JOURNAL OF PROTEOMICS 75 (2012) 4962–4989

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

Review

Imaging mass spectrometry statistical analysis☆

Emrys A. Jones[a,1], *Sören-Oliver Deininger*[b,1], *Pancras C.W. Hogendoorn*[c], *André M. Deelder*[a], *Liam A. McDonnell*[a,*]

[a]*Biomolecular Mass Spectrometry Unit, Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands*
[b]*Bruker Daltonic GmbH, Bremen, Germany*
[c]*Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands*

ARTICLE INFO

Article history:
Received 6 March 2012
Accepted 16 June 2012
Available online 26 June 2012

Keywords:
Imaging mass spectrometry
Data analysis
Biomarker discovery
Molecular histology

ABSTRACT

Imaging mass spectrometry is increasingly used to identify new candidate biomarkers. This clinical application of imaging mass spectrometry is highly multidisciplinary: expertise in mass spectrometry is necessary to acquire high quality data, histology is required to accurately label the origin of each pixel's mass spectrum, disease biology is necessary to understand the potential meaning of the imaging mass spectrometry results, and statistics to assess the confidence of any findings.

Imaging mass spectrometry data analysis is further complicated because of the unique nature of the data (within the mass spectrometry field); several of the assumptions implicit in the analysis of LC–MS/profiling datasets are not applicable to imaging. The very large size of imaging datasets and the reporting of many data analysis routines, combined with inadequate training and accessible reviews, have exacerbated this problem. In this paper we provide an accessible review of the nature of imaging data and the different strategies by which the data may be analyzed. Particular attention is paid to the assumptions of the data analysis routines to ensure that the reader is apprised of their correct usage in imaging mass spectrometry research.

This article is part of a Special Issue entitled: Imaging Mass Spectrometry: A User's Guide to a New Technique for Biological and Biomedical Research.

Contents

1. Introduction 4963
   1.1. The imaging MS experiment 4964
2. Data preprocessing 4965
   2.1. Smoothing and de-noising 4966
      2.1.1. Mass spectral 4966
      2.1.2. Image noise 4966

☆ This article is part of a Special Issue entitled: Imaging Mass Spectrometry: A User's Guide to a New Technique for Biological and Biomedical Research.

* *Corresponding author at:* Biomolecular Mass Spectrometry Unit, Department of Parasitology, Leiden University Medical Center, Albinusdreef 2, 2333ZA Leiden, The Netherlands. Tel.: +31 71 526 8744.
E-mail address: l.a.mcdonnell@lumc.nl (L.A. McDonnell).

1874-3919/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2012.06.014

2.2. Normalization and quantitation . . . 4967
2.2.1. Molecule-specific ion suppression and normalization . . . 4968
2.2.2. Global intensity suppression and normalization . . . 4968
3. Feature extraction . . . 4969
4. Choice of data analysis method . . . 4971
4.1. Data analysis strategy . . . 4971
4.2. Data analysis methodology . . . 4972
4.3. Data analysis — in practice . . . 4973
5. Imaging MS based molecular histology: annotating tissues on the basis of MS profiles . . . 4973
5.1. Principal component analysis (PCA) . . . 4974
5.2. Clustering methods . . . 4975
5.2.1. Hierarchical clustering . . . 4975
5.2.2. k-Means clustering . . . 4976
5.3. Factorization methods . . . 4977
5.3.1. Non-negative matrix factorization (NNMF) . . . 4977
5.3.2. Probabilistic latent semantic analysis (PLSA) . . . 4977
5.3.3. Maximum autocorrelation factor analysis . . . 4977
5.4. Which multivariate method is true? . . . 4977
5.5. Imaging MS based molecular histology of multiple tissues . . . 4978
6. Methods for comparing sample cohorts (biomarker discovery) . . . 4980
6.1. Avoiding bias and randomization . . . 4980
6.2. Univariate comparisons . . . 4981
6.2.1. What is a good biomarker: fold-change and ROC, Gaussian and non-Gaussian distributions . . . 4981
6.2.2. Statistical tests and p-values . . . 4982
6.2.3. Parametric tests (t-test/one-way ANOVA) . . . 4983
6.2.4. Nonparametric tests . . . 4983
6.2.5. Additional considerations for hypothesis testing, paired and unpaired comparisons . . . 4983
6.3. Multivariate comparisons and classifications . . . 4983
6.3.1. PCA and hierarchical clustering for comparison of different datasets . . . 4983
6.4. Classification methods . . . 4983
6.4.1. Discriminant analysis (DA and PCA-DA) . . . 4983
6.4.2. Other classification algorithms . . . 4984
6.4.3. Multivariate analysis vs. univariate markers . . . 4984
7. Concluding remarks . . . 4985
Acknowledgments . . . 4986
References . . . 4986

1. Introduction

MALDI mass spectrometry can generate MS profiles that contain hundreds of biomolecules directly from tissue [1]. Spatially correlated mass spectrometry, imaging MS, reveals how each biomolecular ion varies across tissue samples [2,3]. In current experimental workflows this data cube of position-correlated spectra is aligned with an optical image of the H&E-stained tissue section, stained after the MALDI imaging MS experiment [4]. Advanced data analysis tools enable the datasets to be interrogated to reveal which peptides and proteins have similar spatial distributions, which are correlated with specific histological features, and which can be used to classify the tissue according to tumor type and tumor grade. Subsequent analysis of the tumor specific profiles can even reveal features associated with prognosis or response to treatment [5].

The goal of an experiment, the hypothesis to be tested, should determine the experimental methodology as well as the data analysis technique employed. In reality experimentalists only have access to a limited number of methodologies and tissues. Discussions of mass resolution, spatial resolution, or the need for large clinical cohorts are moot if the infrastructure is not available or if there is a limited number of exquisitely matched patient samples. Data analysis is the one area where the experimentalist is less restrained — commercially available and open source mathematics packages provide a broad array of statistical tools that can be applied to imaging mass spectrometry datasets. A host of multivariate, clustering, factorization, and classification algorithms have been applied. With sufficient training imaging MS experimentalists would be able to apply this rich repertoire of ever-expanding data analysis techniques to their imaging MS datasets.

The exploitation of imaging MS for clinical or pre-clinical applications is exceedingly multidisciplinary, often involving a selection of chemists, biochemists, molecular biologists, clinicians, physicists, but less often highly trained clinical statisticians. Consequently there is a need for training in imaging MS data analysis in order to ensure that statistical analysis tools are used correctly and appropriately. The intention of this review is to provide an accessible and

detailed overview of the data analysis methodologies and techniques currently used in imaging MS research.

1.1. *The imaging MS experiment*

After preparation of the tissue section for imaging MS it is placed inside the mass spectrometer and analyzed using a defined array of positions (for imaging MS instruments based on ultra-rapid lasers and continuous sample movement the pixel array is reconstituted by binning data from a user-defined pixel length). The data is recorded as a sequential list of mass spectra, each of which has an associated pixel position, Fig. 1. Mass spectral images of a specific peak are obtained by projecting the intensity of the peak in each pixel's mass spectrum at the pixel's coordinates. Nevertheless, the data is recorded and typically stored as a list of measurements. More formally, an imaging MS dataset is a 2D matrix of spectrum number vs. *m/z*. Imaging MS datasets are recorded in a format that is highly suited to many statistical data analysis algorithms, which require a 2D data matrix of observations vs. variables. In an imaging MS experiment the observations are the pixel positions and the variables are the *m/z* channels. Most data analysis algorithms treat each pixel's mass spectrum as an independent measurement. Unless explicitly included the spatial relationship between pixels is not included in the analysis (notable exceptions include image smoothing [6–8], maximum autocorrelation factor analysis [9], and spatially-aware *k*-means clustering [10]).

Before progressing onto the different data analysis methods it is important to briefly consider the imaging nature of imaging MS data. The *m/z* precision of the mass analysis, the spatial resolution and the mass resolution are all important aspects that are dependent on the instrument used and the molecular class analyzed, and so influence how it is subsequently processed and analyzed.

- *m/z* precision: an imaging MS experiment of a single tissue section often include >10k pixels, each of which is described by a mass spectrum. The mass precision of the mass spectrometer is a measure of the reproducibility of the mass analysis [11]. When the (mean) average mass spectrum of the imaging MS dataset is calculated, the most common mass spectral representation used for interacting with imaging MS data, the precision determines the width of the averaged peak. Mass spectral peak alignment tools can be used to correct for small drifts in the mass scales of different pixels (whether due to instrumental drift or sample topography [12–15]), thereby increasing the precision of the mass spectral data [16], and thus decreasing the 'apparent' peak width in the mean mass spectral representation. Alignment of each pixel's mass spectrum can improve imaging MS data by increasing the specificity of the mass spectral images [11]. Internal calibration (incl. internal lock mass) can be considered a method of mass spectral alignment [17], though it has been shown that such recalibrations may not correct all mass shifts between different spectra [18].
- Spatial resolution: The spatial resolution of an experiment determines the size of the spatial features that can be resolved [19–22]. Higher spatial resolution analyses require a larger number of smaller-sized pixels, and thus generate more data. The number of spectra increases rapidly at higher spatial resolutions. The increased number of spectra can lead to increased data acquisition times and greater data load, which can undermine efficient data analysis, Fig. 2. The routine high spatial resolution provided by secondary ion mass spectrometry (SIMS) is permitted by the very fast data acquisition rates of SIMS mass spectrometers, and which is slowly being matched by modern MALDI-based systems [22–24].

The spatial resolution of an experiment also has a more nuanced effect on data analysis. The smaller pixels of higher spatial resolution analyses contain less material and so are associated with weaker, and often fewer signals (so-called 'sample volume limited'). Modern day MALDI mass spectrometers can achieve amol, 10^{-18}, sensitivity of (purified) peptides. The highest spatial resolution possible for a peptide present in a 10 μm thick tissue at μM concentration will be 10 μm. At smaller pixel sizes there is simply insufficient peptide available. Note: the increasing number of reports at this resolution, or even higher, is testament to the effectiveness of MALDI sample preparation strategies [19–22]. It has been reported that the single most important factor governing the performance of

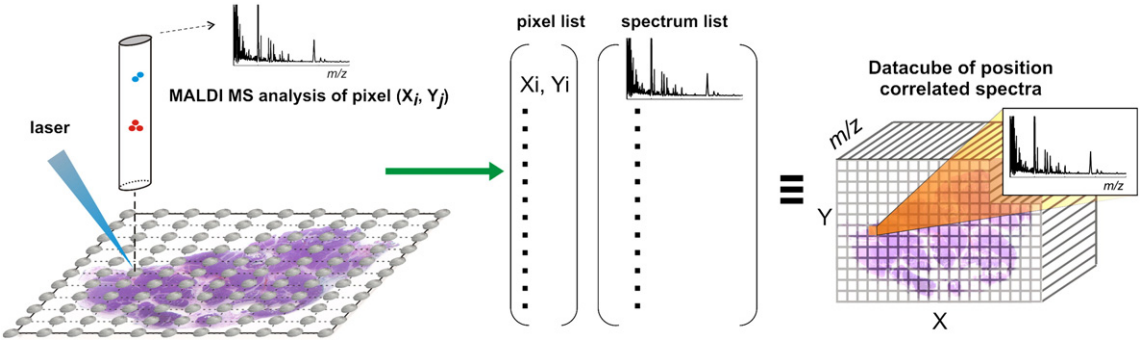

Fig. 1 – Imaging MS experiments typically (but not exclusively) record the data as a list of mass spectra, each of which corresponds to a specific pixel. The image of an MS peak is then created by displaying the peak's intensity in every pixel's mass spectrum at the pixel's coordinates. Such an approach enables the analyst to define measurement regions that trace the edges of irregular shaped tissue sections, thus minimizing measurement time and data load.

multivariate data analysis methods is the signal intensity [25]. From an analytical chemistry perspective, at higher spatial resolution there are fewer ions detected above the limit of quantitation, consequently there is more noise in the dataset.

- Mass resolution: Mass resolution is an important feature because it defines the degree of chemical specificity [26] and, through improving the precision of a mass measurement, helps improve mass accuracy. Unless automated data reduction methods are employed, e.g. feature detection and extraction [27], the large size of higher mass resolution datasets can quickly make them intractable.

Accurate mass measurements allow the analyst to make more confident peak assignments [28]. For example, it has been shown how the percentage of unique, tryptic peptides increases with increasing mass accuracy and increasing *m*/z [29]. Accurate mass measurements at low masses, e.g. metabolites, are especially powerful as they can be used to obtain the elemental composition of the detected ions.

Protein imagining MS experiments are almost universally performed using a linear-ToF mass analyzer because of the need for a wide *m*/z range analyzer, but have the lowest mass accuracy and mass resolution. The *m*/z range of peptides, tryptic peptides, lipids and metabolites are compatible with a much larger range of mass analyzers. Irrespective of which mass analyzer is used for data acquisition imaging MS datasets are typically subjected to a number of preprocessing steps prior to data analysis that address some of the experimental factors outlined here.

2. Data preprocessing

Shin et al. have proposed that MALDI MS analysis may contain three discrete sources of noise: i) electrical noise from the mass spectrometer's components, ii) shot noise due to the discrete nature of ion detection, and iii) a chemical 'background' generated during the MALDI process [30]. A fourth source, not strictly noise, but important for MALDI reproducibility is the matrix preparation: MALDI sample preparations can lead to inhomogeneous distributions of matrix crystals, and thus to position dependent ion signals [31,32]. Many of these (or related) aspects also hold true for NIMS, SIMS, and DESI. Several papers have investigated and characterized the nature of the noise in mass spectrometers [30,33].

Modern MALDI-ToF mass spectrometers average the signals from hundreds of laser shots, distributed across the sample, to increase the signal-to-noise ratios of the detected peaks and to average variations due to instrumental noise and sample heterogeneity. Every pixel in an imaging MS dataset is typically analyzed with several hundred laser shots. Provided matrix inhomogeneities are on length scales smaller than the pixel size then the only spatial variation is that due to the sample (see below). Substantial measurement noise can still

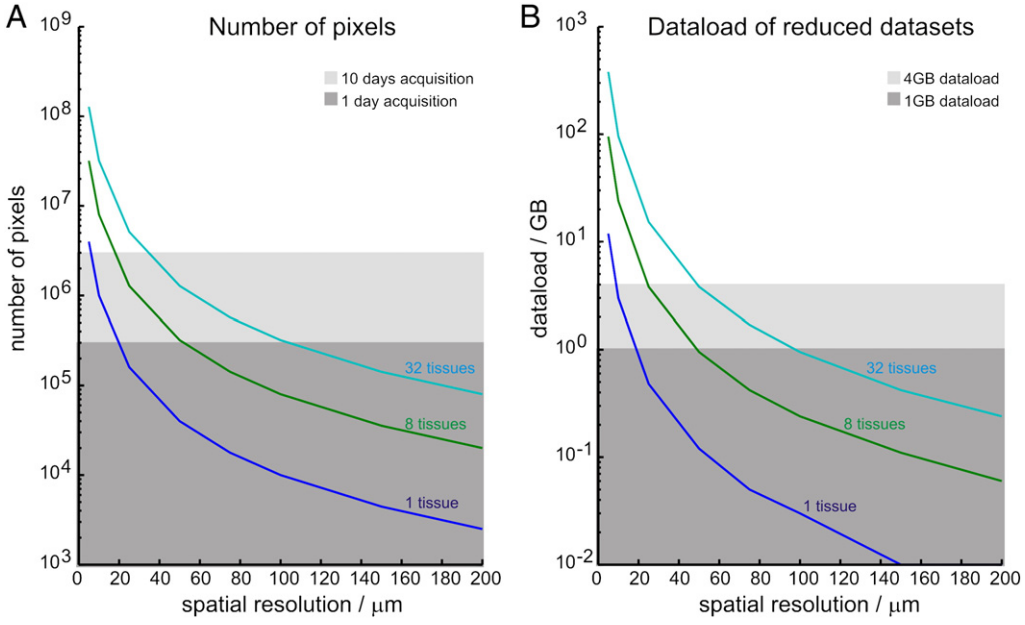

Fig. 2 – Dependence of number of pixels, A), and data load, B), on the spatial resolution and number of tissues in an imaging MS experiment (1 cm^2 tissue section, 3 pixels s^{-1}, 400 MS peaks per pixel, double-precision digits). To guide the reader the number of pixels that can be measured in 1 day and 10 days are included in A) and the data load corresponding to 1 GB and 4 GB are included in B). 4 GB is the maximum RAM that can be assigned by a 32bit operation system (note owing to the memory system requirements the useable memory is less). The data loads correspond to the memory requirements for storing the data. The 1 GB data load represents a more realistic data load of the experiments that maybe analyzed using multivariate techniques on a typical computer workstation.

remain and so some processing steps have proven important to obtaining reliable imaging MS data.

2.1. Smoothing and de-noising

2.1.1. Mass spectral

A peptide or protein MALDI mass spectrum recorded using a time-of-flight mass analyzer often includes distinct peaks due to the MALDI matrix, the most intense of which are normally suppressed using a fast ion-optical gate or gated-detector gain in order to limit the loss of detection sensitivity associated with detector saturation. The spectra also contain high-frequency noise and a broad, non-resolved chemical background. The removal of the noise and background, via smoothing and background subtraction routines, is essential for accurate mass measurements and quantifying MALDI signals [8,11]. In imaging MS the background may exhibit a systematic variation across the tissue and so background subtraction is essential in order to obtain reliable MS images [8], Fig. 3A. The noise and background are very much instrument and application (mass range) dependent, and the algorithms as well as their parameters may need to be tuned for the instrument and the detected mass range. Nevertheless modern mass spectrometers include automated algorithms specifically for these purposes.

2.1.2. Image noise

There can be significant inter-pixel variation, 'salt-and-pepper' noise in imaging MS experiments because of the stochastic nature of the MALDI process. A number of reports have now detailed different methods to reduce the impact of this image noise on the subsequent data analysis.

