Defining a spatially heterogeneous lipid phenotype in hepatocellular carcinoma via MSI.EAGLE analysis

Daniel Boehmler, Aalim Weljie

2024-04-19

Table of contents

Chapter 1

Introduction

Mass spectrometry imaging (MSI) offers a powerful and versatile tool to determine the spatial distribution of biomolecules within tissues. MSI data offers valuable insights into tumor metabolic heterogeneity and microenvironment, mapping neurotransmitters and neuropeptides in brain tissue, and enabling the visualization of drug distribution within organs and tissues.

This tutorial has three main goals:

- The first aims to showcase a small use-case for MSI in characterizing the lipidomic profile of HCC tumors, and identifying spatially distinct phenotypes within the tumor microenvironment.
- The second is to detail the step-by-step process of analyzing MSI data using the MSI.EAGLE Shiny package. In doing so, we will cover setting up the software environment, importing and processing MSI data, peak-picking, background segmentation using both UMAP and SSC methods, targeted lipidomics data analysis, phenotyping through UMAP clustering, and finally statistical analysis of the processed data. Each section will provide instructions, accompanied by reference images, to guide users through the analysis workflow effectively.
- The third is to enhance accessibility to relatively inaccessible MSI data via tools like MSI.EAGLE. Previously, extensive knowlege of how to navigate R was required to perform MSI data analysis. By offering a user-friendly interface, we hope to enable researchers across disciplines to leverage spatial lipidomics analysis without the need for extensive computational expertise.

Chapter 2

Starting MSI.EAGLE

The code chunks below can be run simultaneously as an R script to start the MSI.EAGLE Shiny, but are divided for clarity.

Begin by setting your root project directory and working directory. In this case, they are the same location and are defined as such:

```
rawd = "/path/to/desktop/directory"
wd = rawd
```

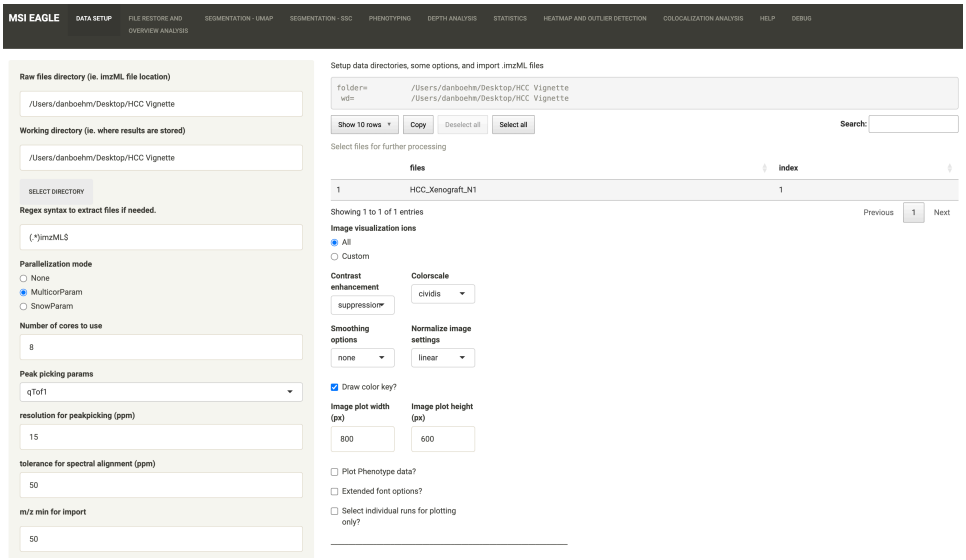
Next, specify the number of cores to utilize. For optimal performance of MSI.EAGLE, it's advisable to use two fewer cores than the maximum available, ensuring efficient operation without compromising your machine's fundamental functions.

```
ncores = as.integer(parallel::detectCores()-2)
```

Finally, load and run the application.

```
library(MSI.EAGLE)
MSI.EAGLE()
```

The start page for MSI.EAGLE should soon appear:


MSI EAGLE
DATA SETUP
FILE RESTORE AND OVERVIEW ANALYSIS
SEGMENTATION - UMAP
SEGMENTATION - SSC
PHENOTYPING
DEPTH ANALYSIS
STATISTICS
HEATMAP AND OUTLIER DETECTION
COLOCALIZATION ANALYSIS
HELP
DEBUG
Raw files directory (ie. imzML file location)
/Users/danboehm/Desktop/HCC Vignette
Working directory (ie. where results are stored)
/Users/danboehm/Desktop/HCC Vignette
SELECT DIRECTORY
Regex syntax to extract files if needed.
(.*)imzML$
Parallelization mode
None
MulticorParam
SnowParam
Number of cores to use
8
Peak picking params
qTof1
resolution for peakpicking (ppm)
15
tolerance for spectral alignment (ppm)
50
m/z min for import
50
Setup data directories, some options, and import .imzML files
folder= /Users/danboehm/Desktop/HCC Vignette
wd= /Users/danboehm/Desktop/HCC Vignette
Show 10 rows
Copy
Deselect all
Select all
Search:
Select files for further processing
files
index
1
HCC_Xenograft_N1
1
Showing 1 to 1 of 1 entries
Previous
1
Next
Image visualization ions
All
Custom
Contrast enhancement
suppression
Colorscale
cividis
Smoothing options
none
Normalize image settings
linear
Draw color key?
Image plot width (px)
800
Image plot height (px)
600
Plot Phenotype data?
Extended font options?
Select individual runs for plotting only?

Chapter 3

Example Data Information

To expidite our workflow in exploring the functionalities of MSI.EAGLE, the provided MSI scan has already undergone the necessary lockmass adjustment, centroiding, and conversion to .ibd and .imzmL file formats. For guidance on these conversion steps with your own data, please refer to the relevant section in the MSI.EAGLE user manual.

The dataset provided for this tutorial comprises hepatocellular carcinoma flank tumors sourced from patient-derived xenografts. These tumors were excised from mice, promptly flash-frozen in liquid nitrogen, and suspended in 1% CMC media.

Slices of 10 μm thickness were prepared, mounted on glass microscope slides, and frozen at -80°C before acquisition.

