

SCIENTIFIC REPORTS



OPEN

Metrics for evaluating the stability and reproducibility of mass spectra

E. S. Zhvansky¹, S. I. Pekov¹, A. A. Sorokin¹, V. A. Shurkhay², V. A. Eliferov¹, A. A. Potapov², E. N. Nikolaev³ & I. A. Popov¹

Received: 13 July 2018

Accepted: 6 December 2018

Published online: 29 January 2019

In this work, we demonstrate a new approach for assessing the stability and reproducibility of mass spectra obtained via ambient ionization methods. This method is suitable for both comparing experiments during which only one mass spectrum is measured and for evaluating the internal homogeneity of mass spectra collected over a period of time. The approach uses Pearson's r coefficient and the cosine measure to compare the spectra. It is based on the visualization of dissimilarities between measurements, thus leading to the analysis of dissimilarity patterns. The cosine measure and correlations are compared to obtain better metrics for spectra homogeneity. The method filters out unreliable scans to prevent the analyzed sample from being wrongly characterized. The applicability of the method is demonstrated on a set of brain tumor samples. The developed method could be employed in neurosurgical applications, where mass spectrometry is used to monitor the intraoperative tumor border.

Mass spectrometry methods^{1–8} are widely used for the analysis of biological and medical samples. Ambient ionization mass spectrometry allows rapid analysis of samples to detect various mixture components, pathogens and disease biomarkers^{9–12}. In particular, it has been shown that brain tissue samples can be classified based on their lipid profiles; this technique is promising in terms of speeding up and improving the accuracy of analysis of biopsy material obtained during neurosurgery^{13–19}. However, in order for mass spectrometry methods to be introduced into routine clinical procedure, all stages of the analysis need to be simplified and automated.

In order to improve molecular profiling methods, one needs to improve the process of analyte microextraction from samples and to optimize the ionization conditions. However, despite the fact that progress has been made in instrumentation to facilitate these improvements, data obtained by molecular profiling is often far from being suitable for subsequent processing, especially when taking into account the need for automated data processing^{20–29}. When applying ambient ionization in projects requiring automated identification (classification) of the studied sample, it is essential to develop an approach for unifying the mass spectra, including improving the stability and reproducibility of measurements. This unification method should aim at eliminating scans with any instabilities and those that do not match the characteristics of the sample^{30–34}. It should be said that if there is heterogeneity in the sample that is characteristic of the sample itself, it must be taken into account when working with the molecular profile; however, ionization artifacts can lead to errors in the classifiers themselves when they are trained on molecular profile databases. ESI and ESI-based methods, such as DESI and spray-from-tissue, are sensitive, for example, to the homogeneity of the solvent flow and the presence of gas bubbles, similar to how MALDI results are sensitive to the uniformity of MALDI application and matrix crystallization. Use of internal standards is the typical approach for spectra quality control and intensity normalization, but it is hardly applicable to many ambient ionization techniques, especially those designed for *in vivo* sampling.

Despite the fact that, when solving issues related to molecular profiling, the majority of attention is usually given to detecting the presence or absence of one or several biomarkers in the sample, the instability of obtained mass spectra and the low reproducibility of the molecular profile (the peak intensity ratios) can make the results of sample analysis unreliable. This is true when automatic identification is applied to the whole mass spectrometric profile, rather than to a single component of the mass spectrum. Due to the instability of the microextraction and ionization processes, the intensity ratios between the peaks in the spectra can vary significantly between successive measurements of one sample; these differences can become even bigger with different samples and

¹Moscow Institute of Physics and Technology, Dolgoprudnyy, Moscow Region, Moscow, Russian Federation. ²Federal State Autonomous Institution «N.N. Burdenko National Scientific and Practical Center for Neurosurgery» of the Ministry of Healthcare of the Russian Federation, Moscow, Russian Federation. ³Skolkovo Institute of Science and Technology, Moscow, Russian Federation. Correspondence and requests for materials should be addressed to E.N.N. (email: ennikolaev@gmail.com)

separate measurements. This can then lead to errors in the automatic analysis of the molecular profile^{35,36}, which takes into account third-party peaks of the complex biological matrix in addition to the characteristic substances. At the moment, there is no solution that would unify the mass spectra, nor is there a criterion for estimating spectrum quality. This makes it impossible to talk about objectivity, both in terms of data for constructing the classifiers and in terms of the results gained from identifying the type of tissue by its mass spectrum with the help of classifiers. Machine-learning methods and methods of statistical analysis usually require that intra-class variability is less than inter-class variability. Presence of highly variable regions in the spectra contributes to the technical variability, thus leading to a higher level in the overall data variability and rendering the application of statistical methods problematic. Therefore, the use of direct mass spectrometry and, especially, the ionization of tissue samples, requires automated methods for monitoring the stability and reproducibility of spectra.

