

# Processing MassTech AP/MALDI Data using MSiReader v2.8

Russ R. Kibbe<sup>1</sup>, Alena N. Joignant<sup>1</sup> and David C. Muddiman<sup>1,2</sup>

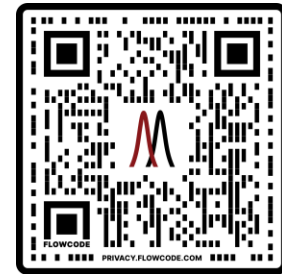
<sup>1</sup>Department of Chemistry, NC State University and <sup>2</sup>MSI Software Solutions, LLC

American Society for Mass Spectrometry  
MassTech Breakfast Presentation

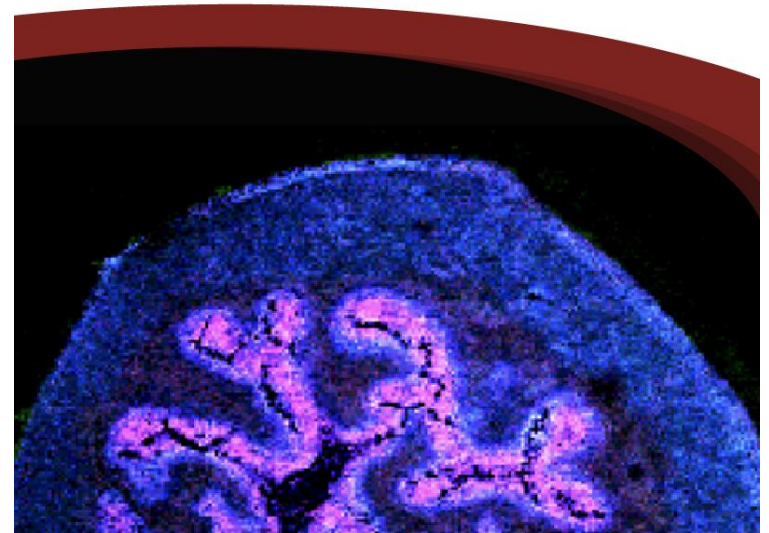
Anaheim, CA  
June 3, 2024

ASMS 2024, Anaheim, CA

[www.msireader.com](http://www.msireader.com)



**MSiReader v2.8 TO SUPPORT  
YOUR MSI RESEARCH NEEDS**



- Intuitive UI for Improved UX
- Computational Efficiency
- User Manual and Video Tutorials

- Batch-mode Data Processing
- BioPharma Mode for HTS/HCS
- Ion Classification Tool
- Mass Correction, QA/QC Tools
- Relative and Absolute Quantification
- Multivariate Statistical Analysis (PCA, t-SNE)
- Save Custom File Formats (for Fasting Loading)

# Introduction to Quality Control & System Suitability Testing

## Quality Control (QC)

*Operational processes and techniques used to evaluate the quality of results by monitoring and controlling existing variability*

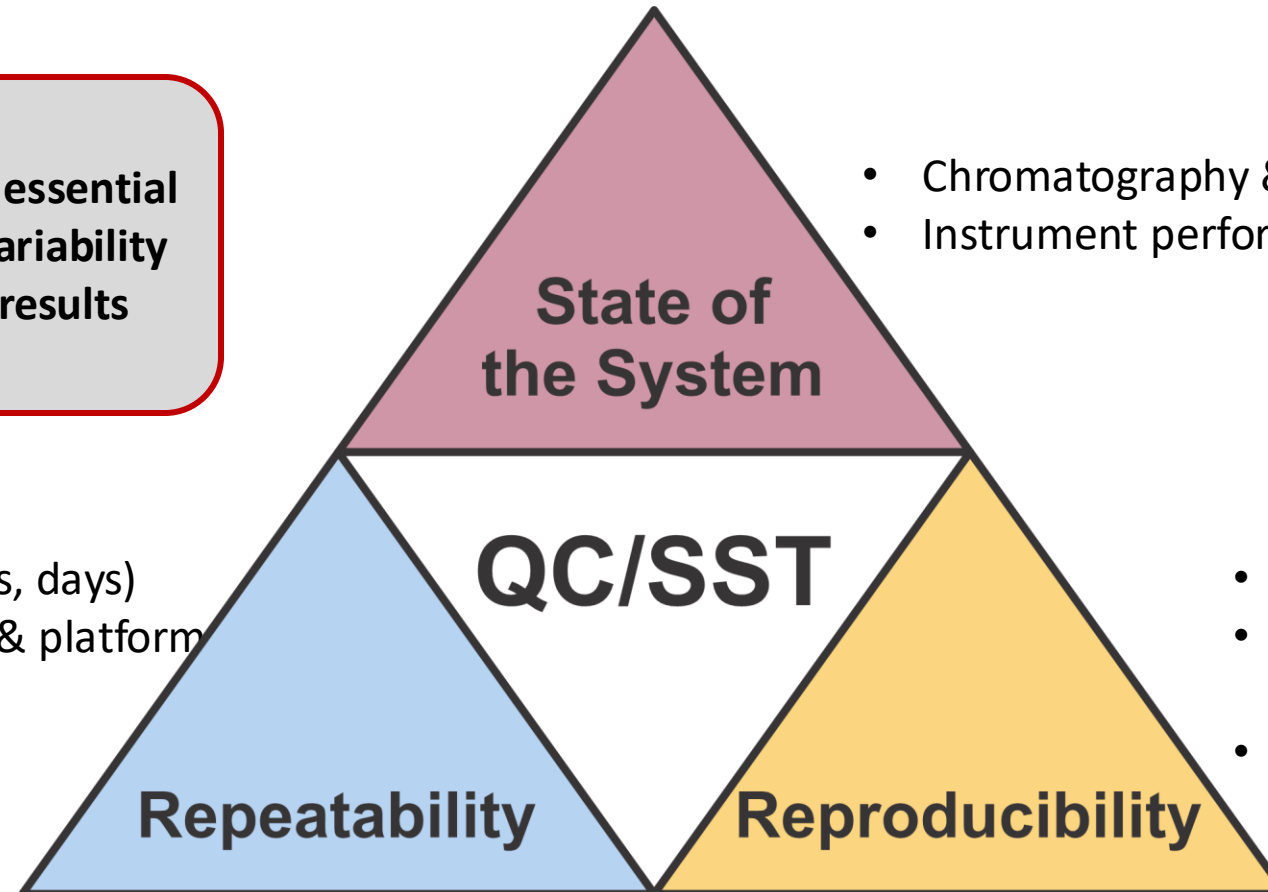
## System Suitability Testing (SST)

*Evaluate analytical performance metrics to qualify the instrument's working state before analyzing biological samples*

