analytical chemistry

Article

pubs.acs.org/ac

Automatic Generic Registration of Mass Spectrometry Imaging Data to Histology Using Nonlinear Stochastic Embedding

Walid M. Abdelmoula,[†] Karolina Škrášková,[#] Benjamin Balluff,[‡] Ricardo J. Carreira,[‡] Else A. Tolner,[§,∥] Boudewijn P. F. Lelieveldt,[†,∇] Laurens van der Maaten,[∇] Hans Morreau,[⊥] Arn M. J. M. van den Maagdenberg,[§,∥] Ron M. A. Heeren,[#] Liam A. McDonnell,*[,‡,○] and Jouke Dijkstra[†]

[†]Division of Image Processing, Department of Radiology, [‡]Center for Proteomics and Metabolomics, [§]Department of Human Genetics, [∥]Department of Neurology, and [⊥]Department of Pathology, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands

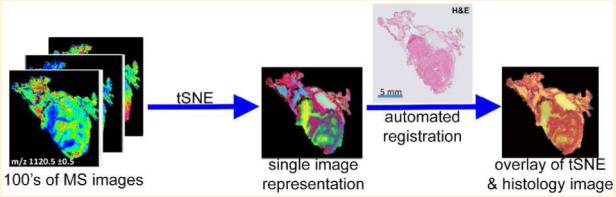
[#]FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

[∇]Intelligent Systems Group, Faculty of EEMCS, Delft University of Technology, 2600 AA Delft, The Netherlands

[○]Fondazione Pisana per la Scienza ONLUS, 56125 Pisa, Italy

Supporting Information

ABSTRACT: The combination of mass spectrometry imaging and histology has proven a powerful approach for obtaining molecular signatures from specific cells/tissues of interest, whether to identify biomolecular changes associated with specific histopathological entities or to determine the amount of a drug in specific organs/compartments. Currently there is no software that is able to explicitly register mass spectrometry imaging data spanning different ionization techniques or mass analyzers. Accordingly, the full capabilities of mass spectrometry imaging are at present underexploited. Here we present a fully automated generic approach for registering mass spectrometry imaging data to histology and demonstrate its capabilities for multiple mass analyzers, multiple ionization sources, and multiple tissue types.



Mass spectrometry imaging (MSI) is a rapidly developing imaging modality that can provide the spatial distribution of hundreds of biomolecules directly from tissue.[1] It has already had a substantial impact in clinical and pharmacological research, uncovering biomolecular changes associated with disease[2] and providing low-cost imaging of pharmaceuticals and their metabolites for drug formulation development.[3] The integration of the biomolecular information obtained by MSI with the anatomical structure provided by histology has proven essential for its clinical and pharmacological application,[4] for example, to identify biomolecular changes associated with specific histopathological entities[2] (e.g., tumors) or to determine the amount of a drug in specific organs/compartments.[3]

The insignificant loss of histoanatomical structures after performance of MSI experiments allows coregistration between MSI data and its histological image.[4,5] To date, this is performed by most researchers either manually or, for data sets acquired on Bruker Daltonics instruments running the FlexImaging MSI data acquisition software, semiautomatically by using fiducial markers. Veselkov et al. recently reported using binary masks of the histological image and MSI data to perform the registration automatically.[6] In this approach the registration algorithm aligns the boundaries of the masks using a global transformation. While this approach is suited to the desorption electrospray ionization based MSI experiments reported in the paper, the significant background in MSI data sets recorded using matrix-assisted laser desorption/ionization[7,8] (MALDI) and secondary ion mass spectrometry[9] (SIMS) make defining the MSI binary mask more problematic.

Furthermore, MALDI and SIMS MSI data sets are frequently acquired from nontransparent mounting substrates (e.g., a gold-coated steel plate or silicon wafer); in such cases the histological images are acquired from proximal tissue sections. Small histological differences between the tissue sections as well as local deformations resulting from their preparation (folds, tears) mean that localized elastic transformations are necessary for their correct registration. A generic registration approach must therefore accurately trace the local differences in tissue structure to make it robust to the background signals present in MALDI and SIMS measurements.

The main challenge is to automatically determine the spatial correspondences between the MSI data and the histological image. The multivariate techniques *k*-means clustering,[10] principal component analysis (PCA),[11] probabilistic latent semantic analysis,[12] and non-negative matrix factorization[13] have all been used to approximately demarcate, on the basis of the MSI signals, different histological regions. These are all

Received: June 12, 2014
Accepted: August 18, 2014

Table 1. Overview of MSI Data Sets Used in This Study[a]

tissue	sample type	ion source	mass analyzer	pixel size (μm)	molecular class measured	histology
thyroid cancer	FFPE	MALDI	TOF/TOF	150	proteolytic peptides	H&E
mouse brain	frozen	MALDI	TOF	100	proteins	Nissl
mouse brain	frozen	MALDI	ion mobility TOF	150	lipids	Nissl
mouse brain	frozen	SIMS	TOF	19.2	metabolites	Nissl

[a]Abbreviations used: FFPE, formalin-fixed and paraffin-embedded; H&E, hematoxylin and eosin. Note: The SIMS data sets were recorded with 0.3 μm pixel size but were rebinned to 19.2 μm for visualization of the entire area in a single 512 × 512 pixel image. All calculations were performed on this rebinned image.

linear dimensionality reduction algorithms that focus on representing dissimilar data points in a lower dimensional space (e.g., the maximization of variance in PCA is determined by the most dissimilar data points in Euclidean space). One of the difficulties of using these methods is selecting the appropriate number of dimensions; a number of papers have shown that the images generated by these methods are dependent on the number of dimensions (components) selected for the analysis.[10] Another is that, by focusing on keeping the most dissimilar data points far apart in the lower dimensionality representation, they can fail to preserve the local structure of the data.[14] In MSI this means that the analysis implicitly focuses on the largest differences in the data set, and can merge regions whose molecular differences are minor in comparison.[15] While this merging may be alleviated by changing the number of dimensions used in the multivariate analysis, the dependence of the images on the number of dimensions (clusters) and the bias toward the largest Euclidian differences in the data set make such techniques suboptimal for summarizing the spatial structures of MSI data sets.