In this paper, one possible approach to the evaluation and quality control of mass spectra, measured using the ambient ionization methods for the task of classifying biopsy material extracted during a neurosurgical operation, is proposed.

Methods

Samples. Tissue samples from primary human brain tumors obtained during the surgery and those removed from patients with a non-tumor pathology (epilepsy) were analyzed. The samples were provided by the N.N. Burdenko NSPCN and analyzed under an approved N.N. Burdenko NSPCN Institutional Review Board protocol. A signed informed consent form, filled out in accordance with the requirements of the local ethical committee specifically noting that all removed tissues can be used for further research, was obtained from all patients before surgery. The study was conducted in accordance with the Helsinki declaration as revised in 2013. All procedures were carried out according to the relevant guidelines and regulations. Samples of tissue removed during surgery were divided into equal parts and half of each sample was subjected to routine hematoxylin and eosin staining and further immunohistochemical analysis, according to which the samples were classified as tumor core (glioblastoma or meningioma) or intact tissue. The other half of each sample was placed in a 0.9% NaCl solution, frozen, and stored at -80°C until analysis. The samples were defrosted at room temperature, split into smaller fragments, and mass spectra of each fragment were measured. Samples used in this paper include two non-tumorous intact brain tissue samples obtained from two patients with drug-resistant epilepsy during surgery; three samples from patients with glioblastoma (samples with necrosis); three samples from patients with fibroblastic meningioma (WHO grade I).

Mass spectrometry. The spectra were measured using a high-resolution Thermo Scientific LTQ FT ULTRA mass spectrometer according to the previously described direct spray-from-tissue ionization method, otherwise known as NESI (Needle ElectroSpray Ionization)³⁷. The tissue sample (about 2 mm^3) was placed at the tip of an injection needle (30 mm in length, 0.6 mm inner diameter). Solvent was then pumped through the needle and flowed around the sample (3–5 ul/min HPLC grade methanol (Merck)). Then, a high voltage (6.0 kV in negative mode) was applied to the solvent stream. The spectra were measured in the range m/z 100–1300, with a mass resolution of 56000 at m/z 800. Each measurement lasted for at least five minutes, what corresponds to about 300 scans per mass spectrum measurement in average.

Data processing. Each mass spectrum was interpreted as an N -dimensional vector. Therefore, the algorithm binned the peaks between m/z 100 and 1300 into 0.01 m/z bins. The binning step was chosen in accordance with the precision of the measurements, which was equal to 2 ppm. This step roughs the data, so the precision of the vector does not exceed the precision of the measurements. A binned spectrum such as this was considered as an N -dimensional vector.

Two characteristics were chosen for comparing the spectra. The first was the Pearson's r coefficient³⁸:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2} \sqrt{\sum(Y - \bar{Y})^2}}, \quad (1)$$

where X and Y are vectors of the binned mass spectra, and $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ — mean of X , X_i - value of i -coordinate of vector X ,

The second characteristic was the cosine measure — the cosine of the spectrum angle³⁸ (angle between two vectors):

$$c = \frac{\sum XY}{\sqrt{\sum X^2} \sqrt{\sum Y^2}} \quad (2)$$

The moving median is able to filter out anomalous values, which is not possible with the moving average algorithm, since it takes into account all values including the outlying ones. Thus to remove the influence of the outliers median filtering (moving median^{39–41}) was applied. For this, each bin in the smoothed spectra was replaced by the median of corresponding bin values of adjacent scans in smoothing window of size N . Smoothing window was shifted by different steps, compressing the information in the spectra. In this work, we have used smoothing windows of size $N = 5, 7, 21$ and 51 and steps of one, half of the window and the whole window size. The Pearson's r coefficient and cosine measure were calculated for the smoothed spectra. All calculations and visualizations were made using a self-written code in the MATLAB software. The source code is available from the authors on request.

The results have been visualized as a correlation matrix or cosine measure matrix. Each pixel of the matrix shows the Pearson's r coefficient between two spectra indexing along the horizontal and vertical axes (Fig. 1).

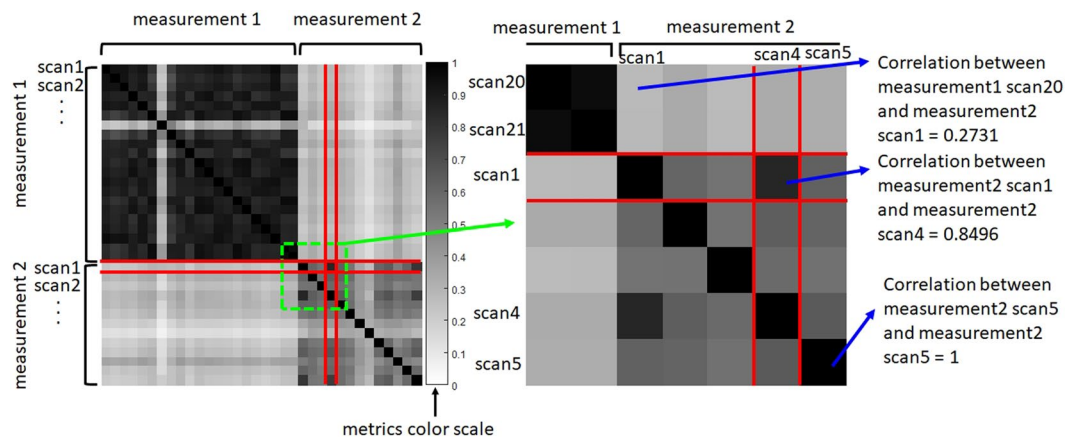


Figure 1. Correlation matrix between two different measurements without smoothing. Dissimilarity metrics of two measurements and inner dissimilarities in each measurement are shown on the left. Scans are grouped by measurements and ordered by time from left to right and from top to bottom within the measurement. The same color scale is used for all following figures in the paper both for Pearson's r coefficients and cosine measures. Selected fragment is shown on the right at the individual pixel scale.

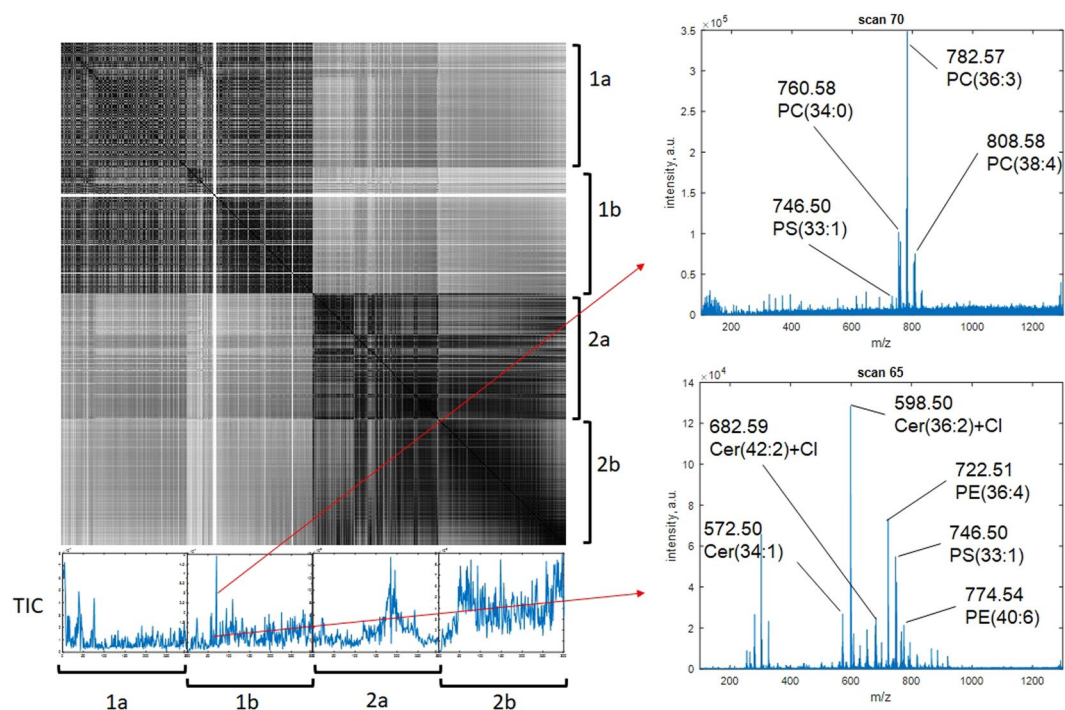


Figure 2. Correlation matrix based on the cosine measure without smoothing. Measurements 1a and 1b correspond to different fragments of one sample classified as a glioblastoma; Measurements 2a and 2b correspond to different fragments of one sample, classified as a meningioma. The total ion current is presented for each measurement, with examples of spectra from two different scans for one of the measurements.

Same colormap and visualization is used for the cosine measure. Scans are grouped by measurements and ordered by the time within each measurement.

Results and Discussion

The correlation matrix (Figs 2, S1) demonstrates that spectra of different fragments of the same tissue sample correlate better with each other than with the spectra of fragments of other tissue samples. There are also white lines on the figure, which are associated with electrospray instability (these lines in the correlogram correspond to anomalous local maxima or minima in the total ion current).

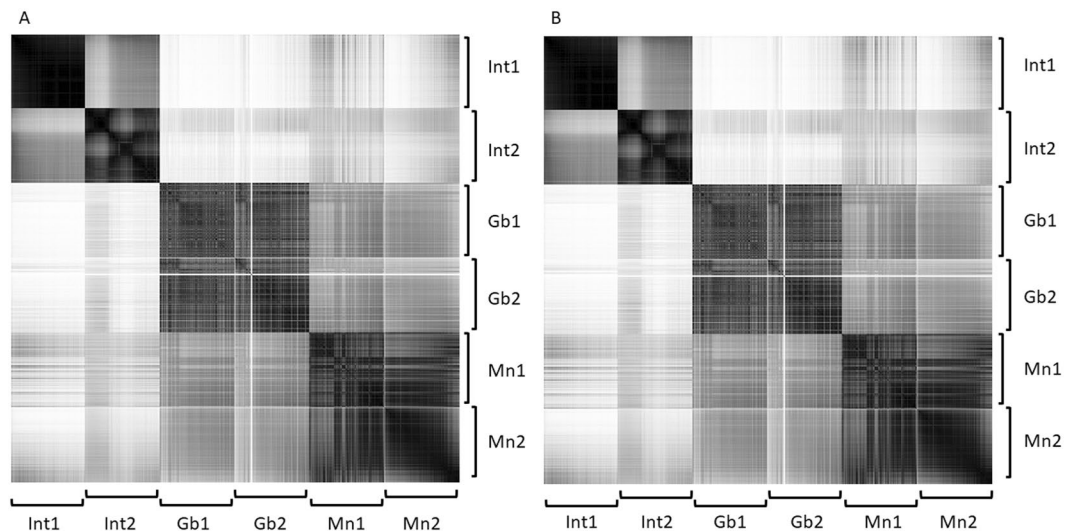


Figure 3. Correlation matrices without smoothing based on Pearson's r coefficients (A) and the cosine measure between vectors (B) for two different samples (from different patients) of brain tissue from non-tumor pathology (Int1 and Int2) and two fragments of one sample classified as a glioblastoma (Gb1 and Gb2) and one classified as a meningioma (Mn1 and Mn2).

Pearson's r coefficient and cosine measure. The Pearson's r coefficient becomes strictly equal to the cosine measure when the mean is equal to 0, what is evident from formulas (1) and (2). Therefore, since the average values of the peak intensities in a spectrum is close to zero, the cosine measure and the Pearson's r coefficient between the two spectrum vectors are very similar (Fig. 3). In general, the Pearson's r coefficient may be negative, although this is unlikely in our case as each scan is a vector of nonnegative values with a mean close to zero. Furthermore, in most scans peaks share their positions so Pearson's r coefficient even if it is negative has a low absolute value, which can be neglected. At the same time, the cosine measure between vectors with nonnegative coordinates will always be greater than zero. Considering the fact that the cosine measure between the spectrum vectors varies between 0 and 1, as do the Pearson's r coefficients between vectors, then it can be regarded as a metrics of scan similarity. Calculating the cosine measure requires slightly less computations due to the fact that, for the numerator and denominator, it is not necessary to calculate the mean value of the vector and then subtract it from each vector; therefore, calculating the cosine measure between vectors is preferable when calculating similarities between vectors of high dimensionality.

The visualization of the correlation matrix clearly demonstrates that brain tissues from different non-tumor pathology patients are significantly different from each other (due to natural biological variability), while also being significantly different from tumor tissues. Meanwhile, the tumor samples are quite similar among each of the patients but visibly differ across the diagnoses (Fig. 3). Furthermore, stable changes in molecular profiles occurring during each measurement could be easily found on the correlation matrix. For example, the part of the matrix corresponding to the second meningioma sample demonstrates that the molecular profile changes at the end of the experiment, what could be explained by partial exhausting of the sample during the investigation. However, the last scans of this sample are in good correlation with the initial scans of the first meningioma sample, meaning that these tumor samples are not homogeneous on the molecular level (as the histology examination describes them). Additionally, this demonstrates that NESI is sensitive to sample orientation as exhaustion from the tissue is changed between the parts coming from one sample. Formally the exhaustion is not observed during 5 minutes of spectra measurement. Thus, detection of such cases is vital for effective tissue classification since features of distinctive tissue parts may not be detected in proper proportions in each moment of the measurement and either splitting of spectra into scans or multilabel classification algorithms should be applied.

The fact that the spectrum is normalized to the total ion current, instead of normalizing to the maximum intensity of the spectrum, affects neither the Pearson's r coefficient or the cosine measure value, as follows from formulas 1 and 2: the cosine measure is independent of the vector length and the Pearson's r coefficient does not depend on the normalization coefficient. Similarly, with a proportional change in spectrum magnitude, no changes are observed in the calculations of Pearson's r coefficients or cosine measures. Thus, this technique allows us to automatically take into account as identical those spectra whose intensities are different due to variations in the total ion current values only.

