)** Histochemical stain of adjacent cross-sections, showing HE (general structure), Miller (elastin and collagen), ORO (neutral lipids), MSB (fibrin and erythrocytes) and PAS-AB (polysaccharides). Scalebars are 250 μm.

MiROM data we confirm the presence of cholesterol crystalline structures by the observation of a signal from vibrational modes associated with cholesterol.

Furthermore, we identify a co-localization of signals at 1053 and 2832 cm^{-1}, which modes we link to carbohydrates and cholesterol, respectively. 1053 cm^{-1} can originate from both carbohydrates and cholesterol. However, there are a few reasons why we believe the signal is, at least partially, related to carbohydrates. First, the shape of the signal at 1053 cm^{-1}, with 'side lobes' as shown in Figure 2d, which is distinctive for carbohydrates[48,187,191]. Secondly, there is a slight shift in the location of the needle structures in the images from 1053 and 2832 cm^{-1}, see Fig. 6.4e. Finally, the PAS-AB staining around the cholesterol clefts is positive for polysaccharides.

In this study we were unable to identify the type of carbohydrates that is co-localized with the cholesterol crystals. Glycolipids have been associated with atherosclerosis[192–194], would exhibit a carbohydrate spectral signature, but the relatively high intensity of the vibrations at 1550 cm^{-1} in these areas makes this unlikely, see Fig. 6.2d. Moreover, in that case we would expect the 1053 cm^{-1} signal to be correlated to the CH_2 symmetric stretching at 2850 cm^{-1} in our spectral cluster analysis, which it is not (Fig. 6.3). Other carbohydrates, like glycoproteins[192,195], have been associated with atherosclerosis, but further investigation will be necessary to identify them. An exploratory analysis with CD 206, macrophage mannose receptor 1, which we performed because of the prominence of mannose-binding monocytes in atherosclerosis[196,197], was not consistently positive for the areas where 1053 cm^{-1} was detected.

We did not find cholesterol crystals and carbohydrates in all cross-sections. Samples originating from P2 and P3 had large areas with cholesterol crystals, also exhibiting sheets of cholesterol. However, in P1 we only found isolated crystals or very small areas with few crystals (Fig. 6.2e). Histology showed that sample P1 was more fibrous in nature and had only small necrotic and/or fibrin areas. Whereas, generally samples P2 and P3 had typical features of vulnerable plaques, like large necrotic core and fresh fibrin.

NMF analysis of the spectral data shows a clustering of lipid related vibrational modes, with contributions from 1171, 1465, 1735 and 2850 cm^{-1}, in component 2. Furthermore, component 1 shows a link between 1053, 1465, 1550 and 2832 cm^{-1}. Component 3 shows a relation between 1053 and 1550 cm^{-1}; this correlation supports the possibility that glycoproteins generate at least part of the signal at 1053 cm^{-1}.

This work represents the first study of MiROM applied to human pathological tissue. It demonstrates the utility in unraveling some of the chemical complexity and may be refined to identify specific features in atherosclerosis, but also in neurodegenerative diseases or cancer, where deposition of aberrant tissue leads to disease and even death. Compared to previous investigations of lipids in atherosclerosis with mid-IR spectroscopic imaging[180–185,198], we add a successful application of multivariate analysis on MiROM both spectral and imaging data of atherosclerotic tissue. The machine classification is consistent with the assigned molecular origin of the defining spectral features. Previous endeavors using multivariate analysis on e.g. diffuse reflectance FT-IR microscopy and ATR-FT-IR data have not been able to show differentiation between normal and atherosclerotic rabbit aorta[198]. Hierarchical cluster analysis was successfully applied on aorta tissue op ApoE\LDLR$^{-\backslash-}$ mice, however, it was only capable of differentiating normal from atherosclerotic tissue and did not find clustering within the plaque itself[185].

6.4.1 Limitations

In this study vibrational modes are assigned to molecular composition based on literature review, and where possible validated using patterns observed in histopathology. The assignment is not unambiguous, and further research is necessary to validate our tentative assignments. However, validation hereof is complex because of the heterogeneity of the tissue. Additionally, molecular assignment of richly detailed spectral features in complex biological tissues is fraught with uncertainty.

Calcification is common in carotid atherosclerosis, but was not present in our data set. Large calcifications severely complicate histologic processing and are likely to introduce artifacts in MiROM because of an uneven surface on the slide. A number of calcium specific spectral features related to (hydroxy)apatites has been identified in previous studies[199]; in this tissue set we could not explore the mid-IR microspectroscopic features of calcification and its interaction with surrounding tissue.

6.5 Conclusions

In conclusion, we have characterized twelve human carotid endarterectomy samples by means of MiROM imaging at a spatial resolution of 25 μm. Additionally, high resolution

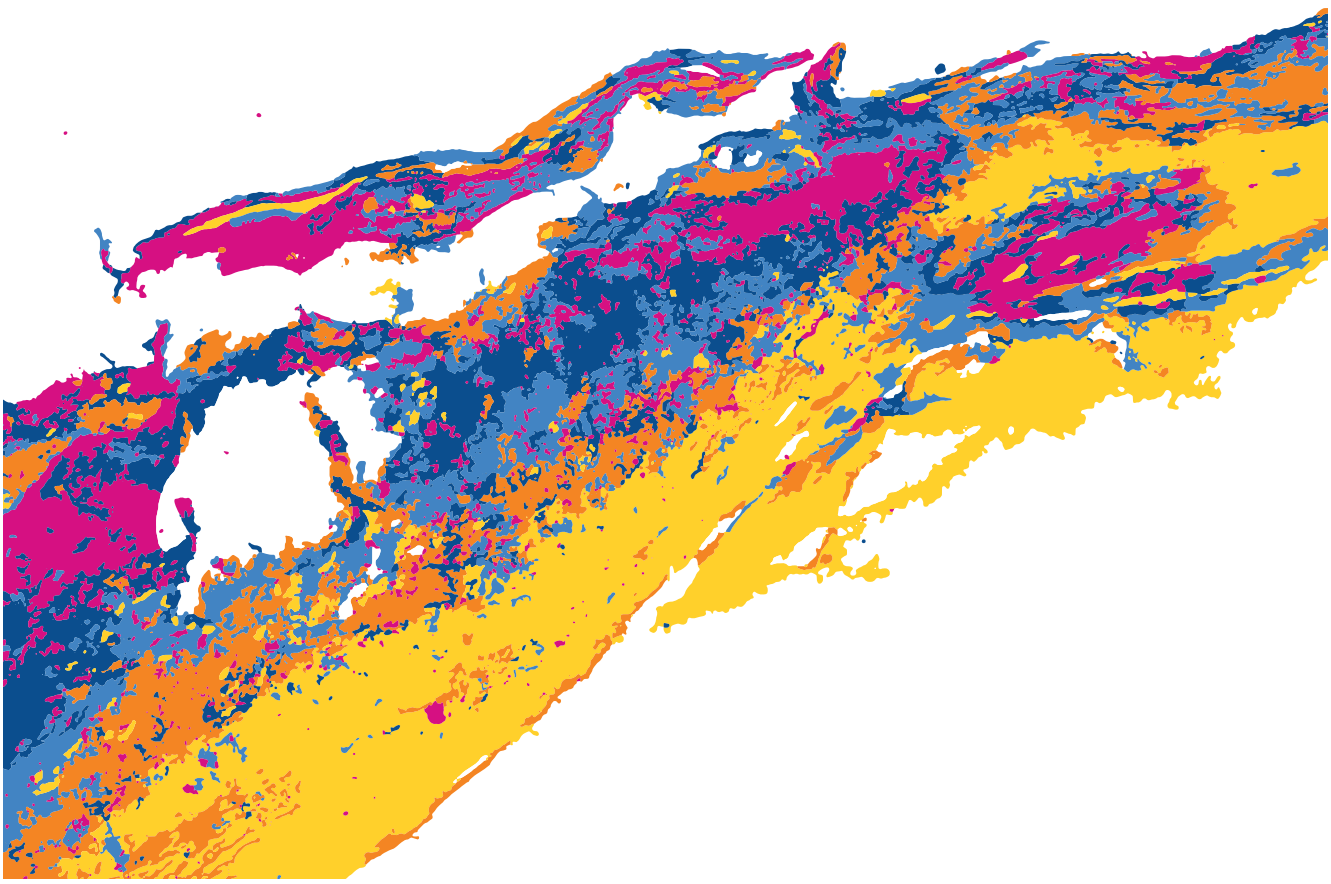
images of 16 ROIs were recorded and spectral information was collected at 37 locations in these plaques. These findings were validated by means of histochemical analysis of adjacent cross-sections. The same patterns were found in the vibrational mode of CH_2 scissoring at 2850 cm^{-1} and the ORO lipid stain. Review of the spectra, supported by NMF decomposition, reveals that 1053 and 2832 cm^{-1} and, 1053 and 1550 cm^{-1} are related as well as all lipid related vibrations, at 1171, 1465, 1735 and 2850 cm^{-1}. Multiwavelength analysis of high-resolution images revealed a co-localization of 1053 and 2832 cm^{-1} signals, for which we presented evidence that they originate from carbohydrates and cholesterol, respectively.

Acknowledgments

This work is supported by:
Stichting Fonds Dr. Catharine van Tussenbroek Fonds
Nederlandse Organisatie voor Wetenschappelijk Onderzoek - project number: 16131

7

Discussion



Atherosclerosis is a multifactorial disease, driven by lipid accumulation and inflammation. It is the largest cause of morbidity and mortality worldwide through strokes and myocardial infarctions, and their clinical sequelae. In a reality where primary prevention by lifestyle changes has limited effectiveness, early identification of patients, and ideally atherosclerotic plaques, at risk of causing events is an important capability in managing the disease burden and loss of life due to atherosclerosis. Early diagnosis with high specificity may enable pre-emptive, targeted medical, surgical, or device-based interventions.

Lipids play a variety of roles in the disease process, as metabolic intermediate, residue, and signaling molecule. In recognition of this central role for lipids, studies looking for biomarkers of atherosclerotic risk have focused on this highly diverse group of molecules. Serum lipid analyses have identified several specific lipids that are associated with elevated risk. Even so, to tailor therapeutic approaches to individual patients, let alone plaques, these associations remain insufficiently specific. Circulating lipids mainly reflect the systemic lipid metabolism, which is related to, and possibly even causative in, the atherosclerotic changes in the artery wall. The actual disease process in the arterial wall, however, remains largely out of view, which obscures a significant segment of the chain of processes that determine the event risk at the individual and plaque level.

Working from the hypothesis that the lipids in atherosclerotic plaque reflect its state of development and may even help in predicting its future, we provide here the foundations for plaque-lipid-based assessment of atherosclerosis. We developed a molecular imaging method for lipids in plaque tissue sections and relate these to optical spectroscopic imaging that may be translated towards *in vivo* detection of these lipids.

To this aim we designed a framework to use matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) to visualize the spatial distribution of more than 90 identifiable, unique lipid species in 12 patients, Chapter 2 and 3. We evaluated various unsupervised machine learning algorithms in terms of performance, providing insight into the highly complex data sets acquired using MALDI-MSI and compared the results to histology, Chapter 3 and 4. We built a photoacoustic microscope to measure vibration modes of lipids, compared this to MALDI-MSI data and validated spectral features with lipid standards. Finding photoacoustic spectra that may be related to plaque instability that can be measured *in vivo*, Chapter 5. We used mid-infrared photoacoustic microscopy to further investigate vibrational modes of different type of metabolites, showing a correlation between carbohydrates and cholesterol crystals, Chapter 6.

