

Overview

- Mass spectrometry imaging collects chemical profiles across the surface of a tissue
- Few statistical methods exist for analysis of such data, and few software provide statistical tools • We provide a unified software framework for clustering and classification based on nearest
- shrunken centroids (Tibshirani et al. 2002) and spatially-aware clustering (Alexandrov & Kobarg 2011) which uses statistical regularization to automatically identify important masses



Figure 1: (A) DESI mass spectra being collected from a painting of a cardinal. (B) Image segmentation that reconstructs the painting via clustering based on the mass spectra. (C) Mass spectrum of a pixel located in the body of the bird, being viewed in the CARDINAL software. (D) Ion image being viewed in the CARDINAL software of the mass feature at m/z649.08, showing the spatial distribution of that mass. GUI allows rich interaction with dataset.

Results

Mouse Brain

- MALDI spectra with intact proteins acquired from a mouse brain
- The mass spectra were TIC normalized, smoothed, baseline corrected, and peak picked
- 2 clusters chosen with an average of 5 (of 31) mass features selected per cluster



Figure 2: (A) Optical image of the brain tissue showing the morphology. (B) The segmentation for initial parameters k = 5, s = 6 yields 2 clusters (red and light blue) showing the structure of the brain. Opacity is used to indicate the probability of a pixel belonging to a cluster. Note that this segmentation has a high degree of uncertainty (seen by the low opacity). (C) The mass feature most associated with the light blue cluster by t-statistic at m/z 9984. (D) The mean mass spectrum of the light blue cluster, showing the 5 (of 31) mass features selected that are important in distinguishing the light blue cluster. (E) The t-statistics associated with the selected masses, showing m/z 9984 as the highest (most co-localized) with the light blue cluster. (F) The resulting number of clusters for different initializations of the sparsity parameter s and the initial number of clusters k. Notice that it stabilizes at 2 or 3 clusters around s = 6.

Discovering Spatio-Chemical Structure in Tissue: CARDINAL Software and Methods for Analysis of Mass Spectrometry Images

¹Department of Statistics, Purdue University, West Lafayette, IN, USA; ²Department of Chemistry, Purdue University, West Lafayette, IN, USA; ³Canary Center at Stanford for Cancer Early Detection, Stanford School of Medicine, Palo Alto, CA, USA; ⁴Department of Computer Science, Purdue University, West Lafayette, IN, USA

Pig Fetus

- DESI spectra acquired from a pig fetus cross-section
- The mass spectra were TIC normalized and peak picked



Figure 3: (A) Optical image of the pig fetus section. (B) The segmentation for initial parameters k = 20, s = 8 yields 11 clusters showing the anatomy of the pig fetus. Opacity is used to indicate the probability of a pixel belonging to a cluster. Note that this segmentation has a low degree of uncertainty (seen by the high opacity). (C) The mass feature most associated with the "heart" cluster (dark blue) by t-statistic at m/z 187.33. (D) The mean mass spectrum of the "heart" cluster, showing the 29 (of 255) mass features selected that are important in distinguishing the "heart" cluster. (E) The t-statistics associated with the selected masses, showing m/z 187.33 as the highest (most co-localized) with the "heart" cluster. (F) The resulting number of clusters for different initializations of the sparsity parameter s and the initial number of clusters k. Notice that it stabilizes after s = 4 and compeletely equalizes after s = 8.

Farmhouse Painting

- DESI spectra acquired from the pigments in an artistic farmhouse painting set a slide
- The mass spectra were TIC normalized and peak picked
- 14 clusters chosen with an average of 60.4 (of 181) mass features selected per cluster









Figure 4: (A) Optical image of the farmhouse painting. (B) The segmentation for initial parameters k = 20, s = 8 yields 14 clusters that reconstructs the painting. Opacity is used to indicate the probability of a pixel belonging to a cluster. Note that this segmentation has a low degree of uncertainty (seen by the high opacity). (C) The mass feature most associated with the "house" cluster (red) by t-statistic at m/z 277.16. (D) The mean mass spectrum of the "house" cluster, showing the 35 (of 181) mass features selected that are important in distinguishing the "house" cluster. (E) The t-statistics associated with the selected masses, showing m/z 277.16 as the highest (most co-localized) with the "house" cluster. (F) The resulting number of clusters for different initializations of the sparsity parameter s and the initial number of clusters k. Notice that it stabilizes after around s = 4.

Kyle D. Bemis¹, Livia S. Eberlin², Christina Ferreira², Stephanie van de Ven³, Parag Mallick³, R. Graham Cooks², Mark Stolowitz³, and Olga Vitek^{1,4}



Classification

- DESI spectra from 8 matched pairs of kidney tumor and normal tissue
- The mass spectra were TIC normalized and binned to unit resolution
- Performed cross-validation with spatial sparse classification based on nearest shrunken centroids using smoothing radius r = 0, 1, 2, 3 and sparsity parameter s = 0, 2, 4, 6, 8, 10, 12
- 8-fold cross-validation used to select the best parameters for prediction
- Cross-validated misclassification rate of 0.078% for smoothing radius r = 2 and sparsity parameter s = 10, using only 151 out of the 849 binned mass features.
- The t-statistics can be used to identify masses strongly indicative of cancer



Figure 5: (A) The H&E stained tissue sections for sample UH9912-01 showing tumor and normal diagnoses. (B) An ion image for sample UH9912-01 at the top-ranked m/z 886.5 showing increased intensity in the tumor tissue. (C) The diagnostic predictions for sample UH9912-01 based on the probabilities from the spatially-aware nearest shrunken centroid classifier. (D) The H&E stained tissue sections for sample UH9911-05 showing tumor and normal diagnoses. (B) An ion image for sample UH9911-05 at the top-ranked m/z 886.5 showing increased intensity in the tumor tissue. (C) The diagnostic predictions for sample UH9911-05 based on the probabilities from the spatially-aware nearest shrunken centroid classifier. Note the classification is very good using only 151 of the 849 mass features.

Conclusions

- There is a strong need for better open-souce software for analysis of MS imaging data • Few existing tools provide methods for statistical analysis or utilize spatial information
- It is useful for interpretation to have a common framework for clustering and classification
- Statistical regularization facilitates idenfication and interpretation important masses
- Sparsity can help guide parameter choices for the number of clusters t-statistics for unimportant masses are shrunken toward 0
- The t-statistics can be used to rank the top masses for either segmentation or classification; the
- Use of transparency for probability in visualization of image segmentations allows quick and straightforward evaluation of segmentation quality and uncertainty

References

spatially aware clustering. Bioinformatics, 27(13), i230 - i238.

gene expression. Proc. Natl. Acad. Sci. U.S.A. 99 6567 - 6572.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 103049.



- Alexandrov, T., & Kobarg, J. H. (2011). Efficient spatial segmentation of large imaging mass spectrometry datasets with
- Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of