

Normalization in MALDI-TOF imaging datasets of proteins: practical considerations

Sören-Oliver Deininger · Dale S. Cornett ·
Rainer Paape · Michael Becker · Charles Pineau ·
Sandra Rauser · Axel Walch · Eryk Wolski

Received: 8 November 2010 / Revised: 7 March 2011 / Accepted: 17 March 2011 / Published online: 12 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Normalization is critically important for the proper interpretation of matrix-assisted laser desorption/ionization (MALDI) imaging datasets. The effects of the commonly used normalization techniques based on total ion count (TIC) or vector norm normalization are significant, and they are frequently beneficial. In certain cases, however, these normalization algorithms may produce misleading results and possibly lead to wrong conclusions, e.g. regarding to potential biomarker distributions. This is typical for tissues in which signals of prominent abundance are present in confined areas, such as insulin in the pancreas or β -amyloid peptides in the brain. In this work, we

investigated whether normalization can be improved if dominant signals are excluded from the calculation. Because manual interaction with the data (e.g., defining the abundant signals) is not desired for routine analysis, we investigated two alternatives: normalization on the spectra noise level or on the median of signal intensities in the spectrum. Normalization on the median and the noise level was found to be significantly more robust against artifact generation compared to normalization on the TIC. Therefore, we propose to include these normalization methods in the standard “toolbox” of MALDI imaging for reliable results under conditions of automation.

Published in the special issue *MALDI Imaging* with Guest Editor Olivier Laprévotte.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-4929-z) contains supplementary material, which is available to authorized users.

S.-O. Deininger (✉) · R. Paape · M. Becker · E. Wolski
Bruker Daltonik GmbH,
Fahrenheitstr. 4,
28359 Bremen, Germany
e-mail: Soeren-Oliver.Deininger@bdal.de

D. S. Cornett
Bruker Daltonics,
40 Manning Road,
Billerica, MA 01821, USA

C. Pineau
Proteomics Core Facility Biogenouest, INSERM, U625,
Campus de Beaulieu,
35042 Rennes, France

S. Rauser · A. Walch
Institute of Pathology, Helmholtz Zentrum München,
German Research Center for Environmental Health,
Ingolstädter Landstraße 1,
85764 Neuherberg, Germany

Keywords Bioanalytical methods · Biological samples · Genomics/proteomics · MALDI imaging · Normalization · Data processing

Introduction

Normalization is the process of multiplying a mass spectrum with an intensity-scaling factor to expand or reduce the range of the intensity axis. It is used to project spectra of varying intensity onto a common intensity scale [1–8].

In matrix-assisted laser desorption/ionization (MALDI) imaging, normalization is used to remove systematic artifacts that affect mass spectral intensity. Artifacts may be a result of matrix crystal distribution at high lateral resolution or ion source contamination that possibly attenuates the total ion count (TIC) as a function of time during which the ion transmission may gradually decrease causing the image brightness to fade along the path of spectra acquisition. Chemical inhomogeneities in micro-environments such as salt or pH gradients, phospholipid

background structure, etc. may further modulate the abundance of all or a selection of signals in MALDI spectra. In addition, it may even be required to compare mass spectra across different imaging datasets in cohort studies, e.g., for biomarker discovery. Some of these artifacts related to sample preparation can be cured by improved matrix application protocols and by extended washing steps that may diminish salt and pH gradients, and these steps deserve attention to achieve well interpretable images.

The question of whether or not MALDI imaging datasets should be normalized, and the optimal model to do so, is subject of intense debate at conferences or MALDI imaging workshops. However, there have been no dedicated studies which evaluate the effects of normalization algorithms on imaging data in detail. Other studies have investigated the topic primarily from a mathematical point of view using artificial (absolutely homogenous) datasets [3]. Although such work is of importance, the discussion requires a new focus on real and non-optimized image datasets in order to draw conclusions for actual measurements. McDonnell et al. observed that normalization has only subtle effects on MALDI imaging results, at least for the datasets analyzed in their study [9]. This observation is based on the analysis of a single dataset and cannot be generalized, as will be discussed in this work.

For this reason, we evaluated a number of different normalization approaches using different MALDI imaging datasets that exhibited, to varying degrees, systematic artifacts that required normalization. We tested how well standard normalization techniques such as TIC and vector norm conditioned these datasets by comparing ion images before and after normalization. We also included the approach of normalizing to a constant noise, as recommended for MALDI imaging previously [3]. Further, we investigated other novel normalization approaches which, to our knowledge, have not been applied to imaging data, such as median normalization or TIC normalization with manual exclusion of mass ranges.

A pledge for normalization

Various effects other than the distribution of endogenous proteins in a tissue sample can influence the intensity of signals in MALDI imaging datasets. Here, two cases have to be considered. First, as in all mass spectrometric analyses of complex mixtures, specific ion suppression can happen. This means that a specific compound may be suppressed by another compound. If this happens, the observed intensities for the signal do not reflect the true concentration differences in the tissue. This effect can be considered an artifact intrinsic to the mass spectrometric measurement process, and it cannot be remedied by application of the spectrum-

wide normalization approaches discussed in this article. On the other hand, there are effects that may lead to a spectrum-wide attenuation of the signals, which can be countered by spectrum-wide normalization. Some of these effects are intrinsic to the tissue, e.g., inhomogeneous distribution of salts, pH gradients, or other endogenous compounds that may influence global ion suppression. For instance, in unwashed tissue samples, the signals of lipids are much more intense than signals of peptides or proteins. This poses the risk that these highly concentrated lipids suppress the formation of peptide and protein ions, a finding that is also supported experimentally [10].

Sample preparation, such as washing steps or matrix coating protocols, can also affect the intensity of protein signals irrespective of protein concentration in the tissue. This is of particular interest, since the MALDI image resolution can be of the same order of magnitude as the matrix layer morphology (such as clusters of matrix crystals). In such a case, higher intensities are expected from larger matrix crystal clusters.

Some of the random effects influencing image data can be minimized by proper spectra normalization. Not applying normalization in such cases inevitably leads to artifacts such as ion images depicting inaccurate ion distributions and, perhaps more importantly, result in incorrect statistical analyses of the data. However, normalization must be applied with the knowledge that actual, biological variability between samples could be wrongly interpreted as a systematic error and consequently obscured by the automatic application of normalizations procedures.

One of the most commonly applied normalization procedures in mass spectrometry is the normalization on the TIC. Here, all mass spectra are divided by their TIC so that all spectra in a dataset have the same integrated area under the spectrum. This normalization approach is based on the assumption that there are comparable numbers of signals present in each spectrum. This assumption is fulfilled in homogenous samples (e.g., in serum biomarker studies), where only a few individual peak intensities change against an otherwise constant background [11]. In MALDI imaging, it cannot always be assumed that this condition is met. Different tissue areas or cell types may be present in a sample and express a heterogeneous set of proteins, resulting in quite different ion distributions. As a consequence, TIC normalization can improve our ability to compare expression levels across samples containing similar cell types, but when comparing widely different tissue types, TIC-corrected expression levels may not be applicable.

Under certain circumstances, normalization may be carried out using one or more selected signals with homogenous distribution. This applies in particular if signals from the MALDI matrix itself are present in the recorded mass range, or if (in drug distribution studies) a

closely related compound is externally applied (e.g., sprayed) onto the tissue. In a discovery approach for proteins or peptides, this is not a commonly used procedure. For this reason, we do not discuss the normalization on standards in this publication.

Materials and methods

Description of the datasets used to evaluate normalization algorithms

The datasets discussed here are to some extent atypical, and some have been specifically selected because they produce artifacts with particular normalization approaches.

Rat brain

The rat brain dataset was acquired over a small region of the hippocampus at a lateral resolution of 20 μm . In this sample, HCCA was used as matrix. The matrix layer is formed by clusters of individual matrix crystals. Many of these crystal clusters were larger than the lateral resolution of the imaging experiment. As a result, a non-normalized image therefore overlays the matrix crystal distribution with the ion abundance across the image (see Fig. 1).

Mouse kidney

The mouse kidney dataset was acquired at a lateral resolution of 200 μm . At this scale, the kidney shows three distinct histological structures: renal medulla, pelvis, and cortex. This dataset contains a substantial amount of noise and was selected because the images show a clear

improvement after normalization, which allows for quantitative assessment of the discussed normalization procedures (see Fig. 2 and “Discussion” section).