**QC and SST protocols are essential for monitoring existing variability to verify experimental results**

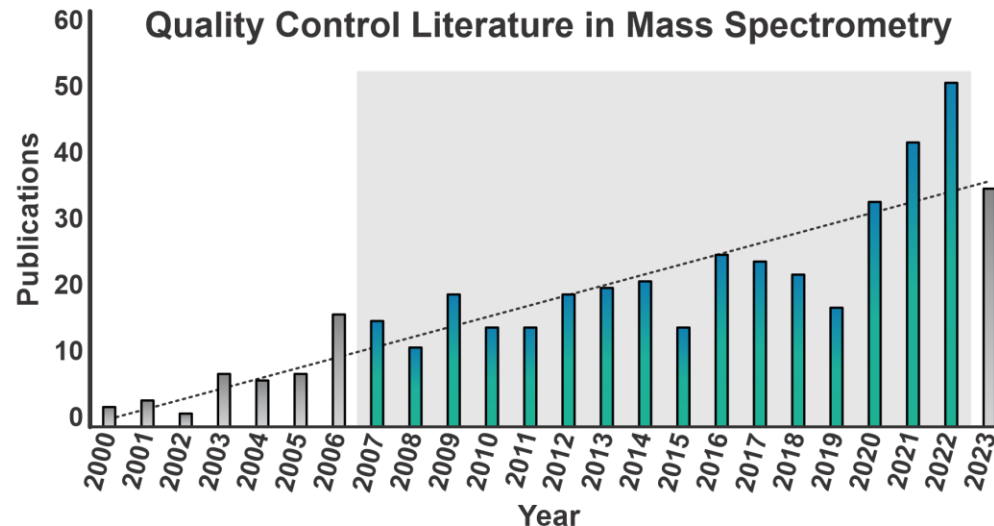
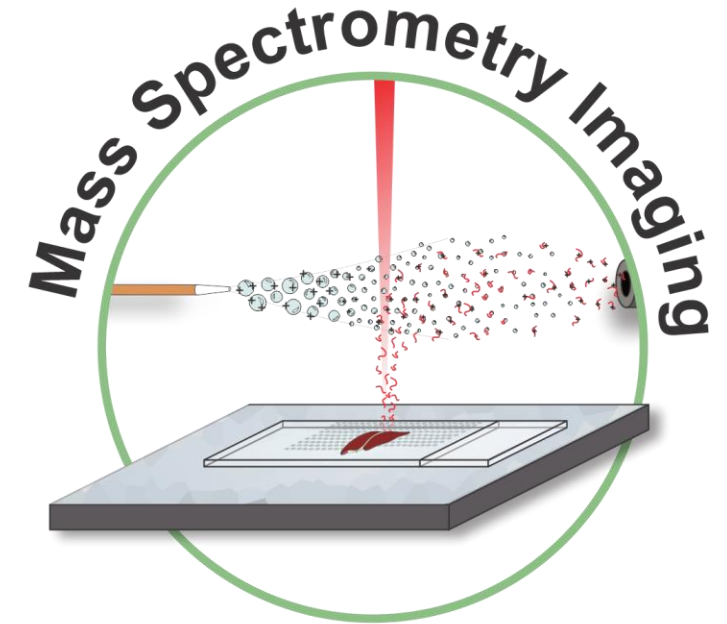
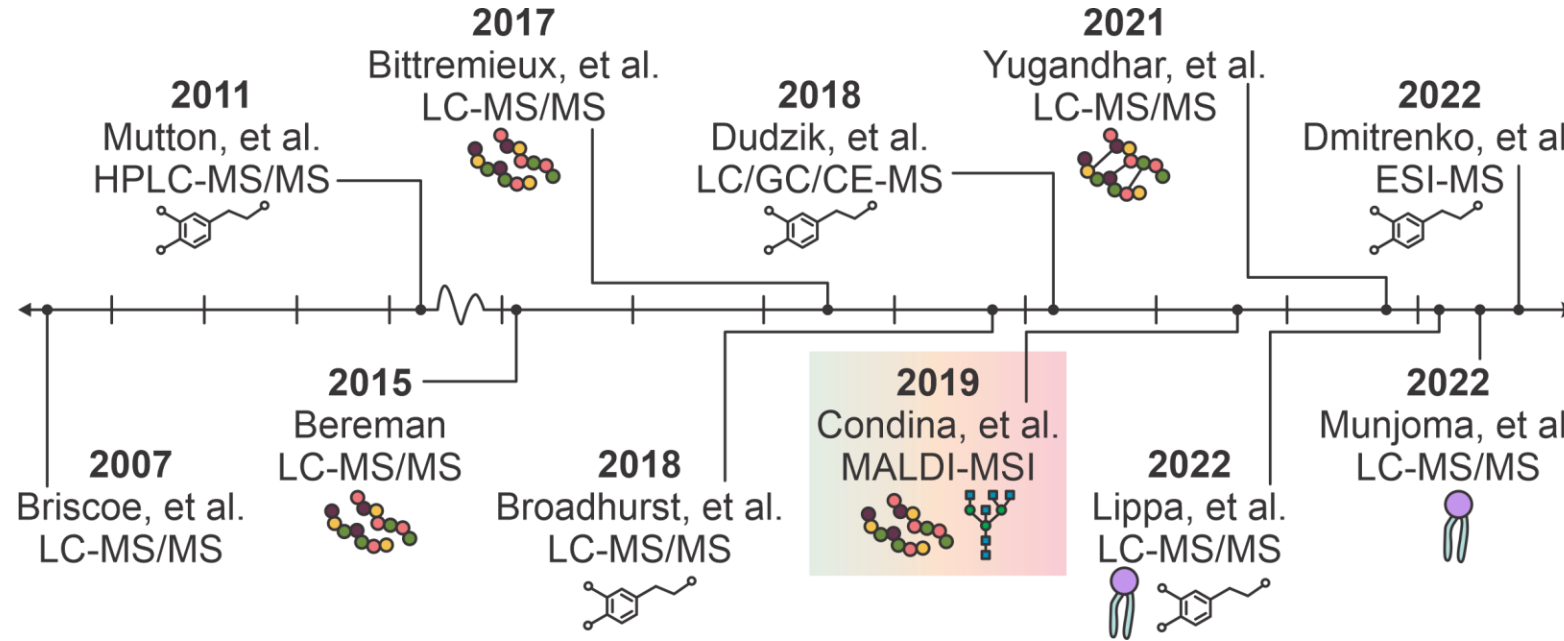
- Short-term (hours, days)
- Same conditions & platform
- Intra-laboratory

- Chromatography & ionization source
- Instrument performance



- Long-term (months, years)
- Different conditions & platform
- Inter-laboratory

# Precedent of QC & SST Across the Field of Mass Spectrometry



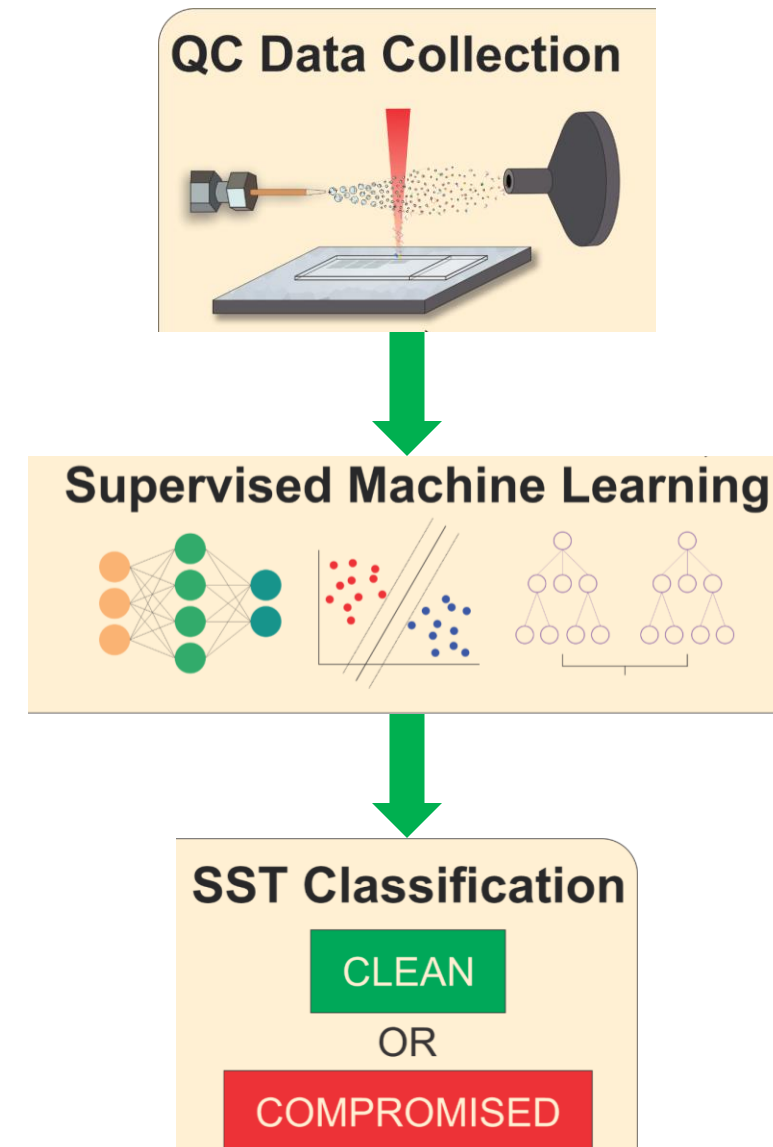
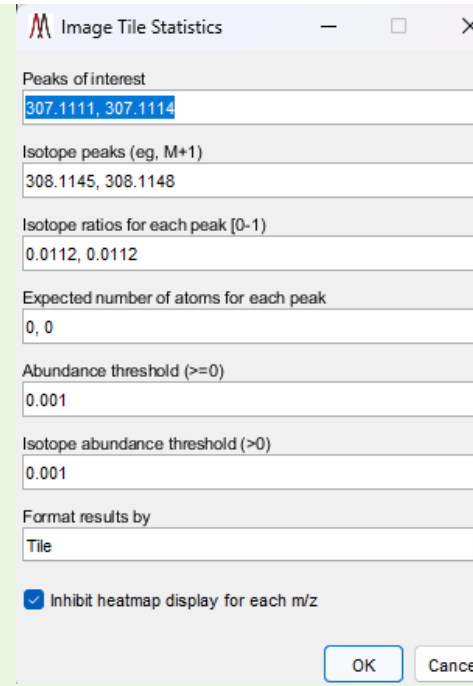
- ✗ No standard practice or protocols
- ✗ Various ionization sources (*e.g.*, MALDI, MALDI-2, DESI, IR-MALDESI, etc.)
- ✗ Different characteristics of the data:
  - Spatial information
  - Primarily MS<sup>1</sup>
- ✗ No commercially available mixtures

# Instrument Classification using Machine Learning

*J. Proteome Res.*, 2024, submitted; ASMS 2024 Posters WP372 and ThP414

1. Collect QC Data (Standard Mix)
2. Use MSiReader to extract QC relevant metrics
  - Auto MSI QC Tool

Monoisotopic and isotope Abundance  
Mass Measurement Accuracy  
Relative Standard Deviation  
Detection Frequency  
Spectral Accuracy
3. Format data and apply class labels
4. Split data into training and testing set
  - Entire set for training and a single ROI for testing for actual implementation
5. Scale data
6. Provide data to supervised ML models and obtain instrument condition classification



# Mass Spectrometry Imaging Mode: Statistical Analysis

**MSi Data Attributes**

Spots per line:  Spot spacing:   $\mu\text{m}$

Number of lines:  Line spacing:   $\mu\text{m}$

Load injection time Filter scans:

Abundance filter  Anchor Threshold:

m/z filter min:  max:

Polarity switch  parity:  Load all scans:

Relative Quantification, PCA, t-SNE\*.imzML (2 files)

**MS Navigation**

m/z:  ...

