analytical chemistry

Subscriber access provided by Helmholtz Zentrum Muenchen - Zentralbibliothek

Article

Spatial autocorrelation in mass spectrometry imaging

Alberto Cassese, Shane Ellis, Nina Ogrinc Poto#nik, Elke Burgermeister, Matthias Ebert, Axel Walch, Arn M. J.M. van den Maagdenberg, Liam A. McDonnell, Ron M.A. Heeren, and Benjamin Balluff

Anal. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.6b00672 • Publication Date (Web): 14 May 2016

Downloaded from http://pubs.acs.org on May 23, 2016

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

Analytical Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Spatial autocorrelation in mass spectrometry imaging

Alberto Cassese[1], Shane Ellis[2], Nina Ogrinc Potočnik[2], Elke Burgermeister[3], Matthias Ebert[3], Axel Walch[4], Arn M.J.M. van den Maagdenberg[5], Liam A. McDonnell[6-8], Ron M.A. Heeren[2], Benjamin Balluff[2,*]

Authors' affiliations

[1] Department of Methodology & Statistics, Maastricht University, 6200 MD Maastricht, The Netherlands

[2] Maastricht MultiModal Molecular Imaging Institute (M4I), Maastricht University, 6200 MD Maastricht, The Netherlands

[3] Department of Internal Medicine II, Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany

[4] Research Unit Analytical Pathology, Helmholtz Zentrum München, 85764 Oberschleißheim, Germany

[5] Departments of Human Genetics & Neurology, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands

[6] Fondazione Pisana per la Scienza ONLUS, 56121 Pisa, Italy

[7]Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands

[8]Department of Pathology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands

* **Corresponding author:**

Dr. Benjamin Balluff, Maastricht University, Maastricht MultiModal Molecular Imaging institute (M4I), Universiteitssingel 50, P.O. Box 616, 6200 MD Maastricht, The Netherlands; Phone: +31 43 388 1251; Email: b.balluff@maastrichtuniversity.nl

ABSTRACT

Mass spectrometry imaging (MSI) is a powerful molecular imaging technique. In microprobe MSI, images are created through a grid-wise interrogation of individual spots by mass spectrometry across a surface. Classical statistical tests for within-sample comparisons fail as close-by measurement spots violate the assumption of independence of these tests, which can lead to an increased false-discovery rate. For spatial data this effect is referred to as spatial autocorrelation.

In this study we investigated spatial autocorrelation in three different matrix-assisted laser desorption/ionization MSI datasets. These datasets cover different molecular classes (metabolites/drugs, lipids, and proteins) and different spatial resolutions ranging from 20 µm to 100 µm. Significant spatial autocorrelation was detected in all three datasets and found to increase with decreasing pixel size.

To enable statistical testing for differences in mass signal intensities between regions of interest within MSI datasets, we propose the use of Conditional Autoregressive (CAR) models. We show that by accounting for spatial autocorrelation, discovery rates (i.e. the ratio between the features identified and the total number of features) could be reduced between 21% and 69%. The reliability of this approach was validated by control mass signals based on prior knowledge.

In light of the advent of larger MSI datasets based on either an increased spatial resolution or 3D datasets, accounting for effects due to spatial autocorrelation becomes even more indispensable. Here we propose a generic and easily applicable workflow to enable within-sample statistical comparisons.

INTRODUCTION

Mass spectrometry imaging (MSI) is a molecular imaging technology that allows visualizing distributions of molecules in surfaces, such as biological tissue specimens or inorganic materials. In microprobe mode, the most common MSI mode, images are created through distinct measurements in a raster grid with a defined mesh-size across the sample's surface using different *in situ* ionization techniques.[1] MSI is applied in many areas, including material science, microbiology, biomedical research, and pharmacological studies.[2]

In all these paradigms it can be of interest to investigate if the abundance of ions significantly differs between two regions of interest (ROI). For this purpose, there exist several techniques ranging from simple box plot evaluation to multivariate methods such as Principal Component Analysis (PCA; Table S-1). Amongst these techniques, statistical hypothesis testing, such as a t-test (Text S-1), offers an objective way to determine significant differences in ion intensities. However it is known that pixels within a MSI dataset are not independent from each other and hence violate the assumption of independence that is made by most statistical tests.[3]

This dependency between pixels can for instance be caused by real biological effects of molecular interaction between nearby cells or by experimental factors such as diffusion of analytes through tissue washes which can introduce artificial spectral correlation between pixels.[4] As a consequence of this spatial correlation between pixels, classical statistical inference may suffer from an increased false positive-rate (i.e. detection of features that are truly not different in the two conditions); especially since the length scales of these effects are becoming more and more accessible with

improvements in spatial resolution of mass spectrometers, where recent commercial systems achieve even subcellular 5 µm pixel sizes.[5]

This inapplicability of classical statistical tests is still an unresolved problem in MSI. But given the rigorousness of statistical testing, there is a strong need in MSI for new statistical procedures that enable within-sample statistical comparisons by accounting for the inherent dependency between pixels.

In this study we show that the failure of classical statistical tests in MSI is due to an effect, which is known as 'spatial autocorrelation'. Spatial autocorrelation refers to the correlation among values of a single variable strictly attributable to their relatively close positions on a two-dimensional surface, thereby introducing a deviation from the independent observations assumption of classical statistics.[6] In statistics, spatial analysis is used to describe and analyze data, which is affected by spatial dependency. The presence of spatial autocorrelation can be assessed using coefficients such as Moran's I (Text S-1).[7] Moreover, spatial regression models can be used to model spatial dependence of correlated values in order to test for differences in their mean.[8-9] Both procedures have in common that values at any given location are analyzed accounting for their dependency on the neighboring values.

This study here addresses for the first time spatial autocorrelation in MSI data by using spatial statistics. First, we investigate the presence of spatial autocorrelation in three representative datasets. Then, we propose Conditional Autoregressive (CAR) models, which have already been applied in biomedical MRI studies, as statistical solution for within-sample statistical comparisons.[10]

EXPERIMENTAL SECTION

Mass spectrometry imaging datasets

All three datasets were acquired using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI).