In this final chapter we discuss the novel insights provided in this thesis, future directions and clinical perspectives of the use spatial metabolomics in treatment strategies of carotid atherosclerosis.

7.1 MALDI-MSI data analysis

Mass spectrometry imaging, especially when applied in untargeted experiments aimed at discovery of chemical patterns associated with (pathological) biological processes, can result in very large, high-dimensional data sets. Additionally, specifics of the ion-

ization technique used, sample preparation and other methodological details impact the chemical sensitivity, linearity, signal-to-noise ratio and other characteristics of the data. These observations imply that standardization of the data acquisition and analysis is paramount for careful interpretation of patterns that appear in the acquired data, and also for an understanding of which features may be omitted, intentionally or not.

In this thesis, Chapter 2, we designed such a pipeline to acquire and process large amounts of MALDI-MSI data in a comprehensive manner. In short, it comprised a recipe for matrix deposition optimized for sensitivity to lipids in arterial specimens, spatial and spectral processing of the data, and feature selection of lipid and tissue specific *m/z* values.

Using this pipeline, we quantified the reproducibility of our measurements by comparing six MALDI-MSI measurements of adjacent tissue sections using the coefficient of variation (CV) as a measure for variability. The inherent reproducibility of the mass spectrometer was quantified by establishing the reproducibility of pure lipid standards spotted on tissue homogenate. We found the reproducibility of count rates for lipid ions in atherosclerotic tissue sections to be 12-25% which is similar to previously reported values[200,201], and corresponding to lipid standards which had a reproducibility of 16%. Variation between adjacent tissue sections is therefore not greater than would be expected when measuring the same tissue repeatedly, and is mostly caused by variation within the measurement.

Understanding variation in the measurement is particularly relevant when applying further downstream data analysis methods, such as multi-variate analysis and unsupervised clustering analysis as have been exploited in Chapter 3 and 4. Data mining techniques enable the visualization of underlying structures within the complex MALDI-MSI data. This is only possible and reliable if within group variability, e.g. variability due to the measurement and artifacts such as noise are sufficiently small[202,203]. Additionally, this indicates that our sample preparation, measurement protocol and data processing pipeline were robust and were not limiting the quality of the data.

Using the framework designed in Chapter 2, we analyzed a data set of 106 tissue sections originating from 12 carotid endarterectomy (CEA) samples in Chapter 3. This experiment represents the largest systematic analysis of MSI applied to atherosclerosis to date. Herein, we analyzed the data through two pathways. First, using non-negative matrix factorization (NMF), an unsupervised machine learning algorithm, which revealed five chemical clusters of lipids. The five clusters found in this analysis mainly show a distinction between different lipid classes having unique spatial distributions. The advantage of using these unsupervised methods is that they do not require any prior knowledge of the data, thus reducing susceptibility to biases introduced by analysis models which may presuppose relations between lipids and certain plaque features.

Additionally, we used the classification by means of matched histology to label spectra, and consequently performed multivariate analysis to identity the molecules that contributed to the separation between plaque components. Histopathology is regarded as the gold standard in tissue characterization[7,61,103]. The results of this analysis are in agreement with outcomes from our NMF analysis. This demonstrates that future work could be analyzed without the need for time and labor intensive histological segmentation process.

NMF is but one unsupervised classification algorithm, while many more exist. These different analysis methods can be grouped in different families according to their mathematical underpinning in matrix factorization, similarity clustering and manifold learning. In Chapter 4 we evaluated the performance of 5 different unsupervised machine learning algorithms[23,24,133], a matrix factorization technique[204], a clustering method[26] and manifold learning methods[27–29]. We compared this to histology based manual segmentation[7,61,103] of adjacent section, and showed that the mass spectrometry analysis reveals patterns in tissue composition that cannot be identified in classic histology interpretation.

Computational costs of manifold methods are relatively high and interpretation in terms of spectral content is complex and non-trivial. The patterns that emerge from analysis results of these methods are, on the isolated data sets that we studied, not very different than the results of NMF and spatial shrunken centroids (SSC). The latter two are simpler, more intuitive, and computationally cheaper methods. Non-linearity, which is inherent to MSI data and would be an argument for using a manifold methods, can be compensated by applying log-based variance-stabilizing normalization[104].

7.2 MALDI-MSI data interpretation

Based on the acquisition and analysis methods, of which we have demonstrated robustness and reproducibility, we can interpret the biological meaning of the observed lipid composition patterns. In Chapter 2 we subjected an autopsy sample with very mild plaque (intimal xanthoma) and two CEA samples to the developed pipeline. The presence of two cholesteryl esters was observed to be significantly different in these disease stages. Cholesteryl oleate (*m/z* 673.59, CE 18:1, $[M + Na]^+$) is more abundant in early stage plaques whereas cholesteryl linoleate (*m/z* 671.57, CE 18:2, $[M + Na]^+$) is more abundant in advanced plaques. This is in agreement with earlier studies[86–93], nonetheless this is the first mass spectrometry imaging study to show this difference.

Differences between cholesteryl oleate (CE 18:1) and cholesteryl linoleate (CE 18:2) were also seen within the data set acquired on advanced atherosclerotic plaques examined in Chapter 3. These molecules showed distinctly different spatial distributions throughout the advanced plaques. CE 18:2, which has been associated with advanced lesions in the literature, is present throughout the tissue section, as would be expected. CE 18:1, typically associated with early-stage plaques, however, is more locally present around the lumen, possibly indicating sites of biological activity related to plaque progression.

A large study, relating >90 identifiable lipid species, imaged in 106 tissue sections, to histological analysis, was performed with the goal to discover lipid biomarkers for plaque phenotype. Results from supervised and unsupervised data analysis were consistent and linked (oxidized) cholesteryl esters (CE) and sphingomyelins (SM), associated with plaque development previously[13,41,69,114–116], was co-localized with necrotic core areas. NMF analysis grouped these families of lipids in the same cluster, based on their spatial patterns. The cluster dominated by triacylglycerols (TAG) was co-localized with fibrin, a protein associated with thrombus formation. We also observed diacylglycerols

(DAG) in these areas. DAG can also form, however, by MALDI in-source fragmentation of TAG or phosphatidylcholines (PC), which makes definite assignment challenging. Spatial patterns of DAG were at least partially distinct from TAG and PC, supporting the observation of DAG in tissue. As discussed in Chapter 2, DAG has been associated with thrombus formation mechanisms.

Lipidomics studies are not new to the field of cardiovascular and atherosclerosis research. Many studies investigating serum lipids found ceramides (Cer) to be a biomarker for risk of plaque rupture[121,122,205,206]. In this thesis we focused on spatial imaging of lipids in the plaque. Combining both may give additional insights in the etiology of atherosclerosis and the interaction between the systemic and local lipid metabolism. In our MALDI-MSI data, we do not observe Cer and, in our non-imaging Lipidyzer data discussed in Chapter 3, Cer made up only 1% of lipids. This finding is in concordance with previous work, documenting an approximately 50 times higher abundance of SM than Cer in advanced plaques[71]. Cers and SMs are structurally related lipids, see Fig. 7.1. Looking at predictors of cardiovascular events, higher risk of ischemic stroke is associated with Cer elevation in serum[121,122], and SM elevation in plaque[13,71,132]. Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes[207]. However, precise mechanisms why the elevation of these two lipid classes increase risk of ischemic stroke are unclear. A question that arises is; are the serum Cers elevated because plaques are overflowing with SMs, or are there more SMs in the plaque because serum is loaded with Cers.

Ceramides originate from *de novo* synthesis, by hydrolysis of SMs catalyzed by sphingomyelinase (SMase), resulting in Cers and PCs[208], see Fig. 7.1. ox-LDL activates SMase which hydrolyses sphingomyelins to ceramides, increasing its levels[71]. The secretion of SMase triggers the aggregation of LDL particles within the arterial wall[206]. Moreover, aggregation prone LDL particles are enriched in SMs and Cers[206], causing a positive cascade effect for the accumulation of SMs and Cers.

Ceramides contribute to plaque inflammation and vascular smooth muscle apoptosis [20,71] and SMs are also associated with inflammation in plaques[71]. Of all circulating lipids, Cer(18:1/16:0) has the strongest association with coronary artery disease[209] and the largest hazard ratio for cardiovascular death[207], and is directly related to the most abundant SM in plaque, SM(34:1), through the SM synthesis pathway. These specific SM species, SM 34:1 but also SM 32:1, are found to be more abundant in plaque tissue compared to plasma, suggesting that these species are synthesized or selectively retained in atherosclerotic plaques[13]. Macrophages try to remove SMs from the plaque by hydrolysing it to Cers after apoptosis[210].

Generation of SMs can occur also from the transfer of the phosphocholine headgroup from PC to Cers, producing DAGs and SMs[208], see Fig. 7.1. Moreover, inhibition of PC may lead to a decrease of SM and an accumulation of Cers[211]. We found PCs to be more abundant around the lumen, Chapter 3, these PC around the lumen may be related to the exchange of SMs and Cers between plaque and plasma. Looking at the spatial distribution of enzymes that catalyze and hydrolyze these lipids may reveal insight towards the pathways of these two lipids and their interaction in atherosclerosis.

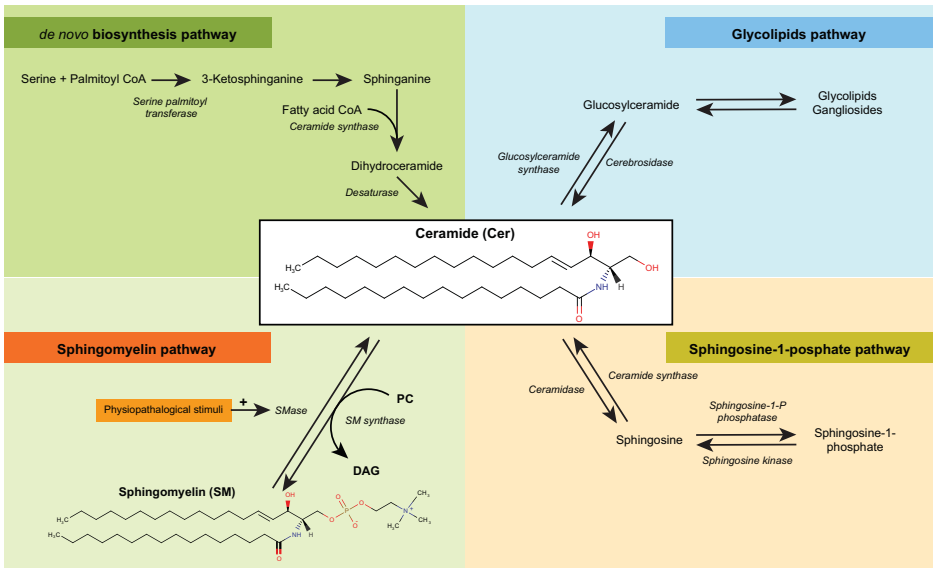


Figure 7.1: **Synthesis pathways of ceramides** showing *de novo* biosynthesis pathway, sphingomyelin pathway, glycolipid pathway and the sphingosine-1-posphate pathway.

