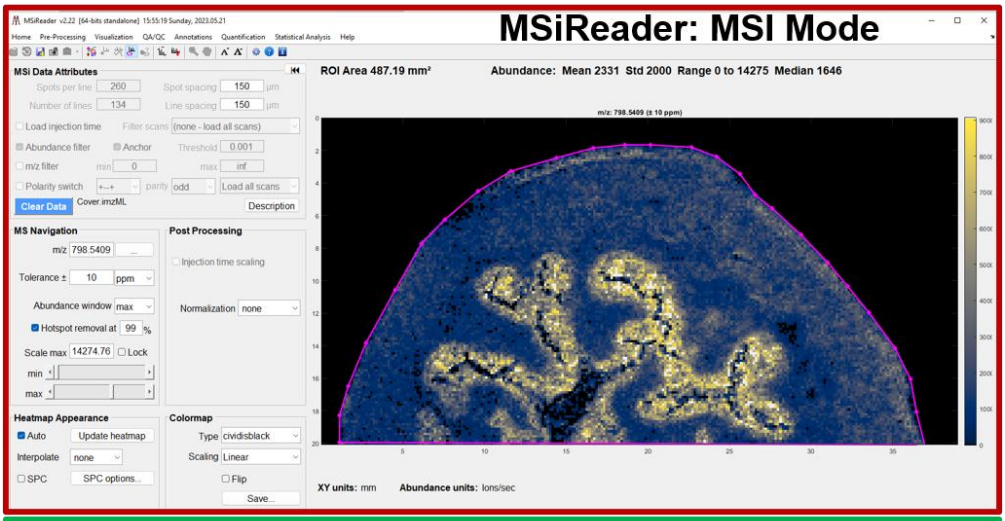
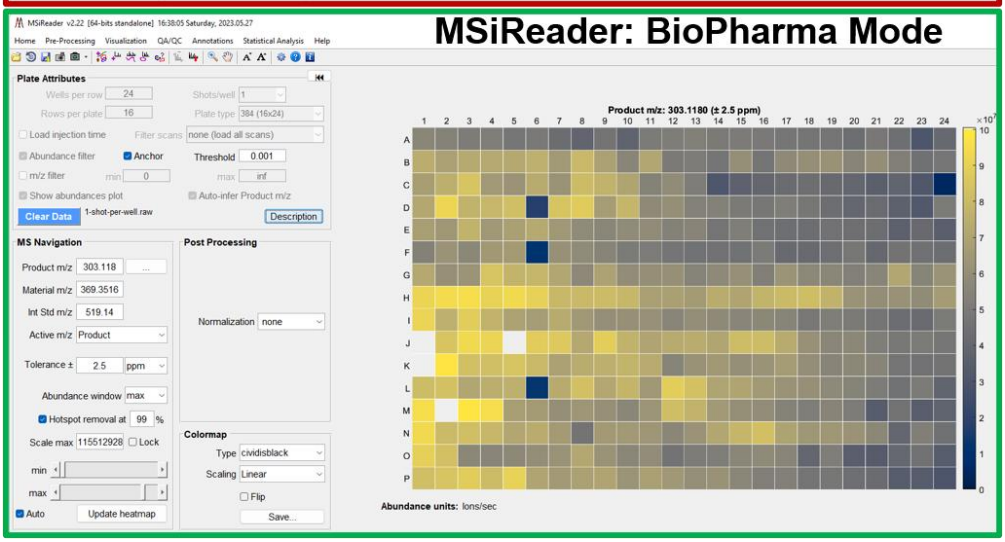
MSiReader v3.03 User Manual:
Freeware, Professional, and BioPharma Versions





MSiReader is for Research Use Only

Not for Use in Diagnostic or Clinical Procedures

Instructions

Please read the user manual in its entirety if you are new to MSiReader v3.03. It will improve your understanding of the software and how to best use it to effectively support your research. Every "§" is a hyperlink to a section within the user manual to assist in navigation. Every *italics* word is a variable in the preferences INI file (**§5**) for which the default can be changed to customize the user's experience. There are ToolTips embedded within the program that are displayed by a mouse over and a brief hover on a GUI object – a description about the tool or data to be entered is provided. The user manual is provided as a PDF file (See "user guide" under the Help pull-down menu in MSiReader) as well as an online PDF and HTML version at

https://msireader.com/docs/MSiReader_UserManual/

Video Tutorials are posted at https://msireader.com/tutorials and **Test Datasets** are posted at https://msireader.com/support

New videos and test data sets are being added regularly so please check the website frequently. MSiReader test datasets are provided so that users can follow the user manual and/or the video tutorials. However, it is highly recommended that you move the test data sets to your local hard drive prior to testing out the software.

IMPORTANT: There are many parts of MSiReader that will allow the user to export/save a new .imzML, a *.mss or a *.mim file (these are MSiReader's custom formats). If the user loads an imzML dataset, they can save it a new imzML file; however, if a user loads a native file (*e.g.*, *.raw file from a Thermo instrument), the user can then save this as a *.mss or a *.mim file. Importantly, once a user does this within MSiReader, these files are no longer usable in third party software. For newly created imzML files within MSiReader, the end-user MUST keep the .ibh file in the same folder as the new .imzML to be read.

Table of Contents

List of Figures

Part I: Getting Started

1 About MSiReader v3.03 Freeware, Professional and BioPharma Versions

Collectively, **Part I:** Getting Started and Freeware (this section), **Part II:** User Guide – Basic Functions, **Part III:** User-Guide – Advanced Functions and **Part IV:** User Guide - BioPharma are included in this user manual, which serves as a comprehensive guide with examples of how to use MSiReader to support your research. **Part I: Getting Started** and Freeware describes computer requirements and installation instructions as well as how you can customize your copy of MSiReader to better suit your specific projects. Moreover, it contains a new mode of MSiReader called Freeware, which are tools that are available to the MSI Community for free even after the trial version expires. **Part II: User Guide – Basic Functions** provides a step-by-step guide for the basic functions in MSiReader; the functionality outlined in the 'User Guide – Basic Functions' are commonly and simple functions. Conversely, the functionality contained within **Part III: User-Guide – Advanced Functions and Part IV: User Guide – BioPharma** are more sophisticated data analysis tools. This portion of the user manual provides the underlying theory and a guide for the advanced functions in MSiReader to support mass spectrometry imaging as well details on the BioPharma mode. Throughout these four sections in this user manual, data files are provided for most of the functions in MSiReader so that the end user can replicate the tool prior to applying it to their own data. The MSiReader data sets can be downloaded **HERE**; in each folder is also a README file to understand how to get started. Please check our website for video tutorials to walk you through different workflows including installation of the software (https://msireader.com/support). The beginning of this user manual also provides the following information: What is MSiReader? Link to the End-User License Agreement (EULA), Copyright Information, Citing MSiReader in your published research and Release Notes for each version. A list of the references (**§9**) cited throughout this manual can be found at the end of the manual.

1.1 What is MSiReader?

MSiReader v3.03 and all accompanying documents are based, in part, on MSiReader v1.03 which was developed from 2012-2022 at North Carolina State University with funding generously provided from NIH (R01GM087964)[1,2]. MSiReader was developed as

a desktