The handle http://hdl.handle.net/1887/45501 holds various files of this Leiden University dissertation

Author: Abdelmoula, Walid M.
Title: Data analysis for mass spectrometry imaging: methods and applications
Issue Date: 2017-01-18
In this thesis we developed a number of automatic methods for multi-modal data registration, mainly between mass spectrometry imaging, imaging microscopy, and the Allen Brain Atlas. We have shown the importance of these methods for performing large scale preclinical biomarker discovery investigations for neurological disorders. We have also proposed a data-driven approach to stratify patients’ tumor tissues into molecularly distinct tumor subpopulations and automatically identify those tumor subpopulations that drive patient outcome. In the following sections we give a summarized overview, discussion, and our future perspective about the developments reported here.

8.1. Technical developments

8.1.1. Co-registration between MSI data and the Allen Brain Atlas

In Chapter 2 we proposed an image-processing pipeline to automatically map the MALDI MSI datasets from different mouse brains to the Allen brain atlas (ABA). The proposed pipeline consists of three key steps: (1) image preprocessing of histological images, (2) automatic selection of target image from the ABA, and (3) non-rigid image registration. The registration results were quantitatively assessed and the overall registration error (30 µm) was below the pixel resolution of the MSI data reported in this study (100 µm). This MSI-ABA automatic registration pipeline enables facile and rapid inter-animal comparisons, quantification of regional biomolecular content, and correlation of the MSI data to the gene-expression image data provided by the ABA.

Manual annotation by visual inspection of stained tissues remains the standard procedure when histology/anatomy is used to guide MSI data analysis. MSI is usually implemented within mass spectrometry laboratories and so such annotation is the main bottleneck for its biomedical application, particularly so when the pre-clinical study requires the analysis of a large number of samples. The development of automatic alignment and annotation tools capable of processing many samples will not only increase the throughput of data analysis, but also facilitate the comparison of MSI data from animal cohorts; i) by extracting data from specific brain regions; ii) by comparing the MSI distributions in different animals using the same coordinate space; and iii) and by enabling the correlation of large MSI datasets with, for instance, the anatomical or gene expression information available in the ABA.

The selection of the best target image from the ABA, i.e. the stereotaxic location of the experimental tissue section in the reference tissue atlas, is of utmost importance to the success of the registration process. The ABA contains 132 coronal and 21 sagittal sections uniformly spaced at 100 µm and 200 µm respectively, spanning the entire mouse brain. The design of the preclinical investigation into cortical spreading depression limited the search space of the ABA to between bregma positions -1.22 and -2.80 mm, or just 14 reference coronal tissue sections. Within this region we determined that the lateral extent of the hippocampus is linearly correlated with the position of the reference tissue section, and could be used to accurately and automatically determine the best match of the experimental tissue section. Similar approaches may be used for experimental tissue sections that have been sampled at different locations but need a different metric to determine the best target match from the ABA. The current work does not correct for
significant differences in the angles of the coronal plane (such as tilted sectioning); such differences would undermine the ability to draw biological conclusions from the MSI experiment.

Mapping the MSI data to the ABA was performed by aligning the experimental histology image to the reference histology image of the ABA, then applying the same transformation model to the MSI data. Therefore, a prerequisite for this MSI-ABA registration pipeline is that the MSI data must already be registered to its histological image. This limited the approach, at that time, to MALDI-MSI data acquired using Bruker Daltonics machines as this commercial vendor had trademarked this approach. The FlexImaging software of Bruker Daltonics provides semi-automatic registration software based on fiducial markers. No generic solution, equally applicable to MSI datasets recorded using any MSI mass spectrometer, was available. Nevertheless, we have tackled this limitation in Chapter 3 to automatically register MSI data to its histological image, and demonstrated its success for registering MSI datasets from different mass spectrometers to their histology images, and to the ABA (Chapter 7).

Whereas the proposed pipeline has been based on using the Allen Brain Atlas as a reference atlas, however, it is also valid to be applied to other reference mouse brain atlases (such as Paxinos atlas). The ABA was chosen because it represents a unique and publically available resource of multi-modal imaging of the mouse brain that includes magnetic resonance imaging data, high-resolution histological images, high resolution neuronal connectivity, and gene expressions of more than 20,000 genes. The proposed pipeline would enable future investigations to analyze neuroproteomics, provided by MSI, in context of these ABA multi-modal and multi-scale resources for better understanding the pathophysiology of neurologic disorders that may eventually lead to developing new therapy.

