

MALDI Imaging Combined with Hierarchical Clustering as a New Tool for the Interpretation of Complex Human Cancers

Sören-Oliver Deininger,^{*,†,‡} Matthias P. Ebert,^{‡,§} Arne Fütterer,[†] Marc Gerhard,[†] and Christoph Röcken[§]

Bruker Daltonik GmbH, Bremen, Germany, Department of Medicine II, Klinikum rechts der Isar, Technische Universität München, München, Germany, and Institute of Pathology, Charité, Berlin, Germany

Received July 29, 2008

Proteomics analyses have been exploited for the discovery of novel biomarkers for the early recognition and prognostic stratification of cancer patients. These analyses have now been extended to whole tissue sections by using a new tool, that is, MALDI imaging. This allows the spatial resolution of protein and peptides and their allocation to histoanatomical structures. Each MALDI imaging data set contains a large number of proteins and peptides, and their analysis can be quite tedious. We report here a new approach for the analysis of MALDI imaging results. Mass spectra are classified by hierarchical clustering by similarity and the resulting tissue classes are compared with the histology. The same approach is used to compare data sets of different patients. Tissue sections of gastric cancer and non-neoplastic mucosa obtained from 10 patients were forwarded to MALDI-Imaging. The *in situ* proteome expression was analyzed by hierarchical clustering and by principal component analysis (PCA). The reconstruction of images based on principal component scores allowed an unsupervised feature extraction of the data set. Generally, these images were in good agreement with the histology of the samples. The hierarchical clustering allowed a quick and intuitive access to the multidimensional information in the data set. It allowed a quick selection of spectra classes representative for different tissue features. The use of PCA for the comparison of MALDI spectra from different patients showed that the tumor and non-neoplastic mucosa are separated in the first three principal components. MALDI imaging in combination with hierarchical clustering allows the comprehensive analysis of the *in situ* cancer proteome in complex human cancers. On the basis of this cluster analysis, classification of complex human tissues is possible and opens the way for specific and cancer-related *in situ* biomarker analysis and identification.

Keywords: Imaging MS • gastric cancer • bioinformatics • clustering • *in situ* proteomics

Introduction

The clinical management of patients with a cancer of the gastrointestinal tract is still a major challenge for physicians and oncologists. More than 500 000 Europeans develop a cancer of the gut every year and the prognosis of most of these cancers is still very poor.^{1–3} The reasons for these poor prognoses are complex. Many cancers of the gastrointestinal tract are diagnosed in advanced stages excluding curative treatment. There are no reliable tumor markers, which may allow early diagnosis or screening of high risk populations. Treatment options are limited in locally advanced and metastatic cancers, and our knowledge of the biology of these malignancies is still limited. To improve the poor prognosis of gastrointestinal cancers, we and others have recently applied

proteomics to search for novel diagnostic and therapeutic targets, including two-dimensional electrophoresis coupled with matrix-assisted laser desorption/ionization mass spectrometry, and surface-enhanced laser desorption/ionization mass spectrometry on serum and tissue samples.^{4–7} With regard to tissue samples, proteomic analysis usually necessitates prefractionation in order to separate tumor from non-tumor, or neoplastic cells from non-neoplastic cells either by macrodissection, laser capture-microdissection, or homogenization and, for example, subsequent fractionation using antibody-coupled magnetic beads.^{4–6} While these technologies have demonstrated their suitability, they still provide only limited information about the spatial variability of the cancer proteome. The protein composition of a malignant tumor varies qualitatively and quantitatively. This stems from the intrinsic heterogeneity of solid tumors (e.g., distinct grades of differentiation; local differences in the metabolic activity of the tumor), the local inflammatory response, neovascularization, the development of a desmoplastic stroma, and so forth. Thus, the tumor proteome is spatially complex and most purification techniques are unable to resolve the corresponding plasticity of the cancer and tissue proteome.^{8–10}

* To whom correspondence should be addressed. Dr. Sören-Oliver Deininger, Applications MALDI-TOF-MS, Fahrenheitstrasse 4, D-28359 Bremen, Germany. Phone: +49-421-2205-403. Fax: +49-421-2205-104. E-mail: sod@bdal.de.

† Bruker Daltonik GmbH.

‡ Technische Universität München.

§ These authors contributed equally.

Charité, Berlin.

A new approach to increase the diagnostic specificity and to characterize the cancer proteome profile *in situ* is MALDI imaging, which allows the detection of proteins in the tissue section and the analysis of their spatial distribution without prefractionation.^{8–10} The exclusion of protein prefractionation and separation usually applied in proteomic research comes at a price: It is only possible to detect rather abundant proteins and peptides. This does not need to be a shortcoming of the technique, since tumor-related proteins are expected to be much more abundant in the tumor itself compared with, for example, serum.