Upon thawing and slide drying, lipids were captured using a Waters G2-XS QTOF with a DESI source. The spray solvent consisted of 98% MeOH containing 50 pg/mL leucine enkephalin, flowing at 2.0 μL/min. Pixel dimensions were 100μm x 100μm, with a raster rate of 200 μm/second. Source and transfer line temperatures were set at 150 °C and 450 °C, respectively. Capillary voltage was maintained at 0.7 kV, and the sample cone at 40 kV. Spectra were acquired in negative ionization mode and spanned m/z 50-1700.

To illustrate average workflow capabilities and throughput, all MSI.EAGLE analyses were conducted on an Apple Macbook Pro featuring an M1 chip and 32 GB of memory, running R version 4.3.2.

Chapter 4

Preliminary Peak-picking and Background Segmentation

Particularly with large data files, it's advisable to initially import only a fraction of the acquired mass range. This approach streamlines the initial peak-picking and background removal procedures, deferring the full analysis to subsequent steps.

This necessity arises from the exponential increase in total mass peaks as the import range expands. For instance, a narrow range such as m/z 200-300 yields approximately 20 picked mass peaks, whereas the broader range of m/z 50-1700 generates about 2000 peaks. Consider that there are 82,152 total pixels in the example data, with each pixel requiring peak picking for the imported range.

Thus, for the range m/z 200-300, there are approximately $20 \times 82,152 = 1.6\text{E}+06$ aggregate peaks, compared to $50 \times 82,152 = 1.6\text{E}+07$ for the range m/z 50-1700, reflecting a significant increase in the number of peaks to be processed.

This growth in data size, and the resulting increase in computational demands, underscores the significance of starting with a narrower mass range. Opting for a focused mass range containing biologically relevant compounds enhances analysis efficiency, particularly during later segmentation steps.

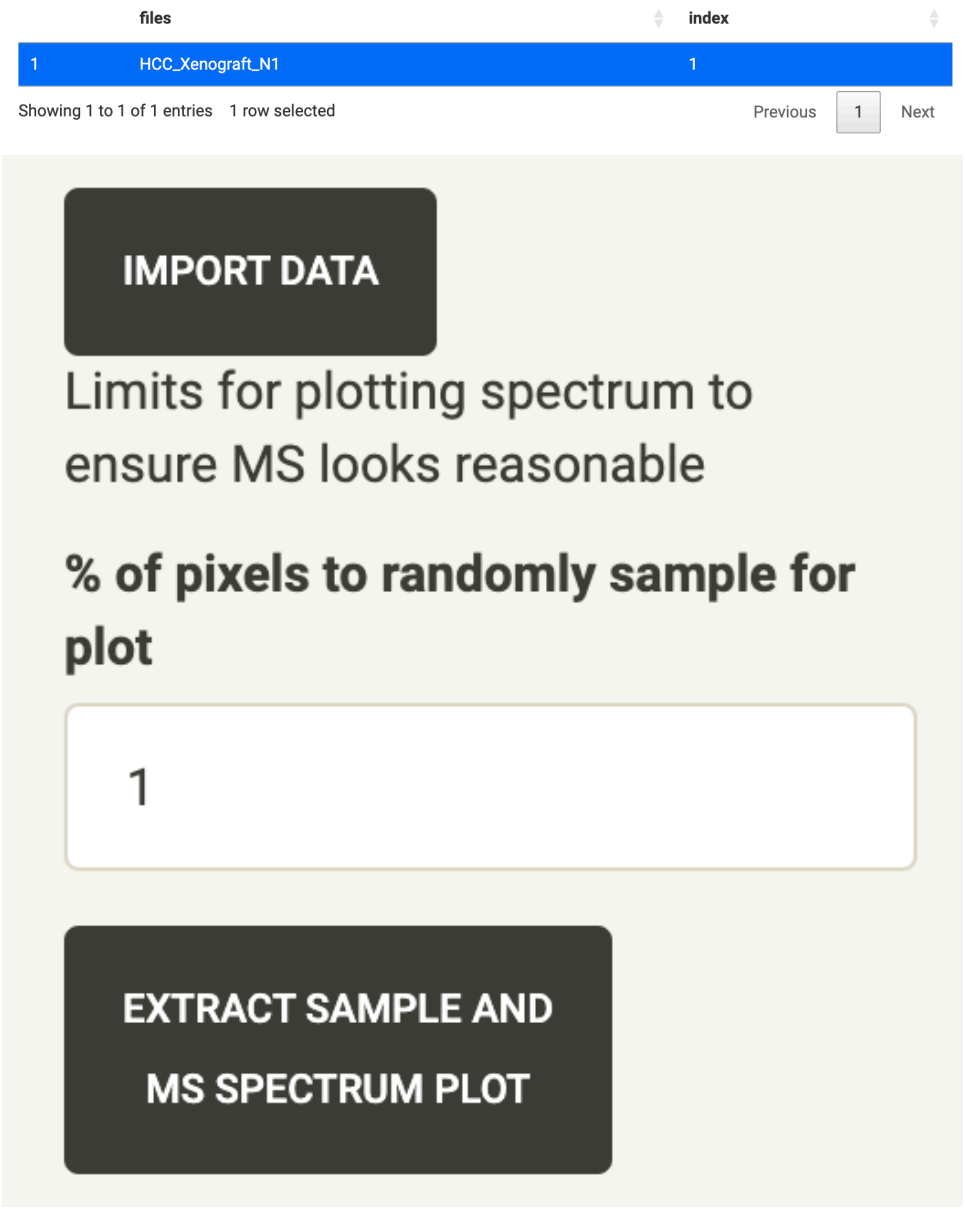
4.1 Initial Data Import

To load the example data into MSI.EAGLE, fill in the parameters below in the "Data Setup" tab, noting that only the min and max m/z values for import differ from the pre-set default values:

Parameter	Fill
Raw files directory	/path/to/desktop/directory
Working directory	/path/to/desktop/directory
Regex syntax	(.*)imzML$
Parallelization mode	MulticorParam
Number of cores to use	8

Parameter	Fill
Peak picking params	qTof1
resolution for peakpicking (ppm)	15
tolerance for spectral alignment (ppm)	50
m/z min for import	200
m/z max for import	300
% of pixels to randomly sample for plot	1

After verifying the import parameters are as-desired, select the data file and import the data by pressing the button shown below.



4.2 Initial Peak-picking

The parameters below should be filled out in the "File Restore and Overview Analysis" tab, being sure to select the option "Peakpick rawfiles".