- Nearest neighboring averaging: McDonnell et al. investigated the effects of mass spectrometric and image processing methods on the correlation coefficients between MS images [8]. A basic near-neighbor (boxcar) smoothing algorithm greatly increased the visual impact of the MS images and reduced the influence of inter-pixel noise on the calculated correlation coefficients, Fig. 3A.
- Hanselmann et al. reported the use of a Vector-Valued Median Filter and Markov Random Field, both of which use the values of each pixel's neighbors, for smoothing the results of a random forest classification algorithm [7]. For both algorithms it was found that smoothing improved the sensitivity and predictive value of the classification.
- More recently an edge-preserving smoothing algorithm was reported for smoothing imaging MS data [6]. It was demonstrated to effectively smooth the images while preserving, with judicious selection of the denoising parameter, the larger organs in a coronal rat brain tissue section. Commendably the

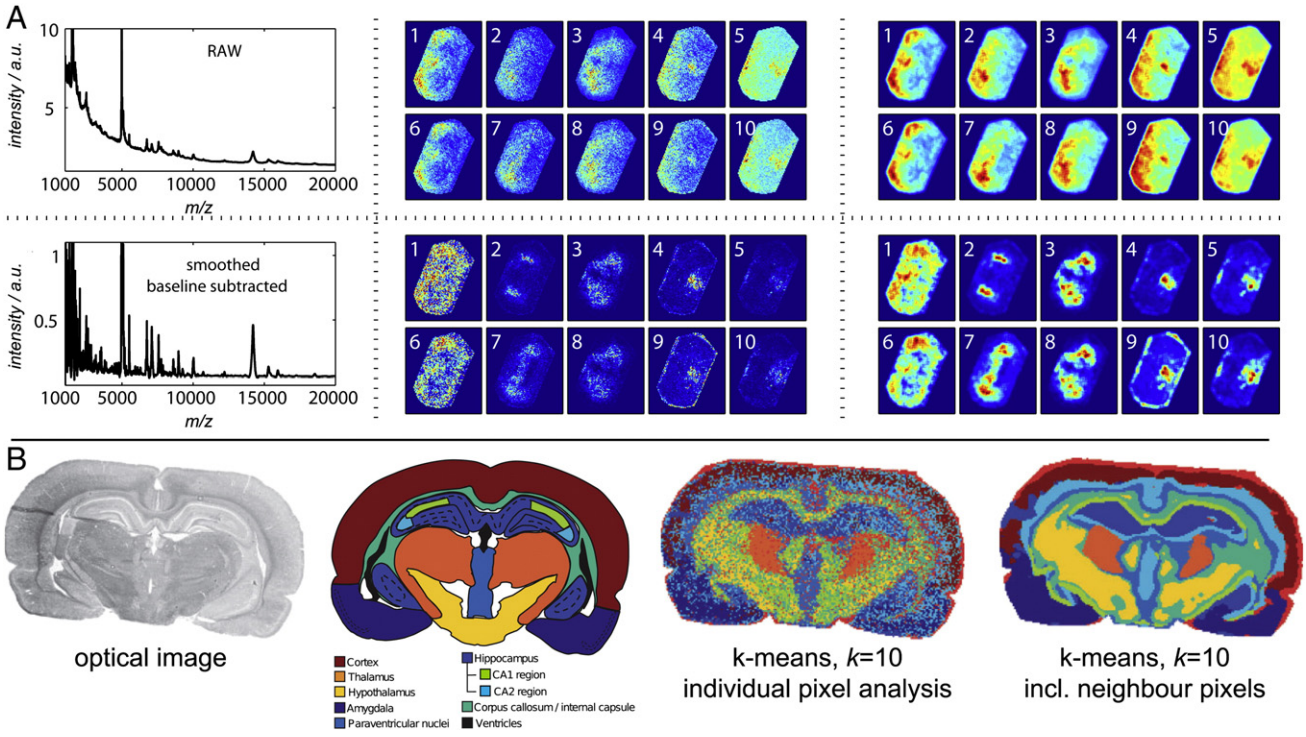

Fig. 3 – A) MALDI imaging MS analysis of a coronal rat brain tissue sections showing the effects of mass spectral and image processing on the mean mass spectrum and a selection of images. The nominal masses of the images are 1 (*m/z* 1499, lipid dimer), 2 (*m/z* 4618), 3 (*m/z* 14 207), 4 (*m/z* 1354), 5 (*m/z* 15 954), 6 (*m/z* 1475, lipid dimer), 7 (*m/z* 6750), 8 (*m/z* 7097), 9 (*m/z* 1234), and 10 (*m/z* 15 275). B) Effect of edge preserving smoothing on the results of a *k*-means cluster analysis of imaging MS data from a rat brain tissue section.

Part A adapted with permission from McDonnell et al. [8] Copyright 2008 American Chemical Society. Part B. adapted with permission from Alexandrov et al. [6] Copyright 2010 American Chemical Society.

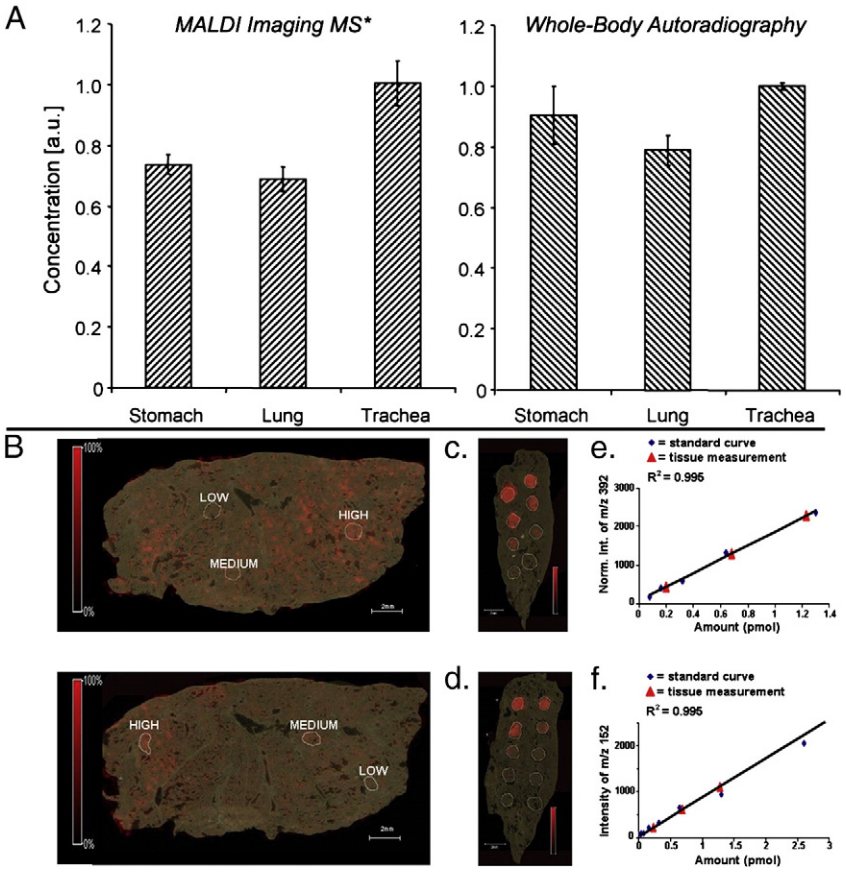

Fig. 4 – Relative quantitation and absolute quantitation using imaging mass spectrometry. A) using previously determined organ-specific MS response factors, whole-body imaging MS of rodents previously administered with pharmaceuticals provided relative quantitation in good agreement with whole-body autoradiography. B) the determination of MS calibration curves for specific pharmaceuticals enables absolute quantitation of administered drugs.
Part A adapted with permission from Stoeckli et al. [107] Copyright 2007 Elsevier B.V. Part B adapted from Nilsson et al. [41].

authors also explicitly demonstrated, by comparison with the H&E stained section of a histologically heterogeneous endocrine tumor, that the optimum smoothing strategy is determined by the length scale of the tissue heterogeneity (and, it follows, the spatial resolution of the experiment), Fig. 3B. This work has been expanded to include a spatial component in the clustering algorithm [10]. Importantly it was reported that this method was computationally less demanding, a significant benefit for the very large datasets created by imaging MS.

- The discrete ion-counting nature of mass spectrometry experiments leads to Poisson statistics, in which the variance of a given peak increases with signal intensity. Consequently intense peaks can dominate the variance based data analysis. Keenen et al. have demonstrated a simple weighting method for accounting for such Poisson noise in SIMS imaging MS datasets [34]. Recently Fonville et al. reported similar findings for MALDI imaging MS of lipids [35].

2.2. *Normalization and quantitation*

There are several aspects that influence the intensity of MS signals besides the concentration of the measured compounds. This is particularly true for the analysis of complex mixtures. Normalization is performed to (try to) account for some or all of these aspects. Normalization is a process in which the intensity of every mass spectrum in the imaging MS dataset is divided by its normalization factor. A common misunderstanding is that normalization is used as a synonym for quantification, which it can be in certain situations but not in others. Some of the effects that can influence the measured intensity of a signal in an imaging MS experiment can be grouped as specific ion suppression effects, which can happen if the ionization of a specific molecule is influenced by the presence of another. There are also global effects that can have an influence on the entire mass spectrum, e.g. it has been shown that the sensitivity of protein spectra is attenuated by the presence of lipids [36,37].

One common normalization technique in targeted analysis (e.g. of drugs) is to use an isotopically labeled standard. If the normalization is performed using such a standard, then the molecule-specific ion suppression effects and global effects are both accounted for; in such a case normalization is effectively (semi-)quantification. If normalization is based on a mass spectrum's mean intensity (or total ion count), then some of the global effects can be accounted for, but molecule-specific ion suppression effects will remain in the data. In this case normalization is not quantification.

2.2.1. *Molecule-specific ion suppression and normalization*

Ionization biases are prevalent in mass spectrometry analysis of complex mixtures [38]; peptide (and protein) purification and separation technologies are routinely used to increase the number of species detected in a mass spectrometry experiment [39]. Such liquid based separation strategies are of limited utility for imaging MS because of the need to retain spatial integrity and the extremely small amounts of tissue analyzed in each pixel: a single 100×100 μm pixel contains just 25 average sized cells (20 μm). The ability of imaging MS to detect hundreds of peptides and proteins directly from a tissue section is testament to the successful on-tissue fractionation that occurs during sample preparation. Nevertheless, even if hundreds of distinct species can be detected significant ionization biases can remain. Stoeckli et al. have demonstrated that if the relative response factors of an analyte in different tissues can be determined then imaging MS can provide relative quantification [40], Fig. 4A shows that good agreement can be obtained with whole-body autoradiography. These experiments concerned the analysis of pharmaceuticals in whole body tissue sections; the relative response factors were determined by homogeneously depositing the pharmaceutical on to whole-body sections from an undosed animal. MALDI imaging MS of the uniformly coated whole body tissue section did not generate a uniform MALDI signal of the pharmaceutical. When the relative response factors were calculated from the relative signal deviations, and then applied to MALDI imaging MS results obtained from a dosed animal, the relative quantitation was consistent with results obtained using whole body autoradiography. Explicit quantitation of pharmaceuticals has been performed by determining tissue specific calibration curves [41], Fig. 4B, which were then confirmed using LC–MS analysis of tissue extracts.

The simultaneous determination of relative response factors for all peptides and proteins detected from tissue is much more challenging (and to these authors' knowledge has not been performed to date); it would require isotopically labeled analogues of all detected peptides and proteins to be added as internal standards as well as a significant increase in the peak capacity of the mass spectrum to resolve every component. The quantitation of all peptide and proteins ions would require calibration curves to be obtained throughout the tissue for every ion. Owing to the lack of practical quantitation strategies peptide and protein imaging MS experiments typically compare the MS signals (after a number of preprocessing and normalization steps [8,11]).

For all these reasons, MALDI imaging MS of peptides and proteins cannot be considered quantitative. Nevertheless there are strong indications that valuable information can be gained from protein imaging experiments at least in a semi-quantitative way: There are many papers that have used MALDI imaging for the discovery of disease markers in tissue. In those cases in which the markers have been identified and independently validated by immunohistochemistry (IHC) the MALDI imaging findings could be biologically linked to the disease. It also should be kept in mind that all other techniques for measuring proteins in tissue, particularly IHC, suffer from poor quantitation.

Molecule-specific normalization using well-chosen isotopically-labeled analogues (i.e. avoid isotopomers in which the functional groups have been labeled) can accurately account for changes in ionization efficiency but great care must be taken to ensure it is added uniformly across the tissue sample.

2.2.2. *Global intensity suppression and normalization*

Molecule-specific ionization biases are distinct from effects that can lead to intensity changes throughout a pixel's mass spectrum (global intensity suppression). These effects can be caused by the chemical background in the tissue, e.g. it has been shown that the presence of lipids has a detrimental effect on the measurement of proteins, and is the reason the sample preparation protocols for protein imaging MS usually include tissue washes using organic solvents [36,37]. It has also been found that the overall intensity of the MALDI imaging MS spectra correlate with the amount of matrix on the tissue, consequently any inhomogeneities during matrix deposition will lead to non-uniform measurement sensitivity across the tissue. It is for these reasons that sample preparation is one of the key aspects of a successful MALDI imaging MS experiments.

MALDI imaging MS experiments require the acquisition of a very large number of individual MS scans (defined as the mass analysis of the ions from a single laser shot). For example a single 2×1 cm tissue section, measured with 50 μm spatial resolution, 300 laser shots per pixel and at an acquisition rate of 2 pixels s^{-1}, would require >11 h continuous measurement time and involve 24 million MS scans. Modern high-quality MALDI instruments have been specifically designed to provide constant laser fluence, ion-transport and ion-detection efficiency for the duration of an imaging MS experiment. Nevertheless, a slow degradation of performance may remain for larger imaging MS studies (the debris produced during the MALDI process can build up during the experiment adversely affecting ion transport) but is more acute for older instruments, which have not been designed for such high intensity applications. Similar considerations for SIMS and DESI imaging would pertain to the stability of the primary ion current (SIMS) or Electrospray (DESI), and sample charging during data acquisition (SIMS).

To account for global intensity changes in an imaging MS dataset one can normalize on spectrum wide parameters such as the mean intensity or total ion count (TIC). There has been ongoing discussion in the field about the applicability of such approaches because of the observation that the normalization can, in some cases, substantially change the image [42]. Substantial changes in distributions associated with TIC normalization indicate significant differences in pixel total-ion-counts throughout the imaging MS dataset; in such circumstances TIC normalization is not valid and the sample preparation/data processing strategy should be reassessed. Normalization to TIC implies that the area under every spectrum in the data-set should be identical, and it is only recommended to be used when the TIC's do not vary significantly.

For MALDI analysis of simple mixtures TIC normalization is effective at improving the quantitative performance of MALDI [11]. TIC normalization is used in imaging to correct for small differences in matrix coverage/laser fluence during data acquisition. When analyzing heterogeneous tissues it is difficult to argue a priori why MALDI spectra should generate a constant number of ions. This is further complicated by the fact that only a fraction of the mass spectrum is actually

recorded, e.g. *m*/*z* 2000–20,000, and efficiently detected [43]. The increasing number of imaging MS studies reporting independently validated biomarkers suggest that there are biomarkers present whose fold-change is greater than any difference in sensitivities due to the different chemical backgrounds of the tissues being compared. It also suggests that improved normalization strategies could lead to improved biomarker discovery capabilities by reducing bias in the measurements.

Deininger et al. have demonstrated that the area under a MALDI-ToF spectrum is mainly determined by the non-specific chemical noise, which makes normalization to TIC quite robust because the contribution of the MS signals is small [42]. When signals with very high intensity are present in localized areas of the tissue then normalization artifacts may occur. If the noise level is directly taken into account for the normalization by normalizing to the noise level or the median intensity, then a very robust normalization can be achieved, Fig. 5.

The normalization strategy is dependent on the instrument configuration. Most peptide and protein imaging MS datasets have been recorded using a MALDI-ToF mass spectrometer using a linear vacuum UV ion source because of the high sensitivity and high mass range of this platform. The MALDI mass spectra recorded with such instruments contain more chemical noise (due to non-specific metastable clusters) than is obtained with other MALDI instruments, such as MALDI-ion trap or MALDI Q-ToF because the metastable clusters do not survive ion transfer to the mass analyzer. Consequently the non-specific noise that is the principal contributor to a spectrum's TIC for MALDI-ToF instruments is absent. Fonville et al. have investigated normalization under such conditions, using data acquired on a MALDI Q-ToF instrument [35]. They investigated the use of the total ion count of the picked peaks and also other normalization methods such as normalization to matrix-related peaks or informative peaks. Informative peaks in this context are those peaks that are expected to carry biological information as established by singular value decomposition. Importantly heuristics were provided for how to select the correct normalization method. The search for a 'best' global normalization method, which can be used for all imaging MS investigations irrespective of the mass analyzer employed maybe an unrealistic goal because, for any normalization approach, examples may be encountered in which certain aspects of the data render them inappropriate. It may be better for the field to develop heuristics based pragmatic approach for normalization, and to independently validate the results of the imaging MS analyses.