Mouse pancreas

The mouse pancreas is an example for a sample in which one highly abundant peak is present in confined tissue areas. The islets of Langerhans in the pancreas are small glands producing and secreting insulin, glucagon, and certain other peptide hormones at a high rate. The intensity of insulin peaks in imaging mass spectra acquired from islets is extremely high, typically 60–125 times higher than other signals in the same spectrum. Some other peptide hormones, such as glucagon, are also present in the mass spectra of the islets of Langerhans. These signals are of much lower intensity than the insulin signal, but still more intense than signals of housekeeping proteins in the dataset. Additionally, this sample was embedded in tissue-freezing medium, a treatment known to have a detrimental effect on mass spectra in MALDI imaging by suppressing ion generation [12]. As a result, this dataset shows a high variance in the intensity of the spectra across the tissue. The lateral resolution in this dataset was 200 μm .

Rat testis

Mammalian spermatogenesis is a highly structured and synchronized process taking place in the seminiferous tubules of the testis and is classically divided into three main phases. In the first (proliferative or mitotic) phase, primitive germ cells (i.e., spermatogonia) undergo a series of mitotic divisions. In the second (meiotic) phase, spermatocytes undergo two consecutive divisions to produce the

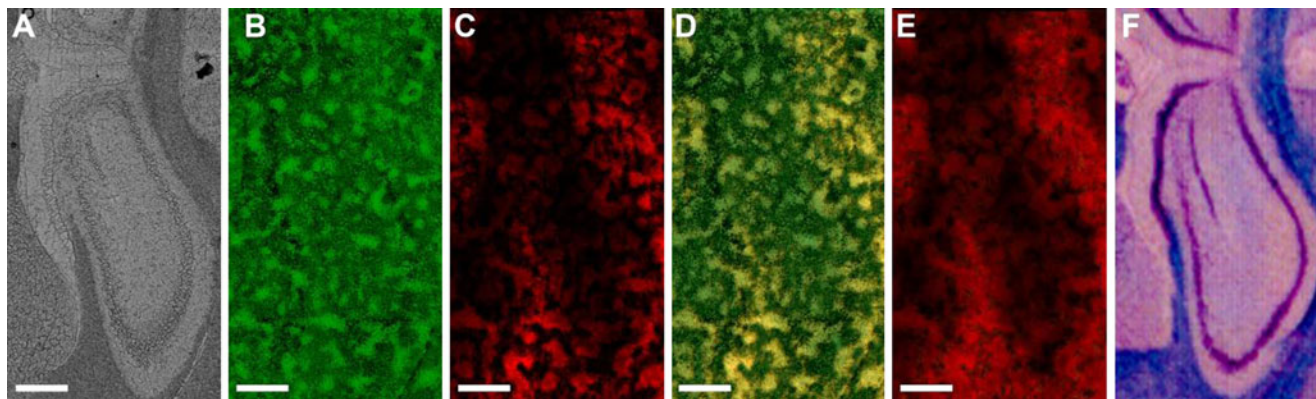


Fig. 1 High resolution imaging of a part of rat hippocampus at 20 μm lateral resolution. *Scale bar*, 500 μm . **A** Optical image of the unstained tissue section prior to the measurement. **B** Optical scan of the matrix morphology (negative image, colored in *green*). **C** Distribution of selected peak m/z 3,530.6 without normalization. **D**

Overlay of **B** and **C**. **E** Distribution of m/z 3,530.6 (from **C**) after normalization on the vector norm. **F** Luxol Fast Blue/Cresyl Violet stain of a similar section, myelin stained in *blue*. Adapted by permission of Mcmillan Publishers Ltd., *J Cereb Blood Flow Metab* 20:563–582, copyright 2000

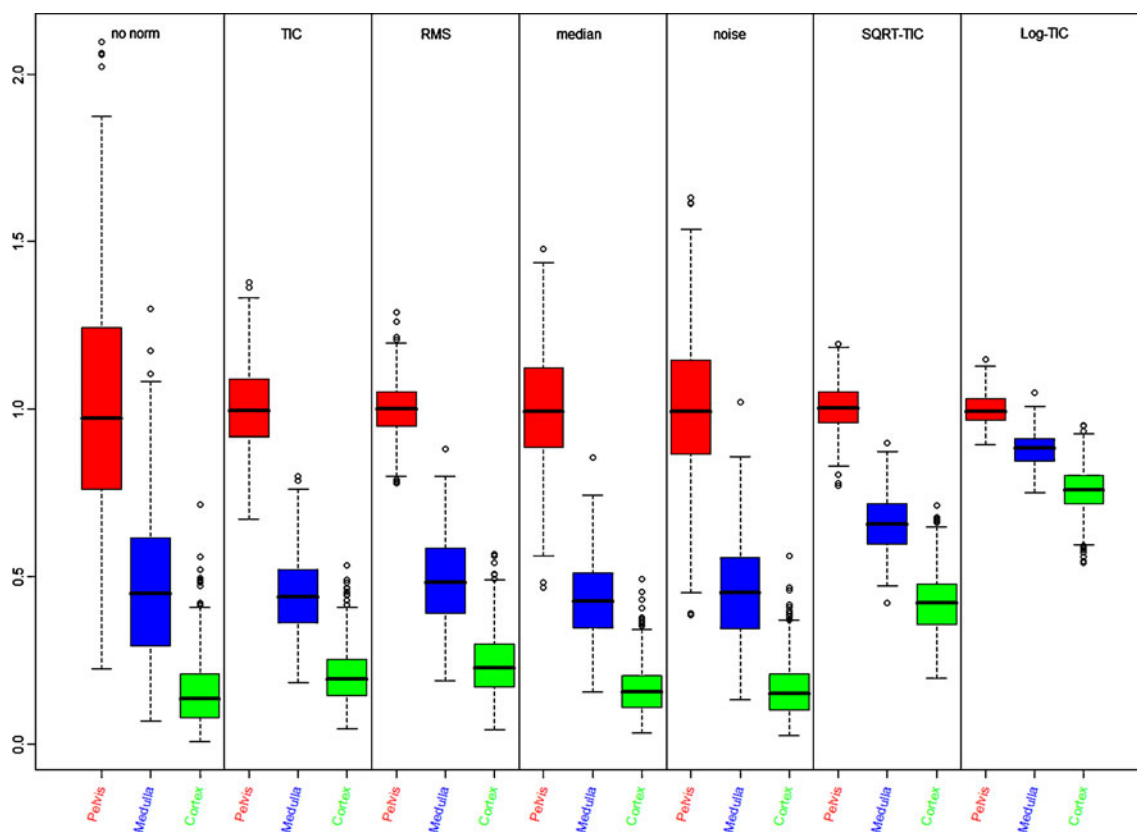


Fig. 2 Intensities of m/z 13,780 in the different regions in the kidney before and after normalization (*red*: pelvis, *blue*: medulla, *green*: cortex). The intensities have been scaled to the mean of the signal in the pelvis region

haploid spermatids. Finally, spermatids differentiate into spermatozoa in the third (spermiogenesis) phase.

An intriguing feature of spermatogenesis is that the developing germ cells form associations with fixed compositions or stages, which constitute the cycle of the seminiferous epithelium. In rats, this cycle is classified into 14 stages, designated I to XIV [13], and occurs along the longitudinal axis of each tubule. Thus, a cross-section of a single seminiferous tubule along its longitudinal axis will display a single cell association or stage. In the rat testis section analyzed here, many tubules have been cross-sectioned, and consequently, different stages are visible. Some of these show a uniquely intense signal at m/z 6,263.

The high spatial resolution (20 μm) needed to resolve substructures in the seminiferous tubules was obtained using HCCA as matrix. This matrix forms small crystals but leads to broad protein signals in linear mode MALDI measurements. The intense peak at m/z 6,263 is not as intense as the insulin peak in the pancreas dataset, but since it is relatively wide, it contributes significantly to the TIC. A histological image of the tissue is shown in Fig. 5.

Importantly, in both the pancreas and the testis datasets, the highly abundant signals are related to real histological structures (islets of Langerhans and specific stages of spermatogenesis in seminiferous tubules). In cases like

these, it is easily possible to mistake a normalization artifact for biologically meaningful information. A peak which is actually present at the same abundance across the entire tissue may wrongly display a localized distribution after normalization. In the testis dataset, this could be misinterpreted as a protein differentially regulated in a particular stage of the seminiferous epithelial cycle.

These two datasets (pancreas and testis) are the most extreme ones we have observed so far with regard to normalization artifacts.

MALDI imaging measurements

Cryosections of the tissues were cut in a cryo-microtome (Leica CM1900-UV) at a thickness of 10 μm and transferred onto conductive indium-tin-oxide-coated glass slides (Bruker Daltonik, Bremen, Germany). The sections were vacuum-dried in a desiccator for approximately 15 min then washed two times in 70% ethanol and once in 96% ethanol for 1 min each. The sections were then dried and stored under vacuum until the matrix was applied.