Begin peak-picking by pressing the button pictured below. This is usually the longest step in processing data with MSI.EAGLE, particularly for large or high-resolution files. A few minutes wait time is typical.

Parameter	Fill
Peakpeaking operation	Peakpick rawfiles (imzML loaded in Data Setup tab)
Start from an existing peak_picked file?	No, fresh pick raw files
% of pixels to randomly sample for peak picking	10
S/N for overview peak picking	100
Minimum peak frequency (0-1)	0.01
Peak picking method (simple, adaptive, mad)	mad

Peakpeaking operation

- ○ Open previously peakpicked file (.rds)
- ◉ Peakpick rawfiles (imzML loaded in Data Setup tab)
- ○ Add two imagesets (same coordinates)
- ○ Add two imagesets (same peaklist)

Start from an existing peak_picked file?

- ◉ No, fresh pick raw files
- ○ From Reference file
- ○ From Mean Spectrum

% of pixels to randomly sample for peak picking

10

S/N for overview peak picking

100

Minimum peak frequency (0-1)

0.01

Peak picking method (simple, adaptive, mad)

simple

START PEAK PICKING

Features selected / filtered (by comma or colon)

SAVE SELECTED RUNS FROM PEAK PICKED FILE

Peakpeaking operation

- ○ Open previously peakpicked file (.rds)
- ◉ Peakpick rawfiles (imzML loaded in Data Setup tab)
- ○ Add two imagesets (same coordinates)
- ○ Add two imagesets (same peaklist)

Start from an existing peak_picked file?

- ◉ No, fresh pick raw files
- ○ From Reference file
- ○ From Mean Spectrum

% of pixels to randomly sample for peak picking

10

S/N for overview peak picking

100

Minimum peak frequency (0-1)

0.01

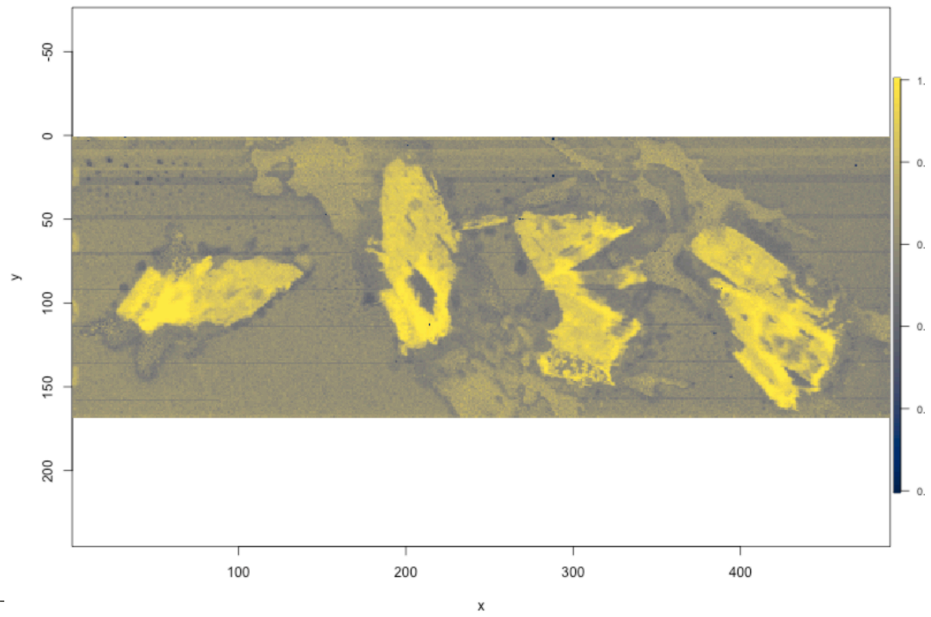
Peak picking method (simple, adaptive, mad)

simple

START PEAK PICKING

Features selected / filtered (by comma or colon)

SAVE SELECTED RUNS FROM PEAK PICKED FILE



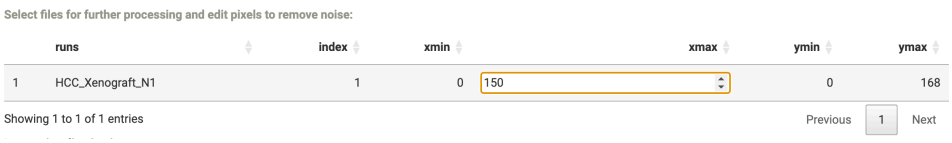
Clearly, the tissue can be discerned over the background signal via masses in the m/z 200-300 range. If the tissue had not be readily distinguishable from the background ionization, this process would have been iterated with broader mass ranges until differentiation was achieved.

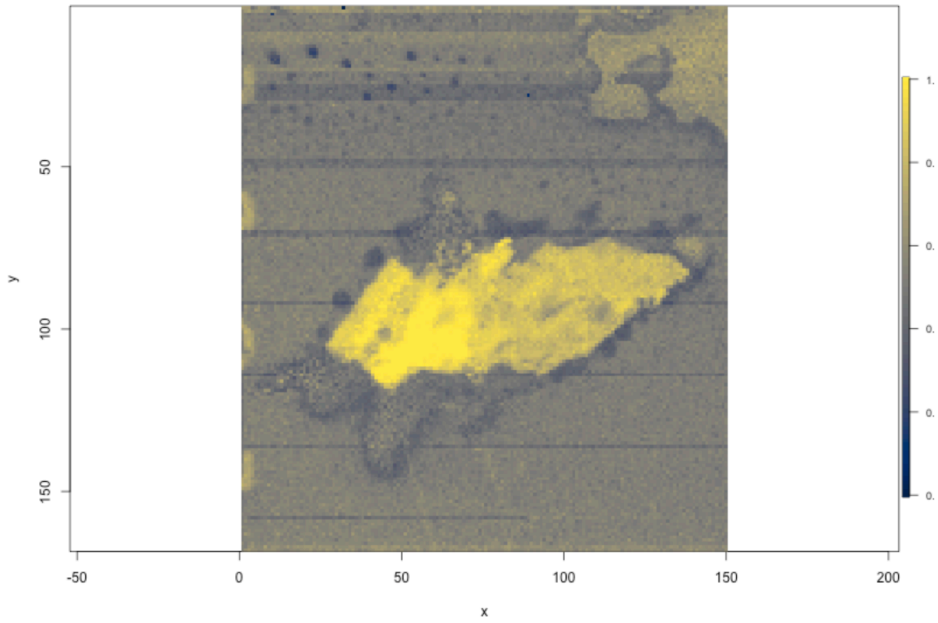
Additionally, we can see that only the first tissue section has the structural integrity to continue with the analysis. The others were poorly sliced tissue sections, so should be cropped. To do so, we can double click the pixel dimension to change (xmax/min or ymax/min), and type the value needed. Here, our aim is to isolate the leftmost section, which is accomplished by setting the xmax value to 150 pixels.