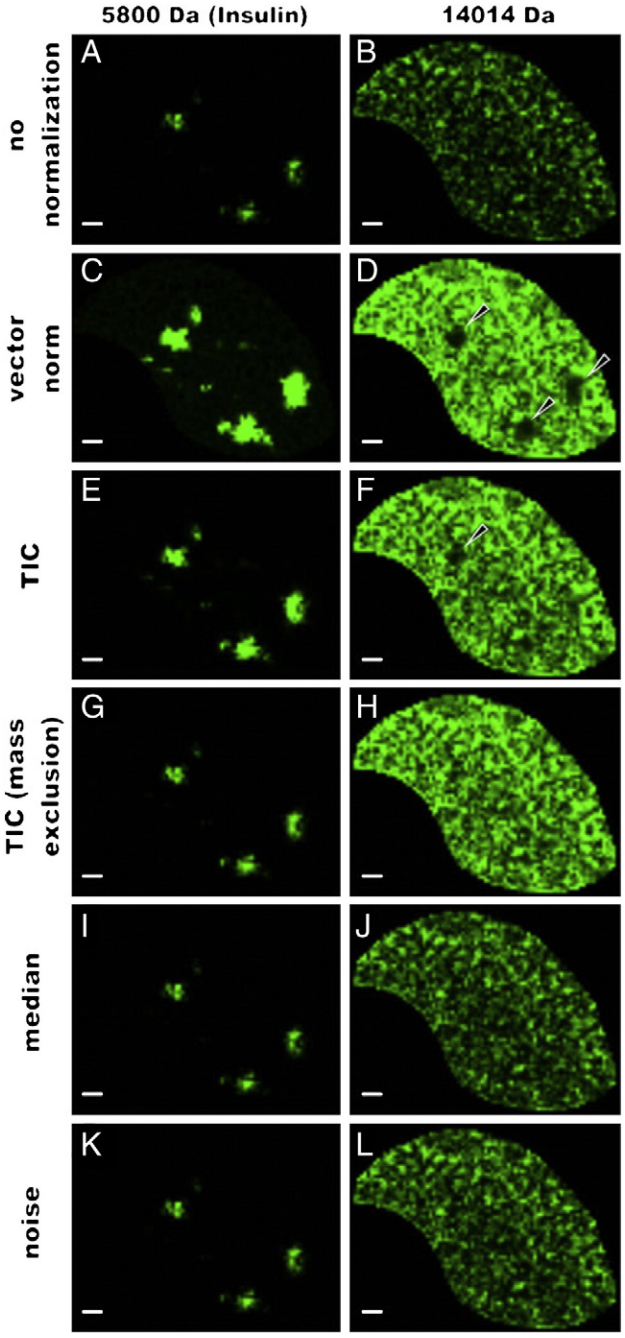

Fig. 5 – MALDI imaging MS of mouse pancreas, displaying the distribution of the intense insulin peak from islets of Langerhans and a ubiquitous peak at *m*/*z* 14,014 for different normalization methods. The high intensity of the insulin peak causes local artifacts in the 14,104 MS image when the vector and TIC normalization methods are used. Excluding insulin peak, or normalizing to the noise/median of each pixel is a more effective normalization strategy. Reproduced with permission from van Deininger et al. [42] Copyright 2011 Springer-Verlag.

3. Feature extraction

The simplest manner of analyzing imaging MS data is the visualization of single MS images and their comparison with optical images of the stained tissue section. Such analyses are sufficient for investigating the distributions of specific ions of interest, e.g. pharmaceuticals and their metabolites. For biomarker discovery or molecular histology experiments the large number of distinct molecular ions detected in a single experiment makes the manual examination of each MS image impractical, especially so if the goal is to identify correlations between the detected ions. Instead powerful statistical tools are used to interrogate the imaging MS datasets.

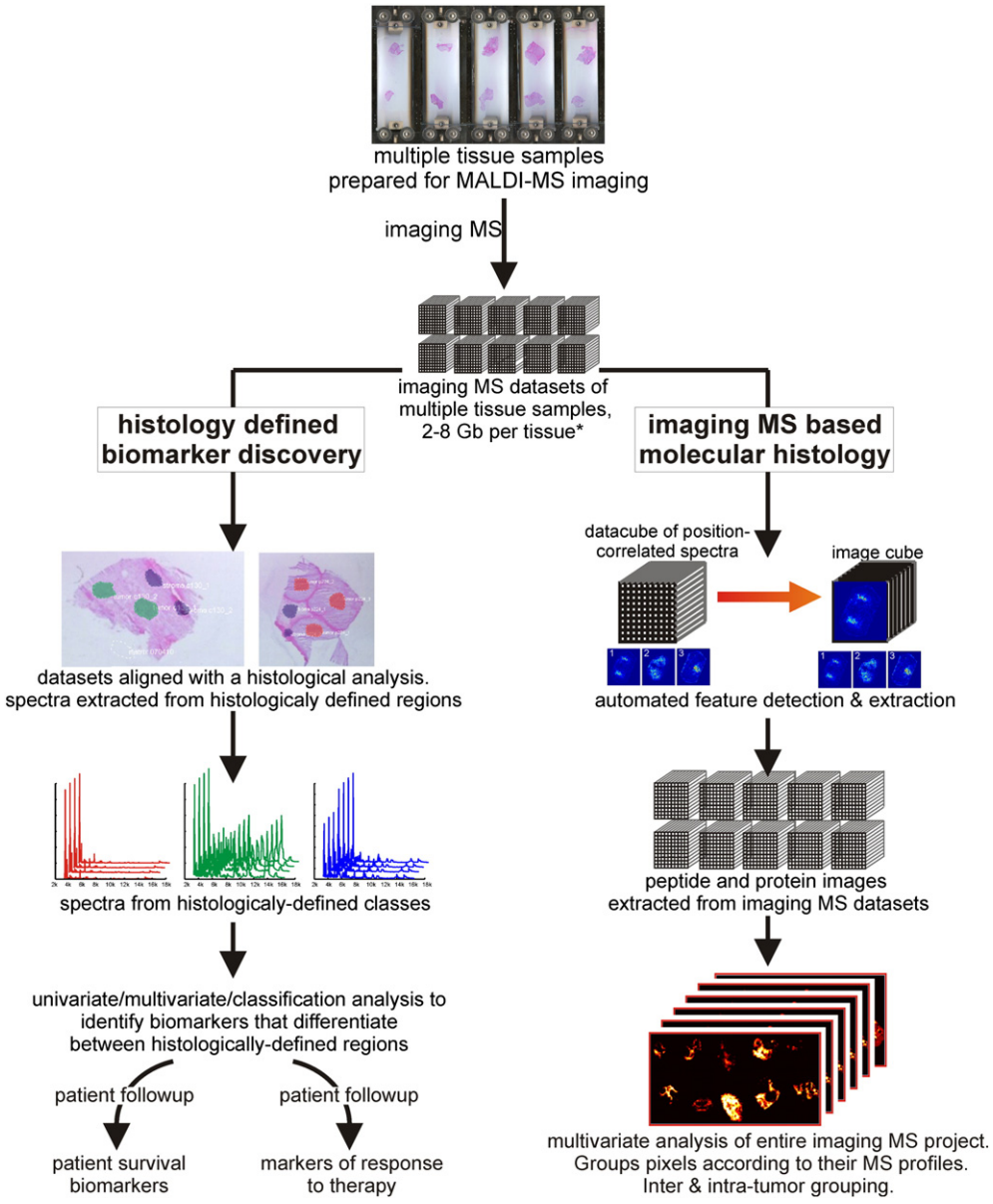

Fig. 6 – Schematic of the workflows for biomarker discovery using histology-defined analysis, in which different classes are assigned to distinct histological features to identify differentially detected biomolecular ions, and imaging MS-based molecular histology, in which the tissue is annotated on the basis of the each pixel's MS spectrum.

The data load and ultra-high dimensionality of the raw or preprocessed (smoothing, background subtraction) imaging MS datasets is unsuited to most data analysis pipelines. The number of channels in each mass spectrum, e.g. 60k for a time-of-flight mass analyzer or 500k for a high mass resolution Fourier transform ion cyclotron resonance mass spectrometer, leads to very large data loads and long processing times. For example the number of floating point operations of a widely used singular value decomposition algorithm, as used in principal component analysis (PCA), is given by *flop's* = $14 \cdot k \cdot N^2 + 8 \cdot N^3$, where k is the number of pixels and N the number of variables, or discrete mass channels in a mass spectrum. PCA of an imaging MS dataset consisting of just 10k pixels and 40k *m/z* channels would require 7.4e14 floating point operations, or more than >2 h using a high performance 3.8 GHz i7 CPU (70 gflops maximum processing speed). Early examples of PCA analysis of imaging MS datasets employed spatial binning, mass spectral binning, and *m/z* range selection, to make the datasets more tractable [44]. However this strategy necessarily compromises mass and spatial resolution to a degree that might be deemed unacceptable, and could lead to the merging of mass spectral peaks.

van der Plas et al. have reported wavelet based smoothing and data reduction of each pixel's mass spectrum [45,46]. A 1D wavelet transform yielded an approximately 8X reduction in channel number but retained almost all spatio-chemical information. The 8× reduction underestimates that which may be obtained, because of the very low mass resolution of the original imaging MS dataset (just 6490 samples spanning the *m/z* range 2800–25,000). 3D wavelet based data reduction exceeding 100× has been reported for hyperspectral imaging from focal plane array detectors [47].

On-the-fly peak detection can be efficiently performed on high mass resolution datasets to reduce the data load. The low mass resolution of the peptide and protein peaks obtained using a linear time-of-flight mass analyzer make automated peak detection less reliable. The low S/N of the peaks in each pixel's mass spectrum (each pixel analyzes a very small number of cells) can exacerbate this problem, consequently on-the-fly detection of the peaks in each pixel's mass spectrum has been limited to Fourier transform-type instruments (Orbitrap, FTICR) and an increasing number of time-of-flight analyzers.

On-the-fly peak detection of each pixel's mass spectrum must balance the use of low peak detection thresholds (to select as many of the low S/N peaks as possible) with the increased amount of noise retained in the reduced dataset. Furthermore, owing to the large number of pixels in an imaging MS dataset, each of which has an associated mass spectrum, the peak detection algorithm should be rapid. Alexandrov et al. reduced the severity of the processing burden by limiting peak selection to 10% of the dataset's pixels [6]. The implicit assumption is that 10% of the pixels provide an accurate representation of the entire imaging MS dataset; this will be true for the broader spatial features but may omit more localized spatial features.

The most common approach for data reduction is to use the (mean) average mass spectrum of the imaging MS dataset as the principal mass spectral interface. Peak detection of the average mass spectrum is adept at highlighting the principal MS features but may omit more localized features (in the average mass spectrum their signals are diluted by the large number of additional pixels in the dataset). McDonnell et al. have reported the use of multiple mass-spectral representations of the imaging MS datasets, including representations that explicitly highlight localized features [27]. The four mass spectral representations are calculated and then a peak-detection algorithm specifically developed for MALDI-TOF of proteins [48], but which is too slow to be applied to individual pixel mass spectra, applied to each representation. Note that the imaging MS software MITICS also calculates multiple mass spectral representations in order to improve peak detection efficiency [49].

The peaks located by the peak-detection algorithms, whether from mass spectral representations or individual pixel mass spectra, are collated into a final feature list which is then used to extract all peaks from every pixel's mass spectrum. Such peak identification and extraction methods reduce imaging MS datasets to the biomolecular peaks present in the dataset, thereby reducing the influence of the chemical background. The reduced data load and number of variables, 300–1000X, enables the efficient application of statistical data analysis algorithms. Feature identification is an essential element of imaging MS data analysis, whether analyzing complete imaging MS datasets or specific regions of interest defined by the histology of the tissue section.

4. Choice of data analysis method

There are several distinct manners of analyzing the data, which can be further subdivided into data analysis strategy and data analysis methodology. The choice of strategy and methodology is ultimately determined by the goals of the experiment.

4.1. *Data analysis strategy*

Imaging MS was originally developed by mass spectrometrists, consequently the first data-analysis methods were MS-centric. As the technique has become more established in the clinical and pharmacological fields a histology-centric view has become increasingly apparent [50–52]. These two viewpoints reflect the different backgrounds of the practitioners and a well-performed experiment requires expertise in mass spectrometry, histology as well as statistics. Three distinct data analysis strategies have emerged, which differ in how histology is used to specify which imaging MS data will be submitted to the subsequent data analysis method and how that data is formatted.

a) Histology-defined. A histopathological examination of the tissue section is performed to annotate specific entities of interest, for example invasive ductal carcinoma cells. The mass spectra from the pixels contained in the annotated regions are then extracted as profiles specific to that histopathological entity, Fig. 6. In this manner the MS profiles of, e.g. HER2 positive invasive ductal carcinoma can be compared with similar profiles from HER2 negative invasive ductal carcinoma to identify biomarkers indicative of HER2 status [50,52]. By comparing tumor-cell specific MS profiles with patient outcome [5,53], or response to therapy [54], it is possible to search for prognostic biomarkers. A histology-defined data analysis strategy subordinates imaging MS to established histopathological methods. In essence imaging MS is used to further increase the specificity of the MS profiles.
   Note: Histology-defined data analysis of imaging MS datasets differs from histology-defined protein profiling [55]. In histology-defined data analysis of imaging MS datasets mass spectra have been acquired from all over the tissue section, consequently the researcher is free to perform retrospective data analysis of any cell type present in the tissue. In histology-defined protein profiling MS profiles are only obtained from those regions previously indicated by the pathologist. The advantage of histology-directed protein profiling is that the experimentalist is more free to optimize the sample preparation of the MALDI experiment (larger matrix droplets) for increased sensitivity, because the spatial resolution requirements are less stringent, and as data is only acquired from specific features of interest the experiments are faster and more focused on the pathological question. Bruand et al. have reported a workflow to identify MS peaks with MS images localized to specific features of interest [56].
b) Histology-directed. Imaging MS experiments have been configured to examine how tumor associated changes in protein content correlate with its histological border. In both malignant fibrous histiocytoma [57] and renal cell carcinoma [58] it was reported that tumor-associated protein signals extended far past the histological tumor border into the surrounding normal tissue.
   The renal cell carcinoma example included an elegant extension of histology-directed analysis for the investigation of tumor borders; the protein intensities were extracted as a function of their nearest-distance to the histological border; the projection of the imaging MS data onto this distance

metric resulted in distance vs. MS intensity profiles that had a much higher spatial resolution than the original imaging MS data, Fig. 7.

Kang et al. have used histology-directed imaging MS to identify protein peaks that were detected at higher intensity in tumor borders [59], it was argued that these regions may provide diagnostically valuable biomarkers as they are the regions of greatest vascularization and growth.

c) Histology-independent analysis. Imaging MS-based molecular histology consists of analyzing entire imaging MS datasets to identify which regions have distinct and/or correlated MS profiles, Fig. 6 (right-hand side). The results of the analysis can then be compared with the tissue's histology. Many examples of imaging MS based molecular histology have concerned rodent brain tissue [6,10,13] or other tissues containing clearly defined histological regions (e.g. viable tumor vs. necrotic [60]). The potential advantage of imaging MS-based molecular histology is that it can complement established histological and histochemical methods by uncovering molecular differences that were previously unknown. Recent notable examples include the differentiation of morphologically overlapping tumors [61], the identification of intratumor heterogeneity [9,61,62], differential MS signatures associated with differentially metastatic human breast cancer xenograft models [60], molecular changes that occur during embryo development [63,64], e.g. Fig. 8, and even how the differential protein profiles revealed by imaging MS can improve the definition of brain anatomy [65]. An emerging application area in imaging MS-based molecular histology is wound repair, in which imaging MS is used to highlight biomolecular ions associated with the repair process [66–68].

4.2. *Data analysis methodology*

The data analysis strategies described above concern which mass spectra are selected from the imaging MS dataset. The researcher then has a number of choices of how one selects which mass spectral peaks to include in the subsequent statistical analysis, and which analysis algorithms will be used.

The first decision to be made is whether a supervised or an unsupervised analysis should be performed. The principal difference between supervised and unsupervised analyses is that the former uses prior knowledge about the samples to generate a model whereas the latter seeks *unknown* latent variables in the data. The most common use of a supervised analysis in imaging MS is to identify profiles or specific biomolecular ions that discriminate pathological entities, e.g. a tumor from benign tissue. If information is known about patient outcome or response-to-treatment a supervised analysis of the tumor specific profiles can be used to identify prognostic markers. In all instances it is essential that the model is validated using an independent validation set of imaging MS measurements (more details in Section 7). In this example the prior knowledge of the supervised analysis is the pathological classification of the patient tissues and the clinical outcomes of the patients. The analysis is defined by the prior knowledge in order to create a model and identify candidate biomarkers. In contrast an unsupervised analysis does not require prior knowledge, instead statistical data analysis tools are used to examine latent correlations within the imaging MS datasets. The principal difference between supervised and unsupervised analysis is that an unsupervised analysis may reveal unexpected/unknown features; for example there are now multiple reports that unsupervised

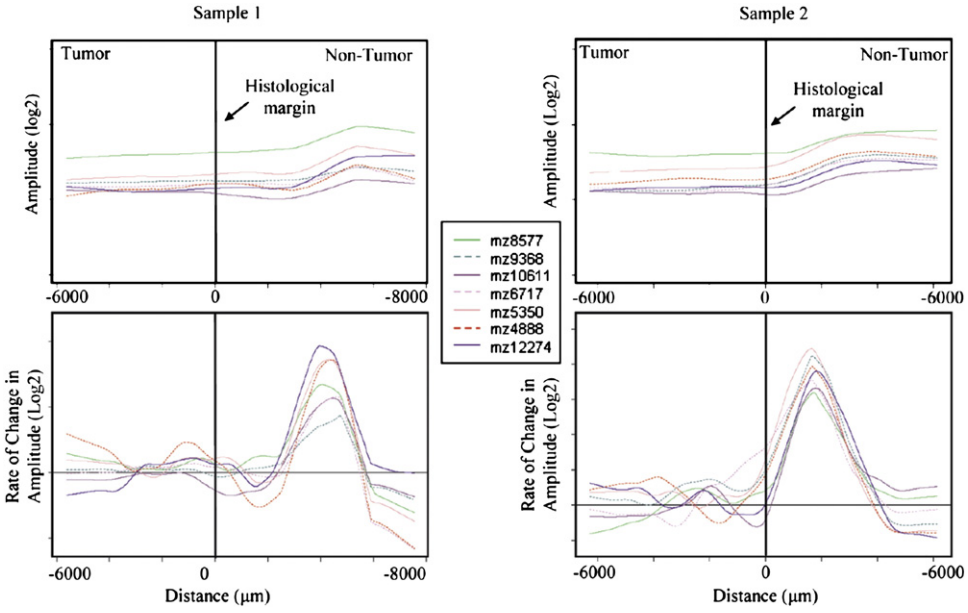

Fig. 7 – A) MALDI Imaging MS analysis of tumor borders of renal cell carcinoma reveal the influence of the tumor may extend several millimeters past the histological tumor border.
Adapted with permission from Oppenheimer et al. [58] Copyright 2010 American Chemical Society.

analysis may reveal intratumor heterogeneity within imaging MS datasets [9,61,62]. Unless previously specified in the input data a supervised analysis would not highlight the heterogeneity. A large number of supervised and unsupervised statistical methods have been applied to imaging MS datasets. In this review we have separated the methods according to how they are applied in imaging MS and their nature. Section 5 provides a detailed overview of the methods used for unsupervised imaging MS based molecular histology; Section 6 describes the methods for comparing sample cohorts and biomarker discovery.

