Coarse-graining of perplexity for the spatial distribution of molecules

博士（医学）Xu Lili

論文題目

Coarse-graining of perplexity for the spatial distribution of molecules

（分子の空間分布に対するパープレキシティの粗視化）

論文の内容の要旨

[Introduction]

Among different molecules in biological tissue, an alternative study approach is to obtain a characteristic quantity such as the Shannon entropy and study its position dependence. An extension of the Shannon entropy was explored for biological diversity. To study biological tissue from an informational point of view, heat maps of the Shannon entropy calculated from the abundance of molecules in each spot have been investigated.

Mass spectrometry imaging (MSI) is a technique with ionizing molecules in a sample to provide both molecule species and positions of those molecules in the sample. At each spot of the sample, hundreds of *m/z* peaks are obtained. In mass spectrometry, the Shannon entropy was calculated for *m/z* spectra. The Shannon entropy has been viewed as a physical quantity which gives information as to how molecules spatially vary. The spatial distribution of the Shannon entropy was studied for the matrix-assisted laser desorption-ionization (MALDI) MSI. Recently, the binning effect of the Shannon entropy, the relation between the Shannon entropy of mass spectra and molecules such as peptides and proteins, and a data targeted extraction method for metabolite annotation were investigated.

In this paper, the characterization of the position dependence of the Shannon entropy is explored. We derive perplexity from entropy and develop an approach to visualize the spatial and mass-spectral diversities by coarse-graining. To this end, a slope k is introduced.

[Materials and Methods]

We used C57BL/6J female mice, with two four-month-old mice kidneys for MALDI MSI observation (using a high-resolution microscopic imaging mass spectrometer, iMScope). Sample preparation steps including mouse sacrifice, sample pre-reservation, and matrix spraying followed our previous paper. The experimental conditions were as follows: negative ion mode, *m/z* range between 550 and 950, laser strength of 45%, and number of irradiations of 100. All experiments in this study were performed in compliance with the licensing instructions from the Institutional Animal Care and Use Committees of Hamamatsu University School of Medicine, Japan (permission code 2015028).

Let us consider coarse-graining for MSI. We selected a square region on the image and divided the square into $N(\varepsilon)$ sub squares with side length ε. Treating these $N(\varepsilon)$ sub squares as new pixels, relative peak intensities in the pixel at (x, y) can be calculated as new perplexity, $PP(\varepsilon)$ (x, y). We considered how $PP(\varepsilon)$ (x, y) behaves as ε varies by coarse-graining.

[Results]

We calculated the Shannon entropy and perplexity on each spot of the kidney MSI data and plotted heat maps. We found there are differences between entropy and perplexity heat maps. However, these differences showed sample dependencies and the difference among regions cannot be distinguished clearly by entropy or perplexity.

To investigate the spatial distribution of perplexity, we picked several points on the image for the kidney. We found a linear trend in the plot of ln ε and perplexity for small ln ε. The unit of ε is the length of the original pixels. Different spots were chosen and the linear dependence was found at each spot with different slopes k. Spots in the same region or proximity showed similar slopes. Thus, slope k can be experimentally determined. We found that the slopes were largest in the renal pelvis area, while slopes were relatively small in the cortex area and smallest in the medulla area. We also found pixels of large k appearing at the vein and artery, where mass spectra inside and outside this area were different.

[Discussion]

We found that the relation between perplexity and ε for the experimental data. Although this relation stems from the statistical nature of the molecules, it is still an open problem of how k reflects the spatial distribution of molecules in the sample. The physical and biological reasons of the power-law behavior need to be clarified in the future. We note that k is positive even when the Shannon entropy for pixels about (x, y) is unchanged if spectral patterns in the neighborhood have a variety. That is, k has different values even if the heat map is homogeneous. In this sense, k is more informative than entropy and perplexity.

[Conclusion]

We have proposed the use of perplexity and introduced k in coarse-graining. We found that the heat map of k visualizes the chemical diversity of surrounding pixels and reveals new structures which are not clearly visible in the Shannon entropy heat map. Although the use of peaks in the mass spectrum as a distribution is not yet established, the spatial distribution of k will help characterize biological tissues.