Fonville et al. recently demonstrated that the nonlinear technique t-distributed stochastic neighbor embedding (tSNE) outperforms linear dimensionality reduction techniques for summarizing MSI data sets.[15] tSNE is a nonlinear dimensionality reduction technique developed by van der Maaten et al. that maps data points from high-dimensional space into a matrix of pairwise similarity in a lower dimensional space.[14] The hallmark that characterizes tSNE is its ability to capture the local structures of high-dimensional data as well as preserving their global features. In MSI this means that relationships characterized by large differences in mass spectral profiles can be visualized concomitantly with those characterized by minor differences (which would be merged by linear techniques such as PCA).[15]

The tSNE representation of MSI data reveals clearly distinguishable anatomical regions that can be treated as landmarks for guiding the coregistration process with histology. Importantly, the tSNE analysis does not require any user input and so can be completely automated. Here we report tSNE-enabled automated alignment of MSI data sets with histology. The method is generic, and we demonstrate its ability on data sets from different organs, different mass spectrometers, and different ionization methods.

■ METHODS

Experimental Data Sets. The automatic alignment routine has been tested on data sets from four different mass spectrometers, representing four different types of MSI experiments, and spanning a wide range of spatial resolution. The algorithm was then validated on a sizable animal cohort of 96 mouse brain coronal tissue sections. Table 1 provides a summary of the MSI data sets. Further experimental details about the MSI data acquisition can be found in the Supporting Information.

Histology Preprocessing. The stained histological images need first to be preprocessed to exclude the background noise, correct for potential image acquisition artifacts (e.g., inhomogeneous lighting and exposure, noise because of dust on the slides), and maximize contrast. We applied the histological preprocessing pipeline proposed by Abdelmoula et al.[16] in which the images were classified into two clusters using k-means ($k = 2$) followed by morphological operations (opening, closing, and region filling with a disk-shaped structural element) to close any potential gaps in the clustered image. The resulting binary mask is then used to separate the tissue from the background.

MSI Preprocessing. *MALDI-TOF—Proteolytic Peptides.* Each pixel's mass spectrum was first processed using FlexAnalysis (version 3.4, Bruker Daltonics); mass spectral smoothing was performed with the Gauss algorithm (width 0.02 m/z and two cycles) and baseline subtraction with the top-hat algorithm. The MSI data were read into MATLAB R2013a (MathWorks, Natick, MA) where they underwent total-ion-count normalization.[7] Peak picking was performed on the global base peak mass spectrum after smoothing, resampling, and baseline subtraction and was performed using an adapted version of the data reduction code previously reported by McDonnell et al.[17] The base peak spectrum displays the maximum intensity detected in the entire imaging data set for every peak and is more effective for detecting peaks with localized distributions.[17] Peak areas were then extracted from every pixel's mass spectrum. This reduced and more computationally manageable representation of a mass spectrum was then placed, on the basis of its original coordinate information, as a pixel into a project-specific data cube[13] and was used for the subsequent registration with histology.

MALDI-TOF—Proteins. The data set was processed identically to the MALDI-TOF—proteolytic peptides data set, except the mass spectral preprocessing parameters were adapted for intact proteins. Here, each pixel's mass spectrum was smoothed using the Savitsky−Golay algorithm with a width of 2.0 m/z and five cycles and baseline subtracted with the top-hat algorithm (10% width).

MALDI Synapt. The data preprocessing was done employing our in-house-developed ChemomeTricks toolbox for MATLAB (MathWorks). In the first step the raw data were converted into a MATLAB format. Mass channels were binned into 0.1 Da wide mass bins. Peak picking was performed on a global mean mass spectrum after smoothing. The peak picking algorithm has been described in detail elsewhere.[18] The created peak list consisted of 1707 mass channels, each of which was defined by its center m/z and an m/z window (peak width at the baseline). The peak list was used to integrate each pixel's mass spectrum.

TOF-SIMS. The data preprocessing was done employing our in-house-developed ChemomeTricks toolbox for MATLAB (MathWorks). Mass channels were binned into 0.05 Da wide mass bins. An average spectrum of all pixels was used for peak picking. Peak picking was performed on a global mean mass spectrum after smoothing as described in detail by Eijkel et al.[18] The created peak list consisted of 1400 selected mass channels. Pixels were spatially binned, resulting in a 256 × 256 pixel data set and a final spatial resolution of 19.2 μm. The peak list was used to integrate each pixel's mass spectrum. Subsequently, a multiorder correction algorithm based on linear discriminant analysis (LDA) was applied to remove MS image distortions caused by the mosaic character of the data acquisition.[19] Finally, the data were recalibrated on gold coating related peaks with well-known m/z values.[20]

tSNE of MSI Data Sets. Each processed MSI data set was unfolded into a set of 1D vectors, $\mathbf{X} = [x_1, x_2, ..., x_k]$, in which each vector x_i represents the normalized mass spectral profile of the ith pixel. tSNE was then applied to find the low-dimensionality representation, in this case a 3D representation, $\mathbf{Y} = (y_1, y_2, y_3)$. The joint probabilities p_{ij} were first calculated to establish the pairwise similarities between data points x_i and x_j for all pairs in the high-dimensional space. Then the joint probabilities q_{ij} were calculated for all pairs y_i and y_j in the low-dimensional space. The optimum low-dimensional representation (i.e., $\mathbf{Y}$) that maximizes the similarities between p_{ij} and q_{ij} was found by minimizing the Kullback−Leibler divergence KL over all data points:

$$\sum_i \mathrm{KL}(P_i \| Q_i) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{1}$$

where P_i and Q_i represent the joint probabilities in the high- and low-dimensional spaces, respectively. The optimization problem was solved using the gradient descent method, yielding an optimum 3D representation of the original hyperdimensional MSI data set. For visualization, each of the three tSNE output dimensions was treated as a separate color channel, and the results were displayed as a 2D RGB (red, green, blue) image.[15] tSNE was performed using the default settings described by van der Maaten et al.[14] and the tSNE Matlab toolbox (http://homepage.tudelft.nl/19j49/t-SNE.html).

Image Registration. The high-resolution histological images and the MSI data were acquired from either the same tissue sections (MALDI data) or adjacent sections (SIMS data). In the former case the histological images and MSI data differ only in their coordinate space and image resolution and thus can be registered using rotation, scaling, and translation (rigid registration). For adjacent sections we also added an elastic deformation step to account for minor differences in brain region size as well as artifacts introduced during sectioning and mounting of the tissue sections.

The registration algorithm transforms a moving image, $I_m(x,y)$, to be spatially aligned with a fixed image, $I_f(x,y)$. The moving image was the gray scale tSNE image and the fixed image the preprocessed histological image. The rigid transform was used to model rotation, scaling, and translation deformations through optimization of the standard registration problem given in the following equation:

$$\hat{\mu} = \arg\min_{\mu} C(I_f, I_m; T_\mu) \tag{2}$$

μ is a vector which contains the transformation parameters that were optimized by minimizing the cost function C with respect to the transformation model T_μ using the adaptive stochastic gradient descent optimizer.[21] The statistical metric mutual information[22] was used as a cost function to assess the registration quality. Mutual information (MI) has demonstrated high efficiency in multimodal data registration, particularly when the intensity distributions of the images differ. MI measures the degree of difference in the intensity distributions between the moving and fixed images through measurement of their marginal and joint entropies:

$$\mathrm{MI}(I_f, I_m) = H(I_f) + H(I_m) - H(I_f, I_m) \tag{3}$$

$H(I_f)$ and $H(I_m)$ represent the marginal entropies of the fixed and moving images, respectively. The best alignment is achieved through the transformation metric in which the joint entropy $H(I_f, I_m)$ is minimal.

For experiments that use the adjacent tissue section for histology, an additional step was incorporated in which the B-spline transform was used to correct any local deformations; mutual information was again the cost function, and the adaptive stochastic gradient descent optimizer was used to achieve the best similarity through optimization of the B-spline parameters. To capture deformations on different length scales, the registration was applied using a multiresolution scheme and implemented using elastix.[23] This elastic registration step is an adaptation of that previously reported for the registration of MSI data sets to the Allen Brain Atlas,[16] in which experimental histological images were registered to the reference histological images contained in the Allen Brain Atlas. In this paper we have adapted the algorithm to directly map the MSI data onto the histological image of a proximal tissue section.

RESULTS AND DISCUSSION

To automatically coregister MSI with histology, we have developed the pipeline shown in Figure 1. The key elements of

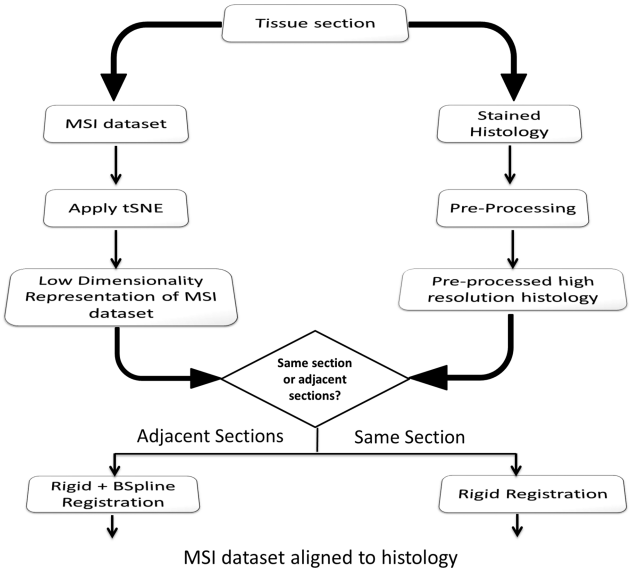


Figure 1. Proposed pipeline to automatically align MSI data to their histological image. The method is generic as it can be applied to different tissues and MSI data sets recorded using different types of mass spectrometers and mass spectrometers equipped with different ion sources.