The first dataset is from a drug imaging experiment at 80 µm lateral resolution in a transgenic gastric cancer mouse model to investigate the distribution of the anti-cancer drug erlotinib (Tarceva®, Roche; 393.17 g/mol) and metabolic changes in the tumor.[11] The experimental details are described in the Supplementing Information (Text S-2).

Two additional MSI datasets from mouse brains were obtained from previously published studies. The first, termed “CSD mouse brain dataset”, is a protein MSI dataset recorded at 100 µm spatial resolution from a coronal brain section of a transgenic mouse model which expresses a $Ca_V2.1$ $\alpha 1$ subunit gene mutation that was found in patients with familial hemiplegic migraine type 1.[12] The mouse was sacrificed after induction of multiple cortical spreading depressions (CSDs) by repeated topical application of KCl through a borehole in the right cortex, which left the contralateral hemisphere cortex unaffected.[13] CSD is considered the electrophysiological correlate of migraine aura.[14-15] The second MSI mouse brain dataset was obtained from a sagittal section of a wild-type (control) mouse brain at high spatial resolution (20 µm) using dithranol for lipid detection.[5]

Details of all three datasets are summarized in Table 1.

Data processing of the drug imaging and CSD mouse brain datasets

Co-registration of the MSI data with the histological image of the tissue sections was done within the FlexImaging 4.1 software (Bruker Daltonics). For both, the drug imaging and CSD mouse brain datasets, regions of interest (ROI) were defined *solely* based on histological features of the tissue, resulting in annotated areas for the tumor and the epithelial layer of the mucosa in the drug imaging dataset (Figure 1A and 1B), and in annotated areas for the left and right cortex (CSD-affected hemisphere) in the CSD mouse brain dataset (Figure 2A).

The ROIs' spectra were then pre-processed in ClinProTools 2.2 (Bruker Daltonics). This included baseline subtraction (TopHat: 10%) and normalization of the spectra to their total ion count (TIC). The resolution was set to 2000 for the drug imaging dataset and to 800 for the CSD mouse brain dataset. Peaks were picked on the total average spectrum with a signal to noise threshold of 5.00 and a relative intensity threshold to the base peak of 1%.

The pre-processed peak intensity table was then further analyzed in the R statistical environment.

Statistical analysis of the drug imaging and CSD mouse brain datasets

All following statistical terms and methods used in this study are described in detail in the Supporting Information Text S-1.

As strong tailing was observed in the intensity distributions for many features in the drug imaging dataset and normal distribution is required for subsequent statistical procedures, it was log-transformed.

For both, the drug imaging and the CSD mouse brain MSI datasets, the processed mass signals were tested for intensity differences between regions of interest using the Student *t*-test (Text S-1). Resulting p-values were subsequently adjusted by the Benjamini-Hochberg procedure, a procedure that aims at controlling the false discovery rate when multiple tests are performed simultaneously (Text S-1).[16] Adjusted p-values ≤ 0.05 were considered significant.

Moran's I is used as an indicator of the presence of a spatial autocorrelation in the data (Text S-1).[17] Values below zero are indicative of a negative autocorrelation, above zeros of a positive autocorrelation, and zero of the absence of a spatial pattern (Figure 1E). Here it was calculated using unit weights for neighbors within a Euclidean distance (i.e. the straight-line distance between two points) of 2 pixel units ('spdep' package). Spatial spectral correlations between pixels were calculated using Pearson's correlation across all features (Text S-1) [17]

Data processing and analysis of the high-resolution sagittal mouse brain dataset

Co-registration of the MSI data with the histological H&E image of the tissue sections was done within the FlexImaging 4.1 software (Bruker Daltonics). Then the data were uploaded to SCiLS Lab 2015a (SCiLS Lab, Bremen, Germany) where it was TIC-normalized and lipid peaks were picked using the mean spectrum (peak width = ± 0.15 Da). The peak intensities per pixel were then exported as imzML file and imported into Matlab R2015a (The MathWorks, Natick, MA, USA) for final data analysis. Spectral correlations between pixels were calculated using Pearson's correlation. Moran's I as indication for spatial autocorrelation was calculated using unit weights for neighbors within a Euclidean distance of 2 pixel units.[17] Simulation of

lower resolution images were done using the *imresize* function of the Image Processing Toolbox and 'nearest' as interpolation method.

Statistical analysis using the Conditional Autoregressive (CAR) model

In order to test whether a feature shows a statistically significant difference in intensity levels between two ROIs, we employed a regression model with Gaussian Conditional Autoregressive (CAR) errors. In this study we used the CAR model implementation in the 'spautolm' function from the 'spdep' package, a well-established toolkit for spatial statistics.[18-19] To statistically analyze features within a MSI dataset, we propose to follow the algorithm described below and summarized in Figure 4.

For a CAR regression model, a neighborhood matrix **W** containing the list of neighbors for each location has to be provided. In **W**, off-diagonal elements $\mathbf{w_{ij}}$ are set to one, which is pixel i is neighbor of j, if the Euclidean distance between the two pixels i and j is smaller or equal than a pre-specified threshold d. As no assumptions can be made *a priori* about the extent of the relevant pixel neighborhood, our algorithm is initiated with d_1=1 (i.e. with a first-order Markov scheme), a commonly used neighborhood structure in spatial statistics.[20] At each iteration, the distance cutoff d is then increased by one unit, which progressively leads to accounting for the influence of a larger number of neighbors. To exclude the influence of tangent ROIs (as observed here for the upper cortex in the CSD mouse brain dataset; Figure 2A), we restricted the neighborhood to pixels from the same ROI.

Then, a CAR model is fit for each feature and the corresponding p-value recorded. The p-value represents the significance of intensity differences between the regions while accounting for spatial autocorrelation. At the end of each iteration,

the list of p-values is compared to the list obtained using the previous threshold through two different statistical tests: a paired *t*-test (Text S-1) and a McNemar's chi-squared test (Text S-1). The first test checks for significant changes between the means of the two lists, hence for global changes. The second is performed on the dichotomized (significant/non-significant) lists of p-values to test whether the two lists have the same behavior in terms of declaring the set of features as significant (i.e. equality of discovery rates in the two lists); thus checking for local, but relevant, changes at the border of significance. The algorithm stops once both tests do not detect a significant ($P > 0.05$) difference between the recent list of p-values (d_i) and the one before (d_{i-1}); in that case, $d_{max}=d_{i-1}$. We found that the two tests complement each other and increase the power of detecting relevant differences between the lists.