4.3. *Data analysis — in practice*

A completely unsupervised data analysis workflow would require no user-input other than selecting the imaging MS dataset. In reality the pre-processing algorithms used for mass spectral smoothing, background subtraction and signal normalization, as well as the feature detection and extraction algorithms, require some user input to match the characteristics of the mass spectra (which are dependent on the mass analyzer employed and the mass range selected for analysis). Once this set-up procedure is complete the algorithms can be automated. (Note: the low mass resolution of the linear time-of-flight mass analyzer used for protein imaging MS can lead to inefficient peak detection and so manual peak selection may be necessary). Formally preprocessing and peak selection is supervised because of the user input during their set-up, even if this was performed using different data (typically standardized methods are developed and then applied without additional 'tinkering').

There are several aspects that determine which input data is provided to the statistical analysis tools used in a biomarker discovery or molecular histology experiment. The selection of regions of interest corresponding to specific cell types, pathohistological entities, or organs determines which pixel's mass spectra are included. The mass spectral processing and feature detection algorithms influence which peaks are input into the data analysis algorithms. Manual inspection of the mass spectra may be required in case of inefficient automated peak detection or to remove noise peaks (electronic and chemical noise). These aspects may be considered data preparation, in that they determine the data input, though formally they are supervised. The question of whether the data analysis algorithm itself is supervised or unsupervised is dependent on whether additional information is used to assign specific groups to the data.

5. Imaging MS based molecular histology: annotating tissues on the basis of MS profiles

Conventional histology is based on the study of the microscopic anatomy of tissue, visualized through the application of different tissue stains such as hematoxylin and eosin (H&E). For example a histopathological examination of soft tissue sarcomas uses cellular phenotype, pleomorphism, and cellularity for tumor classification and differentiation, mitotic rate, and necrosis to grade the tumor [69,70]. Instead imaging MS-based molecular histology partitions the tissue section into regions based on each pixel's MS profile, the goal of which is to identify latent features in the dataset. Such a procedure lends itself to many different multivariate analysis techniques that have been developed for the identification of correlations between high dimensionality datasets. By treating each pixel as an independent sample conventional multivariate techniques can be used where their output is shown not as scatter plots (or heat maps etc.) but as score plots in the spatial domain. Here we provide an overview of the common techniques used in the analysis of imaging MS datasets.

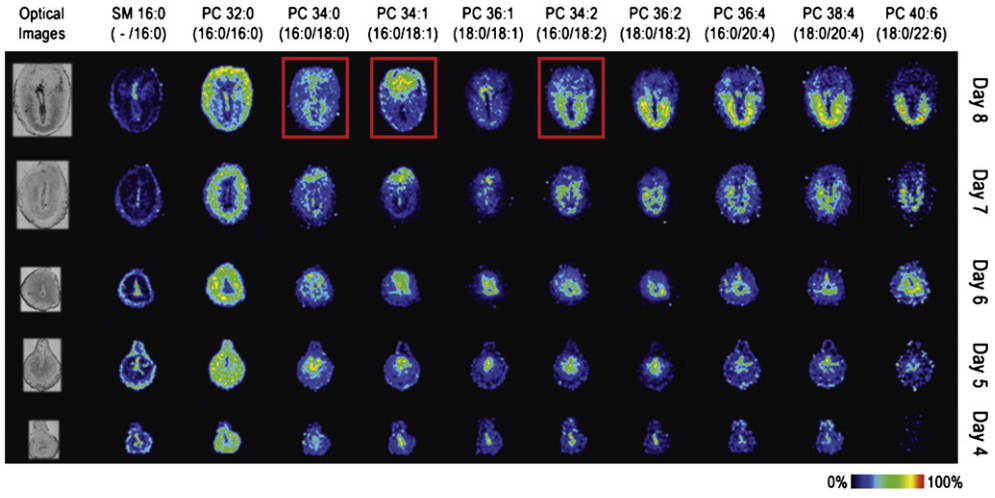

Fig. 8 – MS images of SM and PC phospholipids on days 4–8 of implantation. Each column represents a unique potassiated phospholipid, $[M+K]^+$, and each row represents a different day of pregnancy. Each image is orientated so the mesometrial pole, M, is at the top and the antimesometrial pole, AM, is at the bottom. Note that lipids, that differ by just a single unsaturation can have completely different spatial distributions (indicated with boxes).
Adapted with permission from Burnum et al. [63] Copyright 2009 the American Society for Biochemistry and Molecular Biology, Inc.

5.1. *Principal component analysis (PCA)*

PCA has been, and remains, the most widely used multivariate technique used to analyze imaging MS datasets [13,71]. PCA has a long history of application in the field of secondary ion mass spectrometry [72], principally as a tool to identify patterns in the complex fragment-ion spectra produced by the technique (an extensive list of publications can be found at the multivariate surface analysis website [73]). PCA can be applied to the spectra from specific regions-of-interest and from the entire tissue section [74]; in the former case PCA is used to distinguish groups in the extracted spectra, whereas in the latter PCA is used to trace the distribution of the groups.

The ability of imaging MS to highlight potential groups may not immediately be apparent. The explicit goal of PCA is to reduce the number of variables needed to represent multidimensional data with minimal loss of information. This is achieved by finding relationships between variables (MS peaks) and between samples (pixels). The PCA approach is based upon the variance in the data, removing redundancy through grouping variables with high covariance. In somewhat overly-simplified terms the variance of the data can be

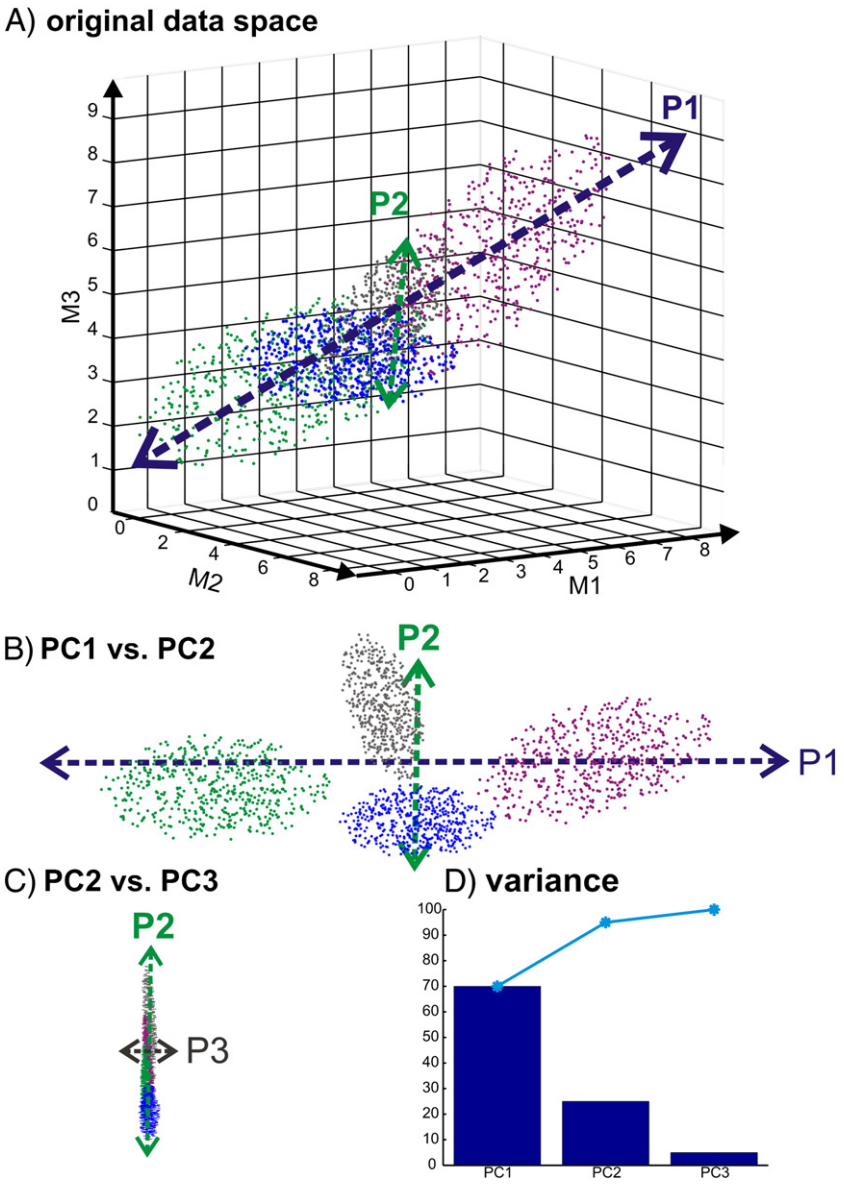

Fig. 9 – Schematic of principal component analysis. A) Distribution of pixel mass spectra in M1, M2, and M3 data space. Each pixel's coordinates correspond to the intensity of the peaks at mass M1 (x-axis), M2 (Y-axis), and M3 (Z-axis). There is an apparent correlation in the data. Rotation and translation of the coordinate system to create a new coordinate system P1, P2 and P3, these are the principal components. B) P1 contains the maximum variance (spread) in the data and P2 the next largest variance subject to the condition that it is orthogonal to P1. C) P3 describes very little variance, and so the same data can be accurately represented by just P1 and P2. D) A plot of variance versus component number can help determine how many principal components should be retained.

deemed its spread and covariance can be considered a measure of similarity (between pixels' mass spectra).

PCA performs a linear orthogonal transformation of the data to maximize variance, resulting in a set of orthogonal principal components that describe the largest variance in the dataset (PC 1), the next largest variance (PC 2), and so on. Fig. 9 shows a schematic representation in 3 dimensions. The X-dimension corresponds to the intensity of the peak measured at mass M_1, the Y-dimension to the intensity of the peak at M_2, and the z-dimension to the intensity of the peak at M_3; the locations of the data points are determined by the intensities of M_1 and M_2 for 5k measurements. The schematic has been designed to include four groups within the data, but which are overlapping. A simple transformation of the axes M_1, M_2 and M_3, a rotation and a translation, can place almost all of the data points along two transformed axes — and that the spread of the data, its variance, is maximum along these transformed axes. P1 is the transformed axis containing most of the variance, and is called the first principal component. Axis P2 maximizes the remaining variance but is orthogonal to P1, and is call the second principal component. It can be seen that just 2 principal components are sufficient to describe most of the variance in the data and distinguish the four groups. PCA performs this calculation, and instead of 3 dimensions it can perform it for very many dimensions. Nevertheless most of the variance in the dataset is held in the first principal components. The number of principal components retained in the transformed data is either (arbitrarily) specified by the user or determined by the amount of data variance retained in the transformed output. The results of PCA are typically presented by showing score plots, spectral loadings and the variance.

- Scores: each pixel's mass spectrum will have a score for each principal component. The scores are the coordinates of the pixel in the transformed axes (re. Fig. 9). Principal component images are obtained by plotting each pixel's score at its pixel location.
- Loadings: represent the contribution of the original coordinate system, (the peaks in the mass spectrum) to the principal components (the new coordinate system). Peaks with a high loading to a principal component contribute a lot to a principal component. When a score-plot image differentiates a tissue into two regions the loadings communicate which peaks in the original mass spectra contributed to the differentiation.
- Variance: reports the proportion of the dataset's variance explained by a principal component. The variance should always be reported, as it communicates the contribution (importance) of the principal component to the spread in the original data.

PCA is a very efficient algorithm and often allows a rapid summarization of potential latent features in the imaging MS dataset, with the majority of a datasets variance (e.g. >90%) described by just a handful of principal component images. The remaining principal components, describing small and decreasing percentages of the dataset's variance, are often discarded. PCA ranks latent features according to their contribution to the dataset's variance, consequently large differences such as substrate vs. tissue or sample preparation artifacts may dominate and more subtle or localized differences, considered less important, can be discarded.

Each principal component output results in an image, consequently regions of the datasets maybe grouped together (as a latent feature) in principal component 2 but not in other principal components. It is therefore imperative that any associations revealed in a specific principal component image are confirmed by examining the images of the peaks that contribute most significantly to that principal component (which are reported in the loading plot). Despite these qualifications, due its low computational burden and rapid computational time PCA is often used for a cursory inspection of the data (e.g. to highlight systematic biases that occur, e.g. due to inhomogeneous matrix application).

PCA is often used as a preprocessing step prior to the application of more advanced data analysis algorithms. In this mode PCA is applied, the principal components retained that describe a defined amount of variance, e.g. 90%. In this manner the uncorrelated noise present in the later principal components is omitted before data analysis. In a somewhat niche application Broersen et al. used the ability of PCA to increase the contrast of the images (by removing uncorrelated noise) to automatically align high spatial resolution imaging MS datasets from adjacent (and partially overlapping) regions of a tissue section [44]. Fletcher et al. recently used a similar approach to improve the alignment of imaging MS datasets of sequential tissue sections for the construction of 3D imaging MS datasets [75].

A number of variants of PCA have been reported, e.g. related to data preparation [25,45,76], visualization [77] and additional optimization steps [71].

5.2. *Clustering methods*

In imaging MS clustering refers to the grouping of pixels on the basis of their MS profiles, to group together pixels with similar MS profiles (similar peptide and protein peaks and similar intensities) and separate pixels with dissimilar MS profiles (different peptide and protein peaks and/or dissimilar intensities).

5.2.1. *Hierarchical clustering*

Hierarchical clustering organizes the pixels into an ordered grouping, referred to as a hierarchical tree or dendrogram, based on the similarity between the pixel's mass spectra. A similarity metric is used to calculate the similarity between the pixels' mass spectra, this is often the Euclidian distance (e.g. the distance between two pixels in the data space, re. Fig. 9) but many other metrics exist [62]. A 'linkage' calculation is also required to determine how, higher up the hierarchical tree, clusters (of mass spectra, corresponding to clusters of pixels) are amalgamated into a higher-order cluster. In imaging MS this is often based on the Euclidian distance between the centroids of the clusters. During the construction of the dendrogram the nearest two clusters (or pixels) are progressively combined into a higher-level cluster. Principal component analysis is typically performed prior to hierarchical clustering owing to its sensitivity to noise; for example McCombie et al. have reported that including too many

principal components (200 vs. 5) led to most pixels being isolated in single-pixel clusters [13].

Hierarchical clustering is extensively used for gene expression analysis. Schwartz et al. have demonstrated that when applied to MS profiles obtained directly from tissues it can be used distinguish the different tumor grades [78]. Partially reflecting its ready availability in a widely used commercial software package hierarchical clustering has proven to be an accessible and interactive data analysis tool for investigating the structure of the imaging MS datasets [62,74,79]. By selecting different branches of the dendrogram the analyst can efficiently investigate the progressive grouping of the data. Fig. 10 shows a hierarchical cluster analysis of an imaging MS dataset of healthy pancreas, pancreatic adenocarcinoma and pancreatic insulinoma. The first junction differentiates healthy pancreas from the two pancreatic tumors. The next junction separates the pancreatic adenocarcinoma from the insulinoma. False color images of the tissues, colored according to which branch a pixel belongs, clearly demonstrate the separation of the tissues according to their phenotype. The advantage of hierarchical clustering is the ability to investigate subgroups within the data. Further exploration of the dendrogram revealed a subgroup of the healthy branch consistent with islets of Langerhans and subgroups within the adenocarcinoma branch clearly revealed the tumor's interface with the surrounding tissue.

5.2.2. *k-Means clustering*

k-Means clustering partitions the imaging MS dataset into a pre-defined number, *k*, of clusters, normally on the basis of the Euclidian distances between the mass spectra. In *k*-means clustering the algorithm is initialized by selecting *k*-locations in the data space, these are the initial *k*-means. An iterative process then ensues, whereby

i) Each pixel's mass spectrum is assigned to the cluster with the closest mean.
ii) The centroid of each of the clusters is defined as the new mean.
iii) Steps i) and ii) are repeated until convergence, when the cluster membership of each pixel does not change with subsequent iterations.

k-Means clustering is a powerful method for quickly assessing the spatio-chemical organization of imaging MS datasets [9,13]. Alexandrov et al. have termed the application of such clustering techniques to imaging MS datasets as 'segmentation' [6]. In their first paper it was demonstrated that an edge-preserving smoothing function (to smooth salt-and-pepper noise) followed by clustering could reproduce most of the tissue's histology if an appropriate number of clusters was selected by the analyst [6]. In place of *k*-means clustering this study was based on high dimensional discriminant clustering, a technique more suited to the high dimensionality of imaging MS data but which is much slower than *k*-means clustering. To circumvent this dimensionality issue it was recently proposed to use the FastMap algorithm to project the imaging MS data onto a lower space while maintaining Euclidian distances between pixels [10]. This enabled *k*-means clustering to be efficiently applied and, importantly, to include a spatial element of the data into the clustering analysis (the data from a pre-defined radius of pixels are clustered together with weightings according to the neighbor's pixel-to-pixel distance or their spectral similarity). The results clearly show that the spatially-adaptive clustering is less prone to inter-pixel noise.

Spatially aware data analysis represents an important development for clustering of imaging MS datasets. However it also raised a number of issues that need to be considered. The results of spatially aware clustering are dependent on the radius of neighboring pixels (that are clustered together), the method used to weight neighboring pixels, the dimensionality of the FastMap space, and the number, *k*, of clusters selected by the analysis. Alexandrov et al. used the tissue's histology to guide the parameter selection (it is commendable that these factors are extensively discussed in the paper and not hidden from the reader). It is these author's opinions that imaging MS based molecular histology will have a greater impact if it can offer additional discriminative

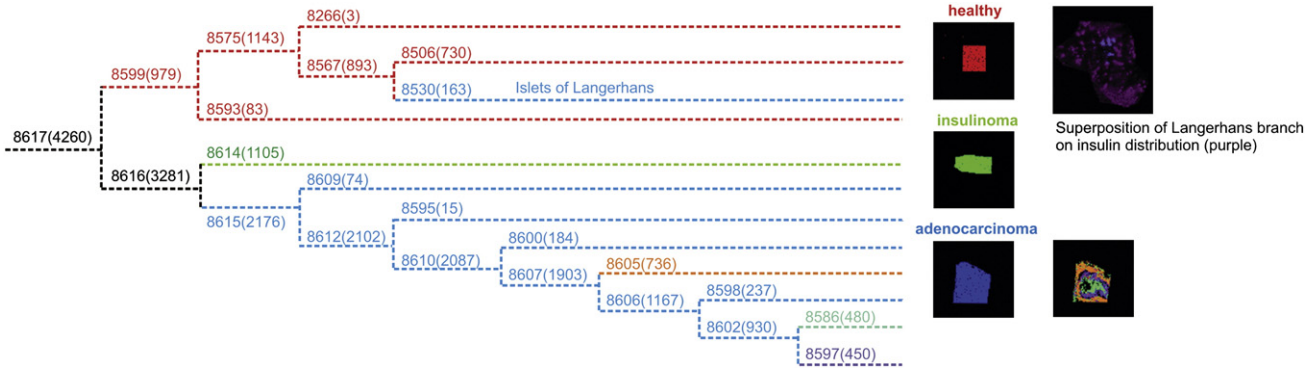

Fig. 10 – MALDI imaging MS analysis of pancreatic adenocarcinoma, insulinoma and healthy tissue. Hierarchical cluster analysis of the combined datasets distinguishes the adenocarcinoma, insulinoma and healthy tissues as well as the substructure in the subsequent branches.
Reproduced with permission from van Remoortere et al. [74] Copyright 2010 Elsevier B.V.

capabilities; if it is reliant on a tissue's histology to define the number of clusters and parameters then its potential application is limited to identifying biomolecular ions that are specific-to/differentially-detected in histological entities. A variant of *k*-means clustering, fuzzy c-means, also partitions the dataset into a number of classes defined by Euclidean distances but allows pixels to occupy multiple classes, enabling the underlying molecular patterns to be identified [80] rather than forcing each pixel into a single class. Along with non-negative matrix factorization and probabilistic latent semantic analysis (described below) fuzzy c-means provided the most accurate assessment of the heterogeneity in imaging MS datasets of myxofibrosarcoma.

Bruand et al. have reported a novel approach to imaging MS analysis based on *k*-means clustering [81]. In order to avoid intense peaks in the mass spectra having excess influence during data analysis the imaging MS dataset is represented in *k*-means space: *k*-means cluster analysis is repeated for $k=1{:}k$, each pixel is then represented by a vector containing its occupancy in each of the *k* experiments. In this workflow *k*-means clustering is used as a data reduction and normalization strategy prior to subsequent data analysis.

5.3. *Factorization methods*

One of the difficulties of using PCA is that it generates pixels with 'negative' as well as 'positive' scores. The score has no direct physical meaning; it is the pixel's value on the new coordinate system (the PC). When analyzing samples containing a simple binary mixture, in which each component had a different but overlapping distribution, van der Plas reported that the dual polarity output of PCA could generate 'ghost' images that were not physically present in the sample [45].

5.3.1. *Non-negative matrix factorization (NNMF)*

The factorization method non-negative matrix factorization avoids the generation of ghost images by constraining the factors to non-zero values (both pixel scores and contribution of each variable, *m/z*, to the 'components'), and thus provide images and spectra that are more realistic and readily interpreted than PCA [9]. In NNMF the unwrapped data matrix X (n, m), where *n* is the number of pixels and *m* the number of peaks in the reduced imaging MS dataset, is approximated by two matrices, W (n-by-k) and H (k-by-m), where k is the user selected number of factors [82]. In this context W contains the images of the k-factors and H the mass spectra of the k-factors. W and H are calculated to minimize the difference between the original data matrix and the product of the two factor-matrices. Typically the Euclidian distance, $\|V-WH\|^2$, is used as the distance metric. The factorization is iterative, starting with random values for both W and H. If the difference has local minima then slight differences in W (distributions in the images) and H (the mass spectra) may be observed, though are typically small [83].

5.3.2. *Probabilistic latent semantic analysis (PLSA)*

A 'web of knowledge' search for the words 'imaging', 'mass', and 'spectrometry' results in a very large number of hits, 2,256,223, 2,428,767, and 740,321 respectively. The combination of words 'imaging mass spectrometry' results in 579. It is the combination of words that define the context and thus the topic. PLSA originates from information retrieval, specifically automated document indexing, to automatically find topics in documents by analyzing the co-occurrence of words. In this sense words are measurable variables and the topics, or combinations of words, are the latent 'semantic' variables. When applied to imaging MS datasets PLSA examines the co-occurrence of MS peaks (the variables) to identify latent variables with distinct MS profiles [84].

PLSA is a statistical mixture model to divulge the latent variables. Importantly it has a solid grounding in probability theory and results in probability distributions of the latent classes throughout the tissue and the variables that contribute to those classes. Like NNMF PLSA is an iterative process that requires the user to specify the number of latent variables. In the first report of the application of PLSA to imaging MS datasets, Hanselmann et al. demonstrated how the Akaike Information Criterion, which measures the fitness of a statistical model to the original data, can be used to calculate the number of latent variables [84].

5.3.3. *Maximum autocorrelation factor analysis*

All of the above algorithms treat each pixel's mass spectrum as independent measurements and so do not take into account any spatial relationships. Alexandrov has developed spatially-aware clustering. The multivariate technique maximum autocorrelation factor (MAF) analysis also incorporates the spatial nature of the data, and has been established longer (in the SIMS community) [76]. The logic behind MAF analysis of images is that most image features are much larger than single pixels and therefore exhibit high autocorrelation between neighboring pixels while noise exhibits low autocorrelation. The first MAF component is the linear combination of original variables that contains the maximum autocorrelation between neighboring pixels. Subsequent components are the linear combinations of the original variables that contain maximum autocorrelation subject to the constraint that they are orthogonal to the previous MAFs. In imaging MS maximization of autocorrelation between adjacent pixels should highlight regions of tissue that exhibit similar MS profiles. Tyler et al. have reported that MAF is good at detecting subtle features in the image data set [76]. MAF was recently included a comparison of multiple multivariate techniques for analyzing MALDI imaging MS data [9].

5.4. *Which multivariate method is true?*

The above text demonstrates that a number of data analysis techniques have been used to analyze imaging MS datasets. This array of techniques provides the user with a veritable data-analysis-toolbox but also raises uncertainty. The different data analysis techniques have different assumptions about the nature of the data, optimize different functions and are based on different algorithms. Consequently their results can differ in both nature (which regions of the imaging MS dataset are distinct) and in order (which output contains a specific region found to be distinct). Fig. 11 shows a comparison of PCA, NNMF, MAF, PLSA and fuzzy c-means analysis of an imaging MS dataset of intermediate grade myxofibrosarcoma. The component images were selected that show the

most similar spatial distributions to target images (which depict heterogeneity in the imaging MS dataset) [9]. In most cases the component images have similar distributions, and where the image-agreement is good the corresponding component mass spectra contain the same peptide and protein ions. It was then found that highlighting those regions of the imaging MS datasets consistently identified by most of the algorithms provided a robust and, importantly, corroborated measure of the variation within the datasets. Most imaging MS-based molecular histology reports have focused on tissues containing well differentiated histologies, enabling histological verification of the regions identified by the analysis [6,7,10,84]. For imaging MS-based molecular histology to complement established histological practice data analysis tools are required that provide additional discriminative capabilities, and so independent performance metrics are required.

5.5. *Imaging MS based molecular histology of multiple tissues*

If multiple tissue sections, mounted on to the same MALDI target plate, are analyzed within a single data acquisition sequence the imaging MS data is a list of pixel-associated-spectra but the pixel coordinates reflect the different locations of the tissues. This approach has been adapted to enable the datasets of multiple tissue sections, recorded on different days (and on different MALDI target plates), to be combined for 3D imaging [85–87], high spatial resolution imaging MS of large samples [44] and simultaneous molecular histology analysis of patient series [9]. Pixel offsets are used to place each tissue's MS data onto a common image plane (2D analysis) or sequential image planes (3D analysis). Note: The increased contrast provided by PCA (through the reduction in uncorrelated noise) has been exploited to improve the alignment of sequential tissues for 3D imaging [75] and overlapping imaging MS datasets for high spatial resolution imaging of large samples [44].

Imaging MS datasets are large and data reduction is necessary to perform a statistical analysis. This is especially true for the simultaneous statistical analysis of multiple tissues' imaging MS datasets. Broersen et al. used mass spectral binning and the selection of a specific *m/z* range (to focus on specific molecular features) as the data reduction method [44] whereas Jones et al. expanded the automated feature identification and extraction strategy [9]. The former is easier to implement but at the cost of a significant loss of mass resolution and retaining much of the non-specific

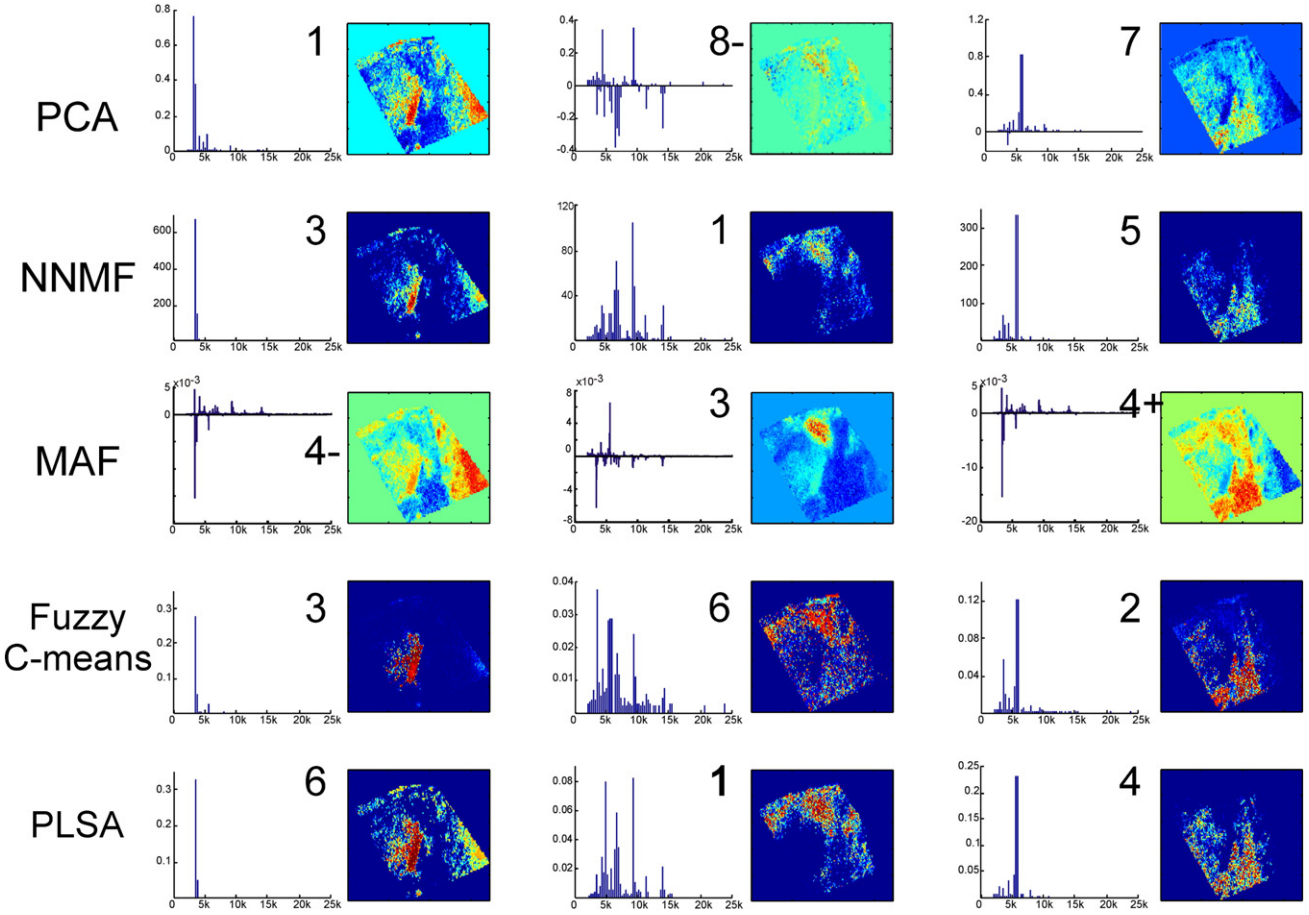

Fig. 11 – Comparison PCA, NNMF, MAF, fuzzy c-means, and PLSA ouputs that highlight similar intratumor heterogeneity in an imaging MS dataset of intermediate grade myxofibrosarcoma. The digit inserted into the mass spectra indicates the component output used. Note: PCA and MAF have negative values in their loading plots, consequently the background surrounding the tissue (defined as zero intensity) can differ in color.
Adapted from Jones et al. [9].

background. Automated feature identification and extraction was achieved by first applying the feature detection routine to each tissue's imaging MS dataset and collating the tissue-specific peak lists into a final project-specific peak list. The peaks present in the project list were then extracted from every tissue's imaging MS dataset. In this manner every MS peak, detected from any tissue, was extracted from every tissue's dataset. Fig. 12 shows a schematic of the workflow and the results of the agreement analysis (PCA, NNMF, PLSA, MAF and fuzzy c-means) simultaneously applied to 4 patient tissue samples. In this instance the 11.5 GB of raw data was reduced to a 61 MB data cube containing the images of 358 *m/z* peaks from all 4 tissues [9].

Smentkowski et al. have reported PCA of a small 3D SIMS imaging MS dataset containing 4 M pixels, 186 MB raw data load [88]. Mass spectral binning and spatial binning in the z-direction were used to reduce the data to 400 *m/z* channels and 30 sequential layers. PCA, by far the fastest multivariate technique (approx. 100 times faster than PLSA [83]) but whose results are among the least accurate [9], of the 3D imaging MS data required 2.5 minute processing time. The simultaneous application of multivariate/clustering/factorization techniques to imaging MS datasets from large patient series of tissues or 3D analyses will require increased processing speed, or improved data analysis routines, to efficiently analyze the large, highly dimensional datasets. Possible

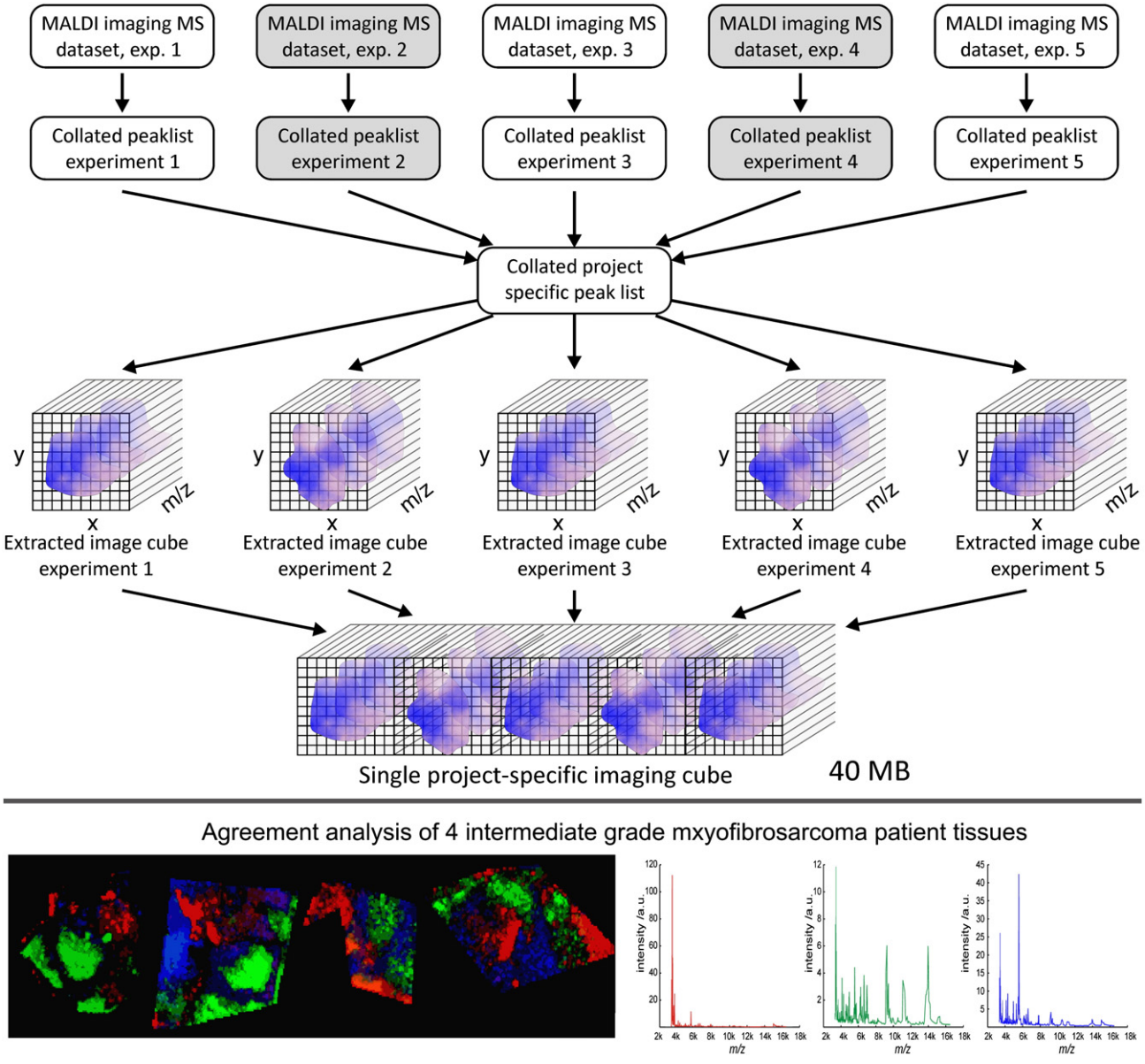

Fig. 12 – Scheme for data reduction and integration of multiple imaging MS datasets. Agreement plots reveal intratumor heterogeneity imaging MS datasets of intermediate-grade myxofibrosarcoma, a tumor known to progress through distinct genetic clones, corroborated by five different multivariate methods. Simultaneous analysis of project, consisting of the intensities of 358 peaks extracted from 31,156 pixels.
Adapted from Jones et al. [9].

solutions include general purpose computing on graphical processing units, the most cost-effective solution for super computer processing capabilities and which were recently reported to be highly suited to imaging MS datasets [83], and the reduced processing required following FastMap projection [10].