Spectra smoothing for measurement artifacts filtration. Mass spectrometric profiling of biological samples is usually based on the principle of assigning one specific profile (a mass spectrum) to a sample. In most protocols, this spectrum is simply the sum of the scans accumulated during one measurement. However, it can be improved by discarding the outlier scans that show various anomalies in the measurement process. For this purpose, the scans should be filtered and averaged only afterwards.

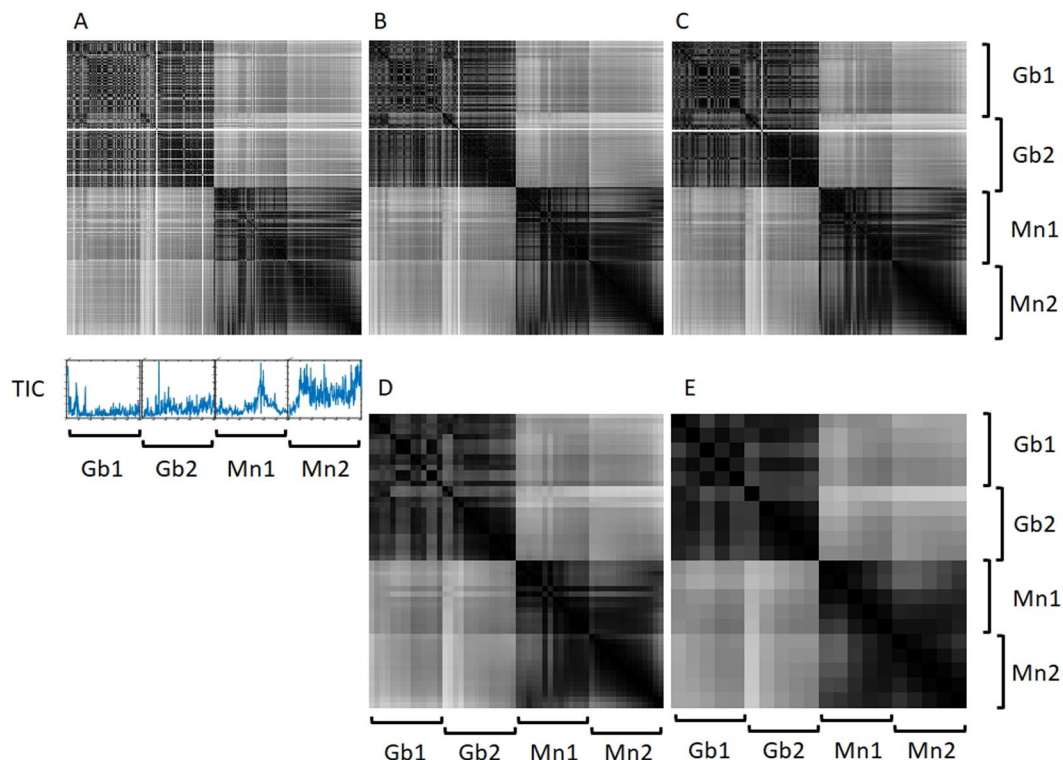


Figure 4. Cosine measure matrices for two fragments of one sample classified as a glioblastoma (Gb1 and Gb2) and two fragments of one sample classified as a meningioma (Mn1 and Mn2). Without smoothing (A) and with moving median smoothing with window size equal to 5 (B), 7 (C), 21 (D) and 51 (E) scans.

For long-term measurements, when multiple mass spectra (scans) are subsequently measured and analyzed, it is necessary to understand whether or not each particular scan should be taken into account. Most statistical and machine-learning algorithms are sensitive to outliers. That is why detection and filtering of outliers is a necessary step for training classifiers. For the detection of outliers the cosine measure between the scan and the smoothed aggregate of all scans in the spectra have to be calculated; then, one can evaluate whether the analyzed scan is a measurement artifact by applying a significance threshold to the calculated value.

The method of median smoothing involves, in a certain sense, adjusting the sensitivity of the obtained spectrum to the signal instability caused, in particular, by the instability of ESI-based sources. This approach makes the stability and reproducibility of this analytical method less sensitive to occasional “bad” scans. In a very unstable spray, a wide smoothing window should be selected for median smoothing to make spectra suitable for automated data processing. In the case when there are no outliers and the stability of the spray and spectra is high enough, the filtering step could be omitted to preserve data for detailed analysis of the ionization process or sample heterogeneity.

Four different windows for the moving median of 5, 7, 21 and 51 scans were considered in order to select the optimal size of the smoothing window (Figs 4, S2). Odd numbers were required to simplify the calculations of median values. The values of 5 and 7 were tested to estimate the number of single spectra with instabilities and to observe their influence on the final result.