At the end of this process, for each feature, a list of p-values (one for each specific threshold up to d_{max}) is compiled. The penultimate step of the workflow consists of selecting, for each feature, the p-value corresponding to the distance d that led to the best-fitted model. This is done by selecting the model with the lowest Akaike Information Criterion (AIC), an estimate of the information loss associated to a model (Text S-1). This AIC-based selection leads to a list containing one p-value per feature. In the last step, the Benjamini-Hochberg correction is applied to this list.[16] Alternatively, the correction is applied to the p-values obtained from the CAR models with the highest threshold d_{max}.

RESULTS AND DISCUSSION

Mass spectrometry imaging (MSI) is a powerful tool to study the spatial distribution of molecules in surfaces. For single-sample analyses it can be of interest to investigate if the abundance of some mass signal significantly differs between some regions of interest (ROI). For this purpose there exist several techniques. Each of these has its strength and disadvantages (Table S-1). For example, visual examination is used in boxplots, and arbitrary thresholds are used in PCA and discriminant analysis to determine the discriminatory power of a feature. However, this is prone to subjectivity of the investigator. In contrast, statistical hypothesis testing offers the most objective way to determine significant differences in ion intensities. It has been reported that for intra-sample statistical comparisons, a very high fraction of seemingly significant differences indicates an unusually high false discovery rate.[21] The reason is that classical statistical tests, including non-parametric tests (i.e. tests that do not make any assumption about the distribution in the population; Text S-1) such as the Mann-Whitney *U* test, fail as the data does not fulfill the assumption of independence between observations.[3] For spatial data, this effect is known as spatial autocorrelation.[22]

In this study, spatial autocorrelation in matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI) was studied in three datasets of different molecular classes and spatial resolution (Table 1). Two datasets illustrate a scenario where intensity differences of mass signals (features) need to be compared on a statistical level between certain regions of interest (ROI) within a sample. The first dataset is from an imaging experiment of the anti-cancer compound erlotinib in a gastric cancer mouse model. The aim of this experiment is to determine differences in drug and metabolite concentrations between the tumor and the

epithelial layer of the stomach mucosa (Figure 1B, 1G, and 1H). The second dataset is from a cortical spreading depression (CSD) experiment in a hemiplegic migraine mouse model. As the CSD is restricted to one hemisphere, it was of interest to study the proteomic changes in the CSD-affected cortex (right hemisphere) compared to the unaffected cortex (left hemisphere) (Figure 2A). The third dataset originates from a high-resolution measurement of a sagittal mouse brain section at 20 µm pixel size, which allows studying the effect of spatial autocorrelation as a function of distance between the measurement points (Figure 3).

MSI data suffers from spatial autocorrelation

An indication of spatial autocorrelation is a large fraction of significant features amongst all tested mass signals when comparing spectral information between intra-sample ROIs using simple *t*-tests. This was 93.4% (99/106) in the drug imaging dataset and 55.4% (62/112) in the CSD dataset (Figures 1F and 2E). Although there is no ground truth available, especially for the CSD mouse brain these numbers seem high compared to only one significant feature reported before.[13] In the drug imaging experiment, the significant difference of the alpha-cyano-4-hydroxycinnamic acid matrix cluster (HCCA, *m/z* 379.1, P=2.5e-4), which was homogenously sprayed on top the tissue, indicated a test susceptible for delivering false-positive results.

This phenomenon is supported by statistical theory, which assumes independence between samples in a classical test. The effect under which such tests lose their accuracy in spatial data is described as spatial autocorrelation. There are statistical measures such as Moran's I to test for the presence of spatial autocorrelation.[17] Examples for the range of Moran's I values are shown in Figure 1E.

For the drug imaging dataset, Moran's test revealed significant spatial autocorrelation for 100% of the features (Table S-2), including erlotinib and the HCCA matrix (Figure 1D). Additionally, a high spatial spectral correlation was observed for each pixel to its 2-pixel unit neighborhood with an average correlation of 0.9 (Figure 1C). Also the CSD mouse brain dataset was found to suffer from spatial autocorrelation: 94% of all features exhibited significant spatial autocorrelation (Table S-3) and the average correlation between neighboring pixels was 0.84 (Figure 2B).

Spatial autocorrelation depends on spatial resolution

We have shown that MSI data can be affected by spatial autocorrelation. The causes on spectral autocorrelation can be of biological and technical nature. Innate to a biological organism, molecules are related to each other on different scales ranging from macroscopic anatomy, over microscopic histology, to nanoscopic inter- and intra-cellular communication. Consequently, the scale of the employed molecular imaging technology determines the degree of observed biological spatial autocorrelation. Since in MSI the scale is microscopic, spectral autocorrelation is mostly determined biologically by histology On top of biological vicinity effects, spectral autocorrelation can be introduced by the technique itself. Examples in MALDI-MSI are analyte diffusion during tissue washes, solvent extraction and crystallization during matrix application, or laser oversampling (i.e. the laser diameter is bigger than the pixel size).[4, 23-24] We hypothesize that dependent on the spatial resolution of the MSI system, the length scales of these effects become more or less accessible and hence spatial autocorrelation stronger or weaker, respectively.

Consequently, we next investigated whether the degree of spatial autocorrelation depends on the selected spatial resolution of a MSI dataset. For this we used a third, 131,082 pixel MSI dataset of a sagittal mouse brain section

measured for the lipid mass range at a spatial resolution of 20 µm (Figure 3A). In order to test the spectral correlation as a function of distance, the correlation for each pixel was calculated to the pixels of layers with different Euclidean distances with respect to the center pixel (Figure 3B).