Tolerance  $\pm$ :  ppm

Abundance window:

Hotspot removal at  %

Scale max:   Lock

min:

max:

**Post Processing**

Injection time scaling

Normalization:

**Heatmap Appearance**

Auto

Interpolate:

SPC

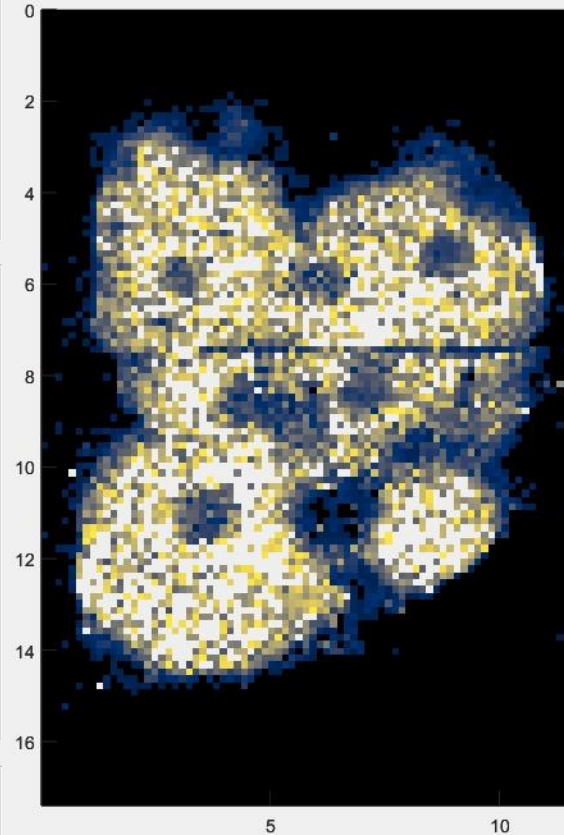
**Colormap**

Type:

Scaling:

Flip color scale

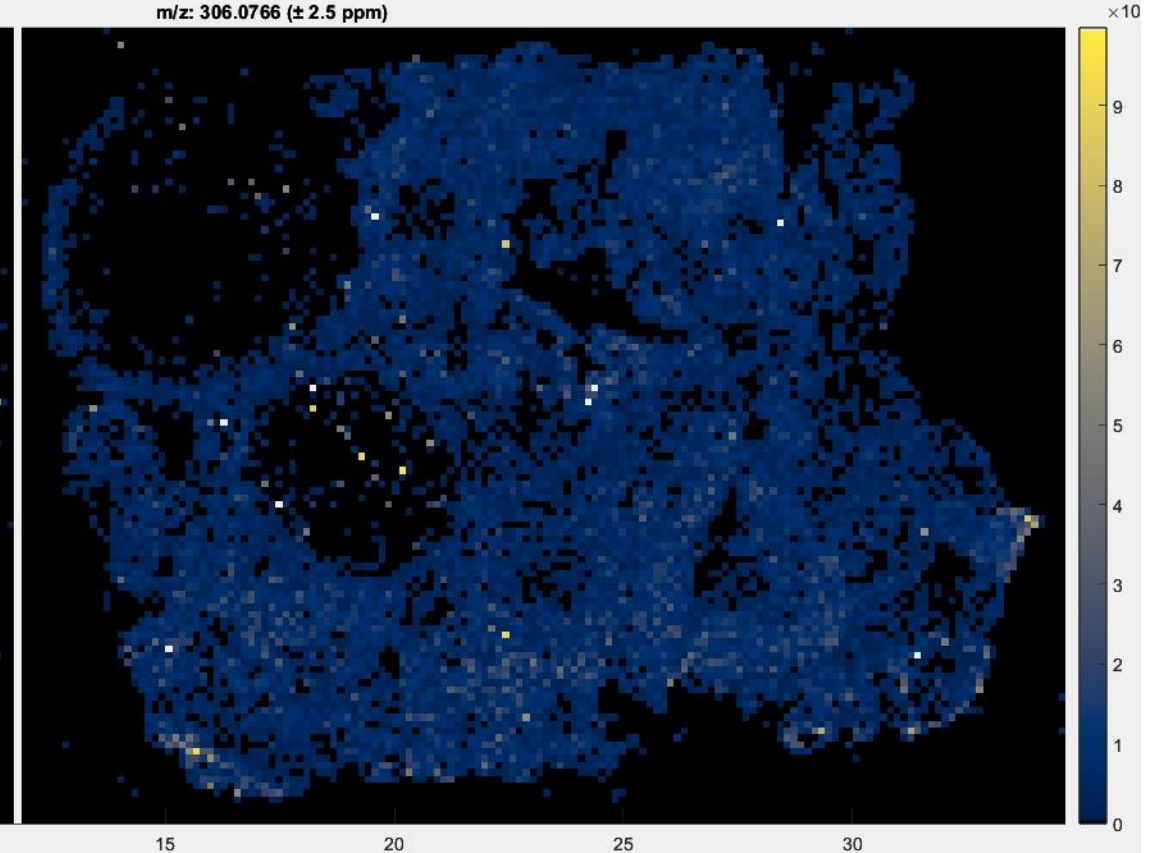
## Cancerous Ovarian Tissue



XY units: mm

Abundance units: ions/sec

## Healthy Ovarian Tissue



$\times 10$



# Data Export to Abundance Matrix

MSiReader v2.70 [64-bits standalone] 11:24:59 Monday, 2024.05.06

Home Pre-Processing Visualization QA/QC Annotations Quantification Statistical Analysis Help

Number of lines: 116 Line spacing: 150  $\mu\text{m}$

Filter scans: none (load all scans)

Load injection time

Abundance filter  Anchor Threshold: 0.001

m/z filter min: 0 max: inf

Polarity switch  $\pm$  parity: odd Load all scans

Clear Data Healthy and Cancerous\*.imzML (2 files) Description

MS Navigation

m/z: 306.0765

Tolerance  $\pm$ : 2.5 ppm

Abundance window: max

Hotspot removal at 99 %

Scale max: 887609.25  Lock

min: max:

Post Processing

Injection time scaling

Normalization: none

Heatmap Appearance

Auto Update heatmap

Interpolate: none

SPC SPC options...

Colormap

Type: cividisblack

Scaling: Linear

Flip color scale Save colormap

XY units: mm Abundance units: ions/sec

3215 Range 0 to 887609 Median 60400

225 Range 0 to 322361 Median 3413

m/z: 306.0765 ( $\pm 2.5$  ppm)

Interrogated Reference

Select scans to export

Two ROIs are active.  
Which scans would you like to process?

- Interrogated ROI Scans
- Reference ROI Scans
- Scans from both ROIs
- All Scans

OK Cancel

# Data Export to Abundance Matrix

Pixel/Voxel abundances at each  $m/z$  in a list

The screenshot shows an Excel spreadsheet with the following structure:

- Columns:** A, B, C, D, E, F (m/z), G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U.
- Row 1:** Headers for Scan, File, Local Scan, X, Y, Z, and Abundance for each m/z value.
- Row 2:** m/z values: 202.071455, 202.0867814, 202.1229963, 202.9866887, 203.0818581, 205.0647607, 207.0115595, 207.113286, 207.1496954, 209.0926756, 209.1289714, 211.1082284, 211.1446996, 213.0335145.
- Row 3:** Abundance values for each m/z.
- Rows 4-37:** Data rows for Scans 3026 through 3718, with columns for File, Local Scan, X, Y, Z, and Abundance for each m/z.

The spreadsheet is titled "Hen\_Ovary\_PeakExport - Excel" and includes a ribbon with tabs for File, Home, Insert, Page Layout, Formulas, Data, Review, and View. The status bar at the bottom shows "Ready" and "100%" zoom.



# Statistical Analysis in MSiReader

## Principal Components Analysis (PCA)

MSiReader v2.70 [64-bits standalone] 14:16:22 Monday, 2024.05.06

Home Pre-Processing Visualization QA/QC Annotations Quantification **Statistical Analysis** Help

PCA  
t-SNE  
Launch Principal Component Analysis tool for either a .txt or .xlsx file. See 57.6.1 and 57.6.2 in the user manual.