The sections were coated with matrix using an ImagePrep (Bruker Daltonik) according to the manufacturer's standard protocols. The brain and testis samples were coated with α -cyano-4-hydroxy-cinnamic acid (Bruker Daltonik),

while the pancreas sample was coated with sinapinic acid (Bruker Daltonik).

All mass spectra were acquired in linear mode on autoflex or ultraflex instruments equipped with smartbeam (pancreas) or smartbeam II lasers (all other samples; Bruker Daltonik). For each pixel, 200 laser shots were accumulated at constant laser energy.

Transformation and normalization

Intensity transformations

If a particular peak can be matched (according to mass) across two or more mass spectra from different tissue areas, this peak's intensity is an estimation of the abundance of the same molecule. However, these estimates may contain errors resulting from noise (e.g., differences due to matrix thickness, ion suppression artifacts, or electronic noise).

The observed error can depend on the observed intensity. Any statistical model would either directly account for the variances or transform the data so that the variances are approximately equal for all peak intensity levels. In an earlier study, we examined which peak intensity transformations lead to equal variance for all intensity levels in MALDI mass spectra [7]. The two transformations examined were the square root or the logarithm of peak intensities. In this work, we employed these two transformations followed by normalization on the TIC of the transformed spectra in addition to using raw peak intensities.

Normalization

For the calculation of experimental normalization approaches, the raw spectra were subjected to a Tophat baseline subtraction [14] and exported as *xy*-values to text files. The calculations were performed on these text files with a custom C# script that calculates the normalized intensities. For selected mass signals, these values were written in a tab-delimited table, with the pixel coordinates as rows and outcome of the normalization as columns. Afterwards, these tables were imported into the flexImaging Software (Bruker Daltonik) to reconstruct the normalized images.

Normalization options

A mass spectrum is a vector of intensity values:

$$\vec{s} = y_1, y_2, \dots, y_n.$$

For the normalization, the mass spectrum is divided by a normalization factor (f):

$$\vec{s}_{normalized} = \frac{1}{f} \vec{s}$$

p -norm, TIC, and vector norm

Both the normalization on TIC and the vector norm are special cases of the so-called p -norm.

$$f = \left(\sum_i |y_i|^p \right)^{\frac{1}{p}}$$

For $p=1$, this normalization will be based on the sum of all intensity values in the mass spectrum (i.e., TIC). For $p=2$, p -norm equals vector norm, which is used in mass spectrometry for library searches [15] or in LC-based metabolomics [16]. For $p \rightarrow \infty$, this formula leads to the maximum norm, in which the normalization is done on the most intense peak of the mass spectrum (this has been used to normalize mass spectra for library searches [15]).

With increasing values for p , higher intensity signals have more impact on the result of the normalization. This effect is also observed for noise spectra. In the maximum norm, the highest intensity value in a noise spectrum will be normalized to the same level as the highest intensity peak in other spectra. Noise spectra are therefore considerably amplified with increasing p and are therefore expected to be least problematic in a TIC normalization approach (lowest exponent $p=1$). If spectra with a different number of data points are to be compared, then the mean intensity or the RMS has to be used instead of the TIC or vector norm, respectively.

TIC and vector norm with manual exclusion of mass ranges

As artifacts in the vector norm or TIC normalization are usually the results of mass signals with high intensity and large areas under the peak in certain regions of the tissue, one way to deal with this problem is to exclude these peaks prior to calculating the normalization factor.

For the calculation of the normalization factors with excluded mass ranges, the intensity values of the mass spectra have been transformed as follows:

$$\tilde{y}_i = \begin{cases} 0, & i_{lower} < i < i_{upper} \\ y_i, & else \end{cases}$$

The boundaries of the exclusion mass range (i_{lower} , i_{upper}) were defined by a detailed inspection of the dataset in order to exclude high intensity/area signals which lead to artifacts in vector norm or TIC normalization. The remaining mass spectrum was normalized as described above.

Median

Since the noise level calculation can be affected by operations like smoothing and especially binning, which are often part of a MALDI imaging workflow, we have also investigated normalization on the median, which should be robust to these preprocessing methods and is expected to be a measure for the intensity of the baseline. Median normalization has previously been used in label-free proteomics approaches [17,18].

The median value of all intensities in the spectrum is used for the calculation of the normalization constant:

$$f = \text{median}(y_i).$$

Noise level

Other normalization approaches are not based on peak areas or intensities but on the variance in the data. These approaches aim for constant variance in the data [19]. Using such an approach, it might be possible to circumvent the inherent dangers of the TIC calculation without the need of user intervention.

Wavelet shrinkage, a signal de-noising technique, is frequently used to smooth chromatographic or mass spectrometric signals [20,21]. It employs the universal thresholding method [22] to derive an estimate of the noise in the spectrum. In this method, the noise level of a signal is estimated from the detail coefficients d of the finest scale.

The detail coefficients of the finest scale can be determined without computing the full wavelet decomposition of the signal. In case of the Haar wavelet decomposition, the detail coefficients are differences of consecutive points in the spectrum and are given by:

$$d_i = y_i - y_{i-1}.$$

The universal thresholding uses an estimate of the common standard deviation of the noise given by the median absolute deviation of the detail coefficients:

$$f = \text{median}(|d_i - \text{median}(d_i)|)$$

We used this estimate of the noise level as normalization factor.

Results

Figure 1A shows an optical image of a transversal rat brain section in the region of the hippocampus. The lateral resolution of the measurement was 20 μm , which means that the matrix structures (especially clusters of crystals)

can be resolved by the laser. The myelinated area appears in a darker shade in the optical image. Figure 1B illustrates this by showing an optical scan of the matrix layer after the MALDI measurement (this is a negative B/W scan converted to black/green). Figure 1C shows the intensity distribution of a protein signal (m/z 14,129) of a myelin basic protein (MBP) 14-kD isoform, outlining the histological structure of the myelinated area. However, it can be seen that the coarse, granular structure of the matrix interferes with this signal. The overlay of the mass signal and the matrix image confirms that the distribution of the observed mass signal is strongly influenced by the matrix crystal clusters (Fig. 1D). After TIC normalization, the distribution of the same mass signal now appears much smoother and in better agreement with the optical image, while the overall distribution of the signal is the same as in the non-normalized image (Fig. 1E). Figure 1F shows a Luxol Fast Blue/Cresyl Violet stain of a similar section. Luxol fast blue stains myelin blue, which shows that the myelin is indeed a smooth structure.

The mouse kidney contains several relatively large and well-differentiated histological regions that are homogeneous when measured at low spatial resolution. This dataset is thus well suited to evaluate the results of different normalization approaches. We selected one mass signal (m/z 13,780) that was present in all three major anatomical regions (medulla, pelvis, and cortex) with a rather homogeneous distribution within each region. The mass spectra from these regions were subjected to the different normalization approaches, and the intensities of this selected mass signal were compared (Fig. 2; the images after normalization and an overview of the selected regions are found in the Electronic Supplementary Material Fig. S1). This allowed comparing the observed intensity ratios and variability of the selected signal. By close inspection of the data, we see that with the exception of the intensity transformations, the variance increases with increasing intensity (heteroscedasticity), and that especially in the non-normalized case, the distribution of the intensities within a tissue is not symmetric. It can be seen that after the square root and log transformation, the quantitative relationships between the different regions are changed, but the data after square root transformation appear homoscedastic.

Table 1 shows the relative intensities of this mass signal in the different regions, after scaling to the mean of the intensity in the pelvis region.

The variances of intensities within the tissue after normalization using TIC, RMS, or median is significantly reduced compared with the non-normalized sample. Furthermore, the overlap between the distributions is reduced (Fig. 2). Hence, normalization helps to better resolve differences between tissues.