The results show a continuous decrease of the overall correlation between pixels with an increasing distance, which corresponds to a lower resolution (Figure 3D). The lower limit of correlation is determined by the average correlation between pixels, which was calculated to be 0.78 based on 10,000 randomly picked pixels. The reason for this high correlation baseline is inherent to the MSI technique, which are the detection of the most abundant and ionization-affine molecules, which result in similar spectral profiles. Another determinant of this lower limit of spatial correlation is the anatomy of the tissue. This means that even with an increasing distance between two pixels the correlation could increase again if the running pixels enter a region that is histologically and molecularly similar to the reference pixel. This effect is depicted in Figure 3C.

This inherent spatial autocorrelation of spectral information is reflected also in the calculation of Moran's I for different simulated spatial resolutions of the MSI dataset where the value of Moran's I decreases with a decreasing spatial resolution (Figure 3E).

While in this study only lower resolutions were simulated by *in silico* down-sampling, Van de Plas et al. proposed a method for augmenting the resolution of MSI images through image fusion with images from high resolution optical microscopy, achieving pixel sizes down to 330 nm.[8] It would be interesting to study the effect of image fusion on spatial autocorrelation since the method uses histology as a guide to make molecular predictions for non-extant MSI pixels. Certainly it would affect a

statistical test because –given a fixed ROI size and homogenous histology– the statistical power (i.e. the ability of a test to detect differences when there are true differences; see Text S-1) is influenced by the number of pixels, where a higher spatial resolution should lead to a higher statistical power and vice versa. However this holds only if the observations are independent to each other. If not, as observed in MSI data, this has to be corrected for when performing intra-sample statistical tests.

Conditional Autoregressive Model (CAR) can correct for spatial autocorrelation in MSI data

In order to test for statistically significant differences in intensity levels between two ROIs, while accounting for spatial autocorrelation of the observations, we employ a Gaussian Conditional Autoregressive (CAR) model (Text S-1).[9 10] It is a supervised method that allows for testing of significant differences in the mean between two or more groups in spatial data and hence their definition beforehand (here the regions of interest) is necessary. In consequence, it cannot be used as other unsupervised methods (e.g. PCA or non-negative matrix factorization; Table S-1) to segment MSI datasets into different clusters based on multivariate molecular profiles. However, it shares with PCA the fact that it is parameter-free. It only assumes Gaussian distribution of values, which can be effectuated by data transformations in most scenarios. An example is the logarithm transformation we used for the drug imaging dataset. Although in this study only two ROIs per dataset were compared, the CAR model can be extended to any number of ROIs.

The only crucial factor is the definition of the neighborhood of each pixel. As no assumptions can be made on the relevant neighborhood *a priori*, we propose to iteratively extend the neighborhood until there is no dataset-wide significant change

in the CAR-corrected p-values (Figure 4). For the CSD mouse brain dataset, the maximum extension was reached at a distance of 6 pixel units and for the drug imaging dataset at 2 pixel units (Figure 5A and 5C). Using these maximum distance cutoffs (d_{max}), significant observations could be reduced from 62 to 10 (83.9% reduction) in the CSD mouse brain dataset, and from 99 to 64 (35.4%) in the drug imaging dataset (Figure 5B and 5D).

As we observed instability for some features with increasing distance – evidence for this effect of an again upwards-correction of CAR corrected p-values can be observed for the average of all features between distance 3 and 6 in Figure 5A, and for individual features in Figure S-1–, the stability of each feature's CAR model was customized for by choosing the model with the lowest Akaike information criterion (AIC) within d_1 and d_{max}. This way, the number of initially significant features condensed to 19 (69.4% reduction) and 78 (21.2%) for the CSD and drug imaging datasets, respectively (Figure 5B and 5D).

We can only speculate about the difference in reduction rates, as the lack of a biological ground truth limits the options to test for validity of the results. A reason could be the well-known increased chemo-temporal dynamics of the metabolome compared to the slower, biosynthesis-driven proteome which would lead to a higher probability of observing significantly altered signals in the lower mass range. The second effect is that this probability has to be multiplied by the isotopic signals observed. These were only observed in the drug imaging dataset due to the higher experimental mass resolution, which was acquired using a reflectron ToF system.

In an attempt to monitor CAR correction, we selected certain features as control based on prior knowledge. For the CSD mouse brain dataset it was known from a previous conducted study on five mice that intensities belonging to a modified

version of histone H4 (here detected at *m/z* 11,342; P=2.6e-10; Figure 2D) were significantly lower in the CSD-affected hemisphere whereas the unmodified version of H4 (here detected at *m/z* 11,304; P=1.3e-04; Figure 2C) was reported non-significant, indicating a modification of H4 as response to CSD induction (Figure 2E).[5] The CAR correction led to full reproducibility of the results in line with the previous multi-sample study, since the modified H4 is still significant (P=0.009; red triangle, Figure 5C) and the unmodified version became non-significant (P=0.280) (blue triangle, Figure 5C).

For the drug imaging dataset the control was the HCCA matrix (*m/z* 379.1; P=2.5e-04 before CAR) as it has been homogeneously applied during sample preparation. As expected, it became non-significant (P=0.068) in the CAR model with the lowest AIC (blue triangle, Figure 5A). Erlotinib (*m/z* 394.2; P= 7.1e-157) remained strongly significant (red triangle Figure 5A) and suggests a significant lower uptake of erlotinib in the tumor compared to the healthy mucosa (Figure 1D and 5A). However, positive results such as from erlotinib should be still taken with caution, as ionization bias can still be a source of spurious differential mass signals, which can be overcome by normalization strategies.[25]

Role of ionization source

As we have shown, the approach is generic for all kinds of MSI data, ranging from metabolites to protein datasets, from high to low spatial resolution experiments. Although all data investigated here originated from MALDI instrumentation, we surmise a direct applicability to other MSI techniques such as desorption electrospray ionization (DESI) and secondary ion mass spectrometry (SIMS) which all provide spatial data. SIMS and MALDI have a discrete acquisition of the pixels, DESI acquires the spectra continuously. As matrix-free method, DESI lacks the analyte

diffusion caused by the matrix application. But similar to MALDI oversampling, signal carryover of analytes between proximate pixels has been reported in DESI and might be an additional source of spatial autocorrelation.[25] If an estimate of the signal carryover or oversampling effect radius can be made the pixel neighborhood of the CAR models could be adjusted accordingly.

Computational perspective

The datasets presented here, with around 2,000 pixels after ROI definition, required only 32 MB of main memory for the distance matrix. However, next-generation MSI instrumentation will deliver datasets with over 100,000 pixels. If only 50% of the pixels go into the comparison of ROIs, the distance matrix will consume 20 GB [space complexity $O=f(n^2)$; with n being the number of pixels]. The 'spdep' package is able to reduce the memory load through a list structure [$O=f(m*n)$; with m being the average number of neighbors for all n pixels and $m<<n$]. List structures, however, come at the prize of a higher time complexity. Hence, to keep future spatial statistics for MSI data feasible on desktop PCs, we forecast the need for accelerated and memory-efficient implementations of neighborhood representations and CAR models using e.g. parallelization through GPUs.[26]

We foresee this especially in the light of the advent of larger MSI datasets based on either an increased spatial resolution or 3D datasets. Particularly for the latter, within-sample comparisons based on voxels are expected to become more commonly performed which will make accounting for effects due to spatial autocorrelation even more indispensable.

CONCLUSIONS

This is the first paper to describe the effect of spatial autocorrelation in MSI data. Besides creating awareness of spatial autocorrelation in MSI data, we propose a generic and easily applicable workflow as a statistical solution to statistically, and therefore objectively, determine significant differences in peak intensities between regions of interest using CAR regression models. The R code of this workflow together with the data and results (Tables S-2 and S-3) are provided as supporting information.

Acknowledgements

This work was supported by the Province of Limburg of the Netherlands. N.O.P., R.M.A.H., and A.v.d.M. acknowledge the support of the FP7 European Union Marie Curie IAPP Program BRAINPATH (No. 612360). E.B. acknowledges the support of the Deutsche Krebshilfe (No. 108287) and the Deutsche Forschungsgemeinschaft (DFG) (BU2285, SFB 824 TP B1).

Conflict of interest

The authors declare no competing financial interest.

Supporting Information

Figure S-1: Examples for stabilization of each feature's CAR model by choosing the model with the lowest Akaike information criterion.

Text S-1: Description of the statistical methods used in this manuscript (in didactical order).

Text S-2: Experimental details on drug imaging dataset

Source code: the statistical approach used in this study to generate the CAR models as source code for the R statistical programming language.

Table S-1: Comparison of commonly used approaches to determine significant intensity differences of an *m/z* species between regions of interest

Table S-2: Drug imaging dataset and results

Table S-3: CSD mouse brain dataset and results

REFERENCES

(1) Addie, R. D.; Balluff, B.; Bovee, J. V.; Morreau, H.; McDonnell, L. A. *Anal Chem* **2015**, *87*, 6426-6433.

(2) Nilsson, A.; Goodwin, R. J.; Shariatgorji, M.; Vallianatou, T.; Webborn, P. J.; Andren, P. E. *Anal Chem* **2015**, *87*, 1437-1455.

(3) Jones, E. A.; Deininger, S. O.; Hogendoorn, P. C.; Deelder, A. M.; McDonnell, L. A. *J Proteomics* **2012**, *75*, 4962-4989.

(4) Altelaar, A. F.; Luxembourg, S. L.; McDonnell, L. A.; Piersma, S. R.; Heeren, R. M. *Nat Protoc* **2007**, *2*, 1185-1196.

(5) Ogrinc Potocnik, N.; Porta, T.; Becker, M.; Heeren, R. M.; Ellis, S. R. *Rapid Commun Mass Spectrom* **2015**, *29*, 2195-2203.

(6) Griffith, D. A., *Spatial Autocorrelation and Spatial Filtering*. 1st ed.; Springer-Verlag Berlin Heidelberg: 2003; p XIV, 250.

(7) Getis, A.; Ord, J. K. *Geographical Analysis* **1992**, *24*, 189-206.

(8) Bivand, R.; Pebesma, E. J.; Gómez-Rubio, V., *Applied spatial data analysis with R*. Springer: New York, 2008; p xiv, 374 p.

(9) Cressie, N. A. C., *Statistics for spatial data, Revised Edition*. Rev. ed.; Wiley: New York, 1993; p 928.

(10) Ge, T.; Muller-Lenke, N.; Bendfeldt, K.; Nichols, T. E.; Johnson, T. D. *Annals of Applied Statistics* **2014**, *8*, 1095-1118.

(11) Thompson, J.; Epting, T.; Schwarzkopf, G.; Singhofen, A.; Eades-Perner, A. M.; van Der Putten, H.; Zimmermann, W. *Int J Cancer* **2000**, *86*, 863-869.

(12) van den Maagdenberg, A. M.; Pietrobon, D.; Pizzorusso, T.; Kaja, S.; Broos, L. A.; Cesetti, T.; van de Ven, R. C.; Tottene, A.; van der Kaa, J.; Plomp, J. J.; Frants, R. R.; Ferrari, M. D. *Neuron* **2004**, *41*, 701-710.

(13) Carreira, R. J.; Shyti, R.; Balluff, B.; Abdelmoula, W. M.; van Heiningen, S. H.; van Zeijl, R. J.; Dijkstra, J.; Ferrari, M. D.; Tolner, E. A.; McDonnell, L. A.; van den Maagdenberg, A. M. *J Am Soc Mass Spectrom* **2015**, *26*, 853-861.

(14) Lauritzen, M. *Brain* **1994**, *117 (Pt 1)*, 199-210.

(15) Somjen, G. G. *Physiol Rev* **2001**, *81*, 1065-1096.

(16) Benjamini, Y.; Hochberg, Y. *Journal of the Royal Statistical Society. Series B (Methodological)* **1995**, *57*, 289-300.

(17) Kertesz, V.; Van Berkel, G. J. *Anal Chem* **2008**, *80*, 1027-1032.

(18) Bivand, R.; Hauke, J.; Kossowski, T. *Geogr Anal* **2013**, *45*, 150-179.