At this stage, it's advisable to save the peak-picked data as an R Data Serialization file (.rds extension). This format maintains an R object, such as the peak-picked data, across R sessions. Thus, one can close MSI.EAGLE, then later return to the "File Restore and Overview Analysis" tab and pick up the analysis by simply opening the previously peak-picked .rds file.

Once you've clicked "Save Selected Runs from Peak Picked Data", verify that the file is named in a recognizable manner and saved within the same working directory as the raw data files.

MAKE SURE TO SAVE YOUR DATA AS A .RDS FILE AFTER PEAK-PICKING




y
50
100
150
-50
0
50
100
150
200
x
1.
0.
0.
0.
0.
0.

Chapter 5

Segmentation for Background Removal

There are two options for segmentation analysis available, UMAP and SSC. In this example we will compare the two and determine which is more effective for this particular sample/analysis type.

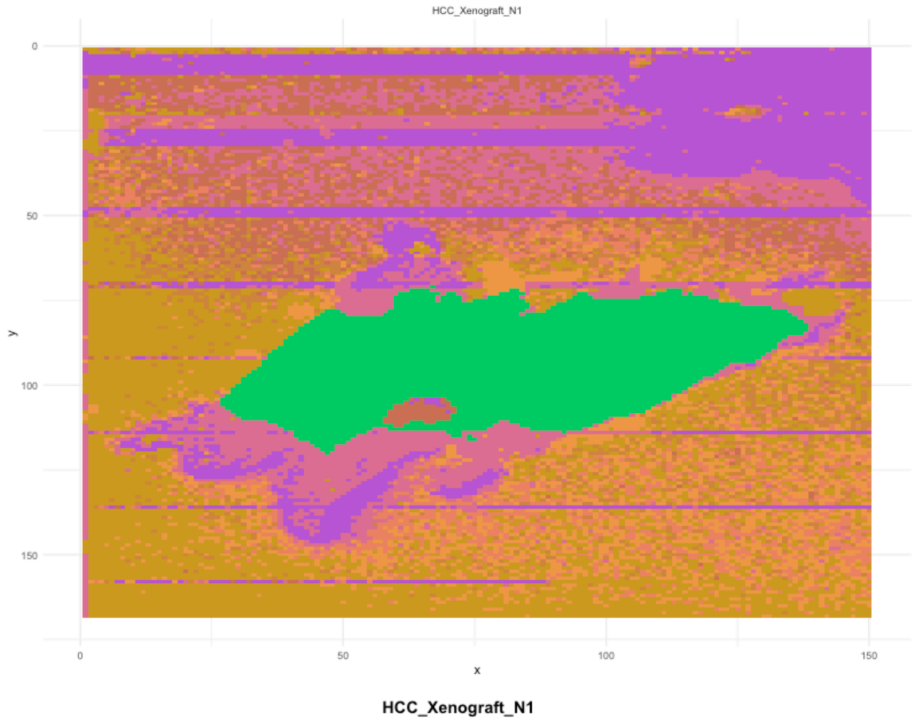
5.1 UMAP Background Segmentation

After moving to the UMAP tab in MSI.EAGLE, verify the default parameters for background removal via UMAP are as listed below:

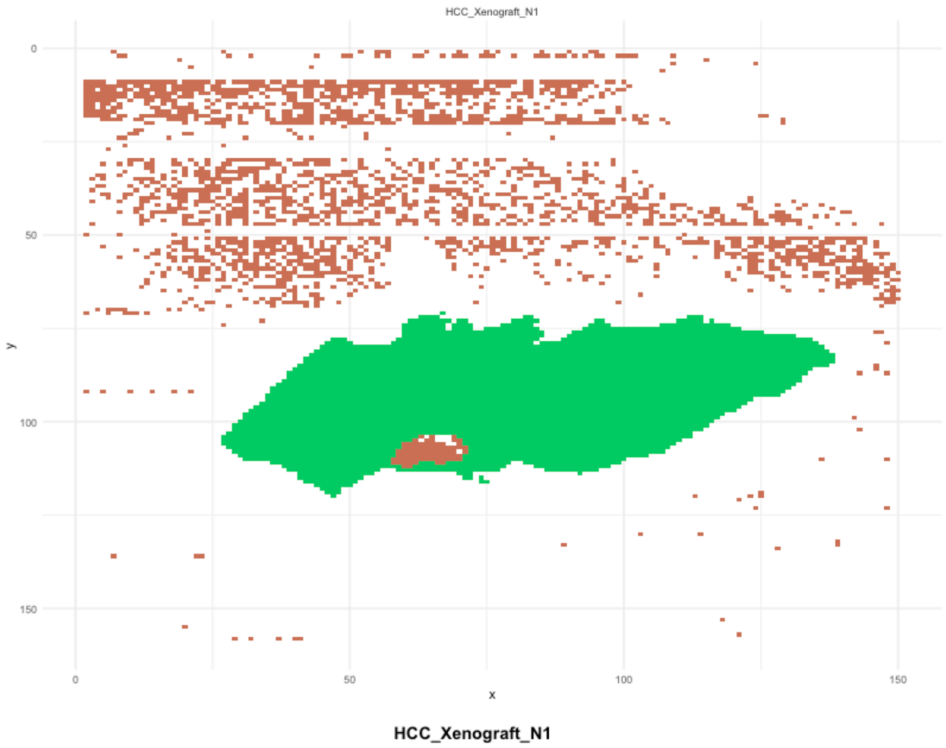
Parameter	Fill
minimum distance for UMAP	0.1
'# of node to search	50
UMAP nearest neighbors	5
Number of trees for UMAP	50
set_op_mix_ratio	1
UMAP PCA components (NULL for no PCA)	
Method for HDBSCAN calculation	R DBSCAN package
tolerance for spectral alignment (ppm)	0.15
DBSCAN eps	50
H/DBSCAN min points / cluster	50
cex, tile plot size adjustment	2
R color reduction quantile (0-1)	0.9

After selecting the data from the peak-picked file, press "Start UMAP" to begin clustering.