Table 1 Mean intensities and standard deviation (in parentheses) of the m/z 13,780 signal in the kidney dataset in the different kidney regions

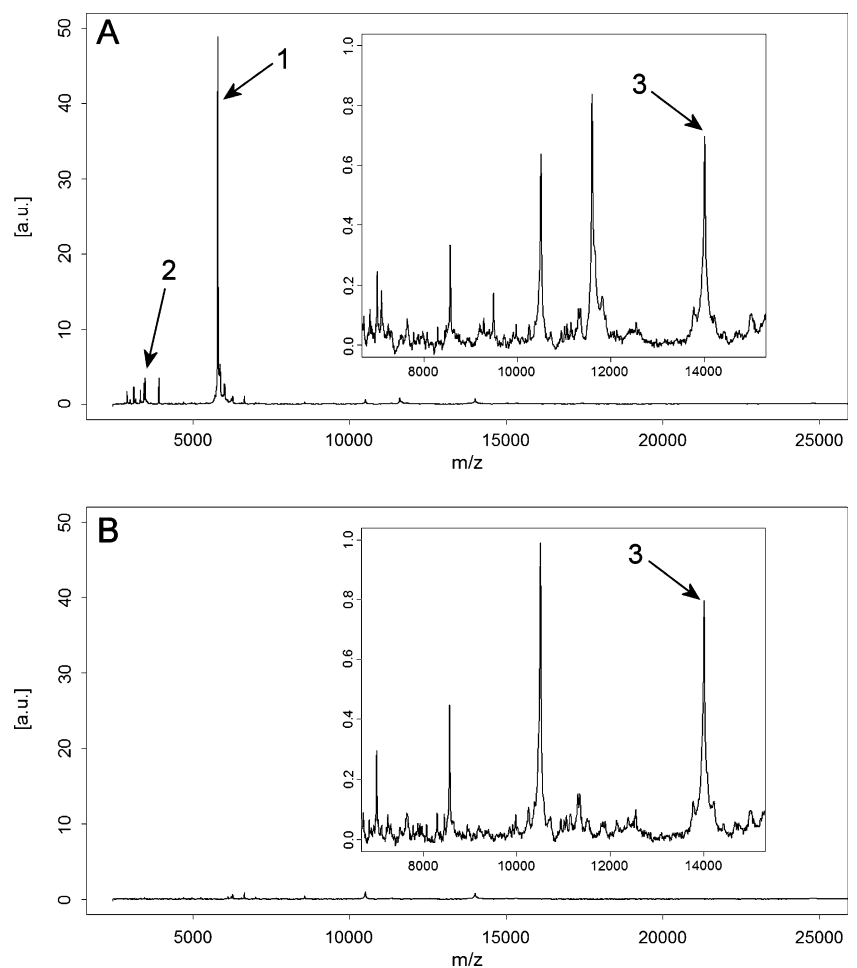
| | No normalization | TIC | Vector norm (RMS) | Median | Noise | Square root | Log |
|---------|------------------|--------------|-------------------|--------------|--------------|--------------|--------------|
| Pelvis | 1 (0.36) | 1 (0.12) | 1 (0.087) | 1 (0.17) | 1 (0.22) | 1 (0.073) | 1 (0.046) |
| Medulla | 0.47 (0.22) | 0.44 (0.11) | 0.49 (0.13) | 0.44 (0.13) | 0.46 (0.14) | 0.66 (0.086) | 0.88 (0.049) |
| Cortex | 0.15 (0.10) | 0.20 (0.081) | 0.24 (0.096) | 0.16 (0.074) | 0.16 (0.081) | 0.42 (0.090) | 0.75 (0.067) |

The intensities were scaled to the mean intensity in the pelvis region to facilitate the comparison of the different normalization approaches

The third dataset we analyzed is a section of the mouse pancreas, where high intensity signals of insulin are present. When comparing averaged spectra from an islet of Langerhans (Fig. 3A) with an averaged spectrum from different region (Fig. 3B), it becomes evident that the intensity of the insulin signal ($\sim m/z$ 5,800) in the islet region is very high compared with other signals. Some other peptide hormones show signals which are still intensive in the islets of Langerhans, while other “non-hormone” signals show similar intensities in both regions (Fig. 3A, B inset). Especially the 14,014-Da signal marked as “3” in both Fig. 3A and B shows similar intensities for both areas.

We compared the distribution of insulin (Fig. 4A, C, E, G, I, and K) and of the ubiquitous 14,014-Da protein (Fig. 4B, D, F, H, J, and L), as visualized using raw data and after normalization. Normalization to the vector norm (Fig. 4C, D) generates obvious changes when compared to the non-normalized images (Fig. 4A, B). Both the spatial distribution and the intensity of the insulin signal appear inflated in the islets, while at the same time, the ubiquitous protein appears to be absent. In contrast, normalization on the TIC (Fig. 4E, F) is in a better agreement with the raw data, only in one islet of Langerhans, an attenuation “hole” (indicated by arrow) appears in the distribution of the 14,014-Da signal. When the TIC with the exclusion of the

Fig. 3 **A** Average mass spectrum of one islet of Langerhans. **B** Average spectrum of a “normal” area on the pancreas. The spectra are on the same absolute scale. *Inserts:* magnified part of the spectrum. *Arrows* indicate (1) insulin signal (m/z 5,800), (2) group of masses related to other peptide hormones (e.g., glucagon), and (3) m/z 14,014 Da signal that shows a similar intensity in both areas



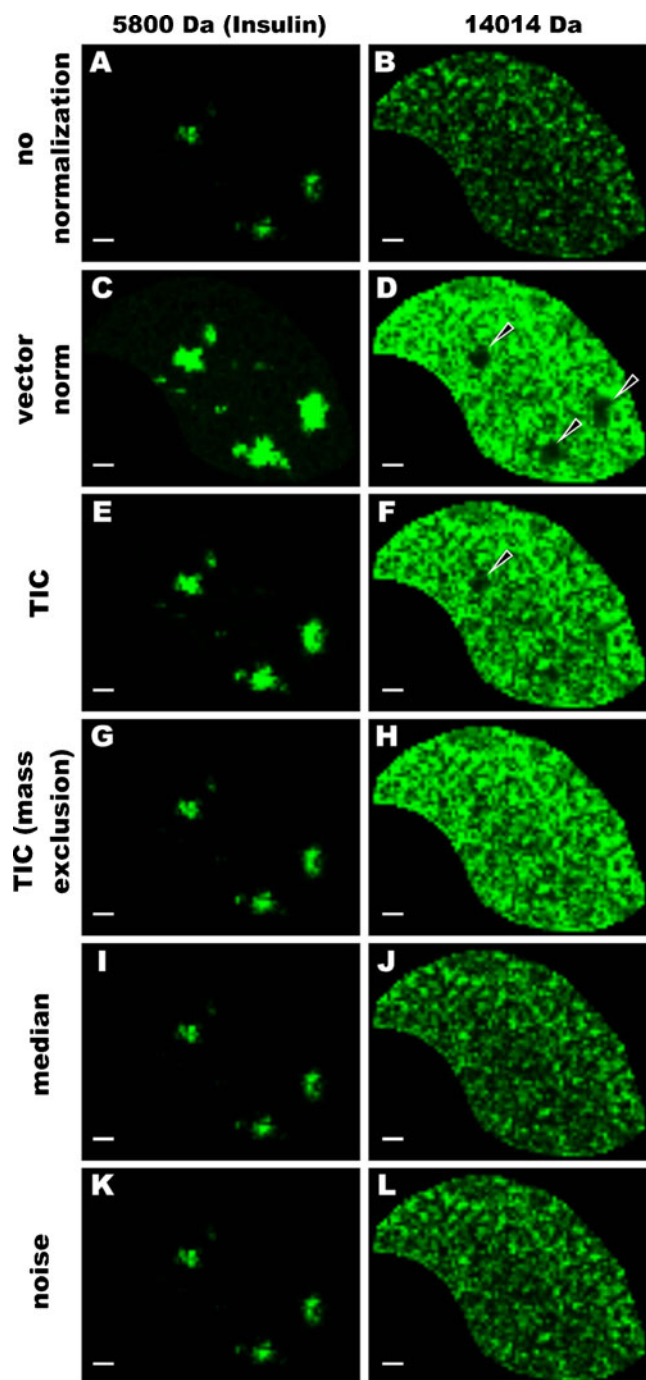


Fig. 4 MALDI images of the insulin signal at m/z 5,800 and the ubiquitous signal at m/z 14,014 in the mouse pancreas after application of various normalization algorithms. For the “TIC with mass exclusion” algorithm, the mass range of the insulin signal was excluded from normalization. *Arrows* indicate artificial m/z 14,014 Da signal attenuation. *Scale bar*, 500 μm . Reconstruction of images was on the highest intensity in a range from m/z 5,788 to m/z 5,812 and from m/z 13,979 to m/z 14,049, respectively. A linear color gradient was used. Full brightness starts at 60% relative intensity

insulin signal is used for normalization (Fig. 4G, H), no holes in the distribution of the ubiquitous 14,014-Da signal are present. Furthermore, the distribution of this 14,014-Da

signal appears to be smoother than in the pre-normalized case. Normalization on the median and the noise level (Fig. 4I, J, K, and L) does not seem to change the visualization when compared to the non-normalized images.