(19) Bivand, R.; Piras, G. *J Stat Softw* **2015**, *63*, 36.

(20) Besag, J. *Journal of the Royal Statistical Society. Series B (Methodological)* **1974**, *36*, 192-236.

(21) Aarts, E.; Verhage, M.; Veenvliet, J. V.; Dolan, C. V.; van der Sluis, S. *Nat Neurosci* **2014**, *17*, 491-496.

(22) F. Dormann, C.; M. McPherson, J.; B. Araújo, M.; Bivand, R.; Bolliger, J.; Carl, G.; G. Davies, R.; Hirzel, A.; Jetz, W.; Daniel Kissling, W.; Kühn, I.; Ohlemüller, R.; R. Peres-Neto, P.; Reineking, B.; Schröder, B.; M. Schurr, F.; Wilson, R. *Ecography* **2007**, *30*, 609-628.

(23) Boggio, K. J.; Obasuyi, E.; Sugino, K.; Nelson, S. B.; Agar, N. Y.; Agar, J. N. *Expert Rev Proteomics* **2011**, *8*, 591-604.

(24) Jurchen, J. C.; Rubakhin, S. S.; Sweedler, J. V. *J Am Soc Mass Spectrom* **2005**, *16*, 1654-1659.

(25) Manicke, N. E.; Kistler, T.; Ifa, D. R.; Cooks, R. G.; Ouyang, Z. *J Am Soc Mass Spectrom* **2009**, *20*, 321-325.

(26) Van de Plas, R.; Yang, J.; Spraggins, J.; Caprioli, R. M. *Nat Methods* **2015**, *12*, 366-372.

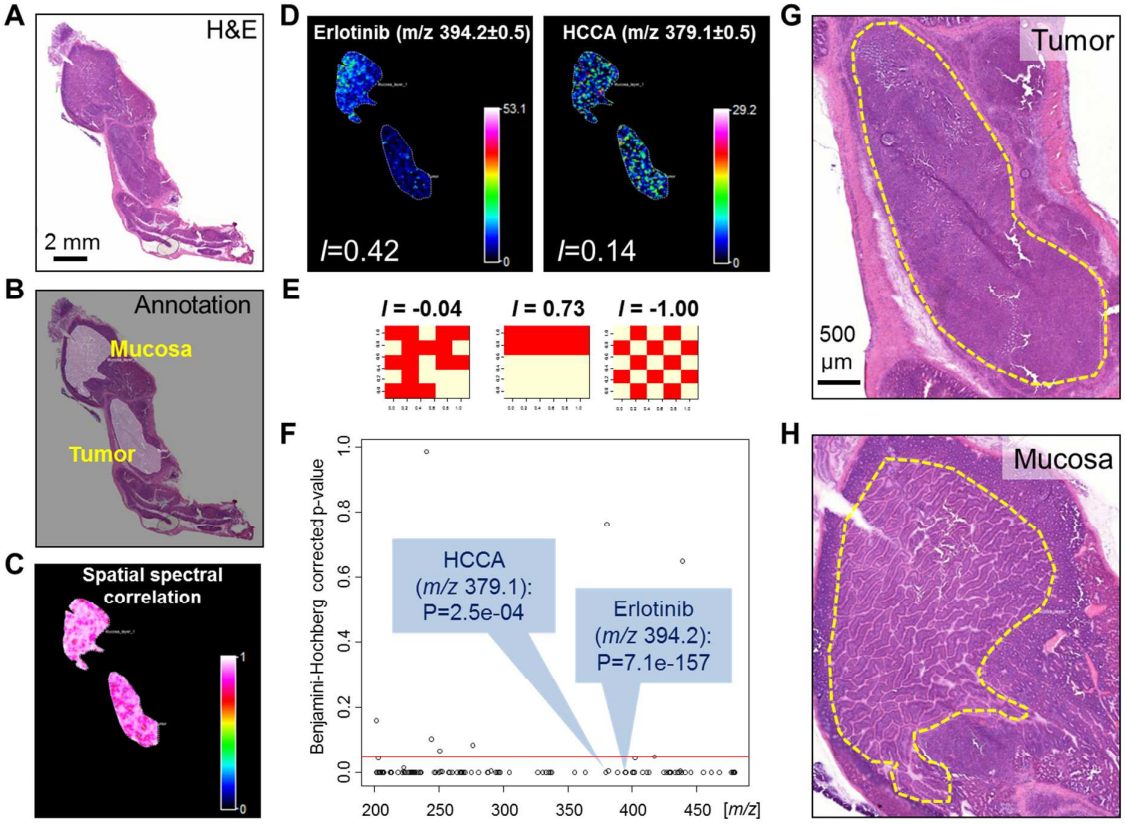
TABLES

Table 1 Dataset descriptions.

	Drug imaging dataset	**CSD coronal mouse brain**	**High-resolution sagittal mouse brain**
ROIs to compare / purpose	Erlotinib signal in tumor vs. mucosa	Biochemical changes in CSD cortex vs. non-CSD cortex	Investigation of spatial autocorrelation as function of distance
Spatial resolution	80 µm	100 µm	20 µm
Pixels	Tumor: 997 Mucosa: 1057	CSD cortex: 967 Control cortex: 919	Total: 131,082
Mass range	200-480	3,000-19,200	300-1,000
Molecular class	Metabolites	Proteins	Lipids
Number of features	106	112	125

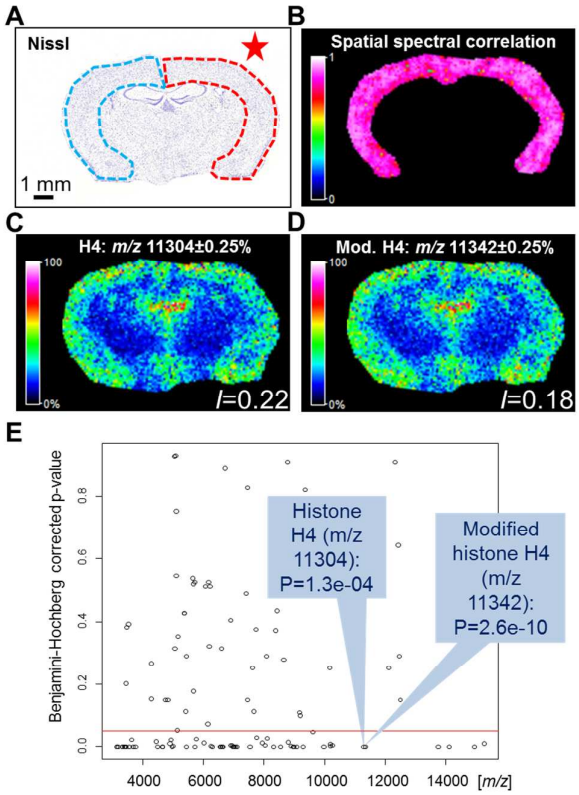
FIGURE LEGENDS

Figure 1 Drug imaging dataset.