**MSi Data Attributes**

Spots per line: 231  
Spot spacing: 100  $\mu\text{m}$   
Number of lines: 116  
Line spacing: 150  $\mu\text{m}$

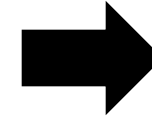
Load injection time  
Filter scans: none (load all scans)

Abundance filter  
 Anchor  
Threshold: 0.001

m/z filter  
min: 0 max: inf

Polarity switch  
polarity: +--+ parity: odd Load all scans

Load Data



MSi PCA

Enter the range of m/z values you want to process. Columns of the input matrix outside this range will be excluded. Multiple range specifiers separated by commas may be entered. Parentheses, brackets or quote marks such as ( ) [ ] { } < > \* may be used to delimit each range but are not required. For example, [250 360], [400 800], [1400 1620]. The full m/z range of the loaded matrix is:

[202.0715 339.3270]

A range can also be entered as a single m/z value (>0). In this case a comparison tolerance is used to select columns of data.

1.000e-05

Abundance data normalization:  
Z-score

Display individual heatmap figures

Display 3D biplot chart (instead of 2D)

OK Cancel

Select an MSiExport data file

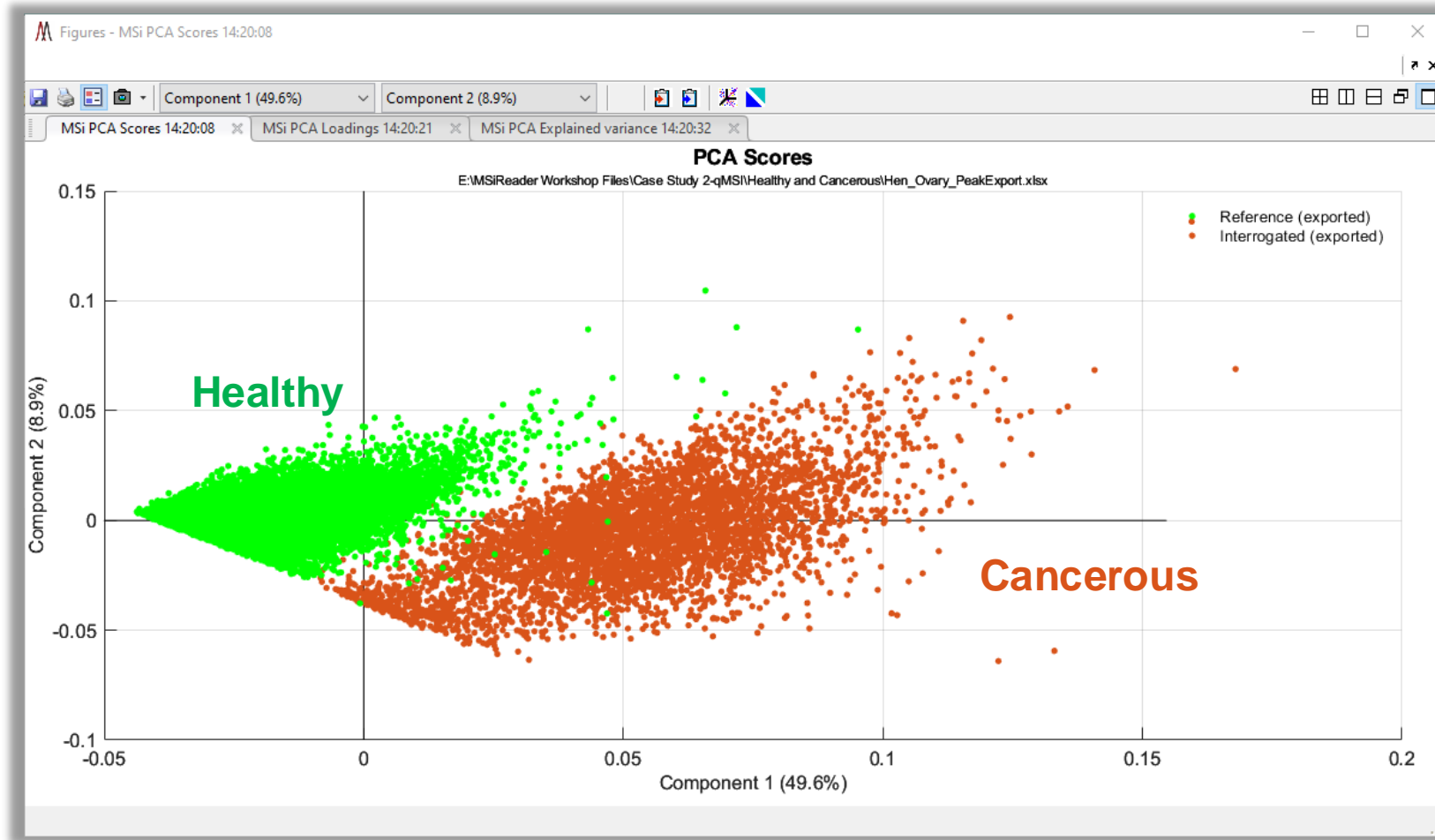
File name: Hen\_Ovary\_PeakExport

Name	Date modified	Type	Size
SSIM	5/6/2024 1:40 PM	File folder	
Cancer_Interrogated	4/24/2024 10:52 AM	Text Document	130 KB
Healthy_Reference	4/24/2024 10:53 AM	Text Document	365 KB
Hen_Ovary_PeakExport	5/6/2024 2:13 PM	Microsoft Excel W...	22,253 KB
Hen_Ovary_QMSI_Cancer_isocount	5/6/2024 1:56 PM	Microsoft Excel W...	1,042 KB
Hen_Ovary_QMSI_Cancer_isocountcarbon	5/6/2024 1:57 PM	Microsoft Excel W...	1,094 KB

Open Cancel

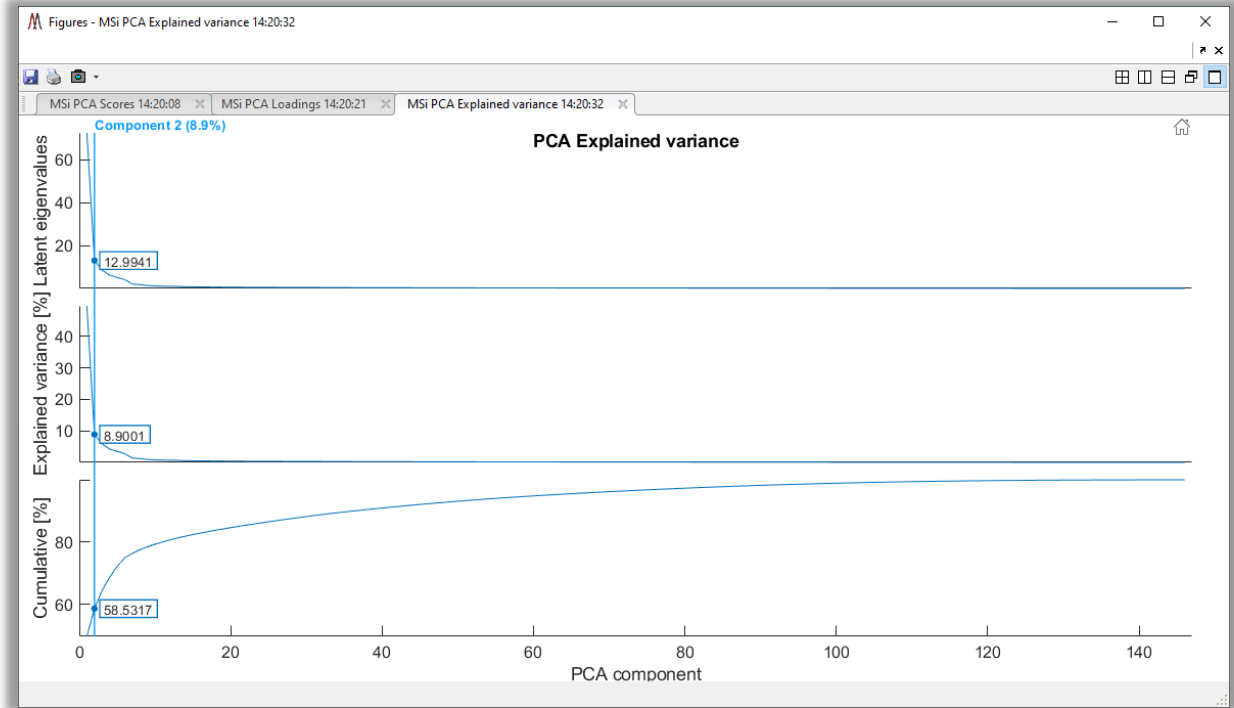
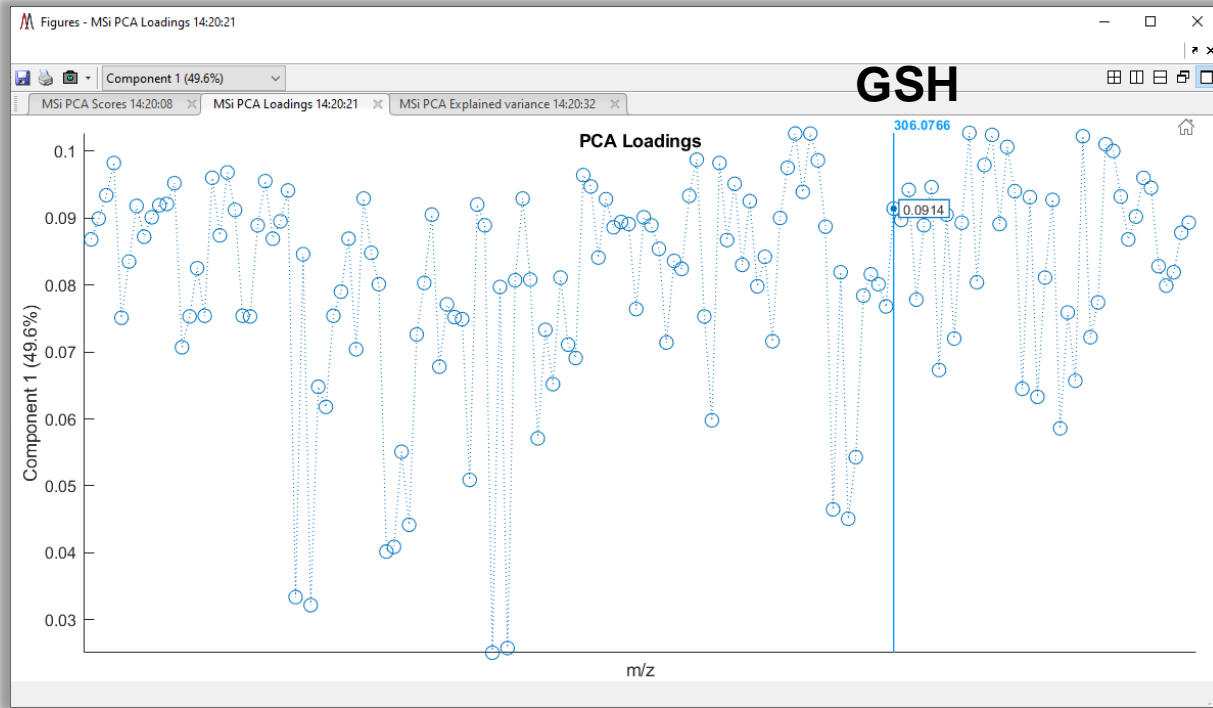
# Statistical Analysis in MSiReader