On the rat testis data, the situation is different. This dataset was acquired at a high spatial resolution of 20 μm , requiring the use of HCCA matrix, which forms relatively small crystals. The drawback of HCCA matrix in linear mode is that it leads to rather broad peaks. In the testis section analyzed, seminiferous tubules and a blood vessel can be seen in the cross-section (Fig. 5). Individual tubules display germ cells at different stages of maturation, which are known to be associated with different protein expression, and thus show different molecular signals. In the region imaged, there is a group of tubule sections characterized by a highly intense signal at m/z 6,263 (see Fig. 6A, B). This signal is not as intense as the insulin signal in the pancreas sample, but due to increased peak width, it is the main contributor to the total area (Fig. 6). Figure 7 shows the distribution of a signal at m/z 4,936 with a homogeneous distribution in the testis tissue, with exception of the blood vessel. The non-normalized image (Fig. 7A) again shows mainly the distribution of the matrix crystals overlaid with the distribution of the signal. After normalization to vector norm or TIC (Fig. 7B, C), the signal seems reduced in certain seminiferous tubules. However, if we applied normalization on the TIC with the exclusion of the aberrant signal, normalization on the noise level, or on the median, no

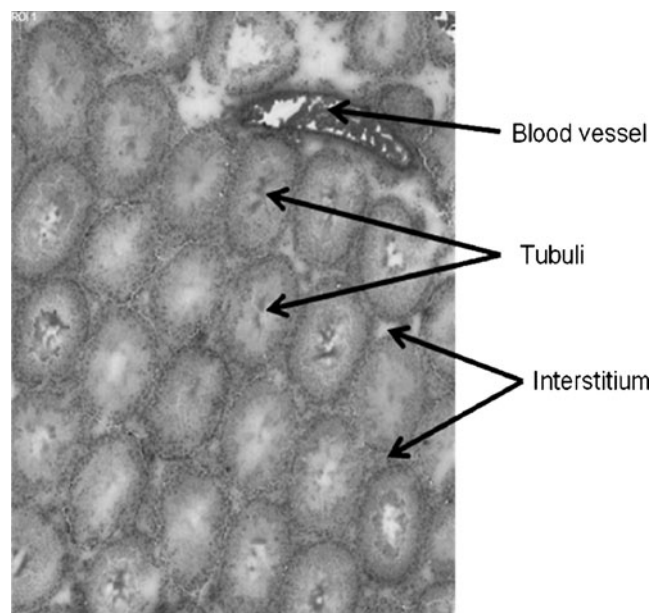
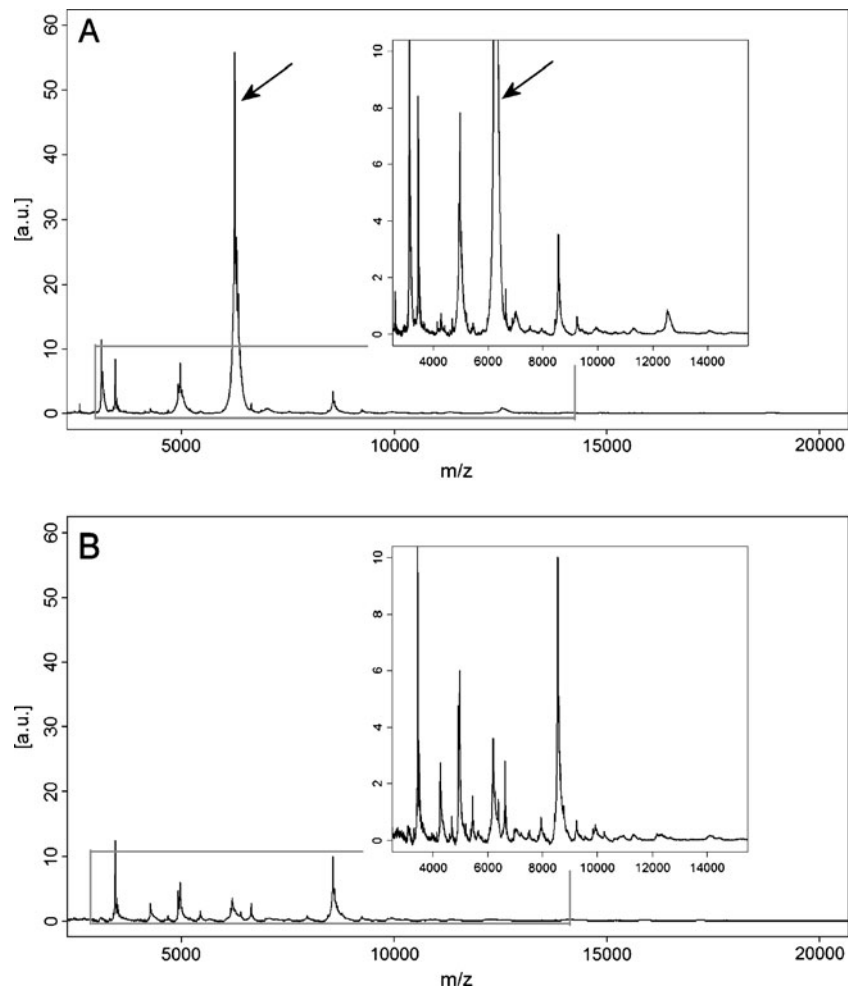


Fig. 5 Microscopic image after H&E staining of the adult rat testis. This image was obtained after the MALDI measurement and shows the same area that is shown in the MALDI images of this dataset in Fig. 7

Fig. 6 Average spectra of 16 individual spectra from the rat testis dataset. **A** From a seminiferous tubule showing the intense signal m/z 6,263 that causes artifacts in normalization (marked with *arrow*). **B** From a “normal” seminiferous tubule



suppression of the signal can be observed in the seminiferous tubules (Fig. 7D–F).

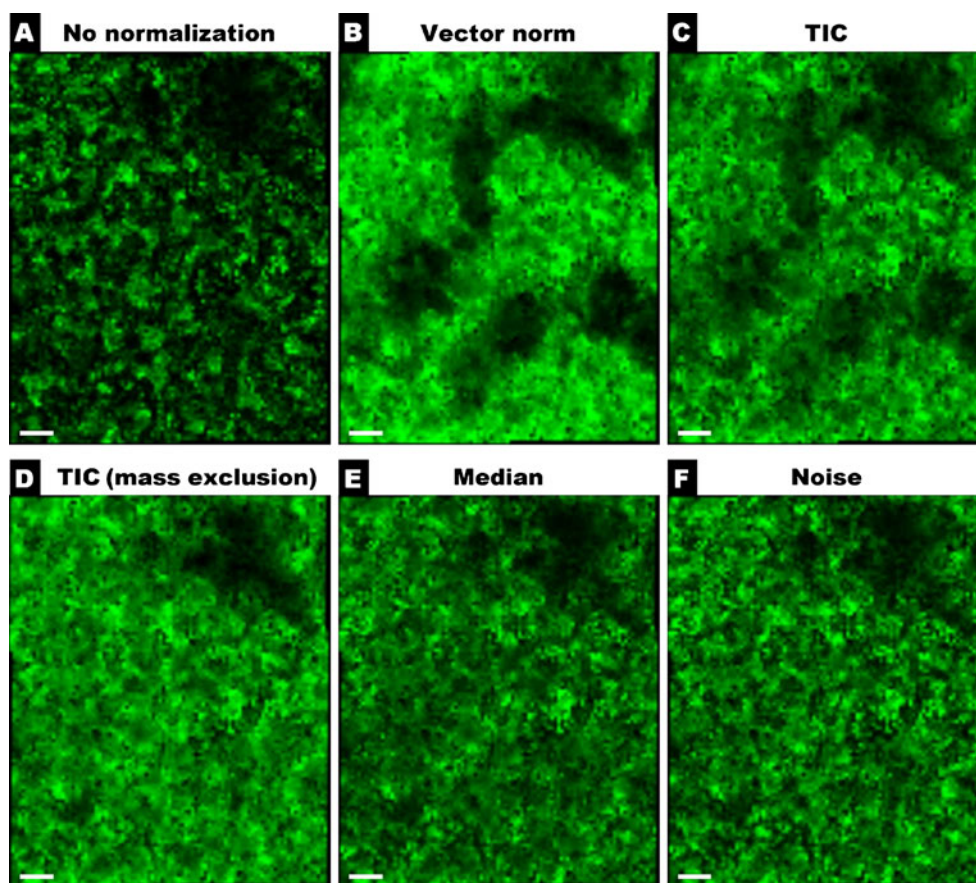
If we compare the images for median or noise level normalization, they look almost identical. The distribution of the selected signal (m/z 4,936) is similar with the non-normalized image but has a less coarse structure caused by the matrix layer. The normalization on the TIC with manual exclusion of the aberrant peak (Fig. 7D) shows the smoothest distribution. This finding is consistent for other masses as well (see Electronic Supplementary Material Figs. S3, S4, S5, S6, S7, S8, and S9).

