**Mass Spectrometry Imaging in Detecting Tumor Heterogeneity**

Ibrahim Elsayed#1, Donia Abd Elsalam#2, Renad Taher#3, Mariem Ahmed#4, and Mustafa Yehia#5

**Abstract:** Tumor subpopulations have molecular phenotypes that drive tumor progression and determine disease outcome which is essential for a more personalized therapy. Mass spectrometry imaging has proven its ability to identify diagnostic and prognostic biomarkers. In this research, we seek to determine tumor subpopulations that affect patient outcomes and the statistically associated subpopulations with poor survival and tumor metastasis. Here we introduce spatially mapped t-distributed stochastic neighbor embedding (t-SNE), a nonlinear visualization of the data that can better resolve the biomolecular intratumor heterogeneity. The outcomes will allow us to uncover subpopulations statistically associated with patient survival in primary tumors of gastric cancer and with metastasis in primary tumors of breast cancer.

**Introduction**

Cancer is considered as one of the most important topics from the 20th century till now and needs a lot of attention, it is mainly an abnormal growth of cells. According to World Health Organization (WHO) 2020 cancer statistics (Fig. 1), it reflects our interest in both breast and gastric cancers. So, we are going to proceed with exploring causes of cancers, ways to discover/treat it and finally moving forward to a new approach called MSI to discover a lot of wonderful things about the field of cancer.

MSI can uncover molecular intratumor heterogeneity. The challenge has been to identify those tumor subpopulations that drive patient outcomes within the highly complex datasets (hyperdimensional data, intratumor heterogeneity, and patient variation). Here we report an automatic, unbiased pipeline to nonlinearly map the hyperdimensional data into a 3D space, and identify molecularly distinct, clinically relevant tumor subpopulations. We demonstrate this pipeline’s ability to uncover subpopulations statistically associated with patient survival in primary tumors of gastric cancer and with metastasis in primary tumors of breast cancer.

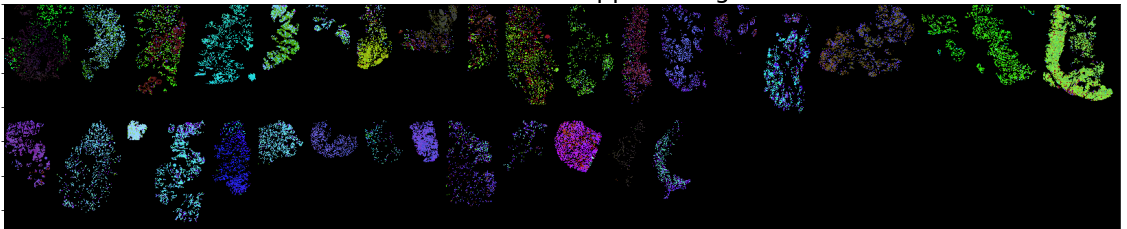
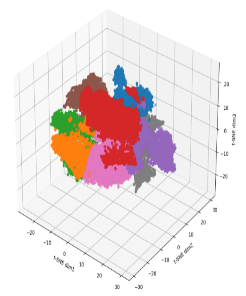
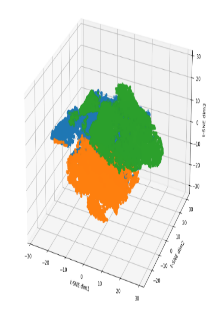
**Materials and Methods**

Tumor-specific signatures obtained by protein matrix-assisted laser desorption MSI analysis of primary tumors of gastric cancer (n = 63) and breast cancer (n = 32) were nonlinearly mapped to a 3D space using t-SNE. Using the perceptually linear LAB color map to color each pixel according to its position in the t-SNE space, a t-SNE colored image can be obtained that depicts regions characterized by similar mass spectral profiles with similar colors. To segment the image into a discrete number of clusters, bisecting k-means and edge correlation algorithms were applied. The resulting clusters, or tumor subpopulations, were then statistically compared with the patients’ clinical data (survival for gastric cancer and lymph node metastasis for breast cancer) to identify the subpopulations statistically associated with patient phenotype. LOPO pixel-based and patient-based classifiers were built to cross-validate the identification of tumor subpopulations and patient outcomes.