## Principal Components Analysis (PCA)



# Statistical Analysis in MSiReader

## Principal Components Analysis (PCA)



PCA loadings and Scree plots can be viewed and interacted with.

# Statistical Analysis in MSiReader

## t-distributed Stochastic Neighbor Embedding (t-SNE)

MSiReader v2.70 [64-bits standalone] 14:16:22 Monday, 2024.05.06

Home Pre-Processing Visualization QA/QC Annotations Quantification **Statistical Analysis** Help

PCA  
t-SNE

Launch t-distributed Stochastic Neighbor Embedding (t-SNE) analysis tool. See §7.6.3 in the user manual.

**MSi Data Attributes**

Spots per line: 231  
Spot spacing: 150  
Number of lines: 116  
Line spacing: 150  $\mu\text{m}$

Filter scans: none (load all scans)

Load injection time

Abundance filter

m/z filter

Polarity switch

Anchor

Threshold: 0.001

min: 0 max: inf

parity: odd

Load all scans

Load Data

Select an MSiExport data file

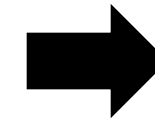
Case Study 2-qMSI > Healthy and Cancerous

Name	Date modified	Type	Size
SSIM	5/6/2024 1:40 PM	File folder	
Cancer_Interrogated	4/24/2024 10:52 AM	Text Document	130 KB
Healthy_Reference	4/24/2024 10:53 AM	Text Document	365 KB
Hen_Ovary_PeakExport	5/6/2024 2:13 PM	Microsoft Excel W...	22,253 KB
Hen_Ovary_QMSI_Cancer_isocount	5/6/2024 1:56 PM	Microsoft Excel W...	1,042 KB
Hen_Ovary_QMSI_Cancer_isocountcarbon	5/6/2024 1:57 PM	Microsoft Excel W...	1,094 KB

File name: Hen\_Ovary\_PeakExport

Exported data file

Open Cancel



MSi tSNE

tSNE algorithm:  
Exact optimizes the Kullback-Leibler divergence of distributions between the original space and the embedded space. Barneshut performs an approximate optimization that is faster and uses less memory when the number of scans is large.

Barneshut

Distance metric:  
Euclidean

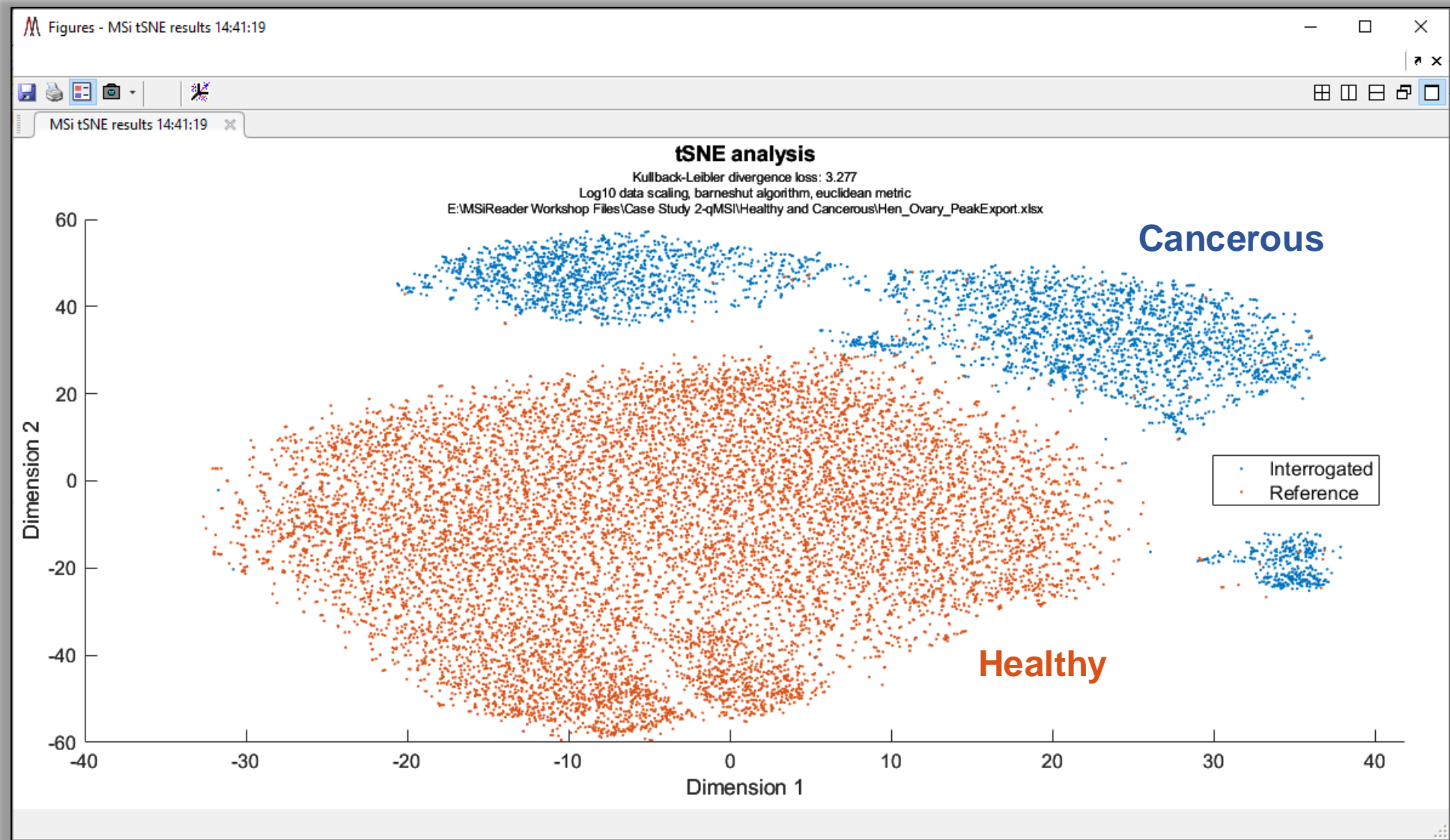
Abundance data normalization:  
Log10 scaling