The highest spatial resolution 3D imaging MS method currently available (but which, unfortunately, is unsuited to biological analysis) is 3D atom probe tomography, which records the *m/z* and 3D spatial origin of individual ions with sub-nm spatial resolution. Keenan et al. have performed multivariate analysis of a 3D atom probe dataset of a nickel-based super alloy polished to a fine tip [89]. Ions up to *m/z* 100 were detected from a sample extending 171 nm in the axial dimension and, at its maximum lateral extent, to a diameter of approximately 144 nm. A total of more than 55 million ions were detected and the data binned to 1 nm spatial resolution. Multivariate analysis was then performed on a selected region of interest (the top 39 nm of the sample, a lateral extent of 100 nm at its base, and contained approximately 7.4 million detected ions).

6. Methods for comparing sample cohorts (biomarker discovery)

A lot of proteomics research is performed using model systems, such as cell cultures or mouse models (e.g. xenograft tumors). MALDI imaging MS is also frequently used in clinical biomarker research (in this context clinical biomarker research means biomarker discovery using human patient samples). There are important differences between studies that are based on model systems and clinical studies. It is often argued that with increasing complexity of the biological system, data analysis becomes more challenging because of the increased biological variability. In that line of thought there is a gradual increase in complexity from cell cultures to primitive organisms like *C. elegans* to mouse to humans, and that the increasing complexity demands more careful experiments and a larger number of samples (to account for the biological variation). In reality this is only partially true.

First, let us consider the assumptions that are made in the analysis of model systems. Assume we have a biological model and we want to test the influence of a particular stimulus, e.g. a stress stimulus to mice bearing specific genetic mutations. The experiment would normally consist of two identical groups; the test group (to which the stress stimulus is applied) and a control group (no stress stimulus applied). Samples of animal tissue or body fluids are then analyzed to detect protein changes associated with the stimulus.

Multiple replicates are performed to assess the random variability in the experiment. If the experiment is replicated on the biological level, those samples are considered biological replicates. There are some important implicit assumptions that are made in such an experimental set-up: it assumes that the experimental stimulus is the only factor that differentiates the two groups and all fluctuations observed within the groups are random. It may be argued that with increasing complexity of the biological system it is more challenging to control all parameters that can influence the organism's proteome, nevertheless the underlying assumptions is still that there is one dominant influence parameter (and the remainder are random). If enough replicates are measured then the random effects are expected to level out, so that the true levels of the measured proteins can be approximated by the mean measured value. In turn statistical methods based on mean values, such as fold-changes, *t*-test or ANOVA, are applicable. This means that the observed fold-changes are considered to be caused by the experimental stimulus (confidence intervals are used to differentiate "real" fold changes from random ones). The coefficients of variations of MS peaks that exhibit true changes are thus expected to be larger than those that remain constant; consequently a measureable difference between the groups would show up in the first few principal components following principal component analysis.

This situation is distinctly different when human patient samples are measured, because genetic variability and environmental influences lead to substantial proteome variability. If a clinical endpoint is to be correlated with proteomic data the variability needs to be considered. For example, if patient survival is to be correlated with a protein expression profile the data must be configured such that the clinical question has a significant influence on the sample, greater than many other individual influences (smoker vs. non-smoker, vegetarian vs. non-vegetarian, males vs. females, individual genotypes, young vs. old…). Careful patient matching is performed to reduce the influence of known sources of variability in the patient samples.

6.1. *Avoiding bias and randomization*

There are many steps in an imaging MS experiment that can influence the result. It has been shown that excessive post-mortem time, the time between animal sacrifice and snap-freezing the tissue, can lead to extensive protein degradation. This was first reported for neuropeptide profiling; without careful experiment design and tissue handling the degradation products masked the endogenous neuropeptide signals [90,91].

In imaging MS Goodwin et al. have demonstrated that the time the tissue section sits at ambient temperature after sectioning, the thawed tissue thus undergoing renewed post-mortem degradation, can have significant effects on the imaging MS datasets of peptides and proteins [92] and pharmaceuticals [93]. As discussed earlier the performance of the mass spectrometer can slowly change during the course of the experiment due to the build-up of matrix debris in the ion source (in modern instruments this is noticeable only for longer experiments). The extraction of biomolecular ions and matrix crystallization that occur during tissue preparation for MALDI imaging MS are dependent on ambient conditions such as humidity and temperature. The title of the review concerning sample preparation in this imaging MS special issue best summarizes the biases that may occur during sample preparation "Sample preparation for mass spectrometry imaging: small mistakes can lead to big consequences" [94]. Nevertheless, it should be emphasized that the increasing number of successful clinical applications clearly demonstrates that sufficiently robust experiments can be performed.

The emergence of dedicated imaging MS training courses and research networks will help ensure that effective protocols are available and can be quickly implemented [95].

A critical element for clinical and pre-clinical research is the avoidance systematic experimental bias. Bias can be introduced from unexpected sources, for example Tempst and workers have demonstrated for serum-based proteomics that sample collection, sample processing, even the lot-number of ostensibly identical commercial peptide purification media can affect the MS profiles [18]. Consequently it is easy to introduce systematic sources of bias. In an investigation using rodent models of a disease there is a risk of introducing systematic bias if the control animals are prepared and measured first, followed by the test animals. Since the biomarker discovery process searches for MS features (or patterns) that differ between the control and test samples any experimental artifacts will be reported as potential biomarkers. Sample randomization is performed such that any potential sources of bias (known and unknown) affect the test and control sample cohorts equally and randomly, so that observed differences can be considered real.

6.2. *Univariate comparisons*

Univariate tests are used to compare spectra from sample cohorts peak-by-peak in order to find specific biomarkers. In most cases univariate tests are applied as part of a histology-defined analysis, to compare comparable histological features, e.g. areas containing a high density of tumor cells to find peaks specific to the tumor (comparison with non-tumor tissue [96]) or to find tumor specific proteins associated with tumor status (e.g. HER2 positive vs. HER2 negative [52]). For each patient tissue (test and control) a histological analysis is performed to select regions-of-interest. The spectra of these histologically-defined regions (also sometimes referred to as "virtual microdissection") are then extracted. A mean average of each patient's extracted spectra is then used for the univariate comparisons.

6.2.1. *What is a good biomarker: fold-change and ROC, Gaussian and non-Gaussian distributions*

Several metrics are used in the search for biomarkers, but their applicability depends on the nature of the underlying data. One of the most common is the fold-change. The calculation of the fold change is simple, but the underlying assumptions are less straightforward: The fold-change is simply the average intensity of the selected MS signal in the test groups relative to the control group.

It is important to be aware of the assumptions underlying a simple fold-change comparison (and any comparison based on mean intensities). If the data has a Gaussian distribution then the data is scattered symmetrically around a central value that is equal to the arithmetic mean. In this case the mean is expected to be equal to the "real" MS response and a comparison of mean intensities is valid. A fold-change comparison can usually be applied to experiments using model systems involving biological replicates.

Gaussian distributions of mean intensities cannot be assumed for clinical samples, and so the arithmetic mean would be less specific (for example of a bimodal or more complex distribution). Nevertheless mean values may still be used if the central limit theorem is satisfied. The theorem states that variables that are influenced by many different factors will show a Gaussian distribution, even if individual factors have a non-Gaussian distribution. For this reason complex variables such as body height usually exhibit a kind of normal distribution. The same can be true for protein signals, if their level is influenced my many individual factors. The central limit theorem also applies to mean values of non-Gaussian distributions if the number of samples is sufficiently large. In this case the statement would be that the mean value can be accurately estimated, even if the distribution is not Gaussian. The number of samples that will be necessary may depend on the characteristics of the dataset, but around 30 samples per group are considered the minimum.

The question of what a good biomarker means needs to be considered. Statistical hypothesis tests can be used to establish if there is a significant difference between two groups (described below). A *t*-test for instance ascertains if the mean value between the two groups is different. It may be the case that while the mean values differ the distributions of the values overlap significantly. Depending on the analytical question this may be an important finding, e.g. if a physiological response in the comparison of two differently treated cell cultures can be observed. However most clinical biomarkers are sought that can correctly classify the samples, e.g. tumor vs. non-tumor. The question is not if there is a difference between the mean values of the two groups but whether or not a classification is possible; and so the important factor is the overlap of the distributions of intensities from both groups. The fold-change or the results provided by statistical hypothesis tests on mean values are inadequate for this task, the correct tool to test the overlap of two distributions is the receiver-operating characteristic (ROC). For this test one group is labeled "positive" (the test samples) and the other is labeled "negative" (the control samples). A discrimination threshold is then varied and the "true positive rate" (TPR) and the "true negative rate" (TNR) are calculated for each threshold value. The ROC curve plots the TPR versus (1-TNR). Fig. 13 shows the data, histograms and ROC curves for three different model datasets that exhibit no separation, moderate separation and high separation. The predictive value of the biomarker is provided by the area under the ROC curve. A value of 1 indicates a perfect separation of the two distributions (a perfect biomarker), a value of 0.5 indicates no separation at all. The supplementary video, roc-curve.wmv, provides a detailed description of how the ROC curve is constructed.

Note: in statistics the true positive rate is usually referred to as the "sensitivity" of a test and the "true negative rate" as the "specificity". These terms also have defined meanings in mass spectrometry, referring to the ability to measure a molecule (the sensitivity of MS detection refers to the amount that can be measured and the specificity refers to the uniqueness of the MS peak assignment). For instance it could be possible to have a biomarker with high (statistical) sensitivity and moderate (statistical) specificity, and high (mass spectral measurement) sensitivity but undermined by poor (mass spectral measurement) specificity. The scope for

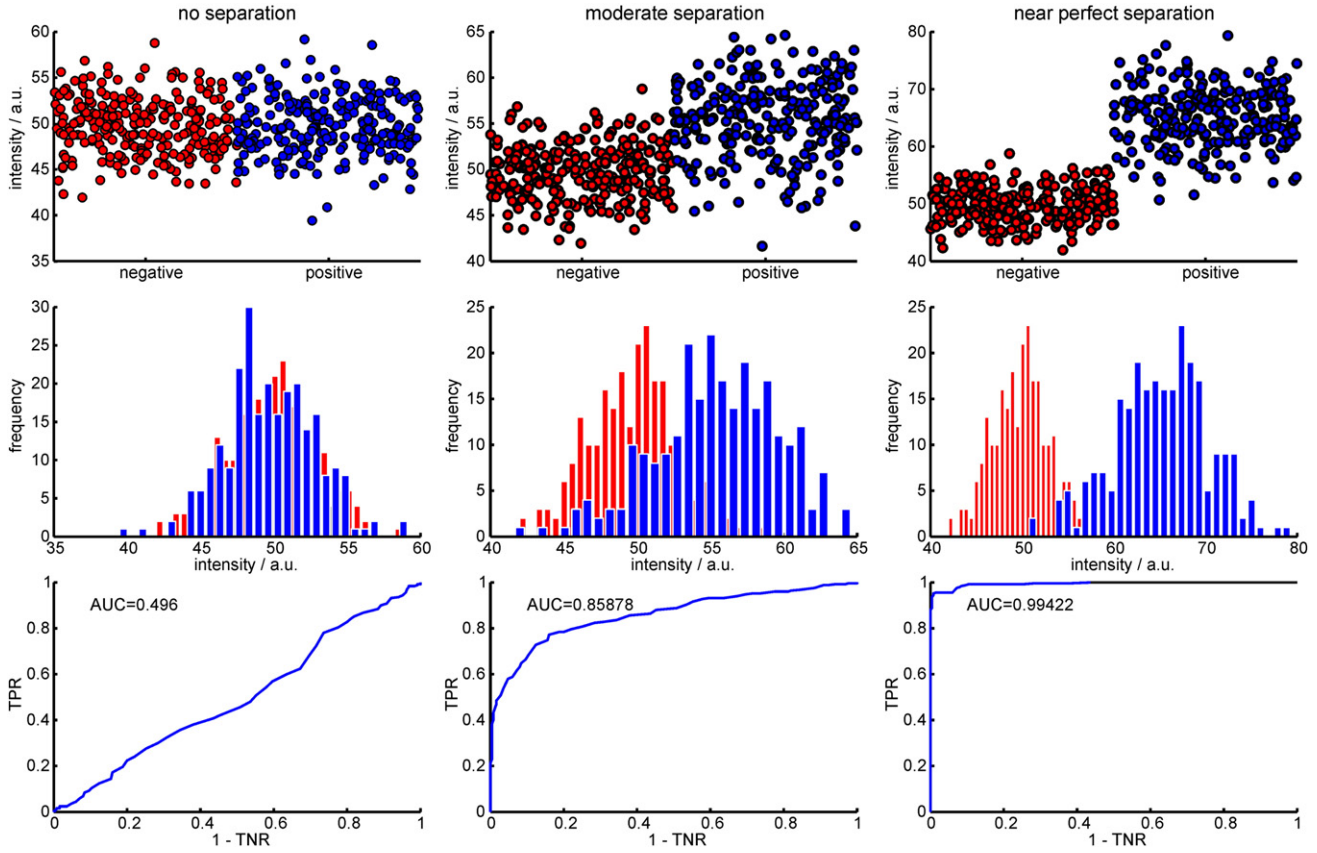

Fig. 13 – Model 'biomarker' discovery datasets in which the candidate biomarker displays no separation, moderate separation and a high degree of separation between the positive (diseased) and negative (control) classes. The top row shows the original data as a scatter plot (x-ordinate randomly dispersed to aid visualization), the middle row shows histograms of the data to show the distribution in intensities, and the final row the ROC curves. The supplementary video provides a detailed overview of how the ROC curves are constructed.

misunderstanding is clear and so contextual clarity is necessary and recommended.

6.2.2. *Statistical tests and p-values*

Statistical tests are used to assess the likelihood that the sample groups being compared are different. The hypothesis we wish to test is that the MS signal is different in the test sample and control sample cohorts. This hypothesis cannot be directly tested, as we do not know if the MS signals are indeed different. Instead, we test the so-called "null-hypothesis", that there is no difference between the two groups. The result of such tests is a p-value, which is a measure of the probability that the observed result is caused by chance. A p-value below 0.05 is often considered statistically significant (corresponding to a false positive rate of 5%).

The hypothesis to be tested should be formulated in advance, ideally before the data has been acquired but at least before the test is applied. A valid hypothesis to be tested in an imaging dataset would be "the insulin levels of normal and obese mice in the islets of Langerhans in the pancreas are different". This poses a common problem in discovery-type applications in which a large number of MS are detected but it is not yet known which are different. Univariate statistical tests are individually applied to each MS signal. If there are 100 MS peaks in the imaging MS datasets the common 0.05 p-value would result in 5 false positive MS signals, just by chance. This is the so-called multiple-test problem, and it requires that the p-values are corrected, for example with the Bonferroni correction [97–99]. Nevertheless as the initial analysis did not begin with a fixed hypothesis, the application of these tests can only generate a hypothesis that "peak *x* is detected at different levels in two groups", so-called biomarker discovery. An independent dataset is necessary for validation.

For all of the statistical tests discussed below (with the exception of nested ANOVA) there is one important requirement, which is that samples (MS measurements) are independent. An imaging MS dataset may contain many thousands of pixels, each with an associated mass spectrum, and so even specific histological entities are described by hundreds of mass spectra. However, mass spectra from a single tissue sample have a common origin and cannot be considered independent. Normally the individual pixel mass spectra are considered replicates of the *same tissue state*. When applying the statistical tests it is important to compare just one value per sample, and not the individual pixels as this would violate the assumption of independent samples. Accordingly the average mass spectrum, of the pixels selected from one tissue, is used for the univariate analysis.

Note: if different tissue samples are placed on the same slide and prepared together they are no longer entirely independent. For practical purposes it may not be suitable to prepare a different target for every tissue sample. In this case extreme care should be taken to guard against the introduction of systematic bias (e.g. test tissues always measured after the control tissues).

6.2.3. *Parametric tests (t-test/one-way ANOVA)*

The *t*-test is probably the most commonly known statistical test, but the underlying assumptions are quite restrictive and often times not applicable to clinical data. The *t*-test tests for the difference in the mean of two distributions, so all limitations discussed in the previous section with regard to the fold-change also apply to the *t*-test.

ANOVA in its simplest form (the so-called "one-way ANOVA") is the equivalent to the *t*-test if more than two classes are to be compared. The null hypothesis tested is that "the mean intensities between all groups are equal", so a significant p-value would indicate the opposite, that at least two groups differ. The one way ANOVA will not give the answer which groups are different, if it is just two or all of them. To make this determination it is necessary to apply the appropriate post-hoc test or to do pairwise t-tests.

There are many different ANOVA designs that are specific to certain statistical problems. ANOVA tests rely on the assumption that each measurement is an independent sample. If we consider pixels from histologically-identical cell types in an imaging MS dataset as technical replicates from the same biological sample they are not independent (and consequently give similar measurement). An alternative test that would be applicable is the so-called nested ANOVA [100], but to the best knowledge of the authors this has not yet been applied to MALDI imaging data.

6.2.4. *Nonparametric tests*

There are nonparametric tests that can be applied to the data if the distribution is non-Gaussian. The most commonly used is the Mann–Whitney U-test (also referred to as Wilcoxon Rank sum tests or simply U-test). This test is based on the rank sum of the observed data: All observed values are sorted, and then numbered. The lowest intensity gets rank one, the second lowest intensity gets rank two and so on. The ranks for the two groups are then summed; the test checks the likelihood that the difference in the sum of the ranks is caused by chance. For this test there are also assumptions, but they are not as strict as those for parametric tests, namely that the distributions should have similar shapes that are only shifted and the samples should be independent. The Kruskal–Wallis test is a similar test that can be applied to multiple classes (and the interpretation of the p-values is similar to one-way ANOVA as discussed above).

6.2.5. *Additional considerations for hypothesis testing, paired and unpaired comparisons*

Consider a series of imaging MS datasets of hepatic carcinoma tissue, in which the tumor is surrounded by cirrhotic tissue. If the goal of the analysis was to identify candidate biomarkers that differentiate hepatic carcinoma from the surrounding cirrhosis then, as the two tissue states are present in every patient sample they cannot be considered truly independent. In this case a paired test should be used, for example a paired *t*-test (parametric), the Friedman or the Wilcoxon signed rank test (both non-parametric).

6.3. *Multivariate comparisons and classifications*

A single MS signal in the imaging MS datasets may not provide the statistical sensitivity and specificity to be an effective biomarker. Multivariate methods are used to improve the discriminative power by searching for new metrics involving multiple MS peaks.