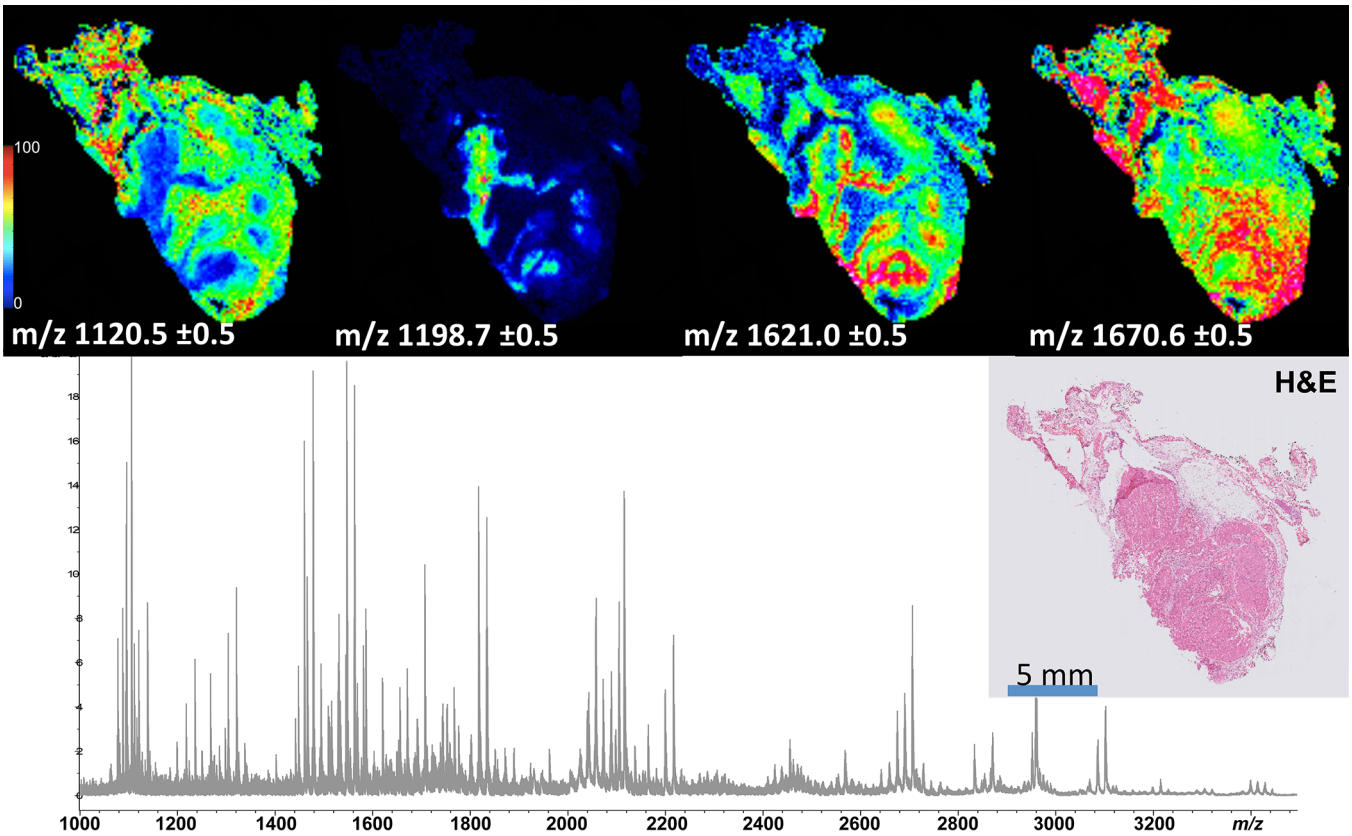


Figure 2. MALDI MSI analysis of a human oncocytic follicular thyroid cancer tissue using on-tissue tryptic digestion and measured using MALDI-TOF/TOF. The MSI data contain hundreds of proteolytic peptide ions, many of which exhibit highly structured distributions (top row). A comparison with the histological image (tissue section H&E stained after the MSI experiment) reveals that many ions are associated with specific histological features.

the pipeline are (i) mapping the MSI data set to a 3D space using tSNE to determine the spatial correspondences that are then used for the registration, (ii) image registration algorithm [for MSI and histology of the same tissue section, a rigid transformation is used; for MSI and histology of adjacent sections, elastic deformation is permitted to account for small differences in the sizes of the histological regions and for small artifacts introduced during the sectioning/mounting procedure (e.g., folds, tears)], and (iii) statistical measurement of MSI and histology fitness—mutual information[22] to overcome the inherent independency of the intensity distributions of the tSNE and histological images.

Figure 2 shows an example of an MSI data set in which the mass spectral signatures are clearly associated with the underlying histology. A thyroid cancer tissue section was first prepared for protein MALDI MSI via on-tissue tryptic digestion and then measured using an UltrafleXtreme MALDI-TOF/TOF instrument. Following MSI data acquisition and removal of excess MALDI matrix, the tissue was hematoxylin and eosin (H&E) stained and a high-resolution optical image recorded. Figure 2 shows the average mass spectrum, the original histological image, and example MS images. It can be seen that the MSI experiment detected a large number of proteolytic peptide ions, many of which were localized to distinct histological regions of the thyroid cancer tissue section. Despite the high contrast of the MSI images, it is far from straightforward to determine which of the distinct MS images best follow the tissue section's histology.

In agreement with Fonville et al.,[15] we found that a 3D representation of the MSI data using tSNE, and visualized as an RGB image, reproducibly produces summary images that exhibit clear correspondences with the tissue section's histology. Accordingly, we surmised that the tSNE map could be used to automatically guide the registration algorithm for finding the optimal transformation to spatially align MSI with histology. The original histological image of the thyroid cancer tissue section was preprocessed to exclude the background, normalize contrast, and exclude potential image artifacts that might bias the registration algorithm[16](Figure 3a). The tSNE representation of the MSI data is shown in Figure 3b; the color coding clearly highlights different histological regions. In this

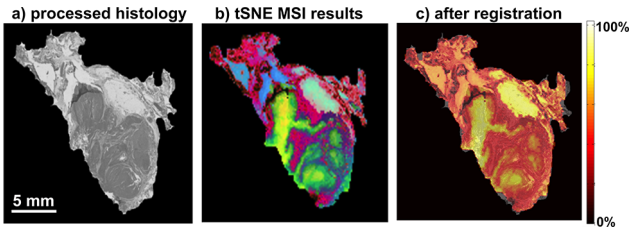


Figure 3. Coregistration of MALDI MSI data and histological image of thyroid cancer tissue: (a) preprocessed histological image; (b) low-dimensionality representation of the high-dimensional MALDI MSI data using tSNE (which is used as the moving image in the registration process); (c) fusion result—overlay of the processed histological image and registered tSNE results.