Re-randomize

OK Cancel

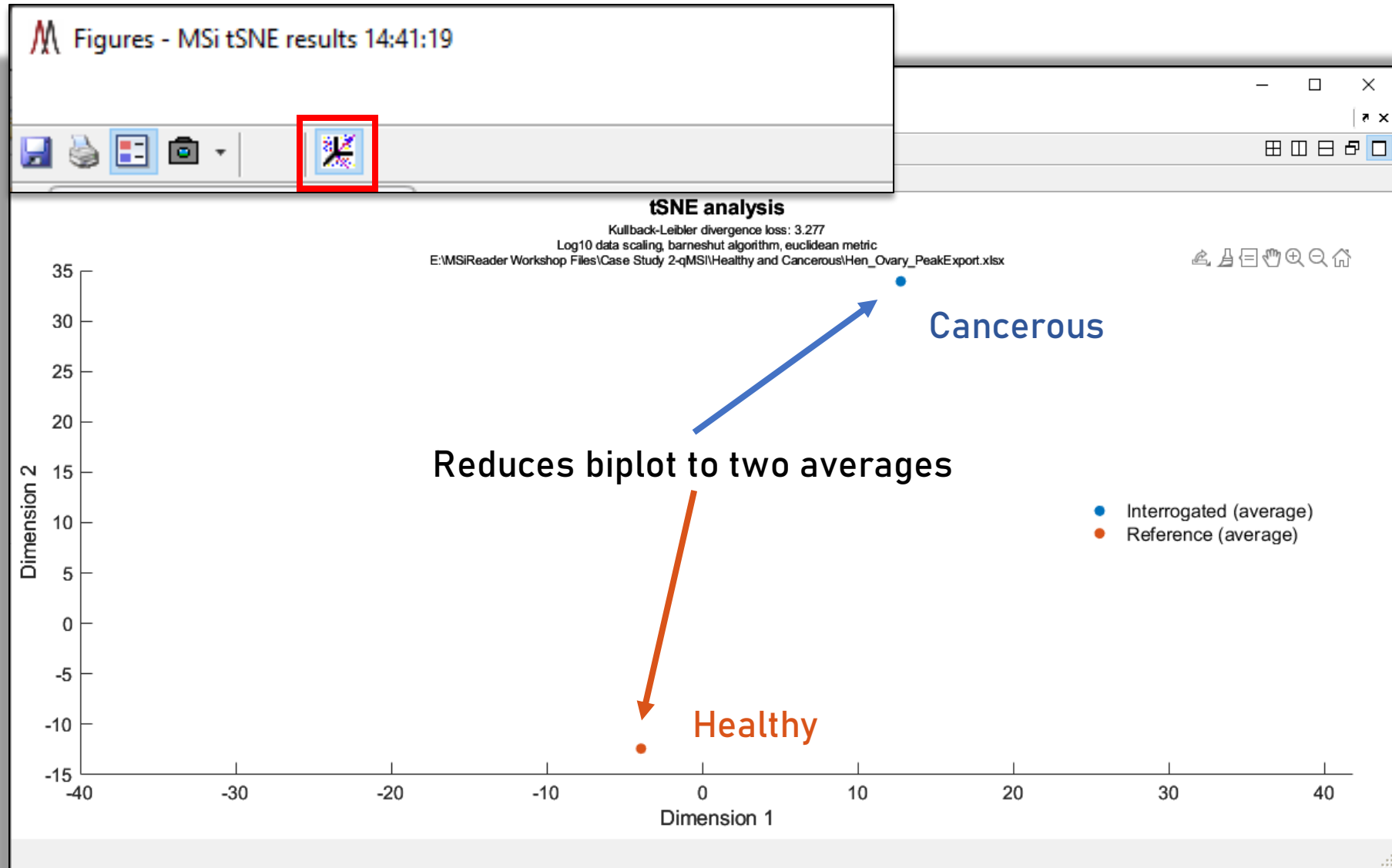
# Statistical Analysis in MSiReader

## t-distributed Stochastic Neighbor Embedding (t-SNE)



# Statistical Analysis in MSiReader

## t-distributed Stochastic Neighbor Embedding (t-SNE)



# AP/MALDI Exploris 480 of Bioactive Peptides and Neurotransmitters

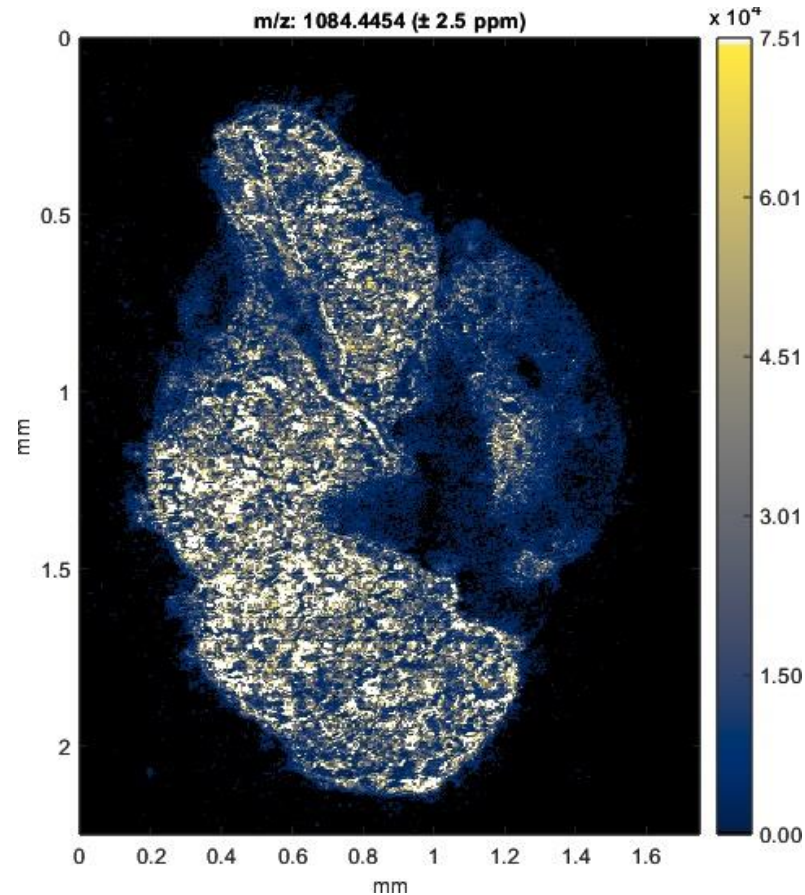


Data Provided by Peter Verhaert and Gilles Frache

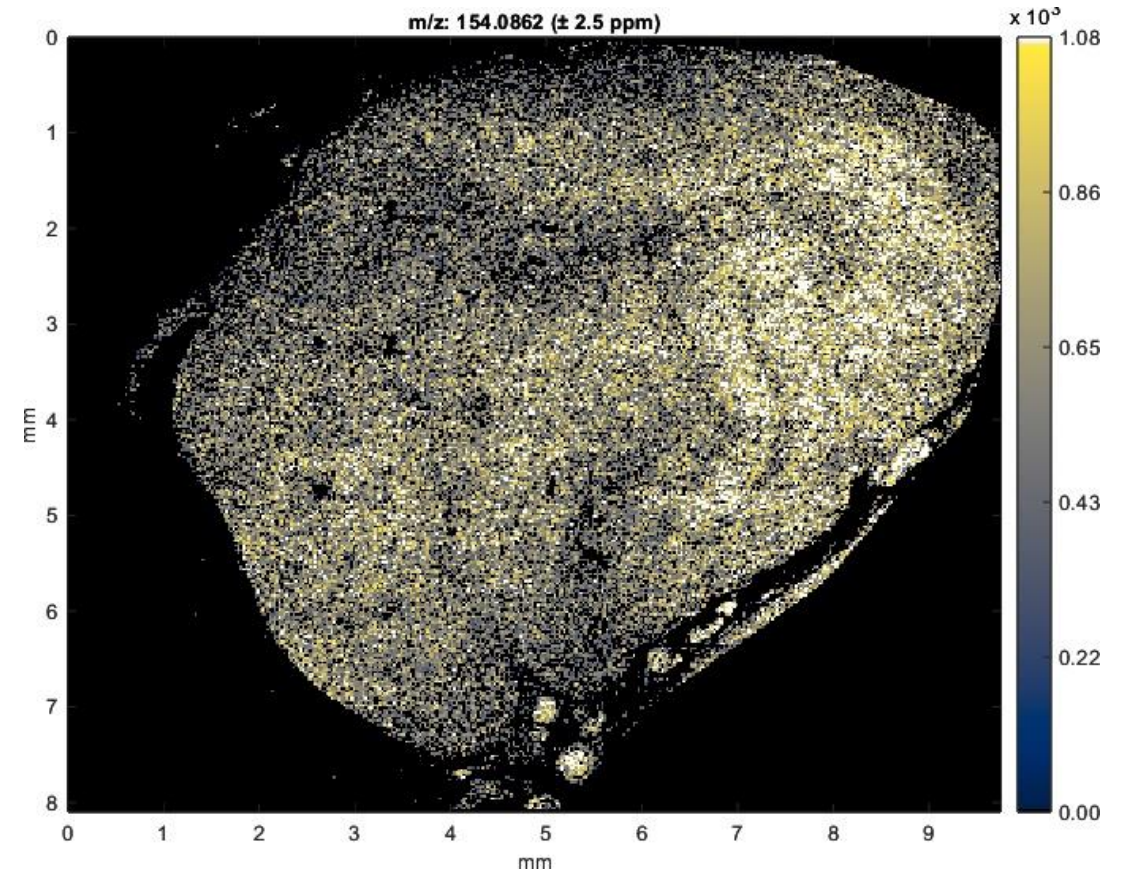
LUXEMBOURG  
INSTITUTE OF SCIENCE  
AND TECHNOLOGY



Vasopressin ( $\pm 2.5$  ppm)  
Adenoma (FFPE)  
5  $\mu\text{m}$  Spatial Resolution