6.3.1. *PCA and hierarchical clustering for comparison of different datasets*

As described earlier PCA is used to reduce the dimensionality of the data by creating new variables which are linear combinations of the original variables (the peaks in the mass spectrum, re. Fig. 9). The algorithm creates new orthogonal variables that describe the largest variance in the data (PC1), the next largest variance (PC2), and so on. Those influences that lead to large variances in the data are kept, and so PCA can be a highly effective tool for investigating which factors have the largest influence on the data. For example, when comparing different patient samples it is quite straightforward to asses if the analysis time, time-of-sampling, or tissue-storage conditions (to name but a few) are highlighted in the score plots of the first principal components. If highlighted by the first principal components then differences in these experimental steps have significantly influenced the data.

Note. As has been described above, only in experiments that compare highly controlled and mainly monogenetic animals may it *be expected* that the difference between the test and control groups will be one of the largest influences in the data. When comparing human (clinical) samples, there is the possibility that inter-patient variability is larger than the inter-group variability.

Hierarchical clustering can also be used to investigate which factors have the largest effect on the data. As with PCA, hierarchical clustering will highlight the largest influences in the data. For human tissues, in which inter-patient variability may otherwise dominate it is common practice to only include the MS peaks in the clustering analysis that show a good separation of the cohorts (as determined by a univariate test). Used in this way the hierarchical clustering algorithm is still unsupervised, but the analysis has been biased by peak selection. It is used to provide an indication if there are real differences between the cohorts that are hidden in the larger inter-patient variations.

6.4. *Classification methods*

6.4.1. *Discriminant analysis (DA and PCA-DA)*

Discriminant analysis is related to PCA but takes the group membership of the data into account. In the case of MS data it attempts to find those linear combinations of MS signals that discriminate between the groups. The resulting discriminant function can then be used to classify new patient samples. DA is often used after a preprocessing step with PCA. In this case

the PCA first reduces the dimensionality of the data (from several hundred MS peaks to a small number of principal components), and the DA is then calculated using the principal components.

Implicit in this approach is the assumption that the differences between the groups should be one of the principal sources of variation in the data. Again, this may be expected for model systems but not necessarily in clinical analyses. There may be real differences between two clinical groups, but which are not among the largest influences in the data and so may be lost during PCA-based data reduction (and so not present for the discriminant analysis). Nevertheless, PCA-DA has been successfully used to classify human cancer samples, which is an indication that the differences found by MALDI imaging can be quite robust [101].

6.4.2. *Other classification algorithms*

There are other classification algorithms that have been used for imaging mass spectrometry data, among them neuronal networks [52], support vector machine [4,51], random forest [7], and genetic algorithm [4]. The recent paper by Walch and workers describing how protein imaging MS can be used to distinguish between six common tumors, as well as identify correlations with tumor depth and grading, provides an excellent example of how classifiers can be applied to imaging MS datasets [102].

It is beyond the scope of this article to give a detailed description on these multivariate algorithms and the interested reader is referred to the excellent review of Hilario et al. [103], here we give some common characteristics. These classification algorithms try to build a statistical model that separates the training cohort. In contrast to a DA the result is not a "simple" function of the form $c_1 * \text{intensity}(1) + c_2 * \text{intensity}(2) \ldots c_n * \text{intensity}(n)$, but the data are fitted by more complex means. A very good explanation of how a random forest classifier works is provided by Hanselmann et al. [7]. It shows how the model is built to match the characteristics of the data, and so these classifiers can find small differences between the cohorts that other approaches may miss, but are also more vulnerable to over-fitting the data (the random differences may be sufficient to generate a classifier). This has an important implication for such classification analysis, namely it is imperative to validate the classifier on an independent sample cohort.

Several papers have used classification algorithms to generate segmentation maps, in which each pixel is assigned to a class (or remains unassigned) and results presented as a false color image [4,51]. This may evolve into a powerful clinical application of imaging MS, it is important though to be aware of one important limitation; the classifier is strictly limited in scope to the classification originally trained (question and dataset). If a model is generated to classify hepatic carcinoma imaging MS data from liver cirrhosis imaging MS data, both obtained using a vacuum UV MALDI-ToF mass spectrometer (linear mass analyzer), it is only applicable to this exact differentiation. If a different tissue type is present in the data (such as a large blood vessel), then the classifier will attempt to classify this additional tissue type. Equally, if changing or upgrading the MS methodology (tissue acquisition, tissue preparation, data acquisition, mass spectral processing), the classification algorithm may need to be re-established. In short human expertise may still be required to judge the applicability and effectiveness of the model.

6.4.3. *Multivariate analysis vs. univariate markers*

A counter-intuitive result is that the multivariate classification models do not necessarily highlight effective univariate markers, or even the best univariate candidate biomarkers that are present in the dataset. The explanation is essentially that multivariate methods seek relationships between variables, which may or may not include the best univariate discriminators. This has been demonstrated with a small simulated dataset that contains 30 samples (15 samples from class A, 15 samples from class B) with 6 peaks, Fig. 14.

It can be seen that peak 1 is the best univariate biomarker, it has the lowest p-values for both the *t*-test and Wilcoxon test and it shows the best univariate separation between the two groups in the area under the ROC curve analysis. Indeed it is visually apparent that this peak displays the largest differences between the two classes; it might even appear as if peak 1 is the only potentially useful biomarker. It might be very surprising then that two classification models can be calculated, which provide a 100% recognition rate for the two groups, but neither of which contain peak 1. The first model contains only peaks 2 and 3, while the other model contains only peaks 4, 5, and 6. From a multivariate perspective such models are perfectly reasonable.

Fig. 15 shows that though peaks 2 and 3 are not good univariate biomarkers, their relationship (their relative intensity) allows a perfect separation between the two classes. Multivariate analysis is concerned with the relationship between variables (MS peaks) and so biomarkers may consist of simple relationships (as shown in Fig. 15) or more complex relationships. It is also noteworthy that the separation is better than that which may be achieved using only peak 1. This is the kind of multivariate classification that is provided by a support vector machine and the quick classifier algorithm.

Fig. 16 shows the intensities of MS peaks 4, 5 and 6 for each individual sample (in imaging MS this would correspond to individual pixels or an average of pixels). Close examination reveals that a perfect classification may be obtained if, for any given sample, any of the three peaks is above a marked threshold. This is the type of multivariate classification generated by a genetic algorithm or a supervised neuronal network. The interested reader is referred to the paper by Hilario et al. for a more extensive discussion about classification [103].

An extrapolation of the simulated results shown in Fig. 16 leads to one of the dangers associated with highly dimensional datasets, namely as more variables (MS peaks) are included in the dataset it may become possible to build perfect classification models based on the random fluctuations of the data, i.e. it becomes possible to get a 100% recognition rate even though there is no real difference between the classes. Some indications for such over-fitting are a large number of MS peaks in the models and very high recognition and cross-validation values. The only way to evaluate if over-fitting has occurred is to validate the model on an independent sample set, and which is why independent validation is essential for multivariate classification analyses.

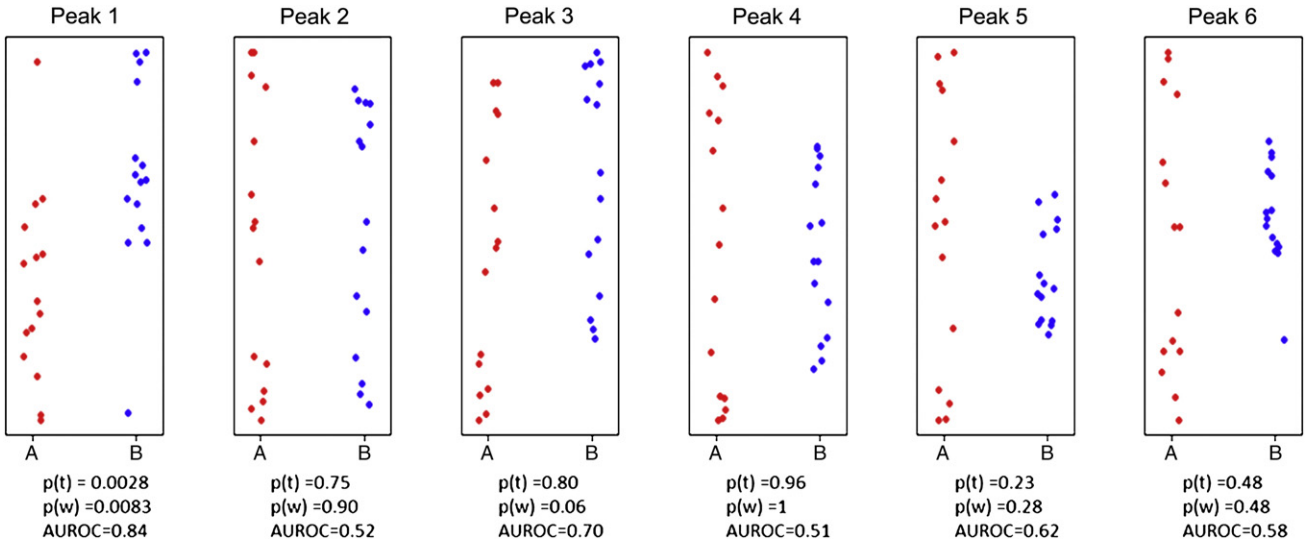

Fig. 14 – Scatter plots for the simulated dataset of six MS peaks. Class A is shown in red and class B is shown in blue, the y-axis shows the intensity of the peaks. The p-value of a *t*-test, p(t), the p-value of the Wilcoxon test, p(w), and the area under the receiver–operator-curve (AUROC) are provided for each variable below their scatter plot.
Reproduced with kind permission of Bruker Daltonics.

In a multivariate classification the entire model is the biomarker, not the individual MS peaks. If the aim of the study is to find good univariate biomarkers (that is peaks with low p-values in statistical tests or a high AUROC) then univariate tests (*t*-test, Wilcoxon–Kruskal–Wallis tests) have to be the statistical method of choice.

7. Concluding remarks

Data analysis is an essential element of an imaging mass spectrometry experiment. In clinical investigations it extracts from the highly complex imaging MS datasets the mass spectral features that are consistently associated with a pathology, whether for disease classification or biomarker discovery. Despite the different analysis strategies that have been developed, MS or histology centric (reflecting the different backgrounds of the practitioners), a consistent paradigm is emerging concerning where imaging MS is having an impact. There is ample and growing evidence that imaging MS can complement established histological and histochemical methods and aid patient stratification. For example imaging MS has been used to identify candidate biomarkers of HER2 status in breast cancer [52] and gastric cancer [50]; candidate biomarkers that discriminate between rare morphologically overlapping/identical tumors [61] or between rare morphologically identical benign/cancerous lesions [104] that can be a challenge for the pathologist; and when combined with patient outcome candidate biomarkers have been found for patient prognosis [5] or response to treatment [54].

The use of imaging MS to help identify new biomarkers that discriminate between known groups within well-defined morphological groups complements established anatomical pathology, and is well-matched with the strengths and

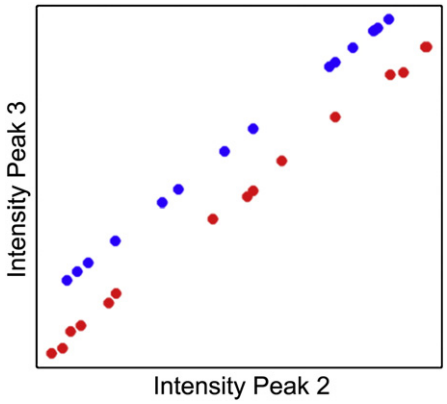

Fig. 15 – Plot of the intensities of peak 2 versus peak 3, measurements from class A are shown in red, from class B in blue.
Reproduced with kind permission of Bruker Daltonics.

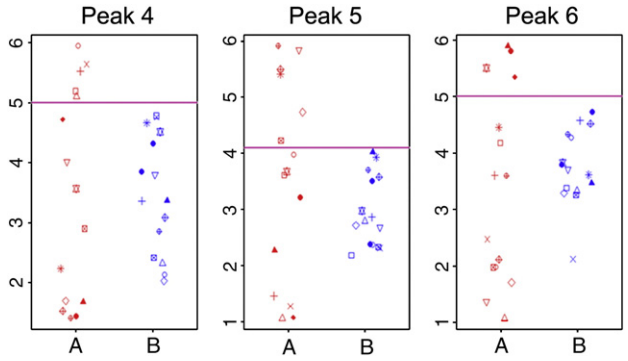

Fig. 16 – The individual samples (MS measurements) of the two groups are indicated by different symbols for peaks 4, 5 and 6.
Reproduced with kind permission of Bruker Daltonics.

limitations of the technique. Each pixel analyzed in an imaging MS experiment contains a very small number of cells; assuming an average cell size of 20 μm a 50×50 μm pixel contains approximately six cells, far smaller than that analyzed in typical proteomics/metabolomics workflows (leading examples based on laser capture microdissection for cellular specificity use cell populations of several thousand [105,106]). Considering the small sample volume analyzed in each pixel, and the complex chemical environment of the tissue sample, it is a testament to the technique that imaging MS is able to detect 100s–1000s of peptides/proteins. Nevertheless, these factors will mean that imaging MS experiments are limited to the more abundant components [74].

The ability to generate rich MS profiles from such small sample volumes, throughout clinical tissue samples, enables imaging MS experiments to be integrated with histology and thereby extract MS profiles specific to particular cell types (within clinical specimens). Patient stratification within well-defined morphological groups is a less ambitious biomarker discovery experiment that is well-matched to the dynamic range of the imaging MS experiment. The examples referred to above clearly indicate that the cell-specific MS profiles may contain biomarkers that subgroup tumors, and thereby may lead to further improvements in patient diagnosis.

In late 2009 the first imaging MS data analysis course, entitled 'Processing & Validation Tools for Imaging Mass Spectrometry' and funded by Nordforsk, was held in Turku, Finland. It became apparent that there is a need in imaging MS research for in-depth training in the proper use and understanding of the statistical analysis techniques that underpin its clinical application. The prime motivation for writing this review was to begin to address this need.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2012.06.014.

Acknowledgments

This work is financially supported by NWO Horizon project 93511027 and the ICT consortium COMMIT project 'e-biobanking with Imaging' and the Cyttron II project 'Imaging Mass Spectrometry.' The authors also acknowledge COST Action BM1104, entitled Mass Spectrometry Imaging: New Tools for Healthcare Research, for enabling the international cooperation of this manuscript.

REFERENCES

[1] Chaurand P, Sanders ME, Jensen RA, Caprioli RM. Proteomics in diagnostic pathology — profiling and imaging proteins directly in tissue sections. Am J Pathol 2004;165:1057–68.
[2] McDonnell LA, Heeren RMA. Imaging mass spectrometry. Mass Spectrom Rev 2007;26:606–43.
[3] Schwamborn K, Caprioli RM. Molecular imaging by mass spectrometry — looking beyond classical histology. Nat Rev Cancer 2010;10:639–46.
[4] Schwamborn K, Krieg RC, Reska M, Jakse G, Knuechel R, Wellmann A. Identifying prostate carcinoma by MALDI-imaging. Int J Mol Med 2007;20:155–9.
[5] Balluff B, Rauser S, Meding S, Elsner M, Schoene C, Feuchtinger A, et al. MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer. Am J Pathol 2011;179:2720–9.
[6] Alexandrov T, Becker M, Deininger S-O, Grasmair G, von Eggeling F, Thiele H, et al. Spatial segmentation of imaging mass spectrometry with edge preserving image denoising and clustering. J Proteome Res 2010;9:6535–46.
[7] Hanselmann M, Kothe U, Kirchner M, Renard BY, Amstalden ER, Glunde K, et al. Toward digital staining using imaging mass spectrometry and random forests. J Proteome Res 2009;8:3558–67.
[8] McDonnell LA, van Remoortere A, van Zeijl RJM, Deelder AM. Mass spectrometry image correlation: quantifying co-localization. J Proteome Res 2008;7:3619–27.
[9] Jones EA, van Remoortere A, van Zeijl RJM, Hogendoorn PCW, Boveé JVMG, Deelder AM, et al. Multiple statistical analysis techniques corroborate intratumor heterogeneity in imaging mass spectrometry datasets of myxofibrosarcoma. PLoS One 2011;6:e24913.
[10] Alexandrov T, Kobarg JH. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. Bioinformatics 2011;27:i230–8.
[11] Norris JL, Cornett DS, Mobley JA, Andersson M, Seeley EH, Chaurand P, et al. Processing MALDI mass spectra to improve mass spectral direct tissue analysis. Int J Mass Spectrom 2007;260:212–21.
[12] Guilliot S, Chaurand P, Stoeckli M, Caprioli RM. Topographic imaging using time-of-flight mass spectrometry. 47th ASMS Conference. Dallas; 1999.
[13] McCombie G, Staab D, Stoeckli M, Knochenmuss R. Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. Anal Chem 2005;77:6118–24.
[14] McDonnell LA, Luxembourg SL, Mize TH, Koster S, Eijkel GB, Verpoorte E, et al. Using matrix peaks to map topography: increased mass resolution and enhanced sensitivity in chemical imaging. Anal Chem 2003;75:4373–81.
[15] McDonnell LA, Piersma SR, Altelaar AFM, Mize TH, Luxembourg SL, Verhaert PDEM, et al. Subcellular imaging mass spectrometry of brain tissue. J Mass Spectrom 2005;40:160–8.
[16] Tracy MB, Chen H, Weaver DM, Malyarenko DI, Sasinowski M, Cazares LH, et al. Precision enhancement of MALDI-TOF MS using high resolution peak detection and label-free alignment. Proteomics 2008;8:1530–8.
[17] Roempp A, Guenther S, Schober Y, Schulz O, Takats Z, Kummer W, et al. Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging. Angew Chem Int Ed 2010;49:3834–8.
[18] Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L, et al. Correcting common errors in identifying cancer-specific serum peptide signatures. J Proteome Res 2005;4:1060–72.
[19] Yang J, Caprioli RM. Matrix sublimation/recrystallization for imaging proteins by mass spectrometry at high spatial resolution. Anal Chem 2011;83:5728–34.
[20] Lagarrigue M, Becker M, Lavigne R, Deininger S-O, Walch A, Aubry F, et al. Revisiting rat spermatogenesis with MALDI imaging at 20 μm resolution. Mol Cell Proteomics 2011;10 [M110.005991].
[21] Römpp A, Guenther S, Schober Y, Schultz O, Takats Z, Kummer W, et al. Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging. Angew Chem Int Ed 2010;49:3834–8.
[22] Luxembourg SL, Mize TH, McDonnell LA, Heeren RMA. High-spatial resolution mass spectrometric imaging of peptide and protein distributions on a surface. Anal Chem 2004;76:5339–44.

