Exploring MS Imaging Data in a semi-supervised and interactive manner

Jocelyne Bruand[1], Theodore Alexandrov[2], Srinivas Sistla[3], Maxence Wisztorski[4], Céline Meriaux[4], Michael Becker[5], Michel Salzet[4], Isabelle Fournier[4], Eduardo Macagno[3], and Vineet Bafna[3]

[1] University of California, San Diego,
jocelyne@ucsd.edu
[2] University of Bremen
[3] University of California, San Diego
[4] FABMS, Université Lille 1, Villeneuve d'Ascq
[5] Bruker Daltonik GmbH

Mass Spectrometric Imaging (MSI), is a molecular imaging technique which allows the generation of topographic 2D maps for a large complement of active molecules in the cell. Most bioinformatics approaches have focused on using MALDI MSI as a tool for the discovery of signature markers of physiology stages. One approach is to computationally segment the tissue section into regions, each characterized by a specific mass spectral signatures. While several methods, such as hierarchical clustering or PCA, have been employed to generate these regions, they are black-box tools with little or no user interaction and thus cannot be used in an exploratory way. Moreover, these methods require prior data reduction. We describe here an interactive algorithm capable of generating image segments.

In previous work, we developed a tool using a supervised approach which returns the spectral signature for a region of interest (ROI) defined on morphological criteria. Here, we expand this tool to provide additional capabilities. By integrating over the m/z values, we determine the likelihood ratio for each MALDI spot belonging to the same cluster as the original ROI, providing a measure of similarity of the spectral signature of each spot to the region. When given several ROIs, the corresponding log-odds images are combined into segmentation maps. We also provide a measure of quality for the segments, allowing the algorithm to run iteratively until quality is satisfactory. The algorithm can also be run on random seed ROIs providing a completely unsupervised mode.

We use our algorithm to create segmentation maps for two MSI datasets: a whole leech embryo dataset and a rat brain dataset. We selected a several partial seed regions for each of the datasets. Our algorithm successfully recovers the molecular signature of each region, as well as finds all other regions with similar signatures. For example, when looking at the leech central nervous system, selecting an anterior ganglion as ROI returns the entire central nervous system, while selecting a posterior ganglion returns a stronger signal in a few of the posterior ganglia. When looking at the molecular signatures, we find that for the m/z values expressed in the CNS, some m/z value show high intensity in only the posterior ganglia while others show a more even intensity distribution throughout the CNS. This is particularly interesting as the leech CNS develops

from the head to the tail, and thus anterior ganglia are a few days ahead in development than the tail ganglia. We are also able to select very specific signal, such as the pores, which only appears in a few MALDI spot. When selecting several region of the cerebral cortex of the rat brain (e.g. retrospenial cortex, somatosensory cortex, auditory cortex, etc) we find that these return similar spectral signatures, as well assign high log-odds scores to most of the cerebral cortex. Initializing random seed ROIs also provided good segmentation maps. While our tool can be used in a fully automatic manner, the user has the ability to input information every step of the way to guide tool, making it fully interactive and exploratory. This involves selecting original regions (or using random ones), selecting specificity to the regions, selecting which regions to keep when calculating segments, among other features.