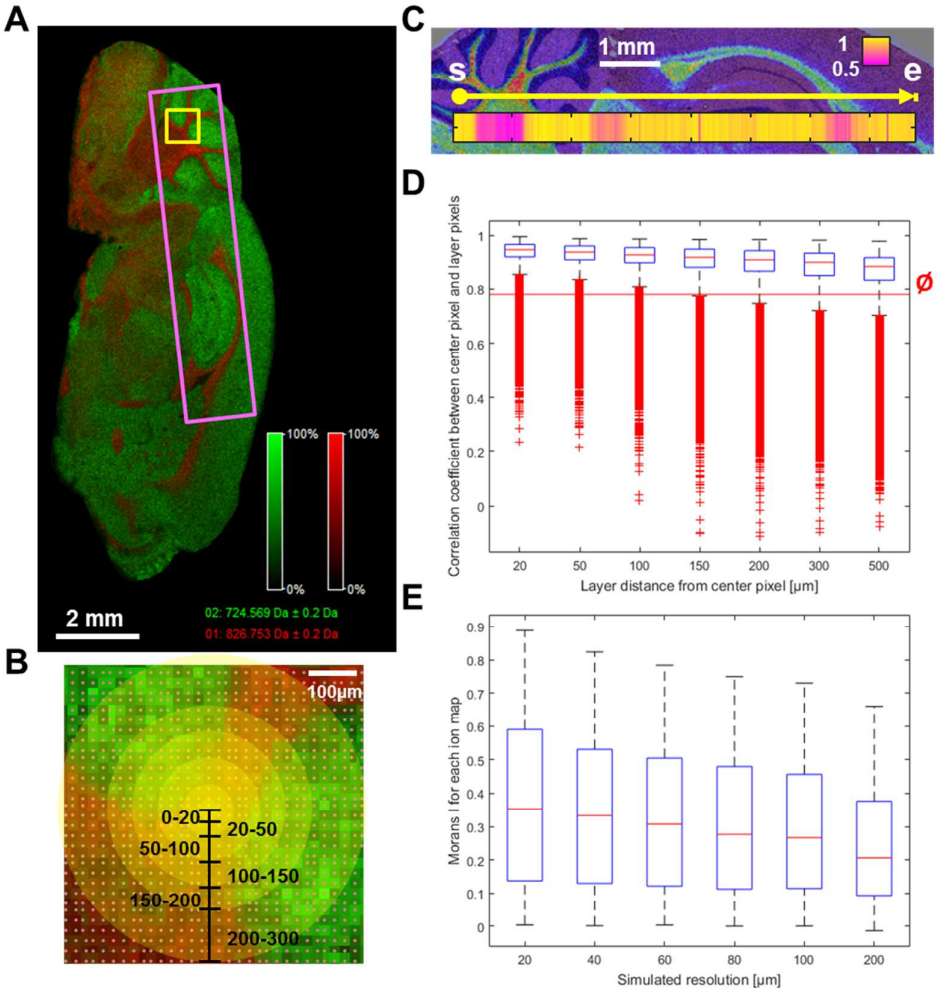
(**A**) Microscopic image of the mouse stomach after hematoxylin and eosin (H&E) staining. (**B**) Histology was used to obtain regions of interest (ROI) such as the tumor area (**G**) and epithelial layer of the mucosa (**H**). Each feature was tested for significant intensity differences between the ROIs. Adjusted p-values of the *t*-tests are shown in (**F**) with insets highlighting the p-values of erlotinib and the alpha-cyano-4-hydroxycinnamic acid (HCCA) matrix. (**D**) Mass spectrometry image of erlotinib and HCCA in both ROIs with level of spatial autocorrelation determined by Moran's I. (**E**) Examples for Moran's I values ranging from perfect dispersion (right), over random dispersion (left), to high spatial correlation (middle). (**C**) Spectral correlation between neighboring pixels confirms presence of spatial autocorrelation.

Figure 2 CSD coronal mouse brain dataset.



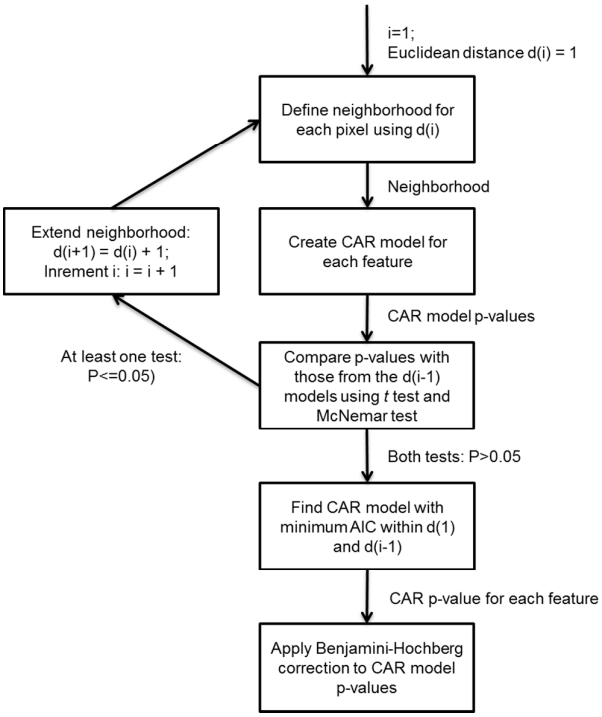
(**A**) Microscopic image of the mouse brain after Nissl staining and histology was used to obtain cortical spreading depression-unaffected (blue) and affected (red) cortex regions. (**E**) Each feature was tested for significant intensity differences between both cortices. P-values are shown in with insets highlighting the p-values of histone H4 and its modified version. Mass spectrometry image of distributions of histone H4 (**C**) and its modified version (**D**) with level of spatial autocorrelation determined by Moran's I. (**B**) Spectral correlation between neighboring pixels confirm presence of spatial autocorrelation.