[23] Hopfgartner G, Varesio E, Stoeckli M. Matrix-assisted laser desorption/ionization mass spectrometric imaging of complete rat sections using a triple quadrupole linear ion trap. Rapid Commun Mass Spectrom 2009;23:733–6.
[24] Spraggins JM, Caprioli RM. High-speed MALDI-TOF imaging mass spectrometry: rapid ion image acquisition and considerations for next generation instrumentation. J Am Soc Mass Spectrom 2011;22:1022–31.
[25] Wickes BT, Kim Y, Castner DG. Denoising and multivariate analysis of time-of-flight SIMS images. Surf Interface Anal 2003;35:640–8.
[26] Taban IM, Altelaar AFM, Fuchser J, van der Burgt YEM, McDonnell LA, Baykut G, et al. Imaging of peptides in the rat brain using MALDI-FTICR mass spectrometry. J Am Soc Mass Spectrom 2006;18:145–51.
[27] McDonnell LA, van Remoortere A, de Velde N, van Zeijl RJM, Deelder A. Imaging mass spectrometry data reduction: automated feature identification and extraction. J Am Soc Mass Spectrom 2010;21:1969–78.
[28] Schober Y, Schramm T, Spengler B, Römpp A. Protein identification by accurate mass matrix-assisted laser desorption/ionization imaging of tryptic peptides. Rapid Commun Mass Spectrom 2011;25:2475–83.
[29] Smith RD. Evolution of ESI-mass spectrometry and Fourier transform ion cyclotron resonance for proteomics and other biological applications. Int J Mass Spectrom 2000;200: 509–44.
[30] Shin H, Mutlu M, Koomen JM, Markey MK. Parametric power spectral density analysis of noise from instrumentation in MALDI ToF mass spectrometry. Cancer Inf 2007;3:219–30.
[31] Bouschen W, Spengler B. Artifacts of MALDI sample preparation investigated by high-resolution scanning microprobe matrix-assisted laser desorption/ionization (SMALDI) imaging mass spectrometry. Int J Mass Spectrom 2007;266:129–37.
[32] Luxembourg SL, McDonnell LA, Duursma M, Guo X, Heeren RMA. Effect of local matrix crystal variations in matrix-assisted ionization techniques for mass spectrometry. Anal Chem 2003;75:2333–41.
[33] Du P, Stolovitzky G, Horvatovich P, Bischoff R, Lim J, Suits F. A noise model for mass spectrometry based proteomics. Bioinformatics 2008;24:1070–7.
[34] Keenan MR, Kotula PG. Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. Surf Interface Anal 2004;36:203–12.
[35] Fonville JM, Carter C, Cloarec O, Nicholson JK, Lindon JC, Bunch J, et al. Robust data processing and normalization strategy for MALDI mass spectrometric imaging. Anal Chem 2012;84:1310–9.
[36] Lemaire R, Wisztorski M, Desmons A, Tabet JC, Day R, Salzet M, et al. MALDI-MS direct tissue analysis of proteins: improving signal sensitivity using organic treatments. Anal Chem 2006;78:7145–53.
[37] Seeley EH, Oppenheimer SR, Mi D, Chaurand P, Caprioli RM. Enhancement of protein sensitivity for MALDI imaging mass spectrometry after chemical treatment of tissue sections. J Am Soc Mass Spectrom 2008;19:1069–77.
[38] Cohen S, Chait BT. Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. Anal Chem 1996;68:31–7.
[39] Domon B, Aebersold R. Options and considerations when selecting a quantitative proteomics strategy. Nat Biotechnol 2010;28:710–21.
[40] Stoeckli M, Staab D, Schweitzer A. Compound and metabolite distribution measured by MALDI mass spectrometric imaging in whole-body tissue sections. Int J Mass Spectrom 2007;260:195–202.
[41] Nilsson A, Fehniger TE, Gustavsson L, Andersson M, Kenne K, Marko-Varga G, et al. Fine mapping the spatial distribution and concentration of unlabeled drugs within tissue micro-compartments using imaging mass spectrometry. PLoS One 2010;5:e11411.
[42] Deininger SO, Cornett DS, Paape R, Becker M, Pineau C, Rauser S, et al. Normalization in MALDI-TOF imaging datasets of proteins: practical considerations. Anal Bioanal Chem 2011;401:167–81.
[43] van Remoortere A, van Zeijl R, van den Oever N, Franck J, Longuespée R, Wisztorski M, et al. MALDI imaging and profiling MS of higher mass proteins from tissue. J Am Soc Mass Spectrom 2010;21:1922–9.
[44] Broersen A, van Liere R, Altelaar AFM, Heeren RMA, McDonnell LA. Automated, feature-based image alignment for high-resolution imaging mass spectrometry of large biological samples. J Am Soc Mass Spectrom 2008;19:823–32.
[45] van de Plas R. Tissue based proteomics and biomarker discovery: multivariate data mining strategies for mass spectral imaging. Leuven: Katholieke Universiteit Leuven; 2010.
[46] van de Plas R, de Moor B, Waelkens E. Discrete wavelet transform-based multivariate exploration of tissue via imaging mass spectrometry. Proceedings of the 23rd annual ACM symposium on applied computing. Fortaleza, Brazil; 2008. p. 1307–8.
[47] Vogt F, Banerji S, Booksh K. Utilizing three-dimensional wavelet transforms for accelerated evaluation of hyperspectral image cubes. J Chemom 2004;18:350–62.
[48] Mantini D, Petrucci F, Pieragostino D, Del Boccio P, Di Nicola M, Di Ilio C, et al. LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. BMC Bioinformatics 2007;8:101.
[49] Jardin-Mathe O, Bonnel D, Franck J, Wisztorski M, Macagno E, Fournier I, et al. MITICS (MALDI Imaging Team Imaging Computing System): a new open source mass spectrometry imaging software. J Proteomics 2008;71:332–45.
[50] Balluff B, Elsner M, Kowarsch A, Rauser S, Meding S, Schumacher C, et al. Classification of HER2/neu status in gastric cancer using a breast-cancer derived proteome classifier. J Proteome Res 2010;9:6317–22.
[51] Groseclose MR, Massion PP, Chaurand P, Caprioli RM. High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. Proteomics 2008;8:3715–24.
[52] Rauser S, Marquardt C, Balluff B, Deininger SO, Albers C, Belau E, et al. Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. J Proteome Res 2010;9:1854–63.
[53] Hardesty WM, Kelley MC, Mi D, Low RL, Caprioli RM. Protein signatures for survival and recurrence in metastatic melanoma. J Proteomics 2011;74:1002–14.
[54] Reyzer ML, Caldwell RL, Dugger TC, Forbes JT, Ritter CA, Guix M, et al. Early changes in protein expression detected by mass spectrometry predict tumor response to molecular therapeutics. Cancer Res 2004;64:9093–100.
[55] Cornett DS, Mobley JA, Dias EC, Andersson M, Arteaga CL, Sanders ME, et al. A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. Mol Cell Proteomics 2006;5:1975–83.
[56] Bruand J, Sistla S, Mériaux Cl, Dorrestein PC, Gaasterland T, Ghassemian M, et al. Automated querying and identification of novel peptides using MALDI mass spectrometric imaging. J Proteome Res 2011;10:1915–28.
[57] Caldwell RL, Gonzalez A, Oppenheimer SR, Schwartz HS, Caprioli RM. Assessment of the tumor protein microenvironment using imaging mass spectrometry. Cancer Genomics Proteomics 2006;3:279–88.
[58] Oppenheimer SR, Mi D, Sanders ME, Caprioli RM. Molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma. J Proteome Res 2010;9:2182–90.

[59] Kang S, Shim HS, Lee JS, Kim DS, Kim HY, Hong SH, et al. Molecular proteomics imaging of tumor interfaces by mass spectrometry. J Proteome Res 2010;9:1157–64.
[60] Amstalden van Hove ER, Blackwell TR, Klinkert I, Eijkel GB, Heeren RMA, Glunde K. Multimodal mass spectrometric imaging of small molecules reveals distinct spatio-molecular signatures in differentially metastatic breast tumor models. Cancer Res 2010;70:9012–21.
[61] Willems SM, van Remoortere A, van Zeijl R, Deelder AM, McDonnell LA, Hogendoorn PCW. Imaging mass spectrometry of myxoid sarcomas identifies proteins and lipids specific to tumor type and grade, and reveals biochemical intratumor heterogeneity. J Pathol 2010;222: 400–9.
[62] Deininger S-O, Ebert MP, Fütterer A, Gerhard M, Röcken C. MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. J Proteome Res 2008;7:5230–6.
[63] Burnum KE, Cornett DS, Puolitaival SM, Milne SB, Myers DS, Tranguch S, et al. Spatial and temporal alterations of phospholipids determined by mass spectrometry during mouse embryo implantation. J Lipid Res 2009;50:2290–8.
[64] Urban PL, Chang C-H, Wu J-T, Chen Y-C. Microscale MALDI imaging of outer-layer lipids in intact egg chambers from *Drosophila melanogaster*. Anal Chem 2011;83:3918–25.
[65] Mathur BN, Caprioli RM, Deutch AY. Proteomic analysis illuminates a novel structural definition of the claustrum and insula. Cereb Cortex 2009;19:2372–9.
[66] Caldwell RL, Opalenik SR, Davidson JM, Caprioli RM, Nanney LB. Tissue profiling MALDI mass spectrometry reveals prominent calcium-binding proteins in the proteome of regenerative MRL mouse wounds. Wound Repair Regen 2008;16:442–9.
[67] Meriaux C, Arafah K, Tasiemski A, Wisztorski M, Bruand J, Wichlacz-Boidin C, et al. Multiple changes in peptide and lipid expression associated with regeneration in the nervous system of the medicinal leech. PLoS One 2011;6: e18359.
[68] Taverna D, Nanney LB, Pollins AC, Sindona G, Caprioli R. Multiplexed molecular descriptors of pressure ulcers defined by imaging mass spectrometry. Wound Repair Regen 2011;19:734–44.
[69] Guillou L, Coindre JM, Bonichon F, Nguyen BB, Terrier P, Collin F, et al. Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma. J Clin Oncol 1997;15:350–62.
[70] Mentzel T, van den Berg E, Molenaar WM. Myxofibrosarcoma. In: Fletcher CDM, Unni KK, Mertens F, editors. World Health Organization classification of tumours pathology and genetics tumours of soft tissue and bone. 2002 ed. Lyon: IARC Press; 2004. p. 102–3.
[71] Klerk LA, Broersen A, Fletcher IW, van Liere R, Heeren RMA. Extended data analysis strategies for high resolution imaging MS: new methods to deal with extremely large image hyperspectral datasets. Int J Mass Spectrom 2007;260: 222–36.
[72] Kargacin ME, Kowalski BR. Ion intensity and image resolution in secondary ion mass spectrometry. Anal Chem 1986;58: 2300–6.
[73] Graham D. Multivariate surface analysis homepage. http://mvsa.nb.uw.edu.
[74] McDonnell LA, Willems SM, Corthals GL, van Remoortere A, van Zeijl RJM, Deelder AM. Imaging mass spectrometry in cancer research: past experiences and future possibilities. J Proteomics 2010;73:1921–44.
[75] Fletcher JS, Rabbani S, Henderson A, Lockyer NP, Vickerma JC. Three-dimensional mass spectral imaging of HeLa-M cells — sample preparation, data interpretation and visualisation. Rapid Commun Mass Spectrom 2011;25: 925–93.
[76] Tyler BJ, Rayal G, Castner DG. Multivariate analysis strategies for processing ToF-SIMS images of biomaterials. Biomaterials 2007;28:2412–23.
[77] Broersen A, van Liere R, Heeren RMA. Parametric visualization of high resolution correlated multi-spectral features using PCA. Eurovis 2007Norrköping; 2007.
[78] Schwartz SA, Weil RJ, Johnson MD, Toms SA, Caprioli RM. Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. Clin Cancer Res 2004;10:981–7.
[79] Bonnel D, Longuespee R, Franck J, Roudbaraki M, Gosset P, Day R, et al. Multivariate analyses for biomarkers hunting and validation through on-tissue bottom-up or in-source decay in MALDI-MSI: application to prostate cancer. Anal Bioanal Chem 2011;401:149–65.
[80] Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern 1973;3:32–57.
[81] Bruand J, Alexandrov T, Sistla S, Wisztorski M, Mériaux Cl, Becker M, et al. AMASS: algorithm for MSI analysis by semi-supervised segmentation. J Proteome Res 2011;10: 4734–43.
[82] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401: 788–91.
[83] Jones EA, RJMv Zeijl, Deelder AM, Andrén PE, Wolters L, McDonnell LA. High speed data processing for imaging MS-based molecular histology using graphical processing units. J Am Soc Mass Spectrom 2012;23:745–52.
[84] Hanselmann M, Kirchner M, Renard BY, Amstalden ER, Glunde K, Heeren RMA, et al. Concise representation of mass spectrometry images by probabilistic latent semantic analysis. Anal Chem 2008;80:9649–58.
[85] Andersson M, Groseclose MR, Deutch AY, Caprioli RM. Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. Nat Methods 2008;5:101–8.
[86] Crecelius AC, Williams B, Cornett DS, Li X, Dawant BM, Bodenheimer RE, et al. Reconstructing and visualizing protein distributions in 3-D by MALDI imaging mass spectrometry. 52nd ASMS conference. Nashville, TN; 2004.
[87] Sinha TK, Khatib-Shahidi S, Yankeelov TE, Mapara K, Ehtesham M, Cornett DS, et al. Integrating spatially resolved three dimensional MALDI IMS with in vivo magnetic resonance imaging. Nat Methods 2008;5:57–9.
[88] Smentkowski VS, Ostrowski SG, Braunstein E, Keenan MR, Ohlhausen JAT, Kotula PG. Multivariate statistical analysis of three-spatial-dimension TOF-SIMS raw data sets. Anal Chem 2007;79:7719–26.
[89] Keenan MR, Smentkowski VS, Ulfig RM, Oltman E, Larson DJ, Kelly TF. Atomic-scale phase composition through multivariate statistical analysis of atom probe tomography data. Microsc Microanal 2011;17:418–30.
[90] Scholz B, Sköld K, Kultima K, Fernandez C, Waldemarson S, Savitski MM, et al. Impact of temperature dependent sampling procedures in proteomics and peptidomics — a characterization of the liver and pancreas post mortem degradome. Mol Cell Proteomics 2011;10.
[91] Svensson M, Boren M, Skold K, Falth M, Sjogren B, Andersson M, et al. Heat stabilization of the tissue proteome: a new technology for improved proteomics. J Proteome Res 2009;8:974–81.
[92] Goodwin RJA, Dungworth JC, Cobb SR, Pitt AR. Time-dependent evolution of tissue markers by MALDI-MS imaging. Proteomics 2008;8:3801–8.
[93] Goodwin RJA, Iverson SL, Andren PE. The significance of ambient-temperature on pharmaceutical and endogenous compound abundance and distribution in tissues sections

when analyzed by matrix-assisted laser desorption/ionization mass spectrometry imaging. Rapid Commun Mass Spectrom 2012;26:494–8.
[94] Goodwin RJA. Sample preparation for mass spectrometry imaging: Small mistakes can lead to big consequences. J Proteomics 2012;75:4893–911 (this issue).
[95] McDonnell LA, Corthals GC, Andrén PE. Going forward: Increasing the accessibility of imaging mass spectrometry. J Proteomics 2012;75:5113–21 (this issue).
[96] Cazares LH, Troyer D, Mendrinos S, Lance RA, Nyalwidhe JO, Beydoun HA, et al. Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase kinase 2 discriminates cancer from uninvolved prostate tissue. Clin Cancer Res 2009;15:5541–51.
[97] Hauskrecht M, Pelikan R, Valko M, James, Lyons-Weiler. Feature selection and dimensionality reduction in genomics and proteomics. In: Dubitzky W, Granzow M, Berrar DP, editors. Fundamentals of data mining in genomics and proteomics. Springer; 2007. p. 149–72.
[98] Manly KF, Nettleton D, Hwang JT. Genomics, prior probability, and statistical tests of multiple hypotheses. Genome Res 2004;14:997-100.
[99] Ting L, Cowley MJ, Hoon SL, Guilhaus M, Raftery MJ, Cavicchioli R. Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling*. Mol Cell Proteomics 2009;8:2227–42.
[100] Karp NA, Spencer M, Lindsay H, O'Dell K, Lilley KS. Impact of replicate types on proteomic expression analysis. J Proteome Res 2005;4:1867–71.
[101] Djidja M-C, Claude E, Snel MF, Francese S, Scriven P, Carolan V, et al. Novel molecular tumour classification using MALDI—mass spectrometry imaging of tissue micro-array. Anal Bioanal Chem 2010;397:587–601.
[102] Meding S, Nitsche U, Balluff B, Elsner M, Rauser S, Schöne C, et al. Tumor classification of six common cancer types based on proteomic profiling by MALDI imaging. J Proteome Res 2012;11:1996–2003.
[103] Hilario M, Kalousis A, Pellegrini C, Müller M. Processing and classification of protein mass spectra. Mass Spectrom Rev 2006;25:409–49.
[104] Lazova R, Seeley EH, Keenan M, Gueorguieva R, Caprioli RM. Imaging mass spectrometry—a new and promising method to differentiate spitz nevi from spitzoid malignant melanomas. Am J Dermatopathol 2012;34:82–90.
[105] Umar A, Kang H, Timmermans AM, Annemieke M, Look MP, Meijer-van Gelder ME, et al. Identification of a putative protein profile associated with tamoxifen therapy resistance in breast cancer. Mol Cell Proteomics 2009;8:1278–94.
[106] Waanders LF, Chwalek K, Monetti M, Kumar C, Lammert E, Mann M. Quantitative proteomic analysis of single pancreatic islets. PNAS 2009;106:18902–7.
[107] Stoeckli M, Staab D, Schweitzer A. Compound and metabolite distribution measured by MALDI mass spectrometric imaging in whole-body tissue sections. Int J Mass Spectrom 2007;260:195–202.