Each MALDI imaging data set is a multidimensional molecular data set that shows the distribution of up to hundreds of molecular signals. Each single data set contains a wealth of information. If larger numbers of samples from different patients are to be analyzed, the time and effort required for these analyses quickly limits the usefulness of the technique. The complexity and amount of the data necessitates a reduction to an informative and accessible volume. Principal component analysis for the analysis of mass spectrometric images is one way to provide an automatic feature extraction. Principal component analysis is routinely used to reconstruct images in secondary ion mass spectrometric imaging.^{11–13} It has also been reported in the context of MALDI imaging as a tool to find characteristic masses for specific tissue regions¹⁴ or as a data pretreatment for multivariate classifications.¹⁵ However, it has not found widespread application in the direct reconstruction of images in MALDI imaging. There are also some limitations to this approach, namely, the scores of spectra can be similar for one principal component, while they can be different for other principal components. Thus, several images need to be generated in order to facilitate a complete interpretation of the data. What is rather desirable is a presentation of the data in a way that similar spectra are shown in the same color. This can be achieved by hierarchical clustering,¹⁶ which was previously employed in MALDI imaging data sets to generate images based on spectra similarity.¹⁵ However, until now, it has not reached widespread use, because a convenient user interface was unavailable to gain full advantage of the clustering results. Here we present hierarchical clustering as a tool for the easy and convenient interpretation of single imaging data sets. Spectra are clustered by similarity in a dendrogram, and specific nodes in the dendrogram can be selected and colors assigned. The spectra that belong to these clusters can be shown in the selected color in the MALDI image. This changes the way MALDI imaging data sets are evaluated. Instead of choosing individual mass signals, branches of the dendrogram are selected starting from the top level until the histology of the section is explained. It also preserves the information on the hierarchy of the spectra as shown in the dendrogram. This reduces the time effort needed for the interactive interpretation of complex imaging data sets to a few minutes and allows the presentation of the relevant information contained in the data set into one single image. In this study, we demonstrate that hierarchical clustering can be applied to MALDI imaging data sets of tissue samples from gastric cancer patients and allows the comprehensive analysis of the complex cancer proteome.

Material and Methods

Tissues. Tissue specimens were obtained from gastric cancer patients and were retrieved from the archive of the Institute of Pathology, Charite, Berlin. All cases were reviewed before study

Table 1. Characteristics of Gastric Cancer Patients

gender	age	tumor type	T	N	M	G	stage
m	60	intestinal	3	3	0	3	4
f	70	intestinal	2a	1	0	2	2
f	82	intestinal	3	3	1	3	4
f	76	intestinal	2b	1	0	3	2
m	77	intestinal	2b	1	0	2	2
f	83	intestinal	3	2	1	2	4
m	59	intestinal	4	1	0	2	4
m	67	intestinal	2b	0	x	2	2
m	76	intestinal	2b	3	x	2	4
f	77	diffus	3	2	x	3	3b

inclusion. Tissue samples used in the present study were obtained from 10 patients (5 men, 5 female), diagnosed with intestinal-type (9 patients), and diffuse-type (1) gastric cancer (Table 1). The age of the patients ranged from 59 to 83 years (mean 72.7 years). Tissue samples (gastric cancer and corresponding non-neoplastic mucosa) were obtained immediately after surgery, snap-frozen in liquid nitrogen, and stored at –80 °C until required. For histological processing and histological verification of gastric cancer, tissue samples were fixed in 10% neutralized formalin and embedded in paraffin. Deparaffinized sections were stained using hematoxylin and eosin (H&E). Following the MALDI imaging analysis, the cryosections were stained by H&E to allow a proper evaluation of the MALDI results.

Tissue MALDI Imaging. Cryosections of tumor and nontumor samples were cut on a cryostat. The sections were transferred to precooled (–20 °C) conductive Indium-Tin-Oxide (ITO) coated glass slides (multiple sections of the same tissue block were placed onto one ITO slide). The glass slide and the tissue were thawed together.¹⁷ Then, the sections were briefly washed in 70% and 100% ethanol, air-dried, and stored at –20 °C until further processing. Prior to matrix coating, the order for the matrix preparation and the measurement was randomized using an Excel spreadsheet.

The MALDI matrix was applied using the imagePrep station (Bruker Daltonics) and the standard protocol provided with the instrument. Sinapinic Acid (SA) at 10 mg/mL in water/acetonitrile 40:60 (v/v) with 0.2% trifluoroacetic acid (TFA) was used as matrix for the MALDI measurement. The MALDI measurement and image analysis was carried out on a linear Autoflex instrument equipped with a smartbeam-laser and with the flexImaging 2.1 and ClinProTools 2.2 software packages (Bruker Daltonik GmbH, Bremen). MALDI measurements were done in linear mode in a mass range of 2000–30 000 Da with a sampling rate of 0.1 GS/s. The lateral resolution for the MALDI imaging was set to 200 μm. A total of 300 laser shots were measured per position.

Following the MALDI analysis, the matrix was washed off the slides with 70% ethanol and the sections were stained with H&E, scanned at low resolution, and coregistered with the MALDI imaging results. Subsequently, the H&E-stained sections were evaluated histologically by an experienced surgical pathologist (C.R.) using a conventional binocular microscope.