This pipeline enabled large-scale preclinical MSI investigations of neurologic disorders in mouse brain. In Chapter 6 we have demonstrated the utility of the MSI-ABA pipeline to uncover biomolecular changes associated with cortical spreading depression in a mouse model of migraine. Similarly, the pipeline can be applied to uncover molecular biomarkers in mouse models of other neurologic disorders such as Alzheimer, Stroke, and Parkinson's disease, etc.

8.1.2. Co-registration between MSI data and Histology
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. We presented, in Chapter 3, an automatic co-registration method to align an MSI dataset to the histological image of the tissue section. We have also demonstrated the generic applicability of the presented approach as it can be applied on data from different organs, different mass spectrometers, and different ionization methods. The registration quality was quantified and the overall registration accuracy was sufficient that any errors were less than the size of a single MSI pixel.

The t-SNE representation played a vital role in MSI-histology registration pipeline by summarizing the spatiomolecular organization of the tissue, which has clear correspondences with the tissue section's histology. The computational complexity of the t-SNE becomes a crucial factor in this process. The original implementation of the t-SNE
has a quadratic computational and memory complexity and thus a limited capacity to handle only up to 10000 data points. Further improvements have been recently made to accelerate the t-SNE based on the Barnes-Hut implementation (BH-SNE) that scales as logarithmically for computation and linearly for memory and thus enables tSNE of MSI data sets to be run much more practically. Current advances in MSI technology that allow 3D MSI data to be acquired [209] would appear to be computationally challenging even for the Barnes-Hut implementation. However 3D MSI data consists of individual MSI datasets of a stack of sequential tissue sections – so the Barnes-Hut implementation could be used to register each tissue section’s MSI dataset to its histological image, following which could then be aligned on the basis of the histological images or MSI datasets of the sequential images.

**8.1.3. HISTOLOGY GUIDED MASS SPECTROMETRY**

Higher spatial/mass resolution leads to increased data size and longer data acquisition times. For clinical applications, which analyze large series of patient tissues, this poses a challenge to keep data load and acquisition time manageable. Therefore, in Chapter 4, we have developed an automated image registration pipeline to register an annotated, histological image of an adjacent tissue section to a lower resolution image of the MSI-prepared (matrix coated) tissue section. Subsequently, the annotation borders are propagated to the low-resolution image, enabling the exclusive analysis of the annotated regions by MSI. This histology-guided MSI (HG-MSI) approach ensures that the MSI data is acquired solely from tissue regions with similar histological makeup. We demonstrated an 80% reduction of data load and acquisition time, thereby enabling high resolution (mass or spatial) to be more readily applied to clinical research.

Delineating ROIs for the HG-MSI approach depends mainly on the level of information expressed in the histological image. While a certain area can appear histologically homogeneous, it can be molecularly heterogeneous. The need for histological specification depends on the application, e.g. the analysis of tumor interfaces, or examining the similarity/differences of well vs undifferentiated tumor cells. It is thus crucial to specify, at the outset, which histological comparisons will/may be made. Accordingly the tool is better suited to those analyses in which the slow scan speed (high mass resolution) or large number of pixels (high spatial resolution) makes it impractical to apply to a patient series of complete tissue sections.

We have demonstrated this HG-MSI approach for the analysis of proteolytic peptides from specific histopathological regions in tumor tissues using an ultra high mass resolution Fourier transform ion cyclotron resonance mass spectrometer from Bruker Daltonics. However, the tool is vendor neutral and is applicable to all MSI methodologies (e.g., DESI, SIMS, etc.); the only requirement is that an optical image of the tissue section is used to define the measurement regions. Similarly the HG-MSI software may be used to better focus high spatial resolution as well as high mass resolution analysis, for all analyte classes and, all tissue types.

**8.1.4. INTRA-TUMOR HETEROGENEITY**

Mass spectrometry imaging may hold great potential in cancer research as it has recently demonstrated its ability to uncover molecularly distinct tumor subpopulations in histo-
logically uniform regions of tumor tissue. In Chapter 5, we presented an automatic data-driven approach, based on spatially mapped t-SNE, to identify molecularly-distinct, clinically-relevant tumor subpopulations in the MALDI-MSI datasets of patient tumor samples. We demonstrated its ability to uncover subpopulations statistically associated with patient survival in primary tumors of gastric cancer and with lymph node metastasis in primary tumors of breast cancer. The identification of these phenotypic tumor subpopulations is of high importance for understanding tumor evolution and may enable the further investigation of the key molecular changes that accompany tumor progression.

The robustness of the findings was assessed using an unbiased cross validation experiment. This allowed investigating the stability of the potential prognostic m/z features that characterize the detrimental subpopulations. We were also able to build a pixel classifier and a patient-based classifier for outcome, which were able to demarcate the intratumor heterogeneity and predict patient survival (primary gastric cancer) or lymph node metastasis status (primary breast cancer).