Analyzing a different signal from the same sample, we observed that normalization can potentially “invert” the intensity ratio of the same mass signal in different regions (Fig. 8). The molecular signal at m/z 6,177 is present only in some seminiferous tubules in the non-normalized image (Fig. 8A, brighter regions indicated by arrows). After normalization on TIC (Fig. 8C) or vector norm (Fig. 8B), the signal shows the highest intensity in the interstitial spaces, but not in the seminiferous tubules as in the non-normalized image. However, by applying TIC normalization with the exclusion of the aberrant signal (m/z 6,263; Fig. 8D) or

normalization to median (Fig. 8E) and noise, respectively (Fig. 8F), the signal is most abundant in the seminiferous tubules but still visible in the interstitium. As described above, normalization using the manually corrected TIC is least affected by the distribution of the matrix crystals and shows the least noisy image. Note that without any normalization, it is not possible to detect the characteristic presence of this signal in the interstitium. (Note: The aberrant m/z 6,263 peak and the m/z 6,177 peak overlap due to the large peak width. A detailed comparison of these peaks is available in the Electronic Supplementary Material Fig. S2.)

Table 2 shows the correlation coefficients between different normalizations, computed from all normalization factors f of all spectra in the testis dataset. Since TIC with manual mass exclusion gave the best result for the testis dataset, this normalization was considered as reference. We observe that the correlation between TIC with exclusion and median or noise level normalization factors is high (>0.91), indicating that these normalization will give similar results, without the requirement for manual interaction with the dataset. The correlation of TIC and TIC with mass exclusion factors is lower (0.88). The normalization factors obtained

Fig. 7 MALDI images of m/z 4,936 from rat testis generated using different normalization approaches. For the TIC with mass range exclusion, the aberrant signal at m/z 6,500 indicated in Fig. 6 was excluded. Scale bar, 200 μm . A linear color gradient was used. Images reconstructed on the highest intensity in the m/z 4,922 to m/z 4,948 range. Full brightness starts at 60% relative intensity



by computing the vector norm of the spectrum have the lowest similarity with all other normalization factors.

One possible problem in all normalization approaches is spectra that do not contain the “full,” non-truncated” noise. Undesired influences during sample preparation or data acquisition can lead to an asymmetric baseline. Partially adhered, uneven tissue surfaces or inhomogeneous matrix deposition can cause changes in the number of ions generated, while incorrect detector/digitizer settings of the MALDI instrument control may cut off the lower part of the baseline. In such cases, only the topmost part of the baseline, possibly only electronic spikes, is recorded. Such spectra negatively interfere with all normalization approaches because they have an erroneously low TIC, vector norm, noise level, and median. This condition artificially increases the intensity of such spectra after normalization. If median or noise level reaches zero, the normalization results for these spectra are undefined because of a division by zero. Therefore, such spectra have to be excluded from the dataset prior to normalization.

Discussion

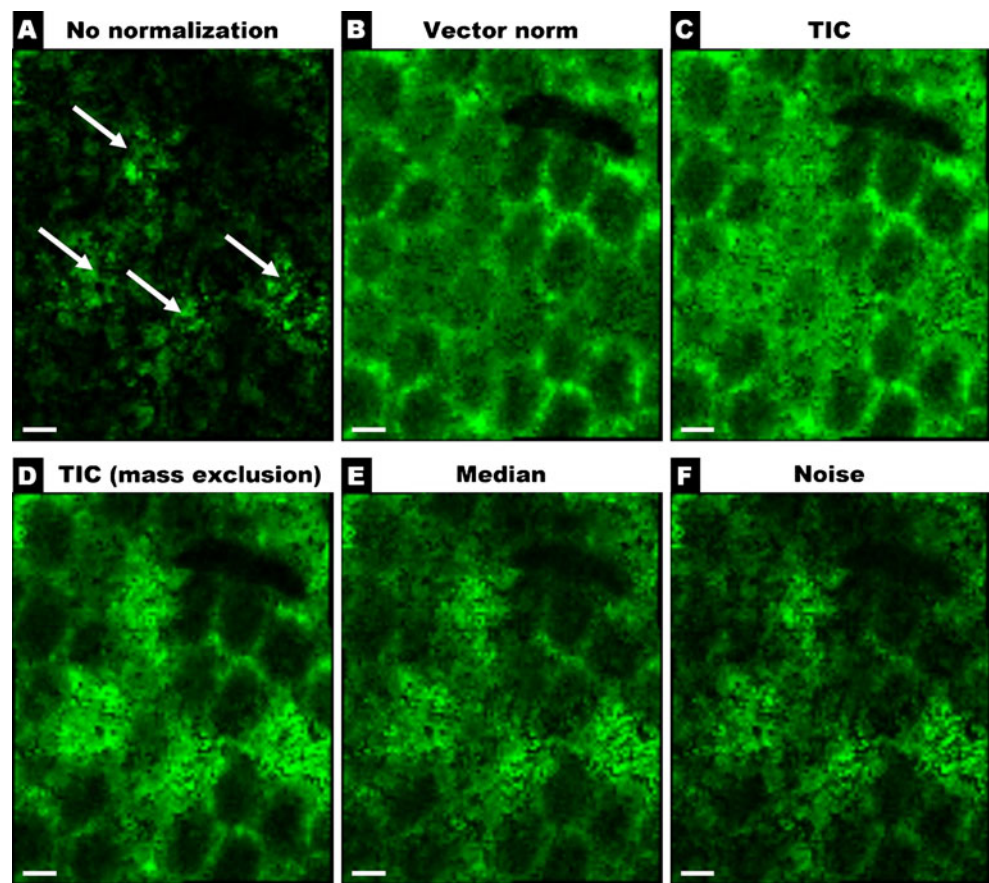
The examples shown make it clear that normalization is necessary to obtain maximum information from certain

datasets, especially if lateral resolution approaches the level of inhomogeneities in the matrix layer, as in the rat brain and the testis datasets presented here. The same may be true if other factors are present that influence the overall intensities of the observed mass spectra, such as different salt or lipid concentrations.

It is necessary to understand that, for all normalization approaches, certain assumptions have to be made about the data. For example, all peak areas are assumed to be similar for normalization on the TIC or overall intensities of peaks should be rather similar when normalizing on vector norm, and the baseline for all peaks should be similar when normalizing on noise level or median.

In mass spectrometry-based serum profiling, where normalization on the TIC is usually used, it is assumed that only a few peaks change throughout the dataset and that the majority of peaks are constant. In the tissue imaging case, this is certainly not true: one can often find completely different protein profiles in different regions of the sample depending on the type of cells or tissues present. Without normalization, assumptions about the data are also made, e.g., that there are no effects such as inhomogeneous matrix layers or disturbing salt or lipid concentrations. The question whether or not normalization is warranted is therefore determined by which of these assumptions is most true.

Fig. 8 MALDI image of m/z 6,177 from the rat testis section in Fig. 5 with different normalization algorithms. For the TIC with mass range exclusion, the aberrant signal indicated in Fig. 6 was excluded. *Arrows* indicate the area of highest intensity in the non-normalized image. *Scale bar*, 200 μm . Images reconstructed on the highest intensity in the m/z 6,165–6,189 range. A linear color gradient was used. Full brightness starts at 60% relative intensity



There are inevitable discussions as to whether normalization is just “cheating,” fabricating images that appear smoother or look nicer. However, it has to be noted that for tissue imaging in contrast to other mass spectrometric profiling techniques, the histology of the underlying samples can provide a measure for judging the effects of data treatment such as normalization. It can easily be evaluated if the resulting images are in agreement with the histology, which thereby provide an independent assessment of the effect of data treatment procedures, if the possibility of artifacts is carefully considered. A quantitative assessment of the normalization is more difficult. Ideally, one would like to compare the result of the MALDI imaging experiment against a known “true” distribution of a measured protein. Unfortunately, there is no method that can yield the “true” quantitative distribution as a reference. The technique that comes closest is immunohistochemistry