Statistical Analysis. Statistical analyses were carried out using the ClinProTools 2.2 Software. For the statistical analyses, the mass spectra were internally recalibrated on common peaks (also known as spectral alignment) and normalized on the total ion count. An average spectrum created from all single spectra was used for a peak picking and to define integration ranges. These integration ranges were used to obtain the intensities or areas on the single spectra. The signal intensities were used

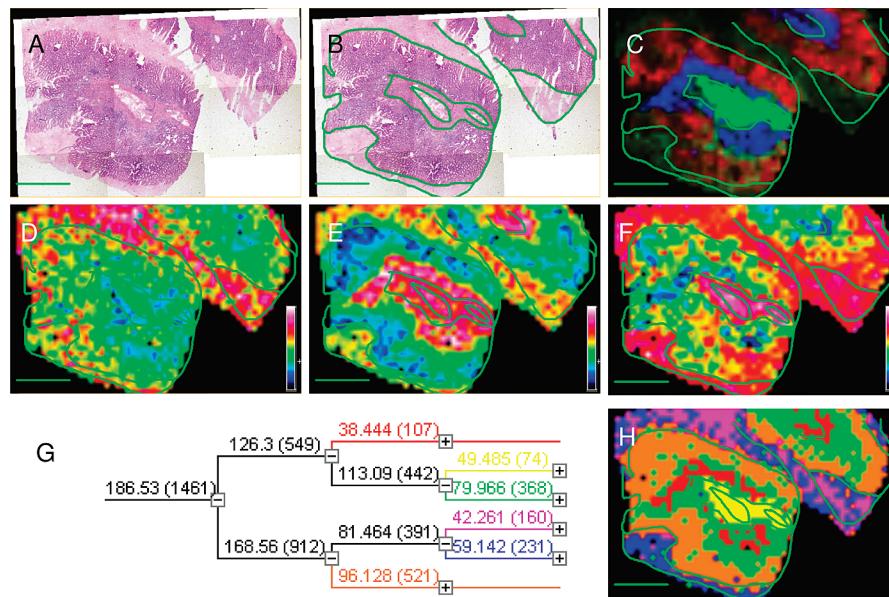


Figure 1. (A) H&E-stained tissue section of a non-neoplastic stomach mucosa. (B) Same section with marked features for orientation. (C) MALDI imaging result shown in false color representation. Colors reflect intensities of selected mass signals; red, 4152 Da; green, 4962 Da; blue, 5697 Da. (D–F) Scores of the first (D), second (E) and third (F) principal component in heat map representation. (G) Top nodes of dendrogram from hierarchical clustering. (H) Reconstructed image from dendrogram. Pixels are colored according to the color of the dendrogram node in (G).

for all calculations. For the principal component analysis and the hierarchical clustering, the individual peak intensities were standardized across the data set. For the PCA analysis and the clustering across different patients, the mass spectra were selected on the tissue by assigning representative areas of tumor and non-neoplastic mucosa. For the statistical calculation, one average spectrum was generated per section. One single data set from a nontumorous section that was rich in necrosis had to be excluded, because it showed very low quality spectra in two measurements.

Results

MALDI Imaging Analysis of Selected Sections. MALDI imaging was done on tissue sections of human gastric cancers obtained from 10 different patients. The analysis of a MALDI imaging data set by molecular mass images, principal component analysis and hierarchical clustering is illustrated in Figure 1. The data set was chosen from non-neoplastic mucosa specifically because it exhibits similar tissue structures in different locations on the same image. Following the MALDI imaging experiment, the same sections were stained with H&E (Figure 1A) and scanned. This allowed a parallel assessment of the histopathological features seen under the light microscope and the proteome profiles obtained by MALDI imaging.^{18,19} Subsequently, non-neoplastic mucosa was marked to allow a better understanding of the molecular images (Figure 1B; note the presence of muscularis mucosa in the center of the section). Subsequently, mass signals were selected by manual inspection of the data set to match the main visible features of the tissue. In Figure 1C, three molecular mass signals are shown in three different colors. Finally, the scores of the first three principal components were visualized in the image (Figure 1D,E). In this context, the PCA is an unsupervised feature extraction and allows the generation of meaningful tissue images without a detailed understanding of the underlying histology. The mass spectra were then subjected to a hierarchical clustering analysis.

Figure 1G depicts the top branches of the clustering tree as they are reported in the flexImaging software. The numbers are the distances from node-to-node. These are usually shown as different lengths of the branches in conventional clustering analyses. However, since this representation here is mainly acting as user interface to highlight certain branches of the clustering on the images, the branches are all shown in the same length to keep the interface simple. The numbers in brackets are the number of mass spectra underneath each branch. Since hierarchical clustering is an agglomerative clustering, always a full tree is created: each branch can be split up and followed down to the individual mass spectra. Each node along this way can be considered a class. When it comes to the evaluation of these clustering results on MALDI imaging data sets, it has to be decided where histological and biological features are classified and at which point random differences between spectra are picked up. In the example shown in Figure 1, the classes highlighted in yellow, red, green and orange clearly denote different tissue classes, while the blue and the pink class seem to denote random differences. The classes were then selected on the dendrogram and a color was assigned. All spectra in the image that belong to this class assume the selected color. The class can now be followed along the dendrogram. The dendrogram can be expanded and the next lower level of tissue classes can be selected. This can be followed until the tissue section is explained sufficiently. This leads to a semisupervised hierarchical clustering of the expression data. The clustering was performed unsupervised, but the selection of the right number of classes was done in accordance with a detailed knowledge of the histology. Since the clustering was done by spectra similarity, the tedious evaluation of hundreds of mass signals was turned into a straightforward clustering analysis. Since there are many different ways to determine the similarity of multidimensional data sets, namely, in how the distances in the multidimensional space are measured and how the distance between clusters are estimated,