One of the greatest challenges in MSI remains the unambiguous identification of molecules detected in the experiment. Often, MS/MS, the gold standard for identification, is only possible for a limited amount of molecules which are abundantly present in the tissue or easily ionized. In this thesis we aimed to tackle this issue by the use of high spectral resolution FTICR-MSI experiment, which allows us to benefit from the open source METASPACE project[111] which mines both spectral and spatial patterns for molecular identification. We combined this with quantitative analysis of a homogenized sample using the standardized Lipidyzer platform. Despite the fact that this identification is not conclusive we did manage to assign identifications to a majority of the detected *m/z* values that remain after preprocessing and filtering, and were able to label >90 unique lipid species in our endeavors. Such numbers are not often reported in MSI studies.

In this thesis we have limited our research to positive ion mode using sublimated DHB as matrix. This choice was made based on previous studies investigating lipids in the cardiovascular field[21], where good results were achieved using this matrix. This limits the lipid species being detected to those that readily ionize in positive mode. In the future the framework designed in this thesis can be easily adjusted to extend this research line towards other matrices and negative ion mode mass spectrometry.

Our studies were biased by the use of CEA samples, which are intrinsically advanced atherosclerotic plaques of symptomatic patients. This was an intentional choice for several reasons. First of all, a biobank of these samples was available. New specimens to extend the data are relatively easy to come by form routine surgery. Second, we limited the study to a specific plaque stage in order to make statistically substantiated conclusions. Increasing the number of variables would only extend an already large and

challenging MSI study, Chapter 3. For future studies it would be interesting to compare CEA samples with autopsy samples. Autopsy-originated plaques will be more diverse in terms of composition and disease stage, but can be classified using conventional histological criteria.

7.3 Photoacoustic microscopy

In Chapter 5 and 6 we explored the photoacoustic (PA) imaging of molecular vibrations using microscopy, with the aim of translation towards *in vivo* imaging of metabolites. The PA signal in the near infrared (NIR) and the mid-infrared (mid-IR) originates from the vibrations of atomic bonds. The signal in NIR, between 1150 and 1240 nm, is mainly a result of the second C-H overtone, asymmetrically stretched, with additional peaks arising due to difference in hyrdocarbon chainlengths[164] and other differences in the local molecular configuration. The signals in the mid-IR are similar to those measured in Raman and FTIR, capable of measuring e.g. lipids, amide I and II bonds, and ester-bonds[48].

In Chapter 5, we built a novel PA microscope capable of spectral imaging of cryo-sections in the NIR, 1150-1240 nm. We used adjacent MALDI-MSI sections as a molecular ground truth and attempted to derive PA spectra associated with features of plaque vulnerability. In our unsupervised analysis of MALDI-MSI data we discovered 5 components explaining the data, Chapter 3. We used a similar approach to analyze the μsPA data and found 3 components, Chapter 5. In MALDI-MSI we have found that lipid classes cluster together due to similar spatial distributions. However, with PA we discovered that each lipid class may show a wide variety of spectra, depending on chain length and number of unsaturated bonds. Moreover, lipids from different classes may have almost identical spectra, making validation of PA spectra with MALDI-MSI complex.

Two lipids that are spectrally indistinguishable, at the sensitivity and resolution available in our instrument, are cholesteryl linoleate (CE 18:2) and sphingomyelin d18:1/16:0 (SM 34:1), see Fig. 7.2. Both lipids have been identified as major contributors to atherosclerosis in the carotid artery[13,66,132]. PA studies identifying lipid in atherosclerosis, have been directed towards the detection of cholesteryl esters[47,50,51,159]. It is possible that they were actually looking at SM 34:1. However, despite the fact that these lipids cannot be discriminated from each other, this spectrum can still be classified as a spectrum associated with necrotic core in advanced atherosclerosis.

In Chapter 6 we moved from the NIR to the mid-IR range where the vibrational modes are more diverse and easier to distinguish from each other. Mid-IR optoacoustic microscopy (MIROM) provides a form of label-free analytic histology that can directly visualize chemical composition. We found that cholesterol crystals were co-localized with a signal assigned to carbohydrates and were able to verify our finding by histology staining of polysaccharides. Cholesterol crystals are observed in very advanced plaque and are frequently seen in ruptured plaques[212]. Cholesterol expands in volume when crystallizing and the sharp tip may destabilize the fibrous cap or provoke an inflammatory response[63,171,190,213,214]. We were unable to identify the type of carbohydrate co-localized with these crystals. A candidate to investigate could be mannose, possibly as

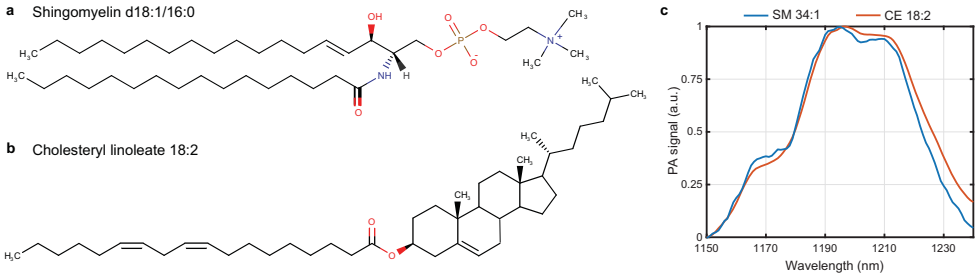


Figure 7.2: **Structural and spectral difference and similarities between SM 34:1 and CE 18:2. a)** Chemical structure of sphingomyelin d18:1/16:0 (SM 34:1) **b)** Chemical structure of cholesteryl linoleate (CE 18:2), **c)** μsPA recorded spectrum of lipid extracts, sphingomyelin d18:1/16:0 (SM 34:1) in blue and cholesteryl linoleate (CE 18:2) in red.

a sugar group on a glycated protein. Receptors for mannose on macrophages, which are activated by cholesterol crystallization[63], have been found to be highly abundant in thin cap fibroatheroma (TCFA)[215,216], and associated with intraplaque hemorrhage and neovascularization[216–218], all features related to high-risk plaques.

Further insight to the relation between carbohydrates and cholesterol crystals, Chapter 6, may give further insights into the etiology of these features associated with high-risk plaques. Investigating their spatial distribution with high resolution and specificity however is not trivial. MALDI-MSI in itself would not be a very suitable technique for looking at the single carbohydrates since many of them have the same *m/z* value. Looking at N-glycans is possible with MALDI-MSI; to do this an enzyme is used to perform on-tissue digestion after which MALDI-MSI is capable of imaging the released species[219,220].

7.4 Conclusion

Coming back to the aim of this thesis, *in vivo* imaging of features related to high-risk plaques. In this thesis we have identified a number of lipids that are characteristic for high-risk plaque phenotype. We have imaged these lipids using μsPA and found that the spectra of two of these lipids are identical. Earlier research has shown the feasibility of spectral imaging of these lipids in the carotid artery by internal illumination via the pharynx and external detection of the signal[51]. The next step in this endeavor should be macroscopic and microscopic imaging of the same carotid artery sample, to see how the spectra translate to larger scale imaging. In addition, investigations should aim at the identification of difference between early and advanced atherosclerotic lesions, in terms of composition, to gain further insights in the first signs of the development of event prone plaques.

Bibliography

1. Holm Nielsen, S., Jonasson, L., Kalogeropoulos, K., *et al.* Exploring the role of extracellular matrix proteins to develop biomarkers of plaque vulnerability and outcome. *Journal of Internal Medicine* **287,** 1–21 (2020).
2. Elffers, T. W. & Jukema, J. W. Het onvoorspelbare beloop van coronaire atherosclerose. *Nederlands tijdschrift voor geneeskunde* **162,** D2109 (2018).
3. Zhang, Y., Koradia, A., Kamato, D., Popat, A., Little, P. J. & Ta, H. T. Treatment of atherosclerotic plaque: perspectives on theranostics. *Journal of Pharmacy and Pharmacology* **71,** 1029–1043 (2019).
4. Gupta, K. K., Ali, S. & Sanghera, R. S. Pharmacological Options in Atherosclerosis: A Review of the Existing Evidence. *Cardiology and Therapy* **8,** 5–20 (2019).
5. Jansen, M. F., Lutgens, E. & Daemen, M. J. A. P. in *Vessel Based Imaging Techniques* 69–91 (Springer International Publishing, 2020).
6. Meikle, P. J., Wong, G., Tsorotes, D., *et al.* Plasma lipidomic analysis of stable and unstable coronary artery disease. *Arteriosclerosis, Thrombosis, and Vascular Biology* **31,** 2723–2732 (2011).
7. Virmani, R., Kolodgie, F. D., Burke, A. P., Farb, A. & Schwartz, S. M. Lessons From Sudden Coronary Death A Comprehensive Morphological Classification Scheme for Atherosclerotic Lesions. *Arteriosclerosis, Thrombosis, and Vascular Biology* **20,** 1262–1275 (2000).
8. Jashari, F., Ibrahimi, P., Nicoll, R., Bajraktari, G., Wester, P. & Henein, M. Y. Coronary and carotid atherosclerosis: Similarities and differences. *Atherosclerosis* **227,** 193–200 (2013).
9. Martinez, E., Martorell, J. & Riambau, V. Review of serum biomarkers in carotid atherosclerosis. *Journal of Vascular Surgery* **71,** 329–341 (2020).
10. Sun, J. & Hatsukami, T. S. in *Vessel Based Imaging Techniques : Diagnosis, Treatment, and Prevention* 111–134 (Springer International Publishing, 2020).
11. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* **69,** 89–95 (2001).
12. Ravindran, K., Brinijiki, W., DeMarco, J. K. & Huston, J. in *Vessel Based Imaging Techniques : Diagnosis, Treatment, and Prevention* 93–109 (Springer International Publishing, 2020).
13. Stegemann, C., Drozdov, I., Shalhoub, J., *et al.* Comparative lipidomics profiling of human atherosclerotic plaques. *Circulation: Cardiovascular Genetics* **4,** 232–242 (2011).

14. Kolovou, G., Kolovou, V. & Mavrogeni, S. Lipidomics in vascular health: current perspectives. *Vascular Health and Risk Management* **11,** 333 (2015).
15. Felton, C. V., Crook, D, Davies, M. J. & Oliver, M. F. Relation of plaque lipid composition and morphology to the stability of human aortic plaques. *Arteriosclerosis, thrombosis, and vascular biology* **17,** 1337–45 (1997).
16. Davies, M. J. Stability and Instability: Two Faces of Coronary Atherosclerosis. *Circulation* **94,** 2013–2020 (1996).
17. Buchberger, A. R., DeLaney, K., Johnson, J. & Li, L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Analytical Chemistry* **90,** 240–265 (2018).
18. Chughtai, K. & Heeren, R. M. Mass spectrometric imaging for biomedical tissue analysis. *Chemical Reviews* **110,** 3237–3277 (2010).
19. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Analytical Chemistry* **69,** 4751–4760 (1997).
20. Berry, K. A. Z., Hankin, J. A., Barkley, R. M., Spraggins, J. M., Caprioli, R. M. & Murphy, R. C. MALDI imaging of lipid biochemistry in tissues by mass spectrometry. *Chemical reviews* **111,** 6491–512 (2011).
21. Mezger, S. T., Mingels, A. M., Bekers, O., Cillero-Pastor, B. & Heeren, R. M. Trends in mass spectrometry imaging for cardiovascular diseases. *Analytical and Bioanalytical Chemistry* **411,** 3709–3720 (2019).
22. Vaysse, P.-M., Heeren, R. M. A., Porta, T. & Balluff, B. Mass spectrometry imaging for clinical research - latest developments, applications, and current limitations. *The Analyst* **142,** 2690–2712 (2017).
23. Verbeeck, N., Caprioli, R. M. & Van de Plas, R. Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. *Mass Spectrometry Reviews* **39,** 1–47 (2019).
24. Alexandrov, T. MALDI imaging mass spectrometry: statistical data analysis and current computational challenges. *BMC Bioinformatics* **13,** S11 (2012).
25. Tu, A. & Muddiman, D. C. Systematic evaluation of repeatability of IR-MALDESI-MS and normalization strategies for correcting the analytical variation and improving image quality. *Analytical and Bioanalytical Chemistry* **411,** 5729–5743 (2019).
26. Bemis, K. D., Harry, A., Eberlin, L. S., Ferreira, C. R., van de Ven, S. M., Mallick, P., Stolowitz, M. & Vitek, O. Probabilistic segmentation of mass spectrometry (MS) images helps select important ions and characterize confidence in the resulting segments. *Molecular and Cellular Proteomics* **15,** 1761–1772 (2016).
27. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research* **9,** 2579–2605 (2008).
28. Fonville, J. M., Carter, C. L., Pizarro, L., *et al.* Hyperspectral Visualization of Mass Spectrometry Imaging Data. *Analytical Chemistry* **85,** 1415–1423 (2013).
29. Thomas, S. A., Race, A. M., Steven, R. T., Gilmore, I. S. & Bunch, J. Dimensionality reduction of mass spectrometry imaging data using autoencoders. *2016 IEEE Symposium Series on Computational Intelligence (SSCI),* 1–7 (2016).