As can be seen from Fig. 4B,C, the windows of 5 or 7 scans do not have any fundamental differences; however, they do not filter the outliers (bands with low values) well enough. This is due to the fact that the ICR scans faster if there is no signal (ions in the Penning trap). Thus a more serious smoothing windows were chosen: 51 and an intermediate value of 21 scans. These values correspond to approximately one minute and 30 seconds of averaging, respectively. There is a big difference between the results obtained with windows of 7 and 21, while there is no significant difference between those with 21 and 51. As expected, as the window size increases, the spectra become more similar to each other. The 21-scan anti-aliasing window removes all single outliers, as does the window with 51 scans, but the second requires fewer calculations and makes little difference in the final correlation matrix.

We also analyzed the influence of the smoothing window step size on the final results, i.e., the first spectrum was obtained by smoothing over the first 51 scans, and then the subsequent spectra were smoothed over 51 scans beginning from (one + step size) scan. This approach allows to reduce the number of calculations, as well as eliminate anomalous scans and spectra, without losing any information that characterizes the studied sample. Three step values were chosen: 1, 25 and 51 (Fig. 5).

From Fig. 5, a step of 1 is optimal from the point of view of the smoothness of the cosine measure matrix and the absence of rough changes from one smoothed scan to the next. However, it does not reduce the number of calculations. Visually, there is a slight difference between the selected steps. A large smoothing step is optimal

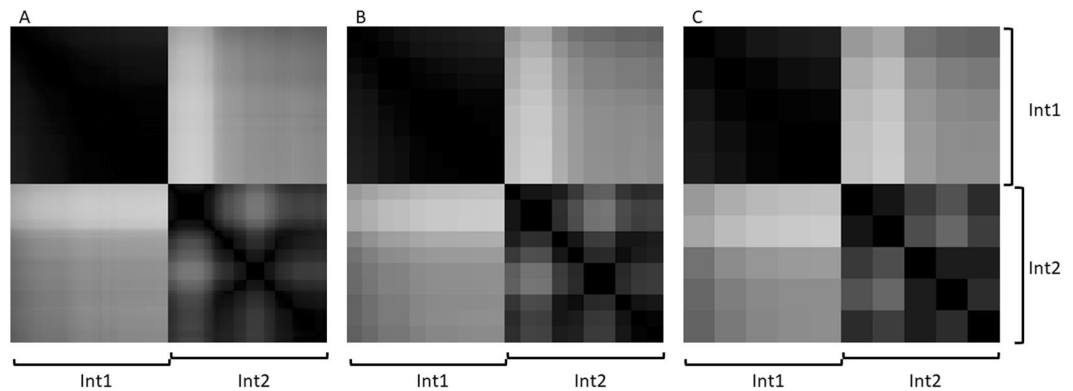


Figure 5. Similarity matrices based on the cosine measure between vectors for two samples classified as a brain tissue from non-tumor pathology (Int1 and Int2), with smoothing over 51 scans using a sliding window shifted by steps of 1, 25 and 51 scans (A–C, respectively).

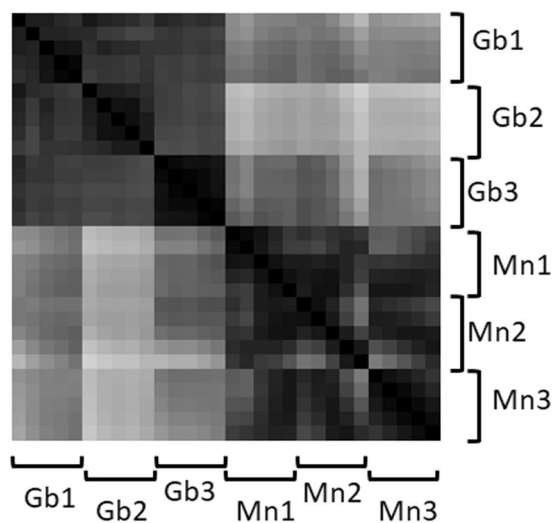


Figure 6. Matrix based on the cosine measure between vectors for three fragments of one sample classified as a glioblastoma (Gb1, Gb2 and Gb3) and three fragments of one sample classified as a meningioma (Mn1, Mn2 and Mn3), with smoothing over 51 scans using a shifting window.

for solving problems related to the global grouping of different samples, while a small smoothing step allows the researcher to obtain more detailed information on spectrum changes during the analysis, which may be important when working with highly heterogeneous samples, or in methodological work to determine the optimal parameters and conditions of ambient ionization.

Thus, depending on the task and experimental conditions, it is possible to select the optimal window and smoothing step values. Therefore, for our data and tasks regarding lipid tissue profiling, the use of a moving median with smoothing window of 51 scans is optimal and corresponds to approximately one minute of actual sample measurement. However, this parameter should be optimized with changes in the experimental conditions.