(IHC). It has to be noted that MALDI imaging and IHC measure different things: In MALDI imaging, the measured signal is specific for one particular isoform of the protein, including a defined set of post-translational modifications (such as truncations). In IHC on the other hand, all forms of the proteins that contain the epitope are detected. As such, IHC and MALDI imaging both provide valuable information, but cannot be compared quantitatively. Besides this fundamental difference between MALDI imaging and IHC, there are other reasons that make IHC difficult to quantify [23] An artificial sample for the same purpose is also almost impossible to prepare. Such a sample would have clearly separable regions, each of which homogeneous in itself, and ideally expressing marker proteins at a known level. Such reference proteins could be introduced by spraying a known protein onto the tissue, but to do this in a homogenous manner is very difficult. Additionally, it

Table 2 Pearson correlation coefficients between the normalization factors of all 11,057 spectra in the rat testis dataset

| | TIC (mass exclusion) | TIC | Vector norm | Median | Noise |
|----------------------|----------------------|-------|-------------|--------|-------|
| TIC (mass exclusion) | 1 | 0.880 | 0.742 | 0.927 | 0.918 |
| TIC | | 1 | 0.956 | 0.845 | 0.824 |
| Vector norm | | | 1 | 0.678 | 0.651 |
| Median | | | | 1 | 0.972 |
| Noise | | | | | 1 |

cannot be ensured that a protein sprayed on top of a tissue behaves similar to the same protein when it is actually expressed in the tissue. Similarly, if the reference protein is deposited on the sample carrier underneath the tissue, it cannot be ensured that it is homogeneously extracted. For example, cracked tissue or blood vessels may provide a direct access of the matrix to the sample carrier.

Since a true quantitative assessment of the normalization seems therefore not feasible, we have chosen a slightly simpler approach: We do not compare the result of the normalization against an unknown “true” distribution, but against the non-normalized result and the histology for comparison. A global change in the coarse “overall” distribution of the signal would then be considered a normalization artifact, and an improved agreement with the histology on the small scale would be seen as a positive effect of the normalization. Please note that this approach does not try to make any statement on whether this reflects the “true” distribution of the protein in question or if there might be specific ion suppression effects for this signal present in certain tissue regions; such a discussion would be outside the scope of this article. This approach has another limitation: The results of our evaluations are strictly speaking only applicable to global intensity differences due to the sample preparation. The applicability to the situation where intrinsic compounds such as lipids or salts cause a global intensity change in the spectra remains to be investigated.

The example in Fig. 1 shows that in this example, the reconstruction of the molecular image for the MBP 14-kDa isoform reflects the structure of matrix layer on top of the myelinated area in this brain. After normalization, the molecular image is in better agreement with the histology that shows a smooth distribution of the myelin.

We used the mouse kidney, which has three major regions, and selected one mass signal that is present in all three regions in a rather homogenous intra-regional distribution. We now conclude that if we do not observe a significant change of the observed intensities for this signal before and after the normalization, then the normalization does not produce an artifact. If on the other hand the signal scatters less inside a homogenous region, then the normalization provides an improvement to the data and is therefore warranted. As seen in Fig. 2 and Table 1, the normalization on the TIC, vector norm (RMS), median, and noise level do not significantly change the quantitative information, but they all provide a lower scattering of the mass signals, especially if the signal is of high or medium intensity. The log or square-root transformations on the other hand change the quantitative relationships significantly, so they cannot be recommended for interpreting MALDI images. It has to be noted, though, that multivariate techniques such as principal component analysis, clustering, or support vector

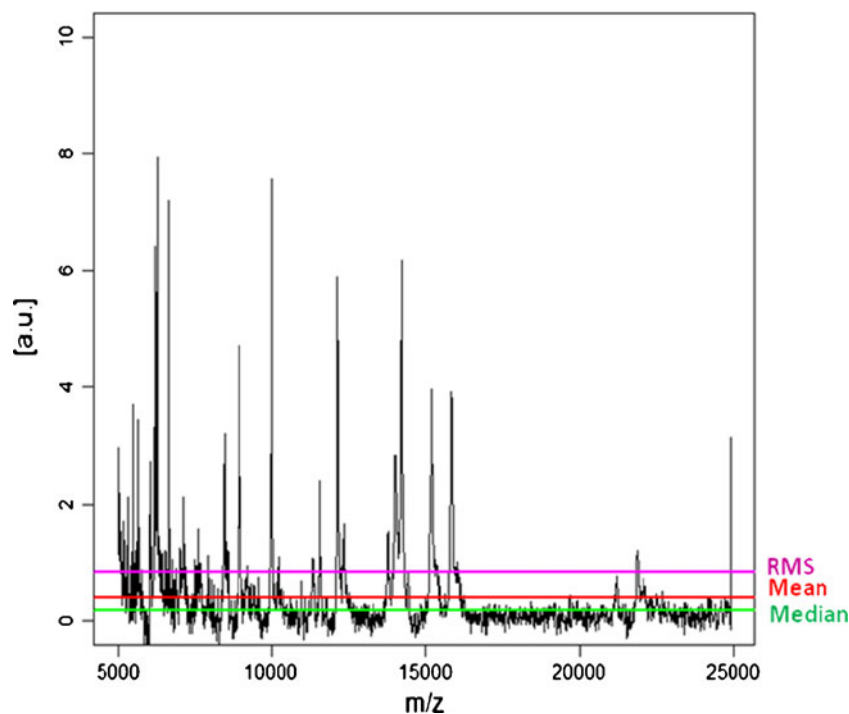
machines become increasingly popular for a concise representation of MALDI imaging data [19,24–28]. For these techniques, homoscedasticity, as well as symmetry and a normal distribution of the variance, is more important than the correct quantitative ratios between mass spectrometric features, so it is common to either scale or transform the peak intensities prior to the calculation [29]. Especially for multivariate treatment of MALDI imaging data, the square root transformation can be considered for the data preparation because for all three intensity groups in Table 1, the variance is approximately constant.

It can also be observed that after normalization on the vector norm (RMS) and TIC, the data scatters less than after median and noise calculation, which is in line with the common expectation that parametric approaches work better than nonparametric ones if they are applicable. It can be concluded that for a dataset like the kidney example shown here, normalization is not an arbitrary manipulation of results but an analytical necessity.

The mouse kidney dataset was relatively straightforward, especially because there are no exceedingly strong signals for any sample regions. This was not the case for the mouse pancreas and the rat testis datasets. Both datasets show unusually intense signals that are confined to specific areas and which lead to artifacts in TIC or vector norm normalization. These artifacts are very dangerous in routine work because they appear to be in agreement with histology. One may easily come to the conclusion that the m/z 14,014 signal in the mouse pancreas is not present in the islets of Langerhans when studying the images after TIC or vector normalization in Fig. 4C and E. Likewise, one may easily conclude studying the images of Fig. 7B and C that the m/z 4,936 signal in the testis is regulated in sperm maturation, while in reality, it is not. With respect to the limitations of detecting the “real” distribution of the selected analytes (discussed above), we judged the presence of artifacts by comparison of normalized with the non-normalized images.

With the limitations discussed above for the applicability of TIC- or RMS-based normalization, we cannot expect those to be valid for many MALDI imaging datasets. However, this is in marked contrast to our experience in routine work, where these procedures do not frequently produce obvious artifacts. This is most likely due to the fact that imaging spectra contain not only signals but also a significant amount of baseline (or noise). Figure 9 shows one spectrum from the kidney dataset with the RMS, average, and median intensity lines. In an example spectrum taken from the kidney dataset, the average line (reflecting a TIC normalization) is only slightly above the baseline and significantly below the intensive peaks (Fig. 9). Therefore, we can conclude that the TIC normalization factor is, to a large part, determined by the

Fig. 9 Typical single mass spectrum from the kidney dataset with lines indicating the RMS intensity (purple), mean intensity (TIC, red), and median intensity (green) of the spectrum



baseline with only a moderate contribution of the peak intensities. This may be the reason why in practice, TIC normalization works well for most datasets. This is supported by the mouse pancreas dataset where some peptide hormones other than insulin are present in the spectra recorded from the islets of Langerhans, which are still more intensive than all other signals in the dataset. Even without the exclusion of those signals from the TIC calculation (only the much more intense insulin signal has been excluded), no visible artifacts appear after normalization on TIC. The importance of the baseline and of the noise is probably surprising in TIC normalization, but it is obvious for median or noise level normalization. Noise in MALDI-time-of-flight (TOF) comes from several reasons, including impurities in the sample, but to a large extent from metastable matrix clusters [30]. It is a reasonable assumption that the noise level is therefore influenced to a large extent by the matrix layer, and consequently, a normalization on the noise can remove the influence of the matrix layer to some extent. It has to be noted that this chemical noise is usually seen as a broad noise band in MALDI-TOF. In instruments that have a longer timescale for the measurement, the metastable matrix clusters are typically not detected, so in, e.g., FTMS or q-TOF MALDI imaging, our observations are no applicable. A software for data reduction that is based on keeping the signals but removing the noise from the dataset is likely not compatible with the normalization approaches discussed; at least a less robust performance has to be expected. The same is true for spectra in which for technical reasons, the baseline does not contain the full noise. It has also

been observed that the normalization on the TIC is less prone to artifacts compared to the normalization on the vector norm (RMS), which can be strongly influenced by peak intensities (Figs. 4A, B and 9, RMS line).