30. Smets, T., Verbeeck, N., Claesen, M., *et al.* Evaluation of Distance Metrics and Spatial Autocorrelation in Uniform Manifold Approximation and Projection Applied to Mass Spectrometry Imaging Data. *Analytical Chemistry* **91,** 5706–5714 (2019).
31. Inglese, P., McKenzie, J. S., Mroz, A., *et al.* Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer. *Chemical science* **8,** 3500–3511 (2017).
32. Abdelmoula, W. M., Pezzotti, N., Hölt, T., Dijkstra, J., Vilanova, A., McDonnell, L. A. & Lelieveldt, B. P. Interactive Visual Exploration of 3D Mass Spectrometry Imaging Data Using Hierarchical Stochastic Neighbor Embedding Reveals Spatiomolecular Structures at Full Data Resolution. *Journal of Proteome Research* **17,** 1054–1064 (2018).
33. Castro-Perez, J., Hatcher, N., Kofi Karikari, N., *et al.* In vivo isotopically labeled atherosclerotic aorta plaques in ApoE KO mice and molecular profiling by matrix-assisted laser desorption/ionization mass spectrometric imaging. *Rapid Communications in Mass Spectrometry* **28,** 2471–2479 (2014).
34. Martin-Lorenzo, M., Balluff, B., Maroto, A. S., *et al.* Lipid and protein maps defining arterial layers in atherosclerotic aorta. *Data in Brief* **4,** 328–331 (2015).
35. Martin-Lorenzo, M., Balluff, B., Sanz-Maroto, A., van Zeijl, R. J. M., Vivanco, F., Alvarez-Llamas, G. & McDonnell, L. A. 30μm spatial resolution protein MALDI MSI: In-depth comparison of five sample preparation protocols applied to human healthy and atherosclerotic arteries. *Journal of Proteomics* **108,** 465–468 (2014).
36. Zaima, N., Sasaki, T., Tanaka, H., *et al.* Imaging mass spectrometry-based histopathologic examination of atherosclerotic lesions. *Atherosclerosis* **217,** 427–432 (2011).
37. Patterson, N. H., Doonan, R. J., Daskalopoulou, S. S., Dufresne, M., Lenglet, S., Montecucco, F., Thomas, A. & Chaurand, P. Three-dimensional imaging MS of lipids in atherosclerotic plaques: Open-source methods for reconstruction and analysis. *Proteomics* **16,** 1642–1651 (2016).
38. Tanaka, H., Zaima, N., Sasaki, T., *et al.* Lysophosphatidylcholine Acyltransferase-3 Expression Is Associated with Atherosclerosis Progression. *Journal of Vascular Research* **54,** 200–208 (2017).
39. Shi, Y., Johnson, J., Wang, B., *et al.* Mass Spectrometric Imaging Reveals Temporal and Spatial Dynamics of Bioactive Lipids in Arteries Undergoing Restenosis. *Journal of Proteome Research* **18,** 1669–1678 (2019).
40. Cao, J., Goossens, P., Martin-Lorenzo, M., Dewez, F., Claes, B. S. R., Biessen, E. A. L., Heeren, R. M. A. & Balluff, B. Atheroma-Specific Lipids in ldlr −/− and apoe −/− Mice Using 2D and 3D Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging. *Journal of the American Society for Mass Spectrometry* **31,** 1825–1832 (2020).
41. Manicke, N., Nefliu, M., Wu, C, Woods, J., Reiser, V., Hendrickson, R. & Cooks, R. Imaging of lipids in atheroma by desorption electrospray ionization mass spectrometry. *Analytical Chemistry* **81,** 8702–8707 (2009).
42. Bodzon-Kulakowska, A., Drabik, A., Mystkowska, J., *et al.* Desorption electrospray ionization-based imaging of interaction between vascular graft and human body. *Journal of Biomedical Materials Research Part B: Applied Biomaterials* **104,** 192–196 (2016).

43. Malmberg, P., Börner, K., Chen, Y., Friberg, P., Hagenhoff, B., Månsson, J. E. & Nygren, H. Localization of lipids in the aortic wall with imaging TOF-SIMS. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1771,** 185–195 (2007).
44. Mas, S., Touboul, D., Brunelle, A., Aragoncillo, P., Egido, J., Laprévote, O. & Vivanco, F. Lipid cartography of atherosclerotic plaque by cluster-TOF-SIMS imaging. *The Analyst* **132,** 24–26 (2007).
45. Bot, M., de Jager, S. C. A., MacAleese, L., *et al.* Lysophosphatidic acid triggers mast cell-driven atherosclerotic plaque destabilization by increasing vascular inflammation. *Journal of Lipid Research* **54,** 1265–1274 (2013).
46. Zhao, T., Desjardins, A. E., Ourselin, S., Vercauteren, T. & Xia, W. Minimally invasive photoacoustic imaging: Current status and future perspectives. *Photoacoustics* **16,** 100146 (2019).
47. Jansen, K., Wu, M., van der Steen, A. F. & van Soest, G. Photoacoustic imaging of human coronary atherosclerosis in two spectral bands. *Photoacoustics* **2,** 12–20 (2014).
48. Pleitez, M. A., Khan, A. A., Soldà, A., *et al.* Label-free metabolic imaging by mid-infrared optoacoustic microscopy in living cells. *Nature Biotechnology* **38,** 293–296 (2020).
49. Jansen, K., van Soest, G. & van der Steen, A. F. Intravascular Photoacoustic Imaging: A New Tool for Vulnerable Plaque Identification. *Ultrasound in Medicine & Biology* **40,** 1037–1048 (2014).
50. Wu, M., Springeling, G., Lovrak, M., *et al.* Real-time volumetric lipid imaging in vivo by intravascular photoacoustics at 20 frames per second. *Biomedical Optics Express* **8,** 943 (2017).
51. Kruizinga, P., van der Steen, A. F. W., de Jong, N., Springeling, G., Robertus, J. L., van der Lugt, A. & van Soest, G. Photoacoustic imaging of carotid artery atherosclerosis. *Journal of Biomedical Optics* **19,** 110504 (2014).
52. Wong, T. T., Zhang, R., Hai, P., Zhang, C., Pleitez, M. A., Aft, R. L., Novack, D. V. & Wang, L. V. Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy. *Science Advances* **3,** e1602168 (2017).
53. Wang, B., Su, J. L., Amirian, J., Litovsky, S. H., Smalling, R. & Emelianov, S. Detection of lipid in atherosclerotic vessels using ultrasound-guided spectroscopic intravascular photoacoustic imaging. *Optics Express* **18,** 4889 (2010).
54. Wang, P., Wang, H.-W., Sturek, M. & Cheng, J.-X. Bond-selective imaging of deep tissue through the optical window between 1600 and 1850 nm. *Journal of Biophotonics* **5,** 25–32 (2012).
55. Iskander-Rizk, S., van der Steen, A. F. W. & van Soest, G. Photoacoustic imaging for guidance of interventions in cardiovascular medicine. *Physics in Medicine & Biology* **64** (2019).
56. Merčep, E., Deán-Ben, X. L. & Razansky, D. Imaging of blood flow and oxygen state with a multi-segment optoacoustic ultrasound array. *Photoacoustics* **10,** 48–53 (2018).
57. Ivankovic, I., Merčep, E., Schmedt, C.-G., Deán-Ben, X. L. & Razansky, D. Real-time Volumetric Assessment of the Human Carotid Artery: Handheld Multispectral Optoacoustic Tomography. *Radiology* **291,** 45–50 (2019).

58. Arabul, M. U., Heres, M., Rutten, M. C. M., van Sambeek, M. R., van de Vosse, F. N. & Lopata, R. G. P. Toward the detection of intraplaque hemorrhage in carotid artery lesions using photoacoustic imaging. *Journal of Biomedical Optics* **22,** 041010 (2016).
59. Seeger, M., Karlas, A., Soliman, D., Pelisek, J. & Ntziachristos, V. Multimodal optoacoustic and multiphoton microscopy of human carotid atheroma. *Photoacoustics* **4,** 102–111 (2016).
60. WHO | Cardiovascular diseases (CVDs). *WHO* (2018).
61. Virmani, R., Burke, A. P., Farb, A. & Kolodgie, F. D. Pathology of the Vulnerable Plaque. *Journal of the American College of Cardiology* **47** (2006).
62. Schaar, J, Muller, J. E., Falk, E., *et al.* Terminology for high-risk and vulnerable coronary artery plaques. *European Heart Journal* **25,** 1077–1082 (2004).
63. Duewell, P., Kono, H., Rayner, K. J., *et al.* NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. *Nature* **464,** 1357–61 (2010).
64. Rader, D. J. & Puré, E. Lipoproteins, macrophage function, and atherosclerosis: Beyond the foam cell? *Cell Metabolism* **1,** 223–230 (2005).
65. Silvestre-Roig, C., de Winther, M. P., Weber, C., Daemen, M. J., Lutgens, E. & Soehnlein, O. Atherosclerotic Plaque Destabilization. *Circulation Research* **114,** 214–226 (2014).
66. Small, D. M. George Lyman Duff memorial lecture. Progression and regression of atherosclerotic lesions. Insights from lipid physical biochemistry. *Arteriosclerosis (Dallas, Tex.)* **8,** 103–29 (1988).
67. Waksman, R. & on behalf of the LRP Investigators. *Assessment of Coronary Near-Infrared Spectroscopy Imaging to Detect Vulnerable Plaques and Vulnerable Patients: The Lipid-Rich Plaque Study* in (TCT2018, San Diego, 24 September 2018).
68. Norris, J. L. & Caprioli, R. M. Analysis of Tissue Specimens by Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry in Biological and Clinical Research. *Chemical reviews* **113,** 2309–2342 (2013).
69. Hutchins, P. M., Moore, E. E. & Murphy, R. C. Electrospray MS/MS reveals extensive and nonspecific oxidation of cholesterol esters in human peripheral vascular lesions. *Journal of Lipid Research* **52,** 2070–2083 (2011).
70. Lehti, S., Sjövall, P., Käkelä, R., Mäyränpää, M. I., Kovanen, P. T. & Öörni, K. Spatial distributions of lipids in atherosclerosis of human coronary arteries studied by time-of-flight secondary ion mass spectrometry. *American Journal of Pathology* **185,** 1216–1233 (2015).
71. Edsfeldt, A., Dunér, P., Ståhlman, M., *et al.* Sphingolipids Contribute to Human Atherosclerotic Plaque Inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology* **36,** 1132–1140 (2016).
72. Zavodni, A. E. H., Wasserman, B. A., McClelland, R. L., Gomes, A. S., Folsom, A. R., Polak, J. F., Lima, J. A. C. & Bluemke, D. A. Carotid Artery Plaque Morphology and Composition in Relation to Incident Cardiovascular Events: The Multi-Ethnic Study of Atherosclerosis (MESA). *Radiology* **271,** 381–389 (2014).
73. Wijeyaratne, S. M., Abbott, C. R. & Gough, M. J. A Modification to the Standard Technique for Carotid Endarterectomy Allowing Removal of Intact Endarterectomy Specimens: Implications for Research and Quality Control of Preoperative Imaging. *European Journal of Vascular and Endovascular Surgery* **23,** 370–371 (2002).