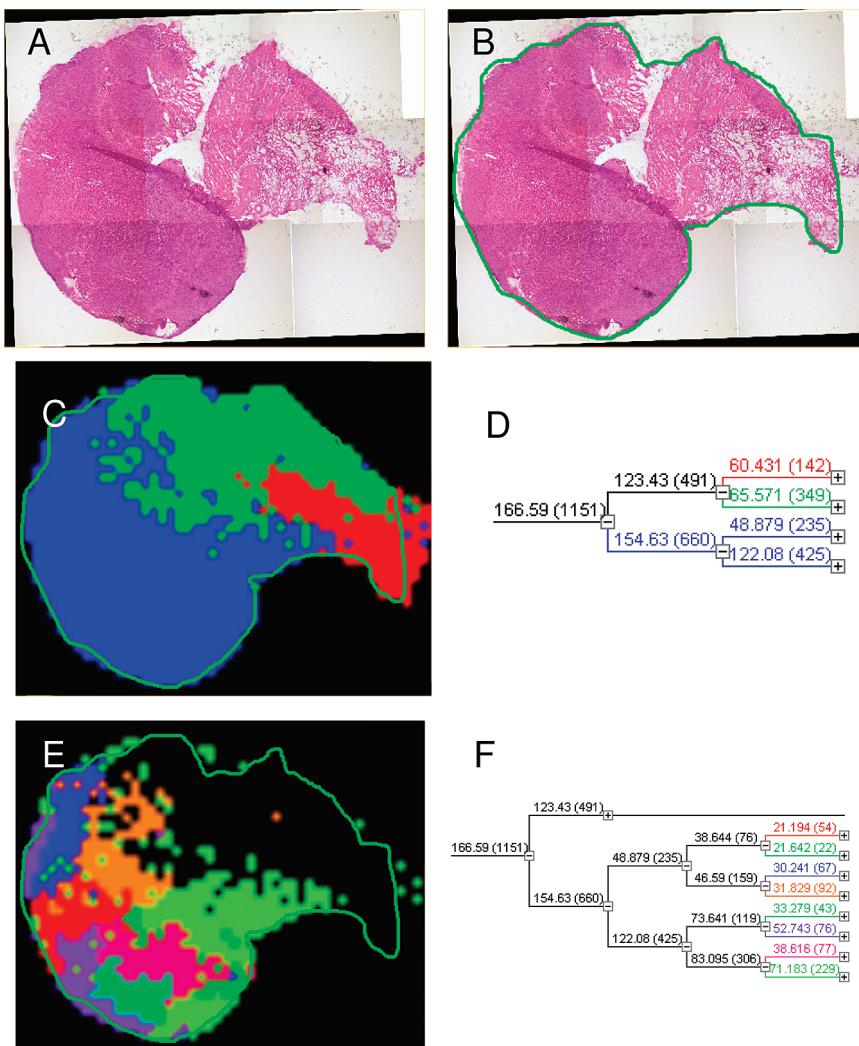


Figure 2. (A) H&E stained tissue section of intestinal type gastric cancer. (B) Same section with marked outline for orientation. (C) Reconstructed image from dendrogram. Pixels are colored according to the color of the dendrogram nodes in (D). (D) Dendrogram used for the reconstruction of (C). (E) Reconstructed image from dendrogram. Pixels are colored according to the color of the dendrogram nodes in (F). (F) Dendrogram used for the reconstruction of (E).

it was possible to generate different dendrograms from the same data set. In our experience, the Euclidean distance metric on PCA-data reduced to 70% explained variance and a Ward linkage gave the most meaningful results.

The application of the clustering to a tumor sample is shown in Figure 2. Here, the top three branches of the clustering show the solid tumor in blue. Expansion and selection of the tumor branches of the clustering tree lead to a more detailed clustering of the spectra in the solid tumor. Interestingly, this did not translate directly into histologically visible distinct tumor areas and illustrates that MALDI imaging detects additional phenotypic differences, for example, tumor subclones.

In further support of the former observation, the results of the clustering were not always in good agreement with the histological appearance. Especially molecular signals that can be attributed to beta-defensins often seem to show distributions in the tissue that not always follow the histology. One example of a section that shows strong beta-defensin signals is illustrated in Figure 3. Figure 3B,C shows the first and second principal component scores. This specimen was obtained from the only study patient with diffuse type gastric cancer. While the first PCA is in good agreement with the histology, the

second PCA shows a strange behavior at the lower left side. The same is picked up in the clustering analysis in the class selected in green, which is already picked up in the second highest level in the dendrogram. A detailed analysis of the molecular signals revealed that this structure was mainly due to the presence of beta-defensins in this sample. This is shown in Figure 3F, which shows the distribution of the molecular species with a mass of 3372 Da in red and 3329 Da in green. This correlates histologically with the ulcerated mucosal surface of the tumor, and is rich in polymorphonucleocytes.