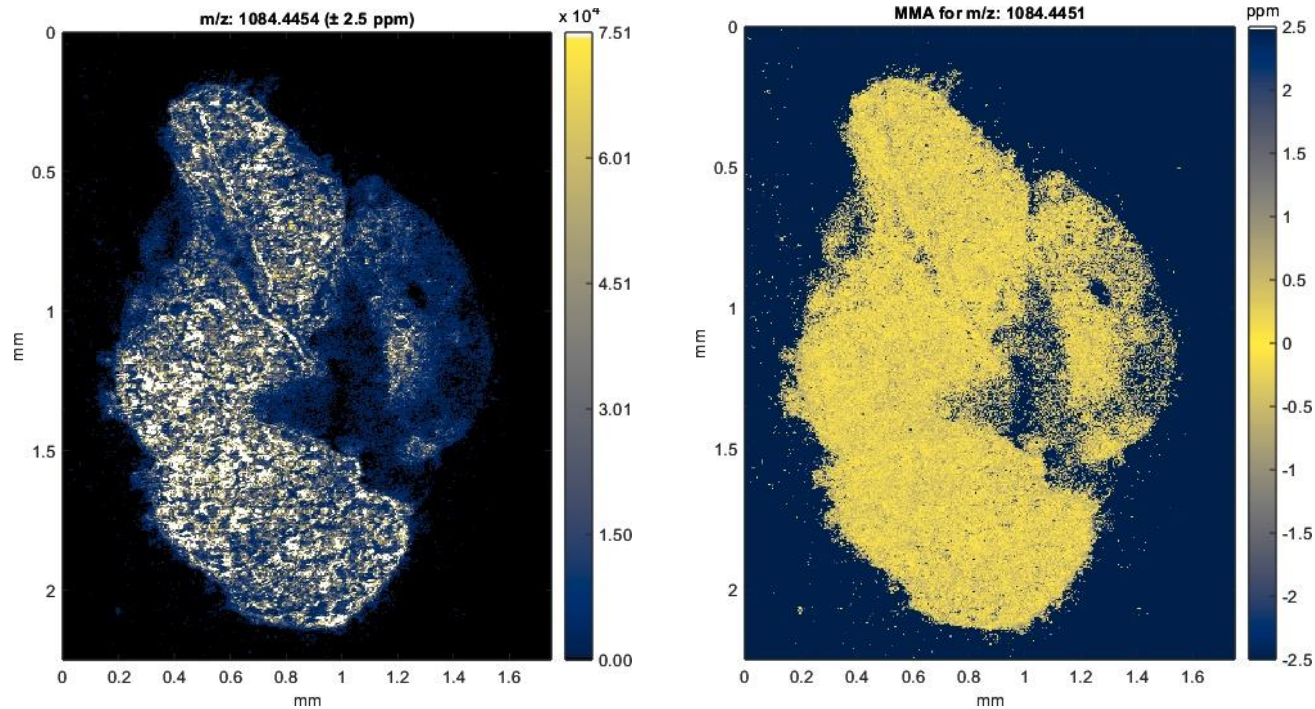
Dopamine ( $\pm 2.5$  ppm)  
Hippocampus (FFPE)  
25  $\mu\text{m}$  Spatial Resolution



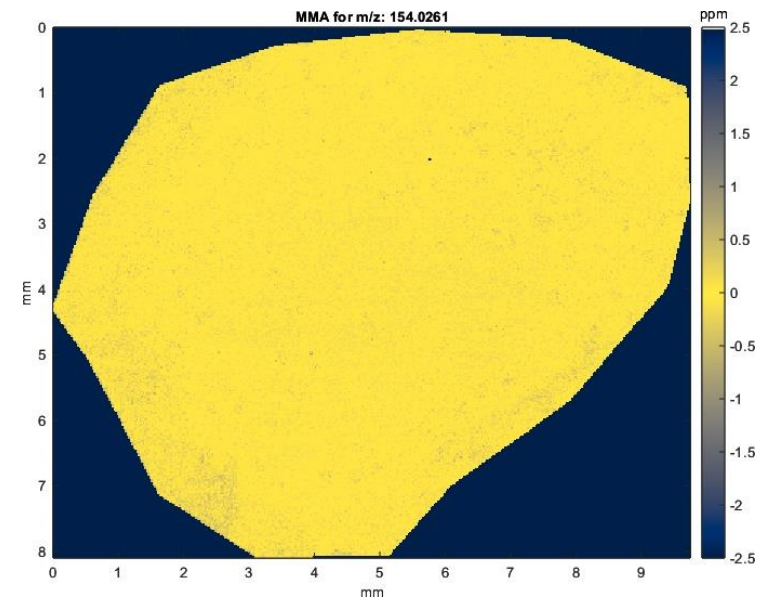
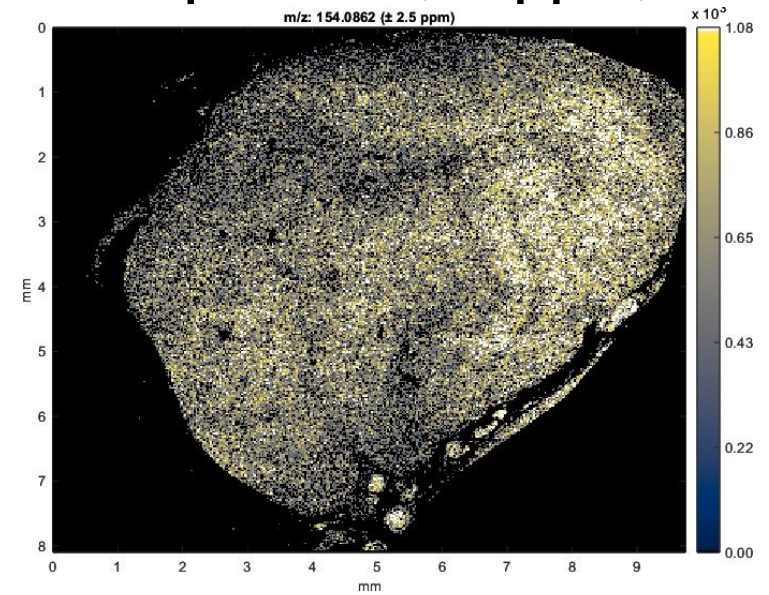
# AP/MALDI Coupled to Exploris 480: Parts-Per-Billion MMA

$$\frac{M_{\text{exp}} - M_{\text{theo}}}{M_{\text{theo}}} \times 10^6 = \text{MMA (ppm)}$$

Vasopressin (< 1 ppm)



Dopamine (< 1 ppm)



AGC Disabled and EasyIC ON




# Co-Localization Algorithm Based on Human Perception (SSIM) Peak List from Clipboard, MSi Peakfinder or METASPACE

*J. Am. Soc. Mass Spectrom.* 2018, 29, 2467-2470

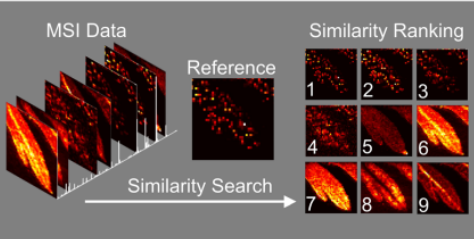
ASMS © American Society for Mass Spectrometry, 2018 CrossMark J. Am. Soc. Mass Spectrom. (2018) DOI: 10.1007/s13361-018-2073-0

APPLICATION NOTE

## Evaluation of Digital Image Recognition Methods for Mass Spectrometry Imaging Data Analysis

Måns Ekelöf,<sup>1</sup> Kenneth P. Garrard,<sup>1</sup> Rika Judd,<sup>2</sup> Elias P. Rosen,<sup>3</sup> De-Yu Xie,<sup>2</sup> Angela D. M. Kashuba,<sup>3</sup> David C. Muddiman<sup>1,2,4</sup> 

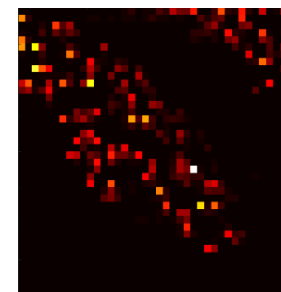
<sup>1</sup>FTMS Laboratory for Human Health Research, Department of Chemistry, North Carolina State University, Raleigh, NC 27695, USA  
<sup>2</sup>Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695, USA  
<sup>3</sup>Division of Pharmacotherapy and Experimental Therapeutics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA  
<sup>4</sup>Molecular Education, Technology, and Research Innovation Center (METRIC), North Carolina State University, Raleigh, NC 27695, USA



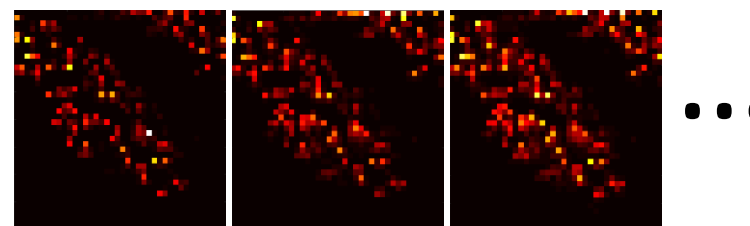
**Abstract.** Analyzing mass spectrometry imaging data can be laborious and time consuming, and as the size and complexity of datasets grow, so does the need for robust automated processing methods. We here present a method for comprehensive, semi-targeted discovery of molecular distributions of interest from mass spectrometry imaging data, using widely available image similarity scoring algorithms to rank images by spatial correlation. A fast and powerful batch search method using a MATLAB implementation of structural similarity (SSIM) index scoring with a pre-selected reference distribution is demonstrated for two sample imaging datasets, a plant metabolite study using *Artemisia annua* leaf, and a drug distribution study using maraviroc-dosed macaque tissue.