74. Dekker, L. J. M., van Kampen, J. J. A., Reedijk, M. L., Burgers, P. C., Gruters, R. A., Osterhaus, A. D. M. E. & Luider, T. M. A mass spectrometry based imaging method developed for the intracellular detection of HIV protease inhibitors. *Rapid Communications in Mass Spectrometry* **23,** 1183–1188 (2009).
75. De Wit, N. M., Snkhchyan, H., den Hoedt, S., *et al.* Altered Sphingolipid Balance in Capillary Cerebral Amyloid Angiopathy. *Journal of Alzheimer's Disease* **60,** 795–807 (2017).
76. Angel, P. M., Bayoumi, A. S., Hinton, R. B., Ru Su, Y., Bichell, D., Mayer, J. E., Scott Baldwin, H & Caprioli, R. M. MALDI Imaging Mass Spectrometry as a Lipidomic Approach to Heart Valve Research. *The Journal of heart valve disease* **25,** 240–252 (2016).
77. Liebisch, G., Binder, M., Schifferer, R., Langmann, T., Schulz, B. & Schmitz, G. High throughput quantification of cholesterol and cholesteryl ester by electrospray ionization tandem mass spectrometry (ESI-MS/MS). *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1761,** 121–128 (2006).
78. McDonnell, L. A., van Remoortere, A., de Velde, N., van Zeijl, R. J. & Deelder, A. M. Imaging Mass Spectrometry Data Reduction: Automated Feature Identification and Extraction. *Journal of the American Society for Mass Spectrometry* **21,** 1969–1978 (2010).
79. Strohalm, M., Hassman, M., Košata, B. & Kodíček, M. mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass Spectrometry* **22,** 905–908 (2008).
80. Jones, J. J., Stump, M. J., Fleming, R. C., Lay, J. O. & Wilkins, C. L. Strategies and data analysis techniques for lipid and phospholipid chemistry elucidation by intact cell MALDI-FTMS. *Journal of the American Society for Mass Spectrometry* **15,** 1665–1674 (2004).
81. Fonville, J. M., Carter, C., Cloarec, O., Nicholson, J. K., Lindon, J. C., Bunch, J. & Holmes, E. Robust data processing and normalization strategy for MALDI mass spectrometric imaging. *Analytical Chemistry* **84,** 1310–1319 (2012).
82. Dill, A. L., Eberlin, L. S., Costa, A. B., Ifa, D. R. & Cooks, R. G. Data quality in tissue analysis using desorption electrospray ionization. *Analytical and Bioanalytical Chemistry* **401,** 1949–1961 (2011).
83. Tillner, J., Wu, V., Jones, E. A., *et al.* Faster, More Reproducible DESI-MS for Biological Tissue Imaging. *Journal of The American Society for Mass Spectrometry* **28,** 2090–2098 (2017).
84. Eisenberg, D. T. A., Kuzawa, C. W. & Hayes, M. G. Improving qPCR telomere length assays: Controlling for well position effects increases statistical power. *American journal of human biology : the official journal of the Human Biology Council* **27,** 570–5 (2015).
85. Fuchs, B., Süß, R. & Schiller, J. An update of MALDI-TOF mass spectrometry in lipid research. *Progress in Lipid Research* **49,** 450–475 (4 Oct. 2010).
86. Brown, J. M., Chung, S., Sawyer, J. K., *et al.* Combined Therapy of Dietary Fish Oil and Stearoyl-CoA Desaturase 1 Inhibition Prevents the Metabolic Syndrome and Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **30,** 24–30 (2010).

87. Degirolamo, C., Shelness, G. S. & Rudel, L. L. LDL cholesteryl oleate as a predictor for atherosclerosis: evidence from human and animal studies on dietary fat. *Journal of lipid research* **50,** S434–9 (2009).
88. Lang, P. D. & Insull, W. Lipid droplets in atherosclerotic fatty streaks of human aorta. *Journal of Clinical Investigation* **49,** 1479–1488 (1970).
89. Lee, E.-S., Shon, H. K., Lee, T. G., Kim, S.-H. & Moon, D. W. The regional ratio of cholesteryl palmitate to cholesteryl oleate measured by ToF-SIMS as a key parameter of atherosclerosis. *Atherosclerosis* **226,** 378–384 (2013).
90. Melchior, J. T., Sawyer, J. K., Kelley, K. L., Shah, R., Wilson, M. D., Hantgan, R. R. & Rudel, L. L. LDL particle core enrichment in cholesteryl oleate increases proteoglycan binding and promotes atherosclerosis. *Journal of Lipid Research* **54,** 2495–2503 (2013).
91. Weinmann, P, Jouan, M, Nguyen, Q. D., Lacroix, B, Groiselle, C, Bonte, J. P. & Luc, G. Quantitative analysis of cholesterol and cholesteryl esters in human atherosclerotic plaques using near-infrared Raman spectroscopy. *Atherosclerosis* **140,** 81–8 (1998).
92. Spector, A. A. & Haynes, W. G. LDL Cholesteryl Oleate: A Biomarker for Atherosclerosis? *Arteriosclerosis, Thrombosis, and Vascular Biology* **27,** 1228–1230 (2007).
93. Rapp, J. H., Connor, W. E., Lin, D. S., Inahara, T & Porter, J. M. Lipids of human atherosclerotic plaques and xanthomas: clues to the mechanism of plaque progression. *Journal of lipid research* **24,** 1329–35 (1983).
94. Bowden, J. A., Albert, C. J., Barnaby, O. S. & Ford, D. A. Analysis of cholesteryl esters and diacylglycerols using lithiated adducts and electrospray ionization-tandem mass spectrometry. *Analytical Biochemistry* **417,** 202–210 (2011).
95. Werner, M. H., Bielawska, A. E. & Hannun, Y. A. Quantitative analysis of diacylglycerol second messengers in human platelets: correlation with aggregation and secretion. *Molecular pharmacology* **41,** 382–6 (1992).
96. Heemskerk, J. W. M., Mattheij, N. J. A. & Cosemans, J. M. E. M. Platelet-based coagulation: different populations, different functions. *Journal of Thrombosis and Haemostasis* **11,** 2–16 (2013).
97. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet (London, England)* **390,** 1151–1210 (2017).
98. Lusis, A. J. Atherosclerosis. *Nature* **407,** 233–241 (2000).
99. Rajamäki, K., Lappalainen, J., Öörni, K., Välimäki, E., Matikainen, S., Kovanen, P. T. & Eklund, K. K. Cholesterol Crystals Activate the NLRP3 Inflammasome in Human Macrophages: A Novel Link between Cholesterol Metabolism and Inflammation. *PLoS ONE* **5** (2010).
100. Eberlin, L. S., Liu, X., Ferreira, C. R., Santagata, S., Agar, N. Y. & Cooks, R. G. Desorption Electrospray Ionization then MALDI Mass Spectrometry Imaging of Lipid and Protein Distributions in Single Tissue Sections. *Analytical Chemistry* **83,** 8366–8371 (2011).

101. Martin-Lorenzo, M., Alvarez-Llamas, G., McDonnell, L. A. & Vivanco, F. Molecular histology of arteries: Mass spectrometry imaging as a novel ex vivo tool to investigate Atherosclerosis. *Expert Review of Proteomics* **13,** 69–81 (2016).
102. Visscher, M., Moerman, A. M., Burgers, P. C., *et al.* Data Processing Pipeline for Lipid Profiling of Carotid Atherosclerotic Plaque with Mass Spectrometry Imaging. *Journal of The American Society for Mass Spectrometry* **30,** 1790–1800 (2019).
103. Stary, H. C., Chandler, A. B., Dinsmore, R. E., *et al.* A Definition of Advanced Types of Atherosclerotic Lesions and a Histological Classification of Atherosclerosis. *Circulation* **92,** 1355–1374 (1995).
104. Veselkov, K. A., Mirnezami, R., Strittmatter, N., *et al.* Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 1216–1221 (2014).
105. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401,** 788–791 (1999).
106. Li, Y. & Ngom, A. The non-negative matrix factorization toolbox for biological data mining. *Source Code for Biology and Medicine* **8,** 10 (2013).
107. Kim, H. & Park, H. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* **23,** 1495–1502 (2007).
108. Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E. & Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics* **20,** 341–351 (2006).
109. Eriksson, L., Trygg, J. & Wold, S. CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics* **22,** 594–600 (2008).
110. Galindo-Prieto, B., Eriksson, L. & Trygg, J. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *Journal of Chemometrics* **28,** 623–632 (2014).
111. Palmer, A., Phapale, P., Chernyavsky, I., *et al.* FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nature Methods* **14,** 57–60 (2016).
112. Steinberg, D. The LDL modification hypothesis of atherogenesis: an update. *Journal of Lipid Research* **50,** S376–S381 (2009).
113. Steinbrecher, U. P., Zhang, H. & Lougheed, M. Role of oxidatively modified LDL in atherosclerosis. *Free Radical Biology and Medicine* **9,** 155–168 (1990).
114. Brown, A. J., Leong, S. L., Dean, R. T. & Jessup, W. 7-Hydroperoxycholesterol and its products in oxidized low density lipoprotein and human atherosclerotic plaque. *Journal of lipid research* **38,** 1730–45 (1997).
115. Suarna, C., Dean, R. T., May, J. & Stocker, R. Human Atherosclerotic Plaque Contains Both Oxidized Lipids and Relatively Large Amounts of α-Tocopherol and Ascorbate. *Arteriosclerosis, Thrombosis, and Vascular Biology* **15,** 1616–1624 (1995).
116. Upston, J. M., Niu, X., Brown, A. J., *et al.* Disease Stage-Dependent Accumulation of Lipid and Protein Oxidation Products in Human Atherosclerosis. *The American Journal of Pathology* **160,** 701–710 (2002).

117. Stegemann, C., Pechlaner, R., Willeit, P., *et al.* Lipidomics profiling and risk of cardiovascular disease in the prospective population-based bruneck study. *Circulation* **129,** 1821–1831 (2014).
118. Smith, E. Intimal and medial lipids in human aortas. *The Lancet* **275,** 799–803 (1960).
119. Duivenvoorden, R., van Wijk, D., Klimas, M., Kastelein, J. J., Stroes, E. S. & Nederveen, A. J. Detection of Liquid Phase Cholesteryl Ester in Carotid Atherosclerosis by 1H-MR Spectroscopy in Humans. *JACC: Cardiovascular Imaging* **6,** 1277–1284 (2013).
120. Jansen, K., van der Steen, A. F., Wu, M., *et al.* Spectroscopic intravascular photoacoustic imaging of lipids in atherosclerosis. *Journal of Biomedical Optics* **19,** 026006 (2014).
121. Cheng, J. M., Suoniemi, M., Kardys, I., *et al.* Plasma concentrations of molecular lipid species in relation to coronary plaque characteristics and cardiovascular outcome: Results of the ATHEROREMO-IVUS study. *Atherosclerosis* **243,** 560–566 (2015).
122. Hilvo, M., Meikle, P. J., Pedersen, E. R., *et al.* Development and validation of a ceramide- and phospholipid-based cardiovascular risk estimation score for coronary artery disease patients. *European Heart Journal* **41,** 371–380 (2019).
123. Crisby, M., Nilsson, J., Kostulas, V., Björkhem, I. & Diczfalusy, U. Localization of sterol 27-hydroxylase immuno-reactivity in human atherosclerotic plaques. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1344,** 278–285 (1997).
124. Anderson, A., Campo, A., Fulton, E., Corwin, A., Jerome, W. G. & O'Connor, M. S. 7-Ketocholesterol in disease and aging. *Redox Biology* **29,** 101380 (2020).
125. Martinet, W., De Bie, M., Schrijvers, D. M., De Meyer, G. R. Y., Herman, A. G. & Kockx, M. M. 7-ketocholesterol induces protein ubiquitination, myelin figure formation, and light chain 3 processing in vascular smooth muscle cells. *Arteriosclerosis, thrombosis, and vascular biology* **24,** 2296–301 (2004).
126. Rask-Madsen, C. & King, G. L. Proatherosclerotic Mechanisms Involving Protein Kinase C in Diabetes and Insulin Resistance. *Arteriosclerosis, Thrombosis, and Vascular Biology* **25,** 487–496 (2005).
127. D'Souza, A., Hussain, M., Howarth, F. C., Woods, N. M., Bidasee, K. & Singh, J. Pathogenesis and pathophysiology of accelerated atherosclerosis in the diabetic heart. *Molecular and Cellular Biochemistry* **331,** 89–116 (2009).
128. Harper, M. T. & Poole, A. W. Diverse functions of protein kinase C isoforms in platelet activation and thrombus formation. *Journal of Thrombosis and Haemostasis* **8,** 454–462 (2010).
129. Konopatskaya, O., Matthews, S. A., Harper, M. T., *et al.* Protein kinase C mediates platelet secretion and thrombus formation through protein kinase D2. *Blood* **118,** 416–24 (2011).
130. Van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology* **9,** 112–124 (2008).
131. Hasegawa, M., Hakamata, H., Matsunaga, I. & Kusu, F. Detection of oxysterols in oxidatively modified low density lipoprotein by MALDI-TOF MS. *European Journal of Lipid Science and Technology* **113,** 423–429 (2011).

132. Ravandi, A., Babaei, S., Leung, R. & Monge, J. Phospholipids and oxophospholipids in atherosclerotic plaques at different stages of plaque development. *Lipids* **39,** 97–109 (2004).
133. Alexandrov, T. Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence. *Annual Review of Biomedical Data Science* **3,** 61–87 (2020).
134. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences* **99,** 6567–6572 (2002).
135. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Class Prediction by Nearest Shrunken Centroids, with Applications to DNA Microarrays. *Statistical Science* **18,** 104–117 (2003).
136. Selesnick, I. Penalty and shrinkage functions for sparse signal processing. *Connexions,* http://cnx.org/content/m45134/1.1/ (2012).
137. Bemis, K. D., Harry, A., Eberlin, L. S., Ferreira, C., Van De Ven, S. M., Mallick, P., Stolowitz, M. & Vitek, O. Cardinal: An R package for statistical analysis of mass spectrometry-based imaging experiments. *Bioinformatics* **31** (2015).
138. Dexter, A., Thomas, S. A., Steven, R. T., *et al.* Training a neural network to learn other dimensionality reduction removes data size restrictions in bioinformatics and provides a new route to exploring data representations. *bioRxiv,* 2020.09.03.269555 (2020).
139. Abdelmoula, W. M., Gimenez-Cassina Lopez, B., Randall, E. C., *et al.* msiPL: Nonlinear Manifold and Peak Learning of Mass Spectrometry Imaging Data Using Artificial Neural Networks. *bioRxiv,* 2020.08.13.250142 (2020).
140. Timmis, A., Townsend, N., Gale, C. P., *et al.* European society of cardiology: Cardiovascular disease statistics 2019. *European Heart Journal* **41,** 12–85 (2020).
141. Virani, S. S., Alonso, A., Benjamin, E. J., *et al.* Heart Disease and Stroke Statistics—2020 Update: A Report From the American Heart Association. *Circulation* **141,** E139–E596 (2020).
142. Alkhalil, M., Chai, J. T. & Choudhury, R. P. Plaque imaging to refine indications for emerging lipid-lowering drugs. *European Heart Journal - Cardiovascular Pharmacotherapy* **3,** 58–67 (2017).
143. Gupta, A. & Marshall, R. S. Moving Beyond Luminal Stenosis: Imaging Strategies for Stroke Prevention in Asymptomatic Carotid Stenosis. *Cerebrovascular Diseases* **39,** 253–261 (2015).
144. Bourantas, C. V., Jaffer, F. A., Gijsen, F. J., *et al.* Hybrid intravascular imaging: recent advances, technical considerations, and current applications in the study of plaque pathophysiology. *European Heart Journal* **38,** 400–412 (2017).
145. Tomaniak, M., Katagiri, Y., Modolo, R., *et al.* Vulnerable plaques and patients: state-of-the-art. *European Heart Journal* **41,** 2997–3004 (2020).
146. Libby, P., Buring, J. E., Badimon, L., Hansson, G. K., Deanfield, J., Bittencourt, M. S., Tokgözoğlu, L. & Lewis, E. F. Atherosclerosis. *Nature Reviews Disease Primers* **5** (2019).
147. Sakakura, K., Nakano, M., Otsuka, F., Ladich, E., Kolodgie, F. D. & Virmani, R. Pathophysiology of atherosclerosis plaque progression. *Heart, lung & circulation* **22,** 399–411 (2013).

148. Bentzon, J. F., Otsuka, F., Virmani, R. & Falk, E. Mechanisms of plaque formation and rupture. *Circulation Research* **114,** 1852–1866 (2014).
149. Furie, M. B. & Mitchell, R. N. Plaque attack: One hundred years of atherosclerosis in the American Journal of Pathology. *American Journal of Pathology* **180,** 2184–2187 (2012).
150. Seneff, S., Davidson, R. M., Lauritzen, A., Samsel, A. & Wainwright, G. A novel hypothesis for atherosclerosis as a cholesterol sulfate deficiency syndrome. *Theoretical Biology and Medical Modelling* **12** (2015).
151. Epstein, S. E., Zhou, Y. F. & Zhu, J. Infection and Atherosclerosis. *Circulation* **100** (1999).
152. Tanganelli, P., Bianciardi, G., Simoes, C., Attino, V., Tarabochia, B. & Weber, G. Distribution of lipid and raised lesions in aortas of young people of different geographic origins (WHO-ISFC PBDAY study). *Arteriosclerosis and Thrombosis* **13,** 1700–1710 (1993).
153. Small, D. M. & Shipley, G. G. Physical-chemical basis of lipid deposition in atherosclerosis. **185,** 222–229 (1974).
154. Moerman, A. M., Visscher, M., Slijkhuis, N., *et al.* Lipid signature of advanced human carotid atherosclerosis assessed by mass spectrometry imaging. *Journal of Lipid Research* **62,** 100020 (2021).
155. Waksman, R., Di Mario, C., Torguson, R., *et al.* Identification of patients and plaques vulnerable to future coronary events with near-infrared spectroscopy intravascular ultrasound imaging: a prospective, cohort study. *The Lancet* **394,** 1629–1637 (2019).
156. Gardner, C. M., Tan, H., Hull, E. L., *et al.* Detection of Lipid Core Coronary Plaques in Autopsy Specimens With a Novel Catheter-Based Near-Infrared Spectroscopy System. *JACC: Cardiovascular Imaging* **1,** 638–648 (2008).
157. Waxman, S., Dixon, S. R., L'Allier, P., *et al.* In Vivo Validation of a Catheter-Based Near-Infrared Spectroscopy System for Detection of Lipid Core Coronary Plaques. Initial Results of the SPECTACL Study. *JACC: Cardiovascular Imaging* **2,** 858–868 (2009).
158. Caplan, J. D., Waxman, S., Nesto, R. W. & Muller, J. E. Near-Infrared Spectroscopy for the Detection of Vulnerable Coronary Artery Plaques. *Journal of the American College of Cardiology* **47,** C92–C96 (2006).
159. Wu, M., Jansen, K., van der Steen, A. F. W. & van Soest, G. Specific imaging of atherosclerotic plaque lipids with two-wavelength intravascular photoacoustics. *Biomedical Optics Express* **6,** 3276 (2015).
160. Jansen, K., Wu, M., van der Steen, A. F. W. & van Soest, G. Lipid detection in atherosclerotic human coronaries by spectroscopic intravascular photoacoustic imaging. *Optics Express* **21,** 21472 (2013).
161. Hobro, A. J. & Smith, N. I. An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. *Vibrational Spectroscopy* **91,** 31–45 (2017).
162. DiDonato, D. & Brasaemle, D. L. Fixation methods for the study of lipid droplets by immunofluorescence microscopy. *Journal of Histochemistry and Cytochemistry* **51,** 773–780 (2003).