Thus far, UMAP segmentation seems to have performed well in separating the ions generated from the tissue (green) from those of the CMC media left on the slide (purple) and the glass slide surface (orange).



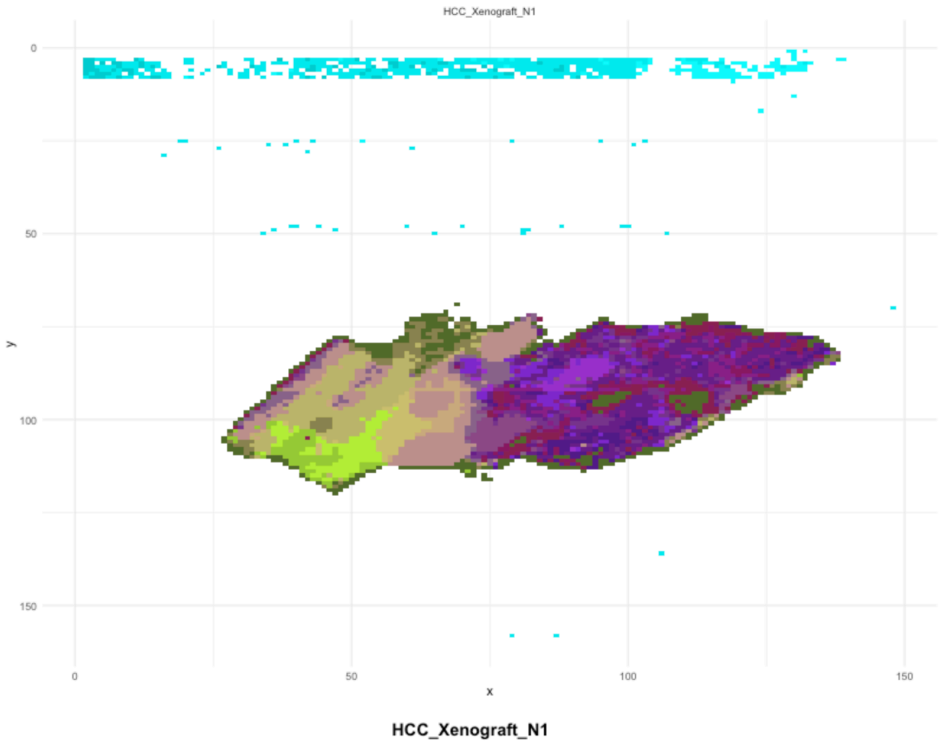
After removing some of the background clusters, we can see that one cluster contains both pixels over the tissue of interest and over the slide. With the listed parameters, this method was not effective in completely segmenting out background pixels.



Since the background was not completely removed, we should re-cluster the pixels. Press “Store Processed

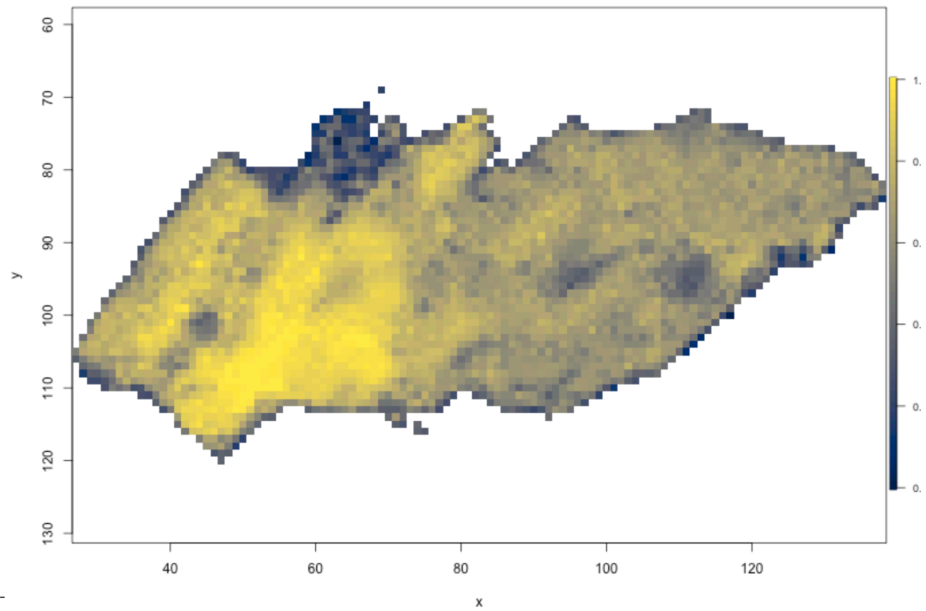
Data" to preserve the background pixels that have already been removed, lower the R color reduction quantile to 0.7 to produce more segmented pixel clusters, then "Start UMAP" again.

This generates a more complex pixel mapping due to the lowered color reduction quantile, which can now distinguish background from sample pixels with finer detail.



HCC_Xenograft_N1

After removing the band from the top of the image, we are left with only the pixels covering the tissue surface. This file should first be stored, then saved as a .rds file for later analysis.



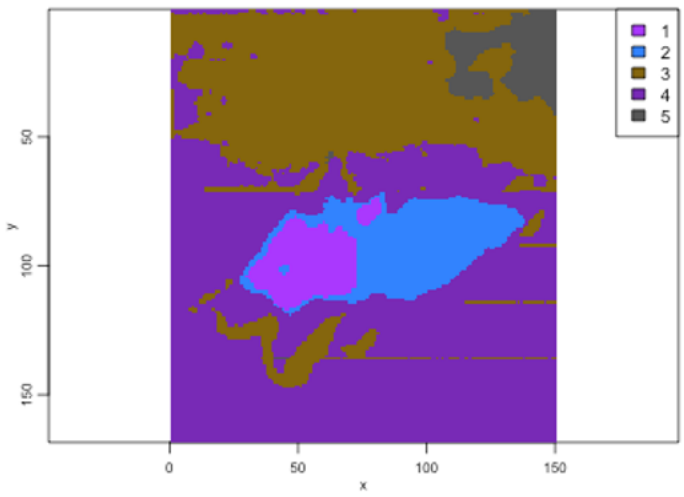
5.2 SSC Background Segmentation

Though UMAP segmentation was effective, comparing it to SSC segmentation is a useful exercise so users can make an informed choice with their own data.