Statistical Comparison between the Sections. A molecular imaging technique like MALDI imaging falls short of its capabilities, if only single sections were to be analyzed. In fact, our aim is to find diagnostic, predictive or prognostic biomarkers. In this study, tumor and tumor-free sections containing mucosa from the same patients were available. If tumor areas and tumor-free mucosa from the same patient were compared, it was always possible to find molecular signals that separated the tumor from the non-neoplastic mucosa. For an in-depth comparison of tumor and non-neoplastic mucosa, on all sections tumor and tumor-free mucosa areas were selected and the spectra for those regions were compared to find specific

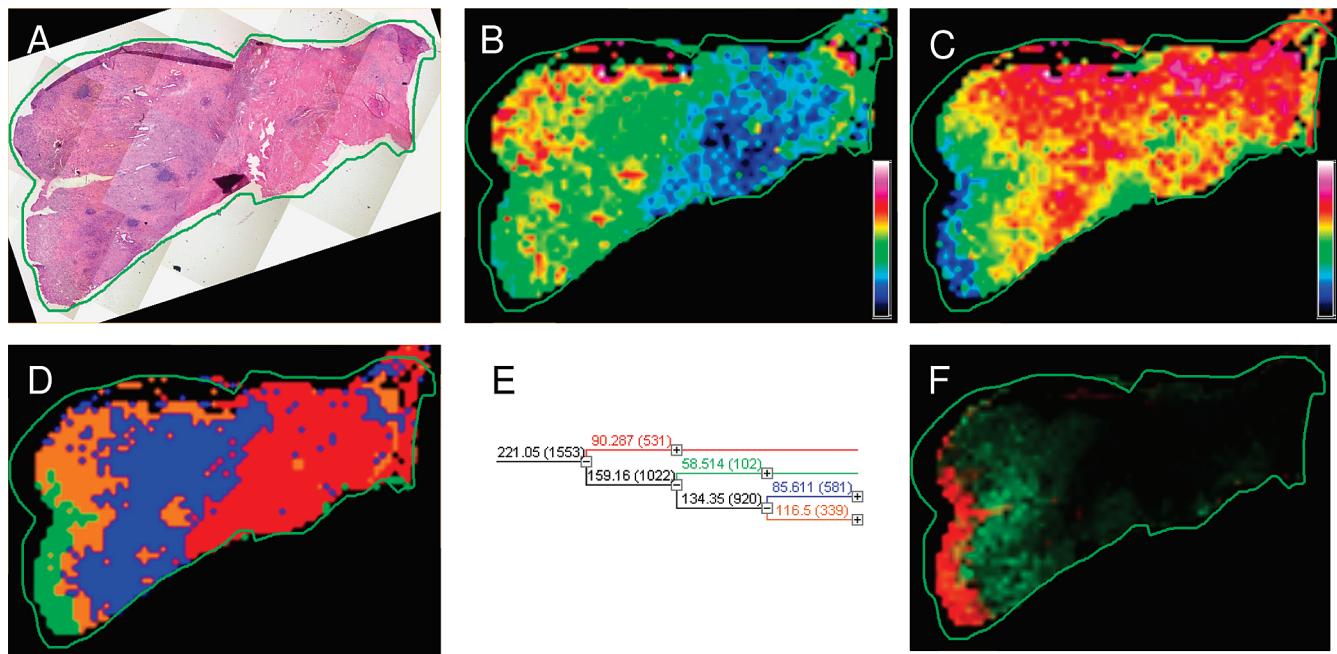


Figure 3. (A) H&E staining section of diffuse type gastric cancer with marked outline. (B and C) Scores of the first (B) and second (C) principal component in heatmap representation. (D) Reconstructed image from dendrogram. Pixels are colored according to the dendrogram in (E). (E) Dendrogram used for the reconstruction of (D). MALDI imaging result shown in false color representation. Colors reflect intensities of selected mass signals; red, 3371 Da; green, 3328 Da.

tumor markers. For a comparison, two tissue classes were defined, tumor and tumor-free mucosa. Figure 4A shows a comparison of the different patients in two classes. Between the dashed lines, a random selection of 50 spectra from the respective area of one sample is shown. Masses that would be present in all spectra of one class would appear in a vertical line in this representation. By visual inspection, no apparent mass seems to be present in all the tumor samples and absent in all mucosa samples. This leads to the interesting observation that for each patient molecular signals can be found to be tumor specific, but if all tumor specimens are compared with all non-neoplastic mucosa specimens, the situation is less clear. However, the statistical tools PCA analysis and hierarchical clustering can also be applied to the data set as a whole. In the PCA analysis, it is clearly seen in the scores plot that, in general, the green points for the non-neoplastic mucosa and the red points for the tumor specimen fall apart and can be separated. The same is seen in the hierarchical clustering, where two of the main clusters contain all tumor spectra, while one cluster contains only non-neoplastic mucosa spectra. The three non-neoplastic tumor spectra that are clustering together with the tumor can be explained and have no functional implications. If the number of samples is comparatively small as in this experiment and the distance between the classes is not large in comparison to the distances within each class, then the clustering is to some extent effected by random (e.g., one additional datapoint could change the dendrogram significantly). It needs to be pointed out that the hierarchical clustering of larger patient cohorts allows to search for patient subclasses that share common meta-information.