**Keywords:** Mass spectrometry imaging, Image recognition, SSIM

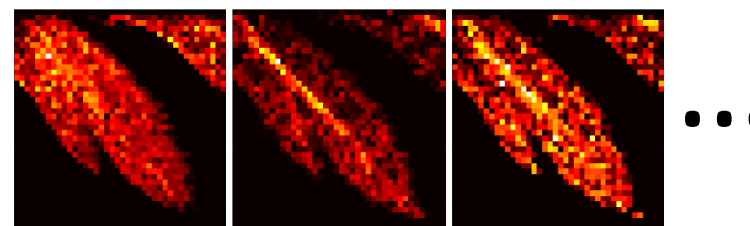
Given a reference image (can also be optical)



Automatically find *similar* images

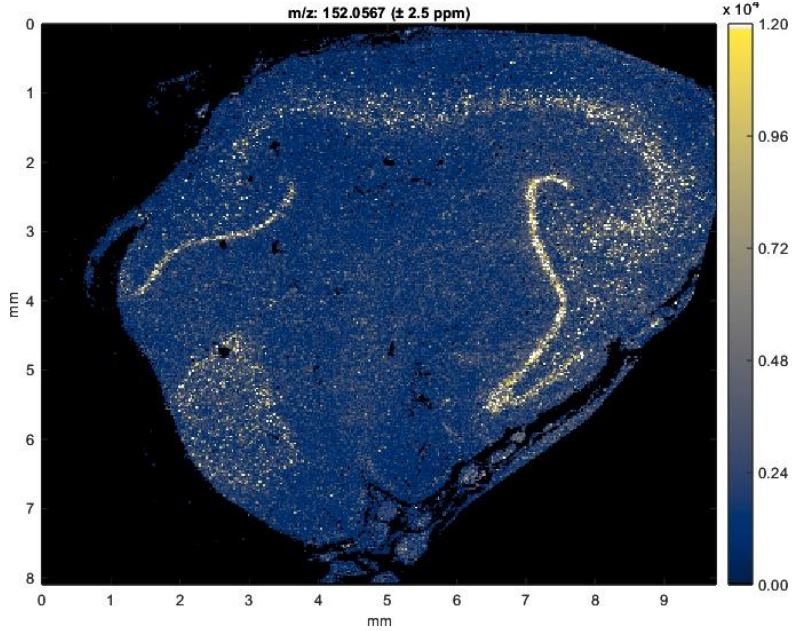


Or perhaps *different* images

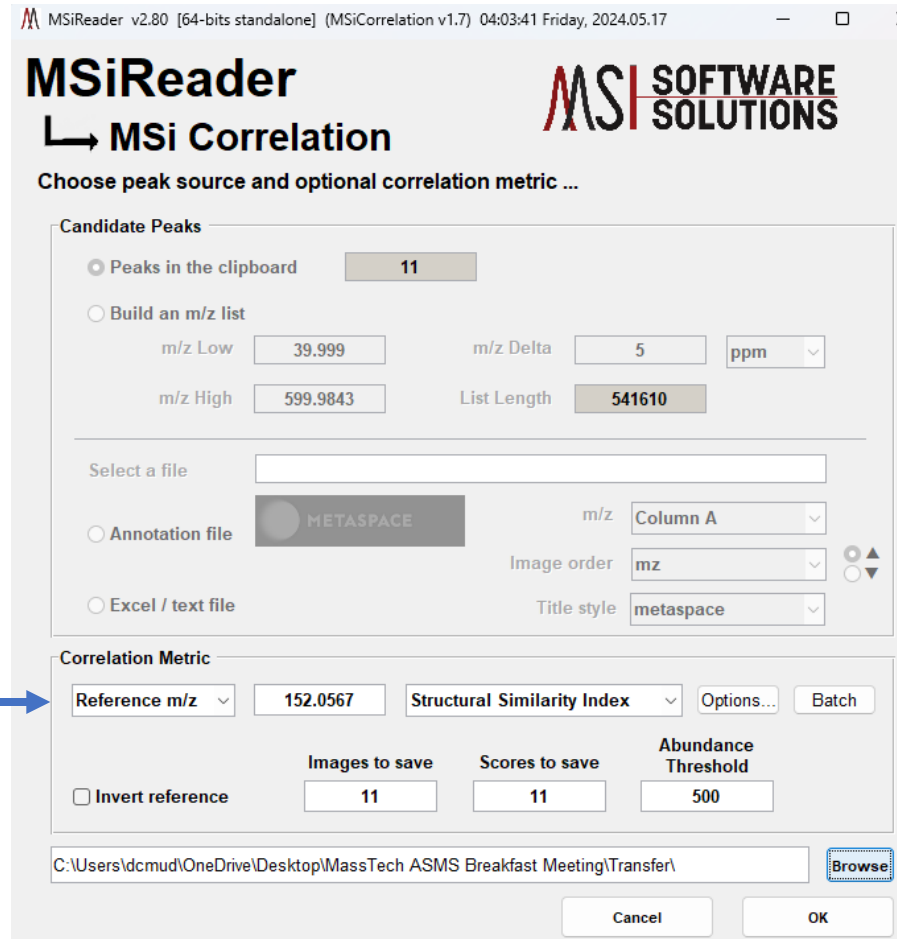


# MSi Correlation Tool using Spectral Similarity Index Matching (SSIM)

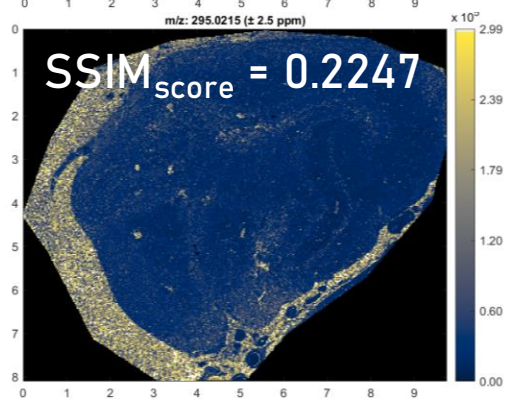
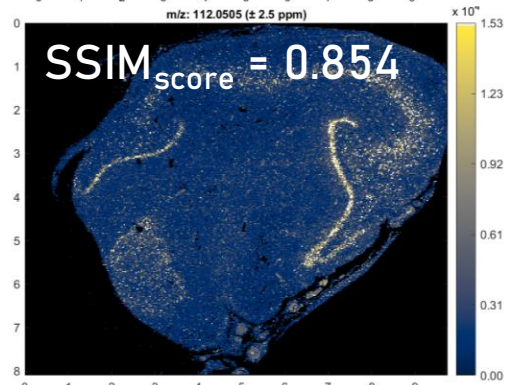
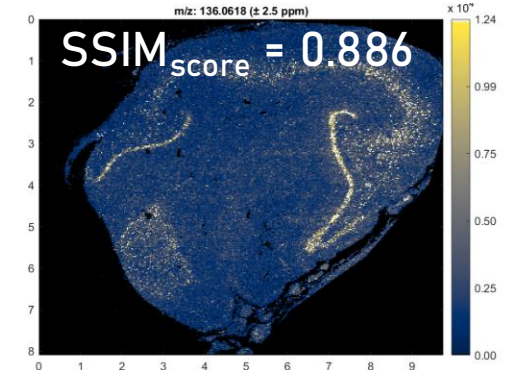
Reference Image:  
Guanine  $m/z = 152.0567$



## MSi Correlation Parameter Selection



Most Similar to Reference



Least Similar to Reference

## SSIM Scores Output

136.061829	0.8861008
112.050537	0.8536921
290.042297	0.2695734
245.044785	0.2680179
70.065178	0.2588272
313.032196	0.2270192
102.127724	0.2264914
176.008057	0.2252148
273.039490	0.2248135
154.026093	0.2247863
295.021484	0.2247147





# Acknowledgements



Venkat Panchagnula



Peter Verhaert



Konstantin Novoselov



Gilles Frache



Nivedita Bhattacharya