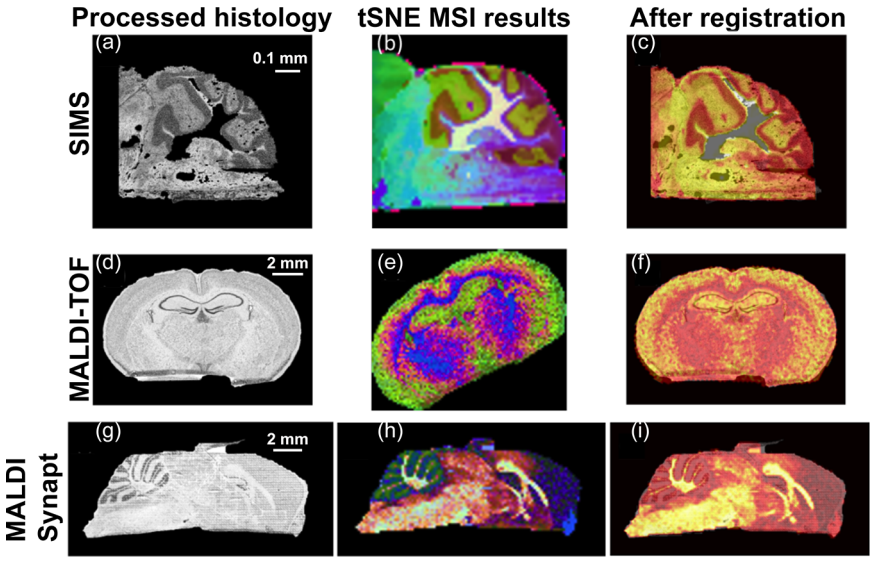


Figure 4. Coregistration of MSI data sets and their histological images. The data sets are from different mice and different mass spectrometers (SIMS, MALDI-TOF, and MALDI-Synapt). tSNE representations of the MSI data sets (second column) show clear spatial correspondences with their associated histological images (first column), enabling registration to be performed successfully (third column; for improved clarity the histological image and tSNE representation are shown in gray scale and hot color scale, respectively).

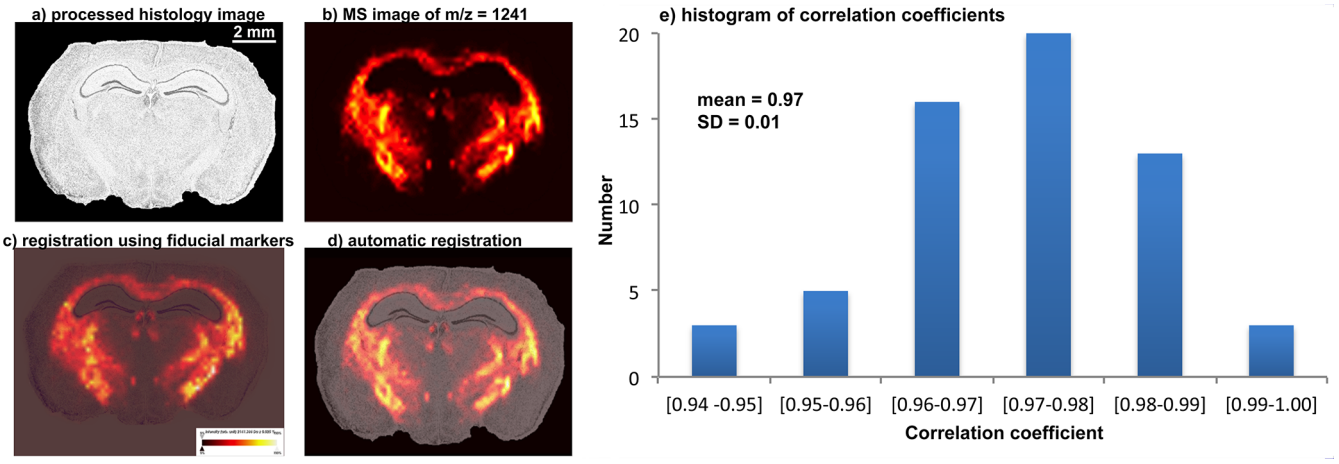


Figure 5. Comparison between semiautomatic and automatic coregistration of mouse brain data sets: (a) preprocessed histology, (b) original spatial distribution of a selected mass (m/z = 1241 Da), (c) fusion result combining the histological image and the MS image (coregistration was performed semiautomatically and was based on manually selected fiducial markers), (d) fusion result combining the histological image and the MS image (coregistration was performed automatically using tSNE), (e) histogram of correlation coefficients between 60 MSI data sets of coronal mouse brain tissue sections automatically registered using the tSNE-based pipeline and semiautomatically registered using FlexImaging. Figure S-4 (Supporting Information) shows the results of the automatic registration for all 60 tissue sections.

example the histology image and MSI data were from the exact same tissue section. The tSNE image could thus be registered using a rigid registration (scale, translation, rotation) and using the mutual information as the registration metric (as mutual information can accommodate the different intensity distributions and color scales of the images). The high accuracy of the registration can be seen in Figure 3c, in which the registered tSNE image (using the hot color map) is placed on top of the gray scale processed histology image.