Metrics application in direct tissue profiling experiments. The developed analysis method was used to estimate the selectivity, reproducibility and stability of lipid profiles measured using the NESI method. For these purposes, the data obtained from two samples with different diagnoses, each of the which had been previously divided into three fragments, was analyzed (Fig. 6). It follows from the constructed matrix that the lipid profiles obtained from different fragments are highly reproducible within a single sample and there are large differences between the samples, but for accurate tissue classification statistical analysis and application of classifiers are still required. For this reason metrics calculation with suitably selected averaging parameters allows to select only reliable scans for further automated analysis.

A method for analyzing the stability and reproducibility of mass spectrometric data, obtained in the result of a consecutive accumulation of a large number of mass spectra, has been developed in this study. This method makes allows to filter data for the purpose of further clustering and classification. A method of multidimensional data analysis has been proposed based on the cosine measure between multidimensional vectors and the Pearson's

r coefficient calculations. The proposed analysis method uses median smoothing and facilitates the identification of groups of scans in one measurement, which can be considered as stable and characteristic of the sample. It also eliminates measurement artifacts from consideration, what is especially important for projects dealing with automatic classification of samples based on mass spectrometric profiling methods. The visualization of cosine measure matrices can be used as a simple method for selection of optimal ionization and measurement conditions, in order to evaluate the stability and reproducibility of signals during experimental work using ambient ionization methods.