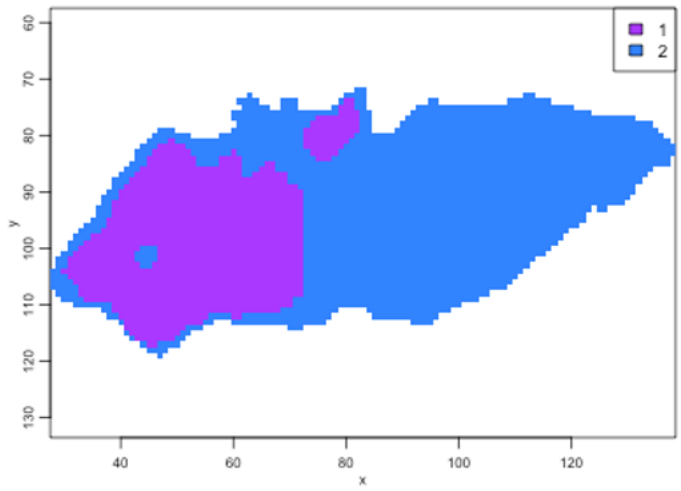
After moving to the SSC tab in MSI.EAGLE, verify the default parameters for background removal via SSC are as listed below, then, press "Start SSC".

Parameter	Fill
radius for SSC	3
k (clusters) for SSC	5
shrinkage for Cardinal SSC	c(5,10)
Method for SSC	adaptive
Colors for ssc visualzation	alphabet

To begin removing background pixels, first select one SSC model from the analysis. In this case, the r = 3 k = 5 s = 5 model was chosen.



By removing all colors except 1 and 2, the background was successfully removed. The data can now be stored, then saved as a .rds file for further analysis.



Conducting both UMAP and SSC segmentation revealed that while both methods were effective in removing background pixels, the SSC method was faster for this particular example.

Chapter 6

Targeted Lipidomics Data Analysis

Next, a known list of ions/compounds will be extracted from the isolated tissue sample.

First, return to the "Data Setup" tab, and import the full mass range acquired, in this case m/z 50-1700. Then, proceed to the "Depth Analysis" tab in MSI.EAGLE.

In the "Imageset with segmented coordinates" field, select and read in the .rds file saved after the background removal step. Select the Targeted option, and import the provided .txt file containing compound IDs and exact mass values with the following structure:

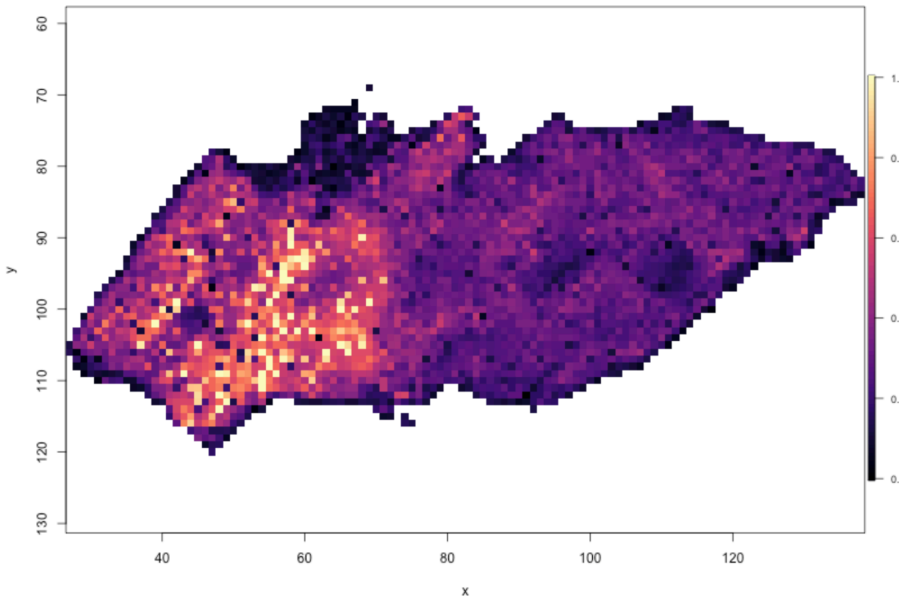
A tab-delimited text file containing two columns, "ID" and "exact.mass", as below:

ID	exact.mass
FA 14:0	227.20184
FA 16:1	253.21738
FA 16:0	255.23317
FA 17:0	269.24875
...	...

The mass list provided in this example is a mix of 206 known lipid species that are detected in negative mode, and is provided in the appendix of this document.

Set the tolerance for peak binning to 15 ppm (subject to change dependent on instrument capability), and begin peak picking. This file should first be stored, then saved as a .rds file for later analysis.

On first observation, there appears to be a delineation in lipid abundance between different tumor regions. To clearly define these regions and perform statistical analyses, segmentation analysis should be carried out to assign a phenotype to each region.