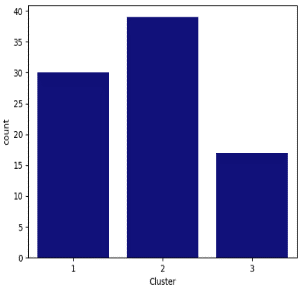
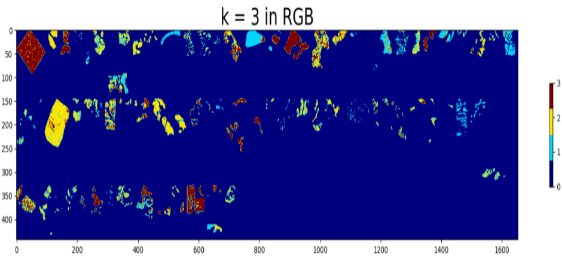
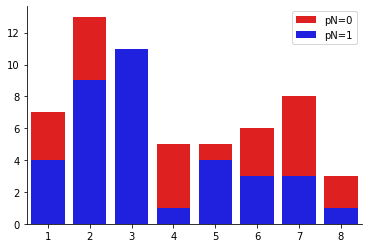
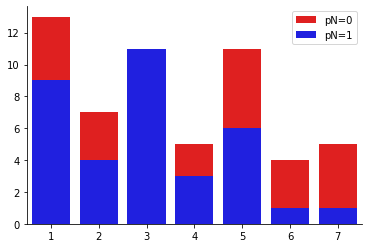
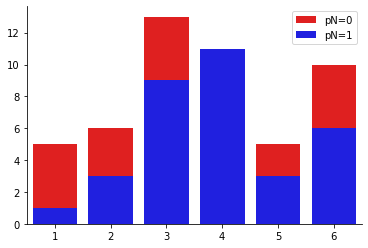
**Results**

**MSI Data**

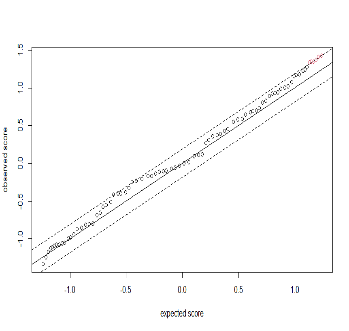
**Dimensionality Reduction (t-SNE)**

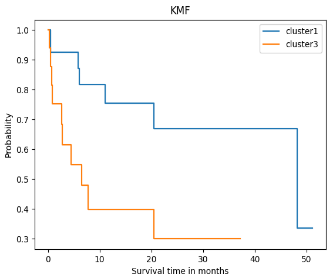
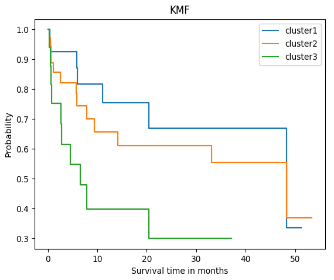


**Clustering (K-means Clustering)**



**Statistical Analysis Significance Analysis of Microarrays (SAM)**





|  |  |
| --- | --- |
| **K Values** | **P-value** |
| 6 | 0.004824 |
| 7 | 0.005738 |
| 8 | 0.003732 |

**Discussion**

**MSI Data**

HE Image is a remarkable tissue section. The recognition of the tumor subpopulations that affect the results of patients is important for better describing the changes in molecules. We used MSI because it has an ability to detect tumor subpopulations in histologically identical regions of tumor tissue.

**Dimensionality Reduction (t-SNE)**

It preserves the global and local similarity structure of the dataspace in the low dimensional representation. We can see samples representing two types of tumor heterogeneity; inter/intratumor heterogeneity of MSI Data. And contribution of the molecular heterogeneity in Data due to intratumor heterogeneity and patient variability.

**Clustering (K-means Clustering)**

Gastric Data; blue color: background, others: cluster. We have a total of 79 patients which is more than our number of patients and that shows us that some patients are assigned to more than one cluster. Breast Data; Cluster 3 is full of metastatic patients only.

**Statistical Analysis**

The greatest significant difference in survival between the subpopulations in clusters 1 and 3 with P-value of 0.02 less than 0.05.

**Significance Analysis of Microarrays (SAM)**

The significance level of SAM output could be found at the high end of SAM plot revealing the expected score along the line of interest.