To provide examples of the general applicability of the approach for different MSI platforms, different ionization methods and different application areas, three mouse brain tissue sections, which were sectioned differently (i.e., coronal and sagittal) and analyzed in different mass spectrometers are shown in Figure 4. The top row shows a high spatial resolution SIMS MSI analysis, using a TOF-SIMS instrument, of the cerebellum region of a sagittal tissue section. The middle row shows a protein MALDI MSI analysis, using linear MALDI-TOF, of a coronal tissue section of a mouse brain. The bottom row shows a lipid MALDI MSI analysis, using a MALDI ion mobility TOF instrument, of a sagittal tissue section of a mouse brain. In each case tSNE of the MSI data reveals clearly distinguishable anatomical features, for example, cerebellar cortex (Figure 4b), corpus callosum (Figure 4e), and cerebellum (Figure 4h). The anatomical landmarks generated by the tSNE representations enable the MSI data sets to be registered to the histology images (Figure 4c,f,i). Overlaying the tSNE images on top of the histology images demonstrates the high alignment accuracy. Additional examples of the registration of SIMS, MALDI-TOF, and MALDI ion mobility TOF are included in Figure S-1 (Supporting Information).

The SIMS MSI and histology data shown in Figure 4 were of adjacent sections, so there were minor differences between the histology image and the MSI data due to the manual nature of mounting the thin tissue sections onto the target plate. In this instance an elastic registration step was necessary to account for the local deformations between the MSI data and the histology image (Figure S-2, Supporting Information).

To quantify the accuracy of the registration, a set of control points were selected in the histological and MSI images. Figure S-3 (Supporting Information) shows the control points selected for coronal mouse brain tissue sections as well as the results of the registration. After registration the errors ranged from under 10 μm for the SIMS data set to approximately 40 μm for the MALDI-TOF analysis of mouse brain tissue sections to 80 μm for MALDI-TOF analysis of tryptic peptides in thyroid cancer tissue. In each case the registration accuracy was sufficient that any errors were less than the size of a single MSI pixel (see Table 1).

The tSNE-based automatic registration algorithm was then compared to the only commercial package currently available and de facto standard for registering histology and MSI data, namely, FlexImaging from Bruker Daltonics. FlexImaging is Bruker's MSI data acquisition and data analysis software and is only compatible with Bruker MALDI mass spectrometers. To record MSI data using FlexImaging, the mass spectrometer's sample stage is first aligned to an optical image of the MALDI-matrix-coated tissue. This alignment is performed by manually selecting features in the matrix-coated-tissue image and manually selecting the corresponding features in the mass spectrometer's sample visualization system. In this manner the mass spectrometer's coordinate system, and thus the MSI data, is aligned to the matrix-coated-tissue image. After MSI data acquisition, the histology image is then registered to the MSI data through the matrix-coated-tissue image by selecting common features in the high-resolution histology image and the matrix-coated-tissue image.

Figure 5a shows the preprocessed high-resolution optical image of a coronal tissue section of a mouse brain and Figure 5b the spatial distribution of a selected mass (m/z = 1241). FlexImaging was then used to align the histology image and the MSI data (Figure 5c), and the tSNE-based automatic registration algorithm was applied to the same data (Figure 5d). Visual inspection of the automatic and semiautomatic coregistration results shows a close consensus in the MS distribution with respect to the tissue's anatomy. To validate the automatic registration algorithm, its results were compared with those from FlexImaging's semiautomatic registration for data sets from 60 coronal mouse brain tissue sections spanning three different molecular classes (20 metabolite MSI data sets, 20 peptide MSI data sets, and 20 protein MSI data sets). Parts a–l of Figure S4 (Supporting Information) visualize the results of the tSNE-based automatic registration algorithm. The Pearson correlation between the automatically registered results and those from the FlexImaging semiautomatic method was then calculated. A histogram of the resulting correlation coefficients, Figure 5e, demonstrates excellent agreement between the two methods, with a mean correlation coefficient of 0.97 and a standard deviation of 0.01. Figure S-4m shows the histogram of the Dice coefficients, another image overlap metric that again confirms the high quality of the automatic registration algorithm.

We have developed an automatic generic technique to coregister MSI data sets to their histological images; we have demonstrated its applicability to MSI data sets measured on different mass spectrometers using different ionization mechanisms and different tissue samples and validated the results using a large series of mouse brain tissue sections. The tSNE representation plays a vital role in the registration by summarizing the spatiomolecular organization of the tissue, which has clear correspondences with the tissue section's histology. While even a single tSNE dimension was sufficient to reveal the spatiomolecular organization, the 3D tSNE map was significantly smoother and so was used here (Figure S-5, Supporting Information).

The computational and memory requirements of the original tSNE algorithm,[14] as used by Fonville et al.,[15] scale with the square of the number of data points. An MSI data set of just 200 × 100 pixels, and 500 detected peaks, contains 10 million data points. Accordingly, tSNE analyses could run very slowly. A new implementation, termed the Barnes–Hut implementation,[24] scales as $N \log N$ for computation and N for memory and thus enables tSNE of MSI data sets to be run much more practically. Freely available code, for many different platforms, is available from the tSNE Web site.[25]

All the experiments referred to here were recorded using MALDI or SIMS, ionization methods that generate a substantial background signal and so are not well suited to previously reported methods based on the rigid registration of binary images.[6] Figure S-6 (Supporting Information) shows a comparison of the registration results for MALDI MSI of a coronal mouse brain tissue section using the binary image registration method with those obtained using tSNE. It is immediately apparent that there is a translation error in the registration performed using binarized images (due to the background in MALDI MSI data sets). Furthermore, high spatial resolution analyses such as those presented in Figure 4a–c often focus on specific regions of tissue rather than the entire section because of the measurement time/memory demands of the experiment. These MSI data sets do not contain the tissue border regions necessary for the binary image registration method.[6] Finally, binary images do not contain the internal structures needed for elastic registration algorithms to align MSI and histological data from adjacent tissue sections (Figure 4a–c; Figures S-1 and S-2, Supporting Information).