y
60
70
80
90
100
110
120
130
40
60
80
100
120
x

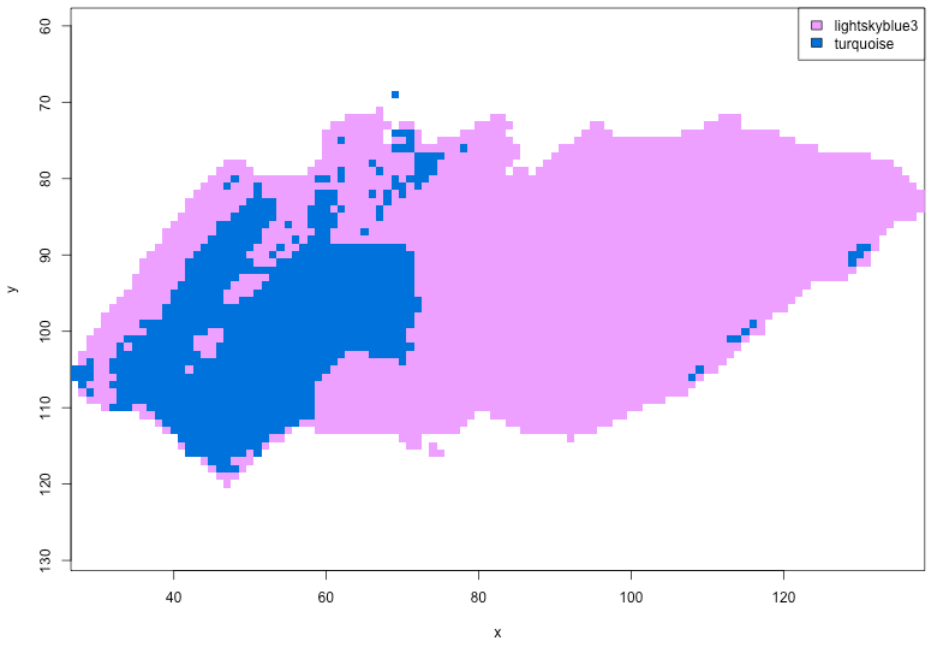
Chapter 7

Phenotyping via UMAP clustering

After restoring the .rds file produced in the Data Analysis tab, move to the Segmentation-UMAP tab. Input the parameters listed in the table below and start the UMAP:

Parameter	Fill
minimum distance for UMAP	10
'# of node to search	50
UMAP nearest neighbors	5
Number of trees for UMAP	50
set_op_mix_ratio	1
UMAP PCA components (NULL for no PCA)	
Method for HDBSCAN calculation	R DBSCAN package
tolerance for spectral alignment (ppm)	0.15
DBSCAN eps	50
H/DBSCAN min points / cluster	50
cex, tile plot size adjustment	2
R color reduction quantile (0-1)	0.97

This will produce a segmented file which can be visualized by selecting the "Plot Phenotype Data" option. This file should first be stored, then saved as a .rds file for later analysis.



Now, move to the "Statistics" tab. Begin by reading in the segmented file that was just produced, and selecting the parameters outlined in the table below. In the "members to include" section, you should have two or more color names which correspond to a region of the UMAP.

Parameter	Fill
Choose variable to test	Rcol_reduced
Choose members to include	your color #1 your color #2
Choose interaction for grouping	x or y
Choose variables for modeling	Rcol_reduced

Conducting a means test on these groups aims to identify ions with significant differences in abundance between the two regions delineated in the UMAP phenotyping step, grouped by their position along the x- or y-axis.

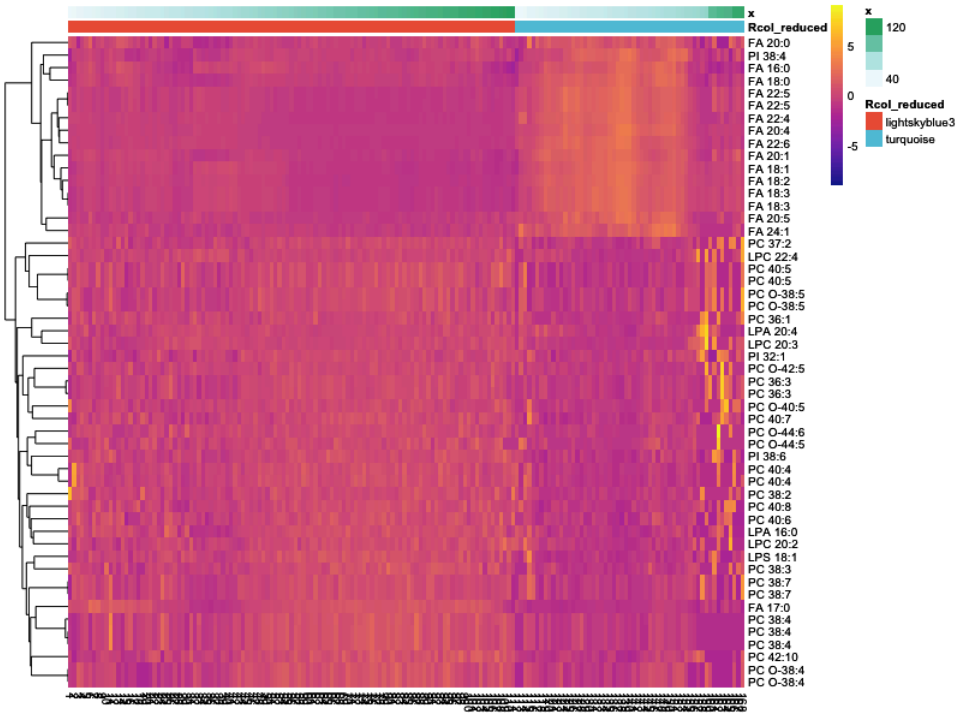
Within the "Output Table and Ion Plots" tab, you can observe statistically significant ($p<0.05$) individual ions and compounds, accompanied by their respective p-values, fold changes, and mean ion intensities. Moreover, this table can be exported for further external analysis.

For visual representation of this data, navigate to the "Heatmap and Outlier Detection" tab and input the parameters detailed in the following table:

Parameter	Fill
Choose variables for heatmap	Rcol_reduced, x
Heatmap parameters	Show only significant peaks, Annotate columns, Scale row, Cluster rows

Parameter	Fill
Choose labels	ID
clustering method (R hclust)	complete
column distance method	euclidean
row distance method	euclidean
Heatmap colors	plasma
Choose filtering column for heatmap	PValue
Filtering stat direction?	Descending
Significance cutoff	1E-9

This will generate a heatmap illustrating the distinct lipid profiles of the two tumor regions. Notably, there's a marked elevation in free fatty acids within the smaller of the two regions of tumor cross-section, juxtaposed with a notable increase in phospholipids in the larger area.



Chapter 8

Conclusions

Delving into the mechanisms responsible for this spatially-resolved phenotype in HCC xenografts goes beyond the scope of this tutorial. Nonetheless, our aim was to first illuminate an interesting use-case for mass spectrometry imaging (MSI), where disease characterization via lipidomics analyses was defined by the incorporation of a novel spatial dimension. Furthermore, by illustrating how researchers can utilize the advanced functionalities and user-friendly interface provided by MSI.EAGLE, we aim to enhance accessibility to MSI data, thereby fostering more thorough and insightful analyses that contribute to the advancement of the field.