163. Mora, J. T., Feng, X. & Gao, L. Photoacoustic Shadow-casting Microscopy (PASM). *Biophotonics Congress: Optics in the Life Sciences Congress 2019,* NT3C.4 (2019).
164. Salzer, R. Practical Guide to Interpretive Near-Infrared Spectroscopy. By Jerry Workman, Jr. and Lois Weyer. *Angewandte Chemie International Edition* **47,** 4628–4629 (2008).
165. Weyer, L. & Lo, S.-C. in *Handbook of Vibrational Spectroscopy* (John Wiley & Sons, Ltd, Chichester, UK, 2006).
166. Prieto, N., Pawluczyk, O., Dugan, M. E. R. & Aalhus, J. L. A Review of the Principles and Applications of Near-Infrared Spectroscopy to Characterize Meat, Fat, and Meat Products. *Applied Spectroscopy* **71,** 1403–1426 (2017).
167. Goddu, R. F. Determination of Unsaturation by Near-Infrared Spectrophotometry. *Analytical Chemistry* **29,** 1790–1794 (1957).
168. Wight, T. N. & Merrilees, M. J. Proteoglycans in atherosclerosis and restenosis: key roles for versican. *Circulation research* **94,** 1158–67 (2004).
169. Mundi, S., Massaro, M., Scoditti, E., Carluccio, M. A., van Hinsbergh, V. W. M., Iruela-Arispe, M. L. & De Caterina, R. Endothelial permeability, LDL deposition, and cardiovascular risk factors—a review. *Cardiovascular Research* **114,** 35–52 (2018).
170. Skålén, K., Gustafsson, M., Knutsen Rydberg, E., Hultén, L. M., Wiklund, O., Innerarity, T. L. & Boren, J. Subendothelial retention of atherogenic lipoproteins in early atherosclerosis. *Nature* **417,** 750–754 (2002).
171. Suhalim, J. L. L., Chung, C.-Y. Y., Lilledahl, M. B. B., Lim, R. S. S., Levi, M., Tromberg, B. J. J. & Potma, E. O. O. Characterization of cholesterol crystals in atherosclerotic plaques using stimulated Raman scattering and second-harmonic generation microscopy. *Biophysical Journal* **102,** 1988–1995 (2012).
172. Baumer, Y., Mehta, N. N., Dey, A. K., Powell-Wiley, T. M. & Boisvert, W. A. Cholesterol crystals and atherosclerosis. **41,** 2236–2239 (2020).
173. Fujiyoshi, K., Minami, Y., Ishida, K., *et al.* Incidence, factors, and clinical significance of cholesterol crystals in coronary plaque: An optical coherence tomography study. *Atherosclerosis* **283,** 79–84 (2019).
174. Falk, E., Nakano, M., Bentzon, J. F., Finn, A. V. & Virmani, R. Update on acute coronary syndromes: the pathologists' view. *European Heart Journal* **34,** 719–728 (2013).
175. Freigang, S., Ampenberger, F., Spohn, G., *et al.* Nrf2 is essential for cholesterol crystal-induced inflammasome activation and exacerbation of atherosclerosis. *European Journal of Immunology* **41,** 2040–2051 (2011).
176. Corr, E. M., Cunningham, C. C. & Dunne, A. Cholesterol crystals activate Syk and PI3 kinase in human macrophages and dendritic cells. *Atherosclerosis* **251,** 197–205 (2016).
177. Geng, Y.-J., Phillips, J. E., Mason, R. P. & Casscells, S. W. Cholesterol crystallization and macrophage apoptosis: implication for atherosclerotic plaque instability and rupture. *Biochemical pharmacology* **66,** 1485–92 (2003).
178. Baker, M. J., Trevisan, J., Bassan, P., *et al.* Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols* **9,** 1771–1791 (2014).

179. Diem, M., Mazur, A., Lenau, K., Schubert, J., Bird, B., Miljković, M., Krafft, C. & Popp, J. Molecular pathology via IR and Raman spectral imaging. *Journal of Biophotonics* **6,** 855–886 (2013).
180. Buschman, H. P., Deinum, G., Motz, J. T., Fitzmaurice, M., Kramer, J. R., van der Laarse, A., Bruschke, A. V. & Feld, M. S. Raman microspectroscopy of human coronary atherosclerosis: biochemical assessment of cellular and extracellular morphologic structures in situ. *Cardiovascular pathology : the official journal of the Society for Cardiovascular Pathology* **10,** 69–82 (2001).
181. Li, C., Ebenstein, D., Xu, C., Chapman, J., Saloner, D., Rapp, J. & Pruitt, L. Biochemical characterization of atherosclerotic plaque constituents using FTIR spectroscopy and histology. *Journal of Biomedical Materials Research* **64A,** 197–206 (2003).
182. Lim, R. S., Suhalim, J. L., Miyazaki-Anzai, S., Miyazaki, M., Levi, M., Potma, E. O. & Tromberg, B. J. Identification of cholesterol crystals in plaques of atherosclerotic mice using hyperspectral CARS imaging. *Journal of Lipid Research* **52,** 2177–2186 (2011).
183. Marzec, K. M., Wrobel, T. P., Rygula, A., Maslak, E., Jasztal, A., Fedorowicz, A., Chlopicki, S. & Baranska, M. Visualization of the biochemical markers of atherosclerotic plaque with the use of Raman, IR and AFM. *Journal of Biophotonics* **7,** 744–756 (2014).
184. Marzec, K. M., Rygula, A., Gasior-Glogowska, M., *et al.* Vascular diseases investigated ex vivo by using Raman, FT-IR and complementary methods. *Pharmacological Reports* **67,** 744–750 (2015).
185. Wrobel, T. P., Mateuszuk, L., Chlopicki, S., Malek, K. & Baranska, M. Imaging of lipids in atherosclerotic lesion in aorta from ApoE/LDLR -/- mice by FT-IR spectroscopy and Hierarchical Cluster Analysis. *Analyst* **136,** 5247–5255 (2011).
186. He, Y., Shi, J., Pleitez, M. A., Maslov, K., Wagenaar, D. A. & Wang, L. V. Label-free imaging of lipid-rich biological tissues by mid-infrared photoacoustic microscopy. *Journal of Biomedical Optics* **25** (2020).
187. Wiercigroch, E., Szafraniec, E., Czamara, K., *et al.* Raman and infrared spectroscopy of carbohydrates: A review. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **185,** 317–335 (2017).
188. Türker-Kaya, S. & Huck, C. W. A review of mid-infrared and near-infrared imaging: Principles, concepts and applications in plant tissue analysis. *Molecules* **22** (2017).
189. Lehti, S., Nguyen, S. D., Belevich, I., *et al.* Extracellular Lipids Accumulate in Human Carotid Arteries as Distinct Three-Dimensional Structures and Have Proinflammatory Properties. *American Journal of Pathology* **188,** 525–538 (2018).
190. Janoudi, A., Shamoun, F. E., Kalavakunta, J. K. & Abela, G. S. Cholesterol crystal induced arterial inflammation and destabilization of atherosclerotic plaque. *European Heart Journal* **37,** 1959–1967 (2016).
191. Khajehpour, M., Dashnau, J. L. & Vanderkooi, J. M. Infrared spectroscopy used to evaluate glycosylation of proteins. *Analytical Biochemistry* **348,** 40–48 (2006).
192. Moore, C. R. & Gilbert, D. B. Protein, glycoprotein and glycolipid profiles of human arterial and venous tissues. *Atherosclerosis* **35,** 267–275 (1980).

193. Mukhin, D. N., Chao, F.-F. & Kruth, H. S. Glycosphingolipid Accumulation in the Aortic Wall Is Another Feature of Human Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **15,** 1607–1615 (1995).
194. Prokazova, N. V., Mukhin, D. N., Orekhov, A. N., *et al.* Neutral glycolipids of atherosclerotic plaques and unaffected human aorta tissue. *European journal of biochemistry* **180,** 167–71 (1989).
195. Mori, T., Sasaki, J., Kawaguchi, H., Handa, K., Takada, Y., Matsunaga, A., Kono, S. & Arakawa, K. Serum glycoproteins and severity of coronary atherosclerosis. *American Heart Journal* **129,** 234–238 (1995).
196. Pilely, K., Rosbjerg, A., Genster, N., *et al.* Cholesterol Crystals Activate the Lectin Complement Pathway via Ficolin-2 and Mannose-Binding Lectin: Implications for the Progression of Atherosclerosis. *The Journal of Immunology* **196,** 5064–5074 (2016).
197. Scott, D. W., Chen, J., Chacko, B. K., Traylor, J. G., Orr, A. W. & Patel, R. P. Role of endothelial N-glycan mannose residues in monocyte recruitment during atherogenesis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **32,** 51–59 (2012).
198. Wang, L., Chapman, J., Palmer, R. A., van Ramm, O. & Mizaikoff, B. Classification of atherosclerotic rabbit aorta samples by mid-infrared spectroscopy using multivariate data analysis. *Journal of Biomedical Optics* **12,** 024006 (2007).
199. Manoharan, R., Baraga, J. J., Rava, R. P., Dasari, R. R., Fitzmaurice, M. & Feld, M. S. Biochemical analysis and mapping of atherosclerotic human artery using FT-IR microspectroscopy. *Atherosclerosis* **103,** 181–193 (1993).
200. Wang, Z., Cai, Y., Wang, Y., Zhou, X., Zhang, Y. & Lu, H. Improved MALDI imaging MS analysis of phospholipids using graphene oxide as new matrix. *Scientific Reports* **7,** 1–9 (2017).
201. Zhu, Z., Shen, J., Wang, D., *et al.* An auxiliary matrix for routine analysis of small molecules and biological macromolecules using matrix-assisted laser desorption ionization mass spectrometry. *Analytical and Bioanalytical Chemistry* **411,** 1041–1052 (2019).
202. Manier, S. K., Keller, A. & Meyer, M. R. Automated optimization of XCMS parameters for improved peak picking of liquid chromatography–mass spectrometry data using the coefficient of variation and parameter sweeping for untargeted metabolomics. *Drug Testing and Analysis* **11,** 752–761 (2019).
203. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. *Current Metabolomics* **1,** 92–107 (2013).
204. Siy, P. W., Moffitt, R. A., Parry, R. M., Chen, Y., Liu, Y., Sullards, M. C., Merrill, A. H. & Wang, M. D. Matrix factorization techniques for analysis of imaging mass spectrometry data. *2008 8th IEEE International Conference on BioInformatics and BioEngineering* **2008,** 1–6 (2008).
205. Gui, Y.-k., Li, Q., Liu, L., *et al.* Plasma levels of ceramides relate to ischemic stroke risk and clinical severity. *Brain Research Bulletin* **158,** 122–127 (2020).
206. Ruuth, M., Nguyen, S. D., Vihervaara, T., *et al.* Susceptibility of low-density lipoprotein particles to aggregate depends on particle lipidome, is modifiable, and associates with future cardiovascular deaths. *European Heart Journal* **39,** 2562–2573 (2018).

207. Laaksonen, R., Ekroos, K., Sysi-Aho, M., *et al.* Plasma ceramides predict cardiovascular death in patients with stable coronary artery disease and acute coronary syndromes beyond LDL-cholesterol. *European Heart Journal* **37,** 1967–1976 (2016).
208. Gault, C. R., Obeid, L. M. & Hannun, Y. A. An Overview of Sphingolipid Metabolism: From Synthesis to Breakdown. **688,** 1–23 (2010).
209. Tarasov, K., Ekroos, K., Suoniemi, M., *et al.* Molecular Lipids Identify Cardiovascular Risk and Are Efficiently Lowered by Simvastatin and PCSK9 Deficiency. *The Journal of Clinical Endocrinology and Metabolism* **99,** E45–E52 (2014).
210. Andrieu-Abadie, N. & Levade, T. Sphingomyelin hydrolysis during apoptosis. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1585,** 126–134 (2002).
211. Cui, Z. & Houweling, M. Phosphatidylcholine and cell death. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1585,** 87–96 (2002).
212. Falk, E. Pathogenesis of Atherosclerosis. *Journal of the American College of Cardiology* **47,** 0–5 (2006).
213. Niyonzima, N., Bakke, S. S., Gregersen, I., *et al.* Cholesterol crystals use complement to increase NLRP3 signaling pathways in coronary and carotid atherosclerosis. *EBioMedicine* **60,** 102985 (2020).
214. Grebe, A. & Latz, E. Cholesterol crystals and inflammation. *Current Rheumatology Reports* **15,** 313 (2013).
215. Tahara, N., Mukherjee, J., De Haas, H. J., *et al.* 2-deoxy-2-[18F]fluoro-d-mannose positron emission tomography imaging in atherosclerosis. *Nature Medicine* **20** (2014).
216. Kim, J. B., Park, K., Ryu, J., *et al.* Intravascular optical imaging of high-risk plaques in vivo by targeting macrophage mannose receptors. *Scientific Reports* **6** (2016).
217. Finn, A. V., Nakano, M., Polavarapu, R., *et al.* Hemoglobin directs macrophage differentiation and prevents foam cell formation in human atherosclerotic plaques. *Journal of the American College of Cardiology* **59** (2012).
218. Chinetti-Gbaguidi, G., Colin, S. & Staels, B. Macrophage subsets in atherosclerosis. **12,** 10–17 (2015).
219. Heijs, B., Holst, S., Briaire-De Bruijn, I. H., *et al.* Multimodal Mass Spectrometry Imaging of N-Glycans and Proteins from the Same Tissue Section. *Analytical Chemistry* **88,** 7745–7753 (2016).
220. Powers, T. W., Jones, E. E., Betesh, L. R., Romano, P. R., Gao, P., Copland, J. A., Mehta, A. S. & Drake, R. R. Matrix assisted laser desorption ionization imaging mass spectrometry workflow for spatial profiling analysis of N-linked Glycan expression in tissues. *Analytical Chemistry* **85,** 9799–9806 (2013).