This automatic histology–MSI registration pipeline will enable joint histology–MSI experiments to be performed irrespective of the ionization method or mass analyzer used to acquire the MSI data. Accordingly, virtual microdissection can be used to extract region-specific mass spectra from disease entities, e.g., tumors, to enable biomarker discovery experiments utilizing the full repertoire of MSI approaches. Furthermore, by combining the automatic histology–MSI registration pipeline with that previously reported by Abdelmoula et al.,[16] MSI data sets of mouse brain tissue sections can be automatically aligned to the Allen Brain Atlas.[26] The Allen Brain Atlas alignment routine requires the MSI data set and its associated histology to already be registered to each other. Previously, this was performed using fiducial markers in the Bruker Daltonics FlexImaging software. However, this limited the approach to MALDI MSI data recorded using instruments from Bruker Daltonics. The generic and automated histology–MSI coregistration pipeline reported here means that all MSI data may be analyzed in the context of the reference atlas and gene expression data contained in the Allen Brain Atlas.

tSNE can also be used as a distinct classification tool.[27] In a process termed "automatic classification of cellular expression by nonlinear stochastic embedding" (ACCENSE), Shekhar et al. utilized tSNE and a density-based partitioning of the tSNE space to demarcate T-cells into groups on the basis of the expression levels of 35 proteins, measured using mass cytometry.[27] The application of a similar density-based partitioning to the results of a tSNE analysis of MSI data would enable the identification of clusters without the need to predefine their number (as is necessary in NMF, PLSA, and *k*-means clustering). It is expected that the combination of automatic MSI–histology alignment reported here and a classifier (whether based on tSNE or another classification algorithm) will enable the automated identification of specific regions/organs of interest and thereby the automated extraction of their mass spectral profile. Such capabilities would greatly facilitate the biomedical application of MSI, whether for clinical biomarker discovery experiments or quantification of the level of a drug in different animal organs.

■ CONCLUDING REMARKS

MSI experiments can now be performed using a diverse array of ionization methods and mass analyzers that offer complementary capabilities. The development of the imzML data standard[28] and open source data analysis tools[29,30] now enable the MSI data from different platforms to be more readily compared and combined, the latter for greater biomolecular depth of coverage. The automated generic MSI–histology registration tool reported here represents an important development in the efforts to increase the impact, accessibility, and intercomparison of MSI data because it delivers one of the principal strengths of MSI for biomedical analysis (the ability to acquire cell/region-specific mass spectra from tissues with complex histologies) for any combination of mass analyzer and ionization method.

■ ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: L.A.Mcdonnell@lumc.nl. Phone: +31 71 526 8744. Fax: +31 71 526 6907.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge financial support from Cyttron II and the Centre for Medical Systems Biology (CMSB) in the framework of The Netherlands Genomics Initiative (NGI) (A.J.M.v.d.M.). R.J.C., B.B., and E.A.T. are funded by the Marie Curie Actions of the European Union (R.J.C., Grant 303344, ENIGMAS FP7-PEOPLE-2011-IEF; B.B., Grant 331866, SITH FP7-PEOPLE-2012-IEF; E.A.T., FP7 Grant PCIG9-GA-2011-294233 Underneath Migraine). A.J.M.v.d.M. acknowledges support from the FP7 European Union EUROHEADPAIN under Grant 602633. A.J.M.v.d.M. and R.M.A.H. acknowledge support from the FP7 European Union Marie Curie IAPP Program BRAINPATH under Grant 612360.

■ REFERENCES

(1) McDonnell, L. A.; Heeren, R. M. A. *Mass Spectrom. Rev.* **2007**, *26*, 606–643.

(2) Schwamborn, K.; Caprioli, R. M. *Nat. Rev. Cancer* **2010**, *10*, 639–646.

(3) Prideaux, B.; Stoeckli, M. *J. Proteomics* **2012**, *75*, 4999–5013.

(4) Rauser, S.; Deininger, S.-O.; Suckau, D.; Höfler, H.; Walch, A. *Expert Rev. Proteomics* **2010**, *7*, 927–941.

(5) Schwamborn, K.; Krieg, R. C.; Reska, M.; Jakse, G.; Knuechel, R.; Wellmann, A. *Int. J. Mol. Med.* **2007**, *20*, 155–159.

(6) Veselkov, K. A.; Mirnezami, R.; Strittmatter, N.; Goldin, R. D.; Kinross, J.; Speller, A. V. M; Abramov, T.; Jones, E. A.; Darzi, A.; Holmes, E.; Nicholson, J. K.; Takats, Z. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 1216–1221.

(7) Norris, J. L.; Cornett, D. S.; Mobley, J. A.; Andersson, M.; Seeley, E. H.; Chaurand, P.; Caprioli, R. M. *Int. J. Mass Spectrom.* **2007**, *260*, 212–221.

(8) McDonnell, L. A.; van Remoortere, A.; van Zeijl, R. J. M.; Deelder, A. M. *J. Proteome Res.* **2008**, *7*, 3619–3627.

(9) Hanselmann, M.; Kothe, U.; Kirchner, M.; Renard, B. Y.; Amstalden, E. R.; Glunde, K.; Heeren, R. M. A.; Hamprecht, F. A. *J. Proteome Res.* **2009**, *8*, 3558–3567.

(10) Alexandrov, T.; Becker, M.; Deininger, S.-O.; Grasmair, G.; von Eggeling, F.; Thiele, H.; Maass, P. *J. Proteome Res.* **2010**, *9*, 6535–6546.

(11) McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R. *Anal. Chem.* **2005**, *77*, 6118–6124.

(12) Hanselmann, M.; Kirchner, M.; Renard, B. Y.; Amstalden, E. R.; Glunde, K.; Heeren, R. M. A.; Hamprecht, F. A. *Anal. Chem.* **2008**, *80*, 9649–9658.