Discussion

The abundant information obtained by MALDI imaging fully reflects and extends the complexity of the underlying histology. Each molecular species in a MALDI imaging experiment can be

used to generate one image. This bears advantages and disadvantages. Advantageous is the close comparison with the histology. It allows an immediate evaluation of the putative significance of a biomarker, that is, the mass of interest can be allocated to histoanatomical structures such as tumor cells, tumor stroma, tumor vessels, and inflammation. Disadvantageous is the enormous amount of information generated and meaningful evaluation of such data sets can become quite tedious if hundreds of molecular images have to be inspected. Furthermore, the interpretation of MALDI imaging experiments require specialists of different areas and are often carried out in collaboration between proteomics or mass spectrometry facilities and clinical researchers. Here often the situation arises that the person responsible for the acquisition of the MALDI images is not able to evaluate the results because of a lack of understanding of the underlying histopathology and clinical background. Still, the mass spectrometrist in the collaboration is required to provide initial results of the experiments. The clinical researcher, on the other hand, wants easy access to the information in the imaging data sets rather than a long and tedious evaluation of all individual mass signals in an imaging data set.

This problem can be solved by a statistical analysis of the imaging data sets with the aim of an automated feature extraction. One well-known way to generate images from multidimensional data sets is the PCAs. Here, the dimensionality of the problem is reduced while as much information as possible is retained. The scores of the first principal components can now be used to generate meaningful images without a detailed understanding of the underlying tissue. Since the first principal components contain the main variance of the data set and the different tissue regions are expected to contribute the main variance in the data set, the PCA-images usually reflect the histology quite well. The PCA allows an easy feature extraction, but it offers little possibility for a more detailed analysis. Another disadvantage of the use of PCA scores from the generation of images is the fact that the spectra

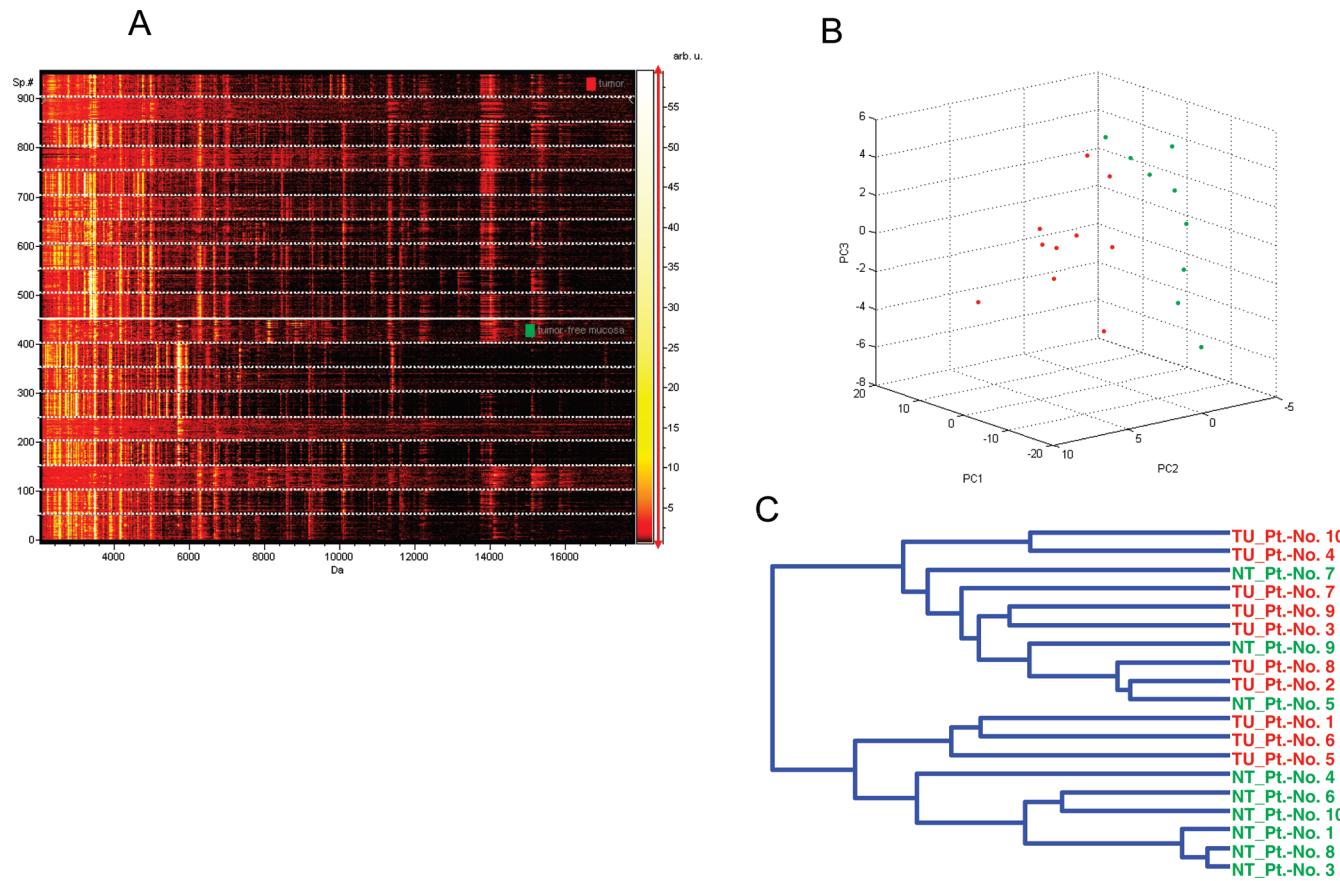


Figure 4. (A) Pseudogel view of 50 randomly selected mass spectra from tumor and tumor-free mucosa areas of each specimen. Spectra from one patient are grouped between dashed lines. (Upper part) Spectra from tumor; (lower part) spectra from tumor-free mucosa. (B) Scores plot of first, second and third principal component. Each data point represents one patient (red, tumor; green, nontumor). (C) Dendrogram of hierarchical clustering of patients. Red, tumor; green, nontumor.

can be similar in one principal component but rather distant in other principal components. To present the full information in the data set, it is therefore necessary to generate several images for the first principal components. The PCA analysis also does not provide a grouping of the spectra into different classes.