Summary

One in six people experience a stroke in their lifetime. Of those people more than 50% lose their independence as a consequence thereof. Stroke also affects people behavior, personality, mental and intellectual capabilities. Current preventive interventions are lifestyle changes and pharmaceutical treatments, and if this is not sufficiently efficient, an operation where the inner lining of the artery is removed. This procedure is not without risk, as tissue or clot fragments may be loosened after which they can lodge themselves in the brain. Because it is difficult to predict the occurrence of stroke, more people are treated than necessary (in hindsight): 6 operations are necessary to prevent 1 stroke. Not very promising numbers. We hope to gain insights in the chances that people will experience a stroke by investigating the molecular composition of atherosclerosis, one of the major causes of stroke, with the goal of improving treatment.

Atherosclerosis is characterized by local deposits of mainly lipids and dead cell debris, in the arterial wall, named plaque. Atherosclerosis starts during adolescence and can develop over decades into a plaque. When this plaque ruptures, the content flows into the bloodstream, causing a transient ischemic attack (TIA) or stroke. However, many plaques never rupture. Timely identification of the ones that will has been a scientific quest in the field of vascular medicine for the past two decades.

Lipids play a significant role in atherosclerosis. In this thesis we investigated the spatial distribution of these lipids. We tried to find a relationship between the molecular composition of plaques and the risk of plaque rupture. To this end we used matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) and photoacoustic imaging.

In **Chapter 2** we designed a data analysis pipeline to process MALDI-MSI data. The methodology uses several steps to filter the data such that only lipid related signals are retained. Furthermore, we quantified the reproducibility of six adjacent tissue sections. The variation of the sections was between 12 and 25%, compared to our lipid standard, where the variation was 16%. From this we conclude that the variation in the data originated from the measurement method.

In addition, we measure atherosclerotic plaques; one early lesion, and two advanced lesions. We found that two cholesteryl esters (CE), linoleate and oleate, were present in varying concentrations, depending on the disease stage of the plaque. Furthermore, we related lipid pattern to histology and found that diacylglycerols were more abundant in thrombotic areas.

In **Chapter 3** we used the pipeline developed in **Chapter 2** to image 12 carotid endarterectomy samples, advanced plaques. We analyzed these data in two ways; (1) by unsupervised clustering and (2) by a histology driven multivariate analysis. Both

methodologies identified that sphingomyelins and oxidized cholesteryl esters were associated with necrotic areas in the plaque. Furthermore, diacylglycerols and triacylglycerols were correlated to fibrin, a protein associated with thrombus formation.

In **Chapter 4** we investigated the use of several types of unsupervised machine learning methods and evaluated their performance and usability for MALDI-MSI data. We compared linear and statistics-based methods to methods that use a neural network, and found that the resulting clustering was similar. Methods that base the clustering on linear or statistical operations were easier to use and interpret: the results were directly related to spectral information.

MALDI-MSI is a method that is provides the molecular composition of tissues with high chemical specificity. However, the technique can only be used *ex vivo*, and therefore cannot be used for diagnostic purposes. Therefore, we moved to photoacoustic imaging in **Chapter 5 and 6**.

We built a photoacoustic microscope capable of measuring photoacoustic spectra in **Chapter 5**. We chose to measure in the wavelength range from 1150-1240 nm, where the photoacoustic signal originates from C-H vibrations of lipids. We compared the photoacoustic signal with MALDI-MSI data from the same cross-sections and validated our results using lipid standards. We found that the photoacoustic spectra of cholesteryl linolate and sphingomyelin 34:1 are very similar. Both these lipids are related to high-risk plaque features, making this spectrum a good candidate for *in vivo* identification of such plaques in patients.

In **Chapter 6** we investigated similar samples in the mid-infrared range, where the spectral diversity is larger, and more specific molecular fingerprints are expected. This enabled us to investigate a wider variety of molecules simultaneously. We found carbohydrates to be co-localized with cholesterol crystals. Both of these molecules have been associated with inflammation, neovascularization and intraplaque hemorrhage.

Finally, we described the most important finding and developments in the thesis in **Chapter 7**. The aim of this research was to image molecules related to high-risk plaque *in vivo*. We identified molecules that are associated with high-risk plaque features and determined the affiliated photoacoustic spectra. To achieve *in vivo* imaging thereof a translation towards macroscopic imaging needs to be made.

Samenvatting

Een op de zes personen krijgt gedurende zijn of haar leven te maken met een beroerte, en meer dan 50% van de mensen kan na een beroerte niet meer zelfstandig wonen. Daarnaast heeft een beroerte ook impact op iemands gedrag, persoonlijkheid, en mentale- en intellectuele capaciteiten. Huidige preventiemethoden zijn aanpassingen van levensstijl of medicijnen, en mocht dit niet werken uiteindelijk een operatie waarbij de verkalkte binnenkant van de ader wordt verwijderd. De operatie zelf is niet zonder risico's, tijdens de operatie kan weefsel of stolsel losraken en in het brein terechtkomen. Omdat het lastig is te voorspellen of er een beroerte gaat plaatsvinden of niet, behandelen we achteraf gezien meer mensen dan nodig; er zijn 6 operaties nodig om 1 beroerte te voorkomen. Kortom, niet erg gunstige getallen. Door te kijken naar de moleculaire samenstelling van aderverkalking, een van de belangrijkste oorzaken van beroertes, hopen we meer inzicht te krijgen in de kansen dat mensen een beroerte krijgen en ze zo gerichter te kunnen behandelen.

Aderverkalking is een plaatselijke ophoping van vetten en dode cellen in de vaatwand, ook wel plaque genoemd. Aderverkalking begint vaak in adolescentie en kan over tientallen jaren uitgroeien tot een plaque. Wanneer deze openbreekt en de inhoud van de plaque in de bloedstroom terecht komt spreken we van een TIA of een beroerte. Echter sommige plaques scheuren nooit. Bijtijdse identificatie van de plaques die wel zullen openbreken, is al tientallen jaren onderwerp voor wetenschappelijk onderzoek.

Lipiden spelen een grote rol in aderverkalking en in dit proefschrift keken we naar de spatiele verdeling van deze lipiden. We probeerden een verband te ontdekken tussen de moleculaire samenstelling van de plaque en de kans op een beroerte. Hiervoor gebruikten we twee verschillende moleculaire beeldvormingstechnieken, matrix geassisteerde laser desorptie/ionisatie massa spectrometrie beeldvorming (MALDI-MSI) en fotoakoestiek.

In **Hoofdstuk 2** ontworpen we een data-analyse methodiek voor het verwerken van MALDI-MSI data. De methodiek maakt gebruik van verschillende stappen om de data zo te filteren dat alleen de signalen afkomstig van lipiden overblijven. We kwantificeerden de reproduceerbaarheid van zes aanliggende coupes. De variatie tussen de coupes was 12-25%, en de variatie van een lipide standaard was 16%. Hieruit concludeerden we dat de variatie in de data voortkomt uit de meetmethode.

Daarnaast maten we in dit hoofdstuk atherosclerose plaques; een beginnende plaque en twee plaques in een vergevorderd stadium. We vonden dat twee cholesteryl esters (CE), linoleate en oleate, in verschillende concentraties aanwezig waren in de verschillende stadia van de plaque. Verder relateerden we de spatiele lipidenpatronen aan histo-

logie en vonden we dat diacylglycerolen meer voorkwamen in gebieden met bloedingen.

In **Hoofdstuk 3** gebruikten we de methodiek uit **Hoofdstuk 2** om twaalf carotis endarterectomie specimenen te onderzoeken. We analyseerden deze dataset op twee manieren; (1) op basis van een datareductie methode en (2) op basis van histologische aangestuurde multivariate analyse. Beide methodes identificeerden dat sphingolipiden en geoxideerde cholsteryl esters in grotere hoeveelheden aanwezig is in necrotische gebieden van de plak. Ook vonden we dat di- en triacylglycerolen geassocieerd zijn met fibrine, een eiwit dat samenhangt met trombusvorming.

In **Hoofdstuk 4** doken we in het gebruiken van verschillende soorten ongecontroleerde automatisch leren methoden en beoordeelden de prestaties en de bruikbaarheid van deze methoden op MALDI-MSI data. We vergeleken lineaire en statistische methoden met methoden die gebruik maken van neurale netwerken, en vonden dat de resultaten van de methoden vergelijkbaar waren. Methoden op basis die werken op basis van lineaire en statistische reductie van dimensie waren echter simpeler in gebruik en gemakkelijker te relateren aan spectrale informatie.

MALDI-MSI is een methode die veel informatie geeft over de moleculaire samenstelling van weefsels, echter is het niet een methode die gebuikt kan worden terwijl de plaque nog in de patiënt aanwezig is. Daarom maakten we in **Hoofdstuk 5 en 6** een overstap naar fotoakoestiek om de moleculaire samenstelling van plaques te beoordelen in de patiënt.

We bouwden een fotoakoestische microscoop die in staat was om spectrale informatie te meten, **Hoofdstuk 5**. Hierbij werd gekozen om te meten van 1150-1240 nm, het domein waarop het fotoakoestische signaal afkomstige is van C-H vibraties van lipiden. We vergeleken de fotoakoestiek spectra met MALDI-MSI data van dezelfde coupes en valideerden het signaal met behulp van lipiden standaarden. Hierbij vonden we onder andere dat een cholesteryl linoleaat en sphingomyelin 34:1 bijna hetzelfde spectrum hebben wanneer je deze fotoakoestisch meet. Beide lipiden staan in verband met risico factoren in plaque en dit spectrum zou dan ook een goede kandidaat zijn voor *in vivo* identificatie van deze plaques in patiënten.

In **Hoofdstuk 6** bekeken we vergelijkbare weefsels in het middel-infrarood domein, waar de spectrale diversiteit groter is, en we de mogelijkheid hadden om meerdere soorten moleculen tegelijkertijd te meten. We vinden een associatie tussen koolhydraten en cholesterol kristallen. Beide moleculen zijn in verband gebracht met ontsteking, neovascularisatie en bloedingen binnenin de plaques.

Tenslotte, worden in **Hoofdstuk 7** de belangrijkste bevindingen en ontwikkelingen uit dit proefschrift besproken. Het doel van dit onderzoek was het *in vivo* in beeld brengen van de moleculen die verband houden met hoog risico plaques. We hebben een aantal moleculen geïdentificeerd die samenhangen met verhoogd risico componenten, en bijbehorende fotoakoestische spectra vastgesteld. Voor *in vivo* beeldvorming zal er nog een vertaalslag gemaakt moeten worden van microscopie naar beeldvorming in patiënten.