(13) Jones, E. A.; van Remoortere, A.; van Zeijl, R. J. M.; Hogendoorn, P. C. W.; Boveé, J. V. M. G.; Deelder, A. M.; McDonnell, L. A. *PLoS One* **2011**, *6*, e24913.

(14) van der Maaten, L. J. P.; Hinton, G. E. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

(15) Fonville, J. M.; Carter, C.; Cloarec, O.; Nicholson, J. K.; Lindon, J. C.; Bunch, J.; Holmes, E. *Anal. Chem.* **2012**, *84*, 1310–1319.

(16) Abdelmoula, W. M.; Carreira, R. J.; Shyti, R.; Balluff, B.; van Zeijl, R. J. M.; Tolner, E. A.; Lelieveldt, B. F. P.; van den Maagdenberg, A. M. J. M.; McDonnell, L. A.; Dijkstra, J. *Anal. Chem.* **2014**, *86*, 3947–3954.

(17) McDonnell, L. A.; van Remoortere, A.; de Velde, N.; van Zeijl, R. J. M.; Deelder, A. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1969–1978.

(18) Eijkel, G. B.; Kükrer Kaletaş, B.; van der Wiel, I. M.; Kros, M.; Luider, T. M.; Heeren, R. M. A. *Surf. Interface Anal.* **2009**, *41*, 675–685.

(19) Gerber, F.; Marty, F.; Eijkel, G. B.; Basler, K.; Brunner, E.; Furrer, R.; Heeren, R. M. A. *Anal. Chem.* **2013**, *85*, 10249–10254.

(20) Altelaar, A. F. M.; Klinkert, I.; Jalink, K.; de Lange, R. P. J.; Adan, R. A. H.; Heeren, R. M. A.; Piersma, S. R. *Anal. Chem.* **2006**, *78*, 734–742.

(21) Klein, S.; Pluim, J. P. W.; Staring, M.; Viergever, M. A. *Int. J. Comput. Vis.* **2009**, *81*, 227–239.

(22) Viola, P.; Wells, W. M., III. *Int. J. Comput. Vis.* **1997**, *24*, 137–154.

(23) Klein, S.; Staring, M.; Murphy, K.; Viergever, M. A.; Pluim, J. P. W. *IEEE Trans. Med. Imaging* **2010**, *29*, 196–205.

(24) van der MaatenL. J. P. Barnes-Hut-SNE. Presented at the International Conference on Learning Representations, Scottsdale, AZ, May 2–4, 2013.

(25) van der Maaten, L. J. P. Manuscript in preparation.

(26) Lein, E. S.; Hawrylycz, M. J.; Ao, N.; Ayres, M.; Bensinger, A.; Bernard, A.; Boe, A. F.; Boguski, M. S.; Brockway, K. S.; Byrnes, E. J.; Chen, L.; Chen, L.; Chen, T. M.; Chin, M. C.; Chong, J.; Crook, B. E.; Czaplinska, A.; Dang, C. N.; Datta, S.; Dee, N. R.; Desaki, A. L.; Desta, T.; Diep, E.; Dolbeare, T. A.; Donelan, M. J.; Dong, H. W.; Dougherty, J. G.; Duncan, B. J.; Ebbert, A. J.; Eichele, G.; Estin, L. K.; Faber, C.; Facer, B. A.; Fields, R.; Fischer, S. R.; Fliss, T. P.; Frensley, C.; Gates, S. N.; Glattfelder, K. J.; Halverson, K. R.; Hart, M. R.; Hohmann, J. G.;

Howell, M. P.; Jeung, D. P.; Johnson, R. A.; Karr, P. T.; Kawal, R.; Kidney, J. M.; Knapik, R. H.; Kuan, C. L.; Lake, J. H.; Laramee, A. R.; Larsen, K. D.; Lau, C.; Lemon, T. A.; Liang, A. J.; Liu, Y.; Luong, L. T.; Michaels, J.; Morgan, J. J.; Morgan, R. J.; Mortrud, M. T.; Mosqueda, N. F.; Ng, L. L.; Ng, R.; Orta, G. J.; Overly, C. C.; Pak, T. H.; Parry, S. E.; Pathak, S. D.; Pearson, O. C.; Puchalski, R. B.; Riley, Z. L.; Rockett, H. R.; Rowland, S. A.; Royall, J. J.; Ruiz, M. J.; Sarno, N. R.; Schaffnit, K.; Shapovalova, N. V.; Sivisay, T.; Slaughterbeck, C. R.; Smith, S. C.; Smith, K. A.; Smith, B. I.; Sodt, A. J.; Stewart, N. N.; Stumpf, K. R.; Sunkin, S. M.; Sutram, M.; Tam, A.; Teemer, C. D.; Thaller, C.; Thompson, C. L.; Varnam, L. R.; Visel, A.; Whitlock, R. M.; Wohnoutka, P. E.; Wolkey, C. K.; Wong, V. Y.; Wood, M.; Yaylaoglu, M. B.; Young, R. C.; Youngstrom, B. L.; Yuan, X. F.; Zhang, B.; Zwingman, T. A.; Jones, A. R. *Nature* **2007**, *445*, 168−176.

(27) Shekhar, K.; Brodin, P.; Davis, M. M.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 202−207.

(28) Roempp, A.; Schramm, T.; Hester, A.; Klinkert, I.; Both, J.-P.; Heeren, R. M. A.; Stoeckli, M.; Spengler, B. *Methods Mol. Biol.* **2011**, *696*, 205−224.

(29) Rübel, O.; Greiner, A.; Cholia, S.; Louie, K.; Bethel, E. W.; Northen, T. R.; Bowen, B. P. *Anal. Chem.* **2013**, *85*, 10354−10361.

(30) Robichaud, G.; Garrard, K. P.; Barry, J. A.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 718−721.