We also provided examples in which the “routine” approach of normalization (on TIC or the vector norm) produces artifacts. These artifacts were based on the inhomogeneous presence of peaks with unusually high intensities (or areas). These artifacts are particularly dangerous for the interpretation of the data because they reflect true histological differences in the tissue and can therefore lead to wrong conclusions such as reporting up- or downregulation of in fact unregulated peaks. As shown, a suitable way to deal with the described normalization artifacts is to exclude signals that cause it from the calculation of the TIC.

As disadvantage of such an approach, it requires manual interaction with the data: First, the user has to be aware of the problem, then he needs to identify the aberrant signal causing it. The presence of the problem can usually be spotted by the appearance of “holes” in the distribution of the noise (as determined when the image is created on a mass range where no signals are present) or of low intensity signals after normalization. The aberrant signals can then easily be identified by evaluating the spectra of these regions.

We have also observed that the normalization on the median and the noise level are robust against the presence of disturbing signals. Although the images produced after normalization on these values look less smooth than when

normalization on the TIC after a manual exclusion, they do not require manual interaction and are therefore more robust. We thus conclude that they should be considered as routine normalization methods for imaging data. Since the computation of the median and the computation of the noise level produce very similar results, we propose the normalization on the median as the most robust approach, since it is influenced less by common processing steps in MALDI imaging such as binning or spectra smoothing.

In many real-world datasets, we have observed that none of the problems discussed for the testis or Langerhans islets datasets do appear, and normalization to the TIC can be applied without restriction. Because TIC-based normalization seems to be superior if applicable, establishing an automated procedure to decide which normalization should be applied would be desirable. One possibility might be to examine the correlation of the normalization factors as was done in Table 2. For the testis dataset, normalization on the TIC with exclusion of the problematic mass resulted in the best visualization of the dataset. The best correlation was observed with the median normalization. We suggest using the correlation of a non-parametric normalization factor (median), with the TIC normalization factor as an indicator for problems with the TIC-based normalization. If this correlation is low, a non-parametric normalization method or TIC with exclusion should be applied; otherwise, the TIC normalization is recommended.

It may be possible to define a “cut-off” value for the correlation between TIC and median that could automatically evaluate the applicability of a standard TIC-based normalization. Defining such threshold needs additional research and the analysis of more problematic datasets and also with different matrix preparations and detection modalities. A direct visual comparison between median and TIC normalization can serve the same purpose until an automatic assessment is available. Once the inapplicability of TIC-based normalization is established, it is usually simple to spot the problematic peaks for a manual exclusion.

Conclusion

In the datasets presented here, conventional normalization on the vector norm or the TIC lead to artifacts, yet normalization was necessary to deal with inhomogeneities of the matrix layer. Although normalization to the noise level or the median could be used to generate normalized images without artifacts, TIC normalization with the manual exclusion of the signals that caused the artifacts produced the best results. However, this approach requires manual intervention by the user.

In any case, caution is advised when applying TIC normalization. We propose the use of the median normalization as an additional tool to spot artifacts. The comparison of the images after TIC normalization and median normalization is a good way to test the applicability of TIC normalization. If this comparison shows ample differences in the resulting images, then TIC is not recommended.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Baggerly KA, Morris JS, Wang J, Gold D, Xiao LC, Coombes KR (2003) A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples. *Proteomics* 3(9):1667–1672
2. Morris JS, Coombes KR, Koomen J, Baggerly KA, Kobayashi R (2005) Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* 21(9):1764–1775
3. Norris JL, Cornett DS, Mobley JA, Andersson M, Seeley EH, Chaurand P, Caprioli RM (2007) Processing MALDI mass spectra to improve mass spectral direct tissue analysis. *Int J Mass Spectrom* 260(2–3):212–221
4. Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78(3):779–787
5. Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L, Fleisher M, Robbins RJ, Tempst P (2005) Correcting common errors in identifying cancer-specific serum peptide signatures. *J Proteome Res* 4(4):1060–1072
6. Wagner M, Naik D, Pothen A (2003) Protocols for disease classification from mass spectrometry data. *Proteomics* 3(9):1692–1698
7. Wolski WE, Lalowski M, Martus P, Herwig R, Giavalisco P, Gobom J, Sickmann A, Lehrach H, Reinert K (2005) Transformation and other factors of the peptide mass spectrometry pairwise peak-list comparison process. *BMC Bioinform* 6:285
8. Wu B, Abbott T, Fishman D, McMurray W, Mor G, Stone K, Ward D, Williams K, Zhao H (2003) Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics* 19(13):1636–1643
9. McDonnell LA, van Remoortere A, van Zeijl RJ, Deelder AM (2008) Mass spectrometry image correlation: quantifying colocalization. *J Proteome Res* 7(8):3619–3627
10. Seeley EH, Oppenheimer SR, Mi D, Chaurand P, Caprioli RM (2008) Enhancement of protein sensitivity for MALDI imaging mass spectrometry after chemical treatment of tissue sections. *J Am Soc Mass Spectrom* 19(8):1069–1077
11. Fung ET, Enderwick C (2002) ProteinChip clinical proteomics: computational challenges and solutions. *Biotechniques Suppl* (34–38):40–31

12. Schwartz SA, Reyzer ML, Caprioli RM (2003) Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. *J Mass Spectrom* 38(7):699–708
13. Leblond CP, Clermont Y (1952) Definition of the stages of the cycle of the seminiferous epithelium in the rat. *Ann NY Acad Sci* 55(4):548–573
14. Lange E, Gropl C, Reinert K, Kohlbacher O, Hildebrandt A (2006) High-accuracy peak picking of proteomics data using wavelet techniques. *Pac Symp Biocomput*: 243–254
15. Crawford LR, Morrison JD (1968) Computer methods in analytical mass spectrometry. Identification of an unknown compound in a catalog. *Anal Chem* 40(10):1464–1469
16. Sysi-Aho M, Katajamaa M, Yetukuri L, Oresic M (2007) Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinform* 8:93
17. Haqqani AS, Kelly JF, Stanimirovic DB (2008) Quantitative protein profiling by mass spectrometry using label-free proteomics. *Meth Mol Biol* 439:241–256
18. Ting L, Cowley MJ, Hoon SL, Guilhaus M, Raftery MJ, Cavicchioli R (2009) Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling. *Mol Cell Proteomics* 8(10):2227–2242
19. Meuleman W, Engwegen JY, Gast MC, Beijnen JH, Reinders MJ, Wessels LF (2008) Comparison of normalisation methods for surface-enhanced laser desorption and ionisation (SELDI) time-of-flight (TOF) mass spectrometry data. *BMC Bioinform* 9:88
20. Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung MC, Kuerer HM (2005) Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics* 5(16):4107–4117
21. Perrin C, Walczak B, Massart DL (2001) The use of wavelets for signal denoising in capillary electrophoresis. *Anal Chem* 73(20):4903–4917
22. Nason GP (2008) *Wavelet methods in statistics* with R. Springer, Berlin
23. Walker RA (2006) Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I. *Histopathology* 49(4):406–410
24. McCombie G, Staab D, Stoeckli M, Knochenmuss R (2005) Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal Chem* 77(19):6118–6124
25. Van de Plas R, Ojeda F, Dewil M, Van Den Bosch L, De Moor B, Waelkens E (2007) Prospective exploration of biochemical tissue composition via imaging mass spectrometry guided by principal component analysis. *Pac Symp Biocomput*: 458–469
26. Deininger SO, Ebert MP, Futterer A, Gerhard M, Rocken C (2008) MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *J Proteome Res* 7(12):5230–5236
27. Yao I, Sugiura Y, Matsumoto M, Setou M (2008) In situ proteomics with imaging mass spectrometry and principal component analysis in the Scrapper-knockout mouse brain. *Proteomics* 8(18):3692–3701
28. Deininger SO, Becker M, Suckau D (2010) Tutorial: multivariate statistical treatment of imaging data for clinical biomarker discovery. *Meth Mol Biol* 656:385–403
29. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ (2006) Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7:142
30. Krutchinsky AN, Chait BT (2002) On the nature of the chemical noise in MALDI mass spectra. *J Am Soc Mass Spectrom* 13(2):129–134