In both the gastric and breast cancer cohorts, the uncovered molecular and phenotypic intratumor heterogeneity was not apparent in the conventional histological images. This opens up much needed new possibilities for assessing the clinical impact of intratumor heterogeneity and clonal evolution in cancer, for instance by using the pixel classification for spatially resolved sample selection for RNA sequencing of tumor subpopulations with different clinical phenotypes.

8.2. PRE-CLINICAL APPLICATIONS

The automatic MSI-ABA registration pipeline presented in Chapter 2 holds great interest in facilitating large-scale MSI investigations in mouse brain models of neurologic disorders. We have demonstrated this in Chapter 6 to uncover the molecular changes associated with cortical spreading depression in a mouse model of migraine. Data cohorts spanning 3 molecular classes, from 32 mice with two different genotypes, were recorded using MALDI-MSI. All MSI datasets were aligned to the ABA using the developments presented in Chapter 2. We could then use the anatomical annotations of the ABA reference atlas to compare the molecular signatures, detected by MSI, in specific anatomical structures of interest (Isocortex, Thalamus, Hippocampus, and Striatum). We found a number of biomolecular features (m/z) that showed significant (p<0.05) changes in these anatomical regions and which were all found in the transgenic mouse that exhibits a stronger CSD phenotype (so molecular changes expected to be greater).

Another pre-clinical application is presented in Chapter 7 in which SIMS-MSI was used to investigate the lipid accumulation in the brain of a transgenic mouse that lacked multifunctional protein 2 (Mfp2). Mfp2 is a protein that is responsible for formation of bile acids and the degradation of pristanic acid and the very-long chain fatty acids (VLCFA). Mice that lack the Mfp2 suffer from severe neuromotor dysfunctions and die before the age of 6 months. It is, therefore, of interest to investigate and anatomically localize which biomolecules are significantly associated with that pathology. We used the developments presented in Chapters 2 and 3 to align the SIMS-MSI datasets to the ABA. The t-SNE registration based pipeline, presented in Chapters 3, was used first to align the SIMS-MSI data to its associated histological image, thus fulfilling the pre-
requisite to run the MSI-ABA registration pipeline. This registration pipeline enabled anatomical localization of the accumulated lipid in Mfp2-deficient transgenic mice to the 4th ventricle. This precise localization of the MSI signal offered deeper insights about the pathological mechanism of the Mfp2 deficiency.

8.3. **FUTURE WORK**

In this thesis we have developed an automatic registration pipeline to align MSI data to the ABA. The aim was twofold: 1) mapping MSI data from different subjects to the same coordinate space would allow their biomolecular profiles to be more readily compared, 2) the anatomical annotations provided by the ABA would enable mass spectrometrists, the principal users of MSI, to interrogate the MSI data in context of their anatomical localization. Nonetheless, this work can be further extended by considering the gene expression maps provided by the ABA. The ABA provides expressions of more than 20000 genes across the mouse brain. Analyzing this gene expression patterns is not trivial as it needs special computational methods to handle such big data. Eventually, integrating MSI data with these gene expressions become now feasible, and would provide more in-depth insights for better understanding the pathophysiology of neurologic disorders, and generally would open the doors for promising perspectives for the future of biology research.

There is current widespread interest in 3D MSI, so that the full biomolecular make-up of the tissues may be investigated [88–90]. Current advances in MSI technology, specifically high speed MSI instruments, allow 3D MSI to be performed in a much shorter time frame. Nevertheless, data reconstruction and processing are still challenging [90, 209]. 3D MSI data reconstruction algorithms should take into consideration the high potential for non-linear deformations between successive tissue sections, and importantly also to define the geometrical constraints that are required to preserve the original shape characteristics and thus avoiding the common related problem of “banana-into-cylinder” effect [44, 210]. The automated and accurate reconstruction of 3D MSI datasets involves two distinct image registration steps: 1) aligning each tissue section’s individual MSI datacube to its associated histological image, for example using the t-SNE based automated registration presented in Chapter 3; and 2) using the histological images to establish spatial correspondences between successive tissue sections, using similar histology based metrics presented in Chapter 2 for alignment with the Allen Brain Atlas, but this time with an additional step of integrating geometrical constraints and this need to be further investigated.

The automatic alignment between MSI and histology can be considered as a basis for MSI image fusion [211], for improving image quality and increasing the interpretability of the MSI data. Recently, Van der Plas et al [39] presented a data fusion pipeline between MSI and histology but this first requires alignment between MSI and histology. The t-SNE based registration pipeline, presented in Chapter 3, would allow such fusion pipeline to be more readily performed, not only on a certain MSI modality but also spanning molecular data from different imaging mass spectrometers.