This can be done by hierarchical clustering. Here, the spectra in a data set are clustered by similarity. The hierarchical clustering is a well-known agglomerative algorithm and a dendrogram always contains classes for every single spectrum. The interesting question here is how many classes can reasonably be defined by the hierarchical clustering. The top branches of the dendrogram can be expected to contain classes for the different tissue types in a section, while the branches lower in the hierarchy will contain more and more random differences between spectra from the same tissue type. Typical mathematical approaches for the evaluation of the number of reasonable classes involve the plot of explained variance in a data set versus the number of classes and the evaluation of the “elbow criterion” in this plot. While this is a suitable way to generate a discrete number of classes and show them on imaging data sets, a crucial information is lost, namely, the hierarchy. We describe here, for the first time, a novel user interface that allows the interactive exploration of the dendrograms and the display of the respective spectra on the MALDI images. This is a semisupervised approach, since now the knowledge of the histology can be used to define the reasonable classes. Effectively, the MALDI imaging is abstracted from the individual ion images to an analysis that is based on spectra similarity.

A detailed analysis of all non-neoplastic mucosa specimens and most tumor specimens showed in general a very good agreement between the PCA/clustering and the obvious histological features. This clearly shows the suitability and practical use of the clustering approach. The statistical calculation per section took on average less than 10 min, but more importantly, this time requires no user interaction. The evaluation of the clustering results on the other hand could be done in minutes per data set. Since usually it can be assumed that the labor time of the histopathologist is the most limited resource in the evaluation of MALDI imaging results, the interactive evaluation of the clustering results is a significant step toward MALDI imaging as a technique for a more routine clinical research. The dendrograms could be expanded until the histology was explained. Of course, at this point one possibility would be that further expansion of the dendrogram would start to show the random differences between the spectra. In this case, one would actually assume that spectra belonging to such branches in the dendrogram would start to show more random distributions. In many cases, however, even those branches showed spectra in close spatial proximity, which seemed not to correlate with distinct histological features. The reason for this cannot be finally determined here. However, the heterogeneity of solid tumors, including gastric cancer, and the clonal expansion of tumor cells may contribute to molecular differences that are invisible in the H&E stained sections but are unravelled by MALDI imaging. Additionally, metabolic differences due to different hypoxia or nutrient levels within a tumor influence mass spectra generated by MALDI imaging. However, apart from pathophysiology,

ological, tumor-intrinsic variability, differences in the sample preparation, for example, due to different salt or lipid content or different drying rates of the MALDI matrix, will certainly affect these clusters as well. As shown here, it is possible to mirror the histological appearance of a gastric cancer specimen by MALDI imaging, but where clustering results extend beyond the histologically visible, which is the most interesting area for the discovery of novel biomarkers, clustering results become less accessible and more difficult to interpret.

Comparing MALDI imaging spectra with the histology of our gastric cancer specimens did not always lead to complete congruent results. There are several reasons why a PCA or clustering result is not in agreement with the histology. One reason could obviously be an effect of an inhomogenous sample preparation. Such effects have been reported in the literature.¹⁵ McCombie et al.¹⁵ discouraged the use of PCA scores for direct generation of images, because of the effect of inhomogeneities of the sample preparation. In recent years, however, the sample preparation in MALDI imaging was significantly improved, and with the imagePrep device used in this study, a homogeneous coating of the sample was regularly achieved. Another explanation for apparent mismatches between the PCA/clustering results and the histology can also be the presence of molecular species in the tissue whose distribution does not follow the histological features. Especially if these compounds contribute intensive signals to the MALDI spectra, they can have a significant influence on the statistical analysis and lead to apparent mismatches between the PCA/clustering images and the histology. In this case, this is not an artifact, but simply due to the fact that MALDI imaging is a molecular imaging technology that also picks up molecular signals, which are not necessarily specific for the tissue. We have found that particularly signals from beta-defensins show this behavior. These signals are associated with inflammatory processes and often appear not to follow the classic histological features. On the other hand, these signals are very intensive in the MALDI spectra. It is also important to note that unfixed tissue is sampled immediately after surgery, and plasma and serum proteins from opened blood vessels may contaminate the tissue specimen. These compounds can even diffuse into the depth of the sample and lead to a nonhistoanatomical distribution of mass spectra.

The same PCA and clustering algorithms can be used to compare different patients. In this case, however, it is necessary to sort out the mass spectra that are typical for the features to be compared. In this study, we used this approach to compare tumor and tumor-free mucosa from different patients. Although it was usually possible in the comparison of tumor-free mucosa and tumor for each patient to find tumor specific masses, this was not possible for the combined data from tumor versus tumor-free across all patients.