References

1. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
2. Dorn, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).
3. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51–78 (2007).
4. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).
5. Van Riper, S. K., de Jong, E. P., Carlis, J. V. & Griffin, T. J. Mass Spectrometry-Based Proteomics: Basic Principles and Emerging Technologies and Directions. *Adv. Exp. Med. Biol.* **990**, 1–35 (2013).
6. Tang, F. *et al.* Rapid *In Situ* Profiling of Lipid C=C Location Isomers in Tissue Using Ambient Mass Spectrometry with Photochemical Reactions. *Anal. Chem.* **90**, 5612–5619 (2018).
7. Ščupáková, K. *et al.* Spatial Systems Lipidomics Reveals Nonalcoholic Fatty Liver Disease Heterogeneity. *Anal. Chem.* **90**, 5130–5138 (2018).
8. Sans, M., Feider, C. L. & Eberlin, L. S. Advances in mass spectrometry imaging coupled to ion mobility spectrometry for enhanced imaging of biological tissues. *Curr. Opin. Chem. Biol.* **42**, 138–146 (2018).
9. Shiea, J. *et al.* Electrospray-assisted laser desorption/ionization mass spectrometry for direct ambient analysis of solids. *Rapid Commun. in Mass Spectrom.* **19**, 3701–3704 (2005).
10. Cooks, R. G., Ouyang, Z., Takats, Z. & Wiseman, J. M. Ambient Mass Spectrometry. *Science* **311**, 1566–1570 (2006).
11. Takáts, Z., Wiseman, J. M., Gologan, B. & Cooks, R. G. Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization. *Science* **306**, 471–473 (2004).
12. Santagata, S. *et al.* Intraoperative mass spectrometry mapping of an onco-metabolite to guide brain tumor surgery. *Proc. Natl. Acad. Sci.* **111**, 11121–11126 (2014).
13. Zhvansky, E. *et al.* High-resolution mass spectra processing for the identification of different pathological tissue types of brain tumors. *Eur. J. Mass Spectrom.* **23**, 213–216 (2017).
14. Calligaris, D. *et al.* Mass spectrometry imaging as a tool for surgical decision-making. *J. Mass Spectrom.* **48**, 1178–1187 (2013).
15. Eberlin, L. S. *et al.* Classifying Human Brain Tumors by Lipid Imaging with Mass Spectrometry. *Cancer Res.* **72**, 645–654 (2011).
16. Eberlin, L. S. *et al.* Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc. Natl. Acad. Sci.* **110**, 1611–1616 (2013).
17. Clark, A. R. *et al.* Rapid discrimination of pediatric brain tumors by mass spectrometry imaging. *J. Neurooncol.* (2018).
18. Pirro, V. *et al.* Intraoperative assessment of tumor margins during glioma resection by desorption electrospray ionization-mass spectrometry. *Proc. Natl. Acad. Sci.* **114**, 6700–6705 (2017).
19. Jarmusch, A. K. *et al.* Lipid and metabolite profiles of human brain tumors by desorption electrospray ionization-MS. *Proc. Natl. Acad. Sci.* **113**, 1486–1491 (2016).
20. Chen, L. C. *et al.* Ambient imaging mass spectrometry by electrospray ionization using solid needle as sampling probe. *J. Mass Spectrom.* **44**, 1469–1477 (2009).
21. Hayashi, Y. *et al.* Intact metabolite profiling of mouse brain by probe electrospray ionization/triple quadrupole tandem mass spectrometry (PESI/MS/MS) and its potential use for local distribution analysis of the brain. *Anal. Chim. Acta* **983**, 160–165 (2017).
22. Pirro, V. *et al.* Analysis of human gliomas by swab touch spray-mass spectrometry: applications to intraoperative assessment of surgical margins and presence of oncometabolites. *Analyst* **142**, 4058–4066 (2017).
23. Jarmusch, A. K., Pirro, V., Logsdon, D. L. & Cooks, R. G. Direct ion generation from swabs. *Talanta* **184**, 356–363 (2018).
24. Balog, J. *et al.* Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Sci. Transl. Med.* **5** (2013).
25. Katajamaa, M. & Orešič, M. Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* **1158**, 318–328 (2007).
26. Mueller, L. N., Brusniak, M.-Y., Mani, D. R. & Aebersold, R. An Assessment of Software Solutions for the Analysis of Mass Spectrometry Based Quantitative Proteomics Data. *J. Proteome Res.* **7**, 51–61 (2008).
27. Scalbert, A. *et al.* Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **5**, 435–458 (2009).
28. Lokhov, P. G., Maslov, D. L., Kharibin, O. N., Balashova, E. E. & Archakov, A. I. Label-free data standardization for clinical metabolomics. *BioData Min.* **10** (2017).
29. Sugimoto, M., Kawakami, M., Robert, M., Soga, T. & Tomita, M. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. *Curr. Bioinform.* **7**, 96–108 (2012).
30. Sorokin, A. *et al.* Feature selection algorithm for spray-from-tissue mass spectrometry. *Eur. J. Mass Spectrom.* **23**, 237–241 (2017).
31. Beckmann, M., Parker, D., Enot, D. P., Duval, E. & Draper, J. High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nat. Protoc.* **3**, 486–504 (2008).
32. Wong, J. W. H., Cagney, G. & Cartwright, H. M. SpecAlign—processing and alignment of mass spectra datasets. *Bioinformatics* **21**, 2088–2090 (2005).
33. Yang, X., Neta, P. & Stein, S. E. Quality Control for Building Libraries from Electrospray Ionization Tandem Mass Spectra. *Anal. Chem.* **86**, 6393–6400 (2014).
34. Wallace, W. E., Ji, W., Tchekhovskoi, D. V., Phinney, K. W. & Stein, S. E. Mass Spectral Library Quality Assurance by Inter-Library Comparison. *J. Am. Soc. Mass Spectrom.* **28**, 733–738 (2017).
35. Golf, O. *et al.* XMS: Cross-Platform Normalization Method for Multimodal Mass Spectrometric Tissue Profiling. *J. Am. Soc. Mass Spectrom.* **26**, 44–54 (2014).
36. Kirwan, J. A., Broadhurst, D. I., Davidson, R. L. & Viant, M. R. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Anal. Bioanal. Chem.* **405**, 5147–5157 (2013).
37. Kononikhin, A. *et al.* A novel direct spray-from-tissue ionization method for mass spectrometric analysis of human brain tumors. *Anal. Bioanal. Chem.* **407**, 7797–7805 (2015).
38. Kim, S. & Zhang, X. Comparative Analysis of Mass Spectral Similarity Measures on Peak Alignment for Comprehensive Two-Dimensional Gas Chromatography Mass Spectrometry. *Computational and Mathematical Methods in Medicine* **2013**, 1–12 (2013).
39. Arce, G. & McLoughlin, M. Theoretical analysis of the max/Median filter. *IEEE Trans. Acoust., Speech, Signal Processing* **35**, 60–69 (1987).
40. Allen, R. L. & Mills, D. W. *Signal analysis: time, frequency, scale, and structure* (IEEE Press, Piscataway, 2004).
41. Pitas, I. & Venetsanopoulos, A. N. *Nonlinear Digital Filters. Principles and Applications* (Kluwer, Norwell, 1990).

Acknowledgements

The study was supported by Russian Scientific Foundation grant #16-15-10431.

Author Contributions

E. Zhvansky and S. Pekov wrote the main text of the manuscript. E. Zhvansky made all of the calculations for the work, developed data treatment and data storage approaches. A. Sorokin and E. Zhvansky prepared the methodics description. E. Zhvansky and A. Sorokin developed the plan and methodics of the research. V. Shurkhay provided tissue samples from the operation room for research and was responsible for obtaining clinically relevant data for the research. V. Eliferov measured all mass spectra of the samples. A. Potapov was the neurosurgeon who removed the pathological tissues. E. Nikolaev, I. Popov and A. Potapov organized the interaction between the mass spectrometry laboratory and the neurosurgical clinic. E. Nikolaev and I. Popov organized the whole process of the mass spectrometry part of the research.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-37560-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019