Figure 3 High-resolution sagittal mouse brain dataset.



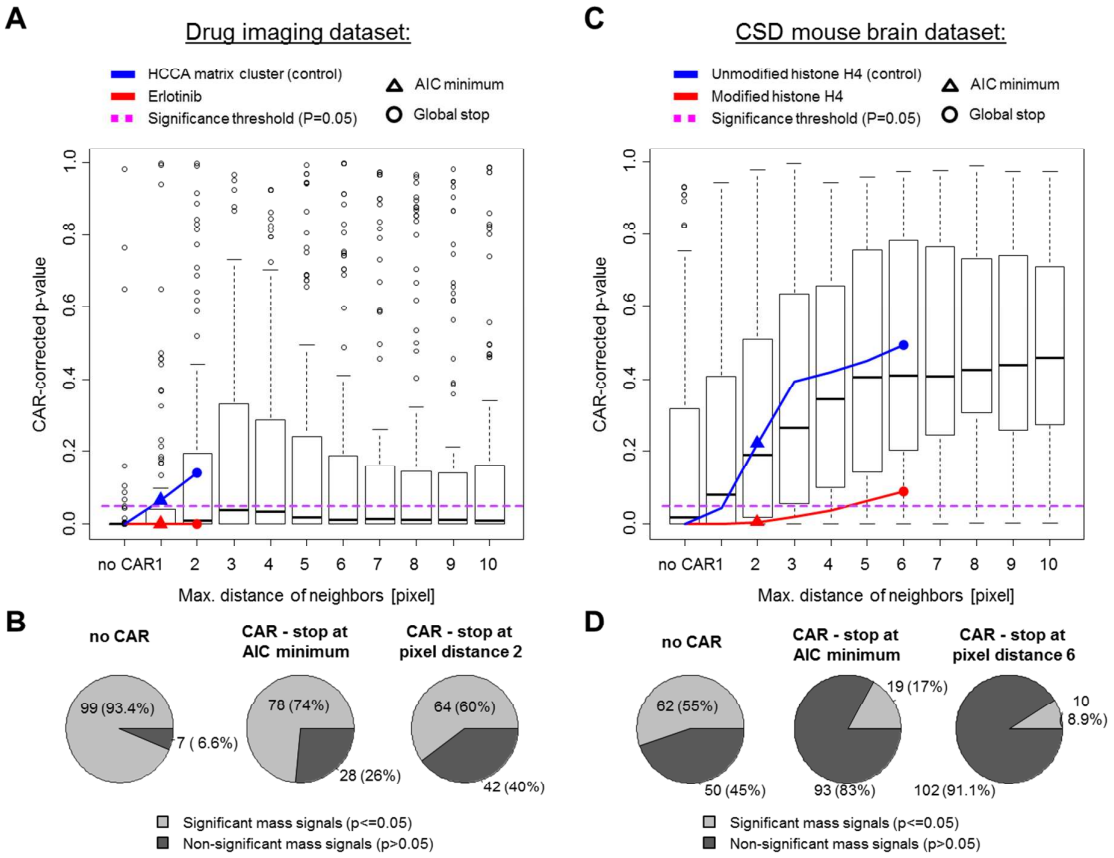
(**A**) Merged mass spectrometry image of two lipid signals in the sagittal mouse brain dataset. The yellow square is magnified in (**B**) which illustrates the different pixel layers used to calculate the spatial spectral correlation for each pixel, as shown in (**D**). (**C**) Spatial spectral correlation depends on anatomical features. This is highlighted by the correlation coefficients (shown in the color bar) between all pixels that lie on the trajectory from s to e (yellow line) with respect to the first pixel s. (**E**) Boxplots show Moran's I values after downscaling each MS image to simulated pixels sizes of 20, 40, 60, 80, 100, and 200 µm.

Figure 4 Within-sample statistical tests workflow.



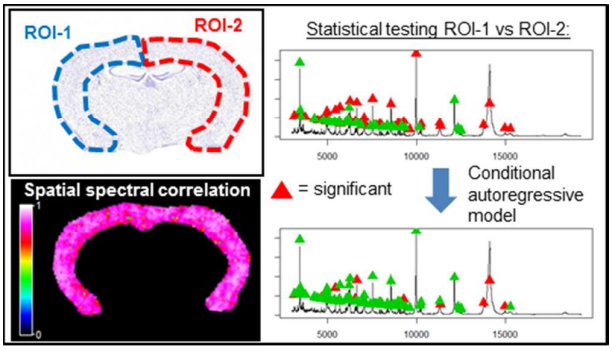
Abbreviations used: CAR, conditional autoregressive model; AIC, Akaike information criterion.

Figure 5 Conditional autoregressive model-corrected p-values.



Boxplots of conditional autoregressive (CAR) model-corrected p-values are shown as a function of the maximum distance of the neighborhood pixels for the drug imaging dataset (**A**) and the cortical spreading depression (CSD) mouse brain dataset (**C**). The course of individual features are shown in red and blue where triangles indicate the p-value obtained from the CAR model with the lowest Akaike information criterion and dots indicate the p-value obtained after a global stop of the algorithm due to non-significant changes from one distance to the following one. Reduction rates in significant features for both approaches are visualized as pie charts for the drug imaging dataset (**B**) and for the CSD mouse brain dataset (**D**).

For TOC only


ROI-1
ROI-2
Spatial spectral correlation
Statistical testing ROI-1 vs ROI-2:
= significant
Conditional autoregressive model
5000
10000
15000