A PCA analysis or clustering shows that the classes can be separated by an unsupervised analysis. The clustering approach also offers a potential way to define otherwise undetected subclasses of patients. For such an approach, a significantly larger set of patients needs to be compared and as much available meta-information as possible needs to be taken into account. This finding is particularly interesting because the separation between the classes is obviously of a multivariate nature that cannot be easily attributed to one mass and clearly shows why multivariate classifications may be necessary if the ultimate goal is a software

assisted classification of unknown tissue. The next logical step would be to define software models to classify tumor from nontumor specimens. However, the heterogeneity of the selected samples does not seem to make this suitable in this case. For this, either a significantly higher number of cases would be necessary or a better preselection of the different specimens, such as only solid tumors with a certain area to minimize the sample-to-sample variation with regard to patient gender, tumor type (intestinal vs diffuse) and extralesional tissue (with and without gastritis, antral vs fundic vs cardiac mucosa). Since we were able to separate tumor from non-neoplastic mucosa by an unsupervised method, this clearly highlights that a software based classification should be achievable.

Acknowledgment. M. Ebert is supported by a grant from the DFG, the Deutsche Krebshilfe (107885) and the MolBioMed Programme (EndoMed 01EZ0802) of the BMBF, Germany. C. Röcken is supported by the Berliner Krebshilfe e.V. (Berlin, Germany).

References

- Boyle, P.; Autier, P.; Bartelink, H.; Baselga, J.; Boffetta, P.; Burn, J.; Burns, H. J.; Christensen, L.; Denis, L.; Dicato, M.; Diehl, V.; Doll, R.; Franceschi, S.; Gillis, C. R.; Gray, N.; Griciute, L.; Hackshaw, A.; Kasler, M.; Kogevinas, M.; Kvinnslund, S.; La, V. C.; Levi, F.; McVie, J. G.; Maisonneuve, P.; Martin-Moreno, J. M.; Bishop, J. N.; Oleari, F.; Perrin, P.; Quinn, M.; Richards, M.; Ringborg, U.; Scully, C.; Siracka, E.; Storm, H.; Tubiana, M.; Tursz, T.; Veronesi, U.; Wald, N.; Weber, W.; Zaridze, D. G.; Zatonski, W.; zur, H. H. *Ann. Oncol.* **2003**, *14*, 973–1005.
- Boyle, P.; d'Onofrio, A.; Maisonneuve, P.; Severi, G.; Robertson, C.; Tubiana, M.; Veronesi, U. *Ann. Oncol.* **2003**, *14*, 1312–1325.
- Ferlay, J.; Autier, P.; Boniol, M.; Heanue, M.; Colombet, M.; Boyle, P. *Ann. Oncol.* **2007**, *18*, 581–592.
- Ebert, M. P.; Meuer, J.; Wiemer, J. C.; Schulz, H. U.; Reymond, M. A.; Traugott, U.; Malfertheiner, P.; Röcken, C. *J. Proteome Res.* **2004**, *3*, 1261–1266.
- Ebert, M. P.; Lamer, S.; Meuer, J.; Malfertheiner, P.; Reymond, M.; Buschmann, T.; Röcken, C.; Seibert, V. *J. Proteome Res.* **2005**, *4*, 586–590.
- Ebert, M. P.; Kruger, S.; Fogeron, M. L.; Lamer, S.; Chen, J.; Pross, M.; Schulz, H. U.; Lage, H.; Heim, S.; Roessner, A.; Malfertheiner, P.; Röcken, C. *Proteomics* **2005**, *5*, 1693–1704.
- Ebert, M.; Xing, X.; Burgermeister, E.; Schmid, R.; Röcken, C. *Expert Rev. Anticancer Ther.* **2007**, *7*, 465–469.
- Chaurand, P.; Sanders, M. E.; Jensen, R. A.; Caprioli, R. M. *Am. J. Pathol.* **2004**, *165*, 1057–1068.
- Cornett, D. S.; Reyzer, M. L.; Chaurand, P.; Caprioli, R. M. *Nat. Methods* **2007**, *4*, 828–833.
- Walch, A.; Rauser, S.; Deininger, S. O.; Hofler, H. *Histochem. Cell Biol.* **2008**.
- Aoyagi, S.; Kawashima, Y.; Kudo, M. *Nucl. Instrum. Methods Phys. Res., Sect. B* **2005**, *232*, 146–152.
- Lockyer, N. P.; Vickerman, J. C. *Appl. Surf. Sci.* **2004**, *231*, 377–384.
- Wagner, M. S.; Castner, D. G. *Langmuir* **2001**, *17*, 4649–4660.
- Trim, P. J.; Atkinson, S. J.; Princivalle, A. P.; Marshall, P. S.; West, A.; Clench, M. R. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1503–1509.
- McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R. *Anal. Chem.* **2005**, *77*, 6118–6124.
- Jain, A. K.; Murty, M. N.; Flynn, P. J. *ACM Comput. Surv.* **1999**, *31*, 264–323.
- Schwartz, S. A.; Reyzer, M. L.; Caprioli, R. M. *J. Mass Spectrom.* **2003**, *38*, 699–708.
- Crecelius, A. C.; Cornett, D. S.; Caprioli, R. M.; Williams, B.; Dawant, B. M.; Bodenheimer, B. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1093–1099.
- Schwamborn, K.; Krieg, R. C.; Reska, M.; Jakse, G.; Knuechel, R.; Wellmann, A. *Int. J. Mol. Med.* **2007**, *20*, 155–159.

PR8005777