Chapter 9

Appendix

Table 9.1: Lipids Inclusion List

ID	exact.mass
FA 14:0	227.2018
FA 16:1	253.2174
FA 16:0	255.2332
FA 17:0	269.2487
FA 18:3	277.2172
FA 18:3	277.2174
FA 18:2	279.2334
FA 18:1	281.2494
FA 18:0	283.2643
FA 20:5	301.2176
FA 20:4	303.2331
FA 20:1	309.2804
FA 20:0	311.2957
FA 22:6	327.2330
FA 22:5	329.2485
FA 22:5	329.2486
FA 22:4	331.2643
FA 22:0	339.3271
FA 24:1	365.3425
LPA 16:0	409.2351
LPA 18:2	433.2363
LPA 18:1	435.2520
LPE O-16:1	436.2837
LPA 18:0	437.2681
LPE 16:0	452.2779
LPE 16:0	452.2782

ID	exact.mass
LPA 20:4	457.2363
LPA 20:3	459.2505
LPE O-18:2	462.2985
LPE O-18:1	464.3146
LPE 18:2	476.2785
LPE 18:2	476.2788
LPE 18:1	478.2938
LPE 18:0	480.3091
LPE 18:0	480.3095
LPA 22:6	481.2359
LPE O-20:1	492.3449
LPE 20:4	500.2786
LPE 20:3	502.2933
LPG 18:2	507.2720
LPG 18:1	509.2889
LPC 14:0	512.2994
LPS 18:1	522.2827
LPS 18:0	524.2999
LPE 22:5	526.2947
LPC 15:0	526.3149
LPC O-16:1	526.3521
LPE 22:4	528.3096
LPC 16:1	538.3154
LPC 16:0	540.3307
LPC 16:0	540.3309
LPC 17:1	552.3314
LPC O-18:1	552.3669
LPC O-18:1	552.3671
LPC 17:0	554.3463
LPC O-18:0	554.3834
LPC 18:2	564.3304
LPC 18:2	564.3309
LPC 18:1	566.3463
LPC 18:1	566.3466
LPC 18:0	568.3621
LPC 18:0	568.3622
LPI 16:0	571.2899
LPC 19:0	582.3760
LPC 19:0	582.3782
LPC O-20:0	582.4139
LPC 20:4	588.3300
LPC 20:4	588.3310
LPC 20:3	590.3467

ID	exact.mass
LPC 20:2	592.3622
LPC 20:1	594.3771
LPC 20:0	596.3932
LPI 18:1	597.3040
LPI 18:0	599.3203
LPC 22:6	612.3310
LPC 22:5	614.3459
LPC 22:5	614.3460
LPC 22:4	616.3627
LPC O-24:2	634.4457
LPC O-24:1	636.4618
LPC 24:0	652.4555
PE O-34:3	698.5128
PE O-34:2	700.5289
PE 34:2	714.5074
PE 34:1	716.5232
PE O-36:6	720.4961
PE O-36:5	722.5123
PE O-36:4	724.5276
PE O-36:3	726.5438
PE O-36:2	728.5596
PE O-37:5	736.5270
PE 36:4	738.5073
PE 36:4	738.5077
PE 36:3	740.5228
PE 36:2	742.5383
PE 36:1	744.5549
HexCer 34:1-O2	744.5613
PE O-38:7	746.5131
PE O-38:6	748.5265
PE O-38:6	748.5270
PC 30:0	750.5283
PE O-38:5	750.5428
PE O-38:5	750.5429
PE O-38:5	750.5438
PE O-38:4	752.5590
PE O-38:3	754.5761
PE 38:6	762.5070
PC O-32:1	762.5646
PE 38:5	764.5226
PE 38:5	764.5237
PC O-32:0	764.5797
PE 38:4	766.5381

ID	exact.mass
PE 38:3	768.5548
PE O-40:8	772.5282
PC 32:2	774.5281
PE O-40:7	774.5435
PC 32:1	776.5436
PE O-40:6	776.5593
PC 32:0	778.5603
PE O-40:5	778.5747
PC O-34:3	786.5646
PC 33:2	788.5441
PS 36:1	788.5445
PC O-34:2	788.5807
PC O-34:2	788.5813
PE 40:6	790.5383
PC O-34:1	790.5949
PE 40:5	792.5533
PE 40:4	794.5702
PC 34:4	798.5273
PC 34:3	800.5433
PC 34:3	800.5443
HexCer 38:1-O2	800.6254
PC 34:2	802.5609
PC 34:1	804.5755
PC 34:0	806.5911
PI 32:1	807.5031
PS 38:4	810.5275
PC O-36:5	810.5651
PC 35:4	812.5436
PC O-36:4	812.5806
PC O-36:3	814.5947
PC O-36:3	814.5948
PC O-36:3	814.5961
PC 35:2	816.5751
PC 35:2	816.5756
PC O-36:2	816.6114
PC 36:6	822.5289
PC 36:5	824.5436
PC 36:5	824.5441
PC 36:5	824.5447
PC 36:4	826.5586
PC 36:4	826.5604
PC 36:3	828.5754
PC 36:3	828.5762

ID	exact.mass
HexCer 40:1-O2	828.6566
PC 36:2	830.5914
PC 36:1	832.6069
PI 34:2	833.5187
PC O-38:7	834.5637
PI 34:1	835.5353
PC 37:6	836.5436
PC O-38:6	836.5804
PC O-38:5	838.5955
PC O-38:5	838.5972
PC O-38:4	840.6110
PC O-38:4	840.6113
PC 37:3	842.5912
PC 37:2	844.6072
PC 38:7	848.5437
PC 38:7	848.5442
PC 38:6	850.5593
PC 38:6	850.5598
PC 38:5	852.5759
PC 38:5	852.5763
PC 38:4	854.5904
PC 38:4	854.5911
PC 38:4	854.5914
PC 38:3	856.6069
PI 36:4	857.5181
PC 38:2	858.6219
PI 36:3	859.5335
PI 36:2	861.5503
PI 36:1	863.5644
PC 39:6	864.5749
PC O-40:6	864.6101
PC O-40:5	866.6277
PC O-40:4	868.6436
PC 40:8	874.5613
PC 40:7	876.5743
PC 40:6	878.5897
PC 40:6	878.5912
PC 40:5	880.6063
PC 40:5	880.6070
PI 38:6	881.5185
PC 40:4	882.6227
PC 40:4	882.6230
PI 38:5	883.5335

ID	exact.mass
PI 38:4	885.5505
PI 38:3	887.5651
PC O-42:5	894.6594
PC 42:10	898.5605
PI 40:6	909.5477
PC O-44:6	920.6752
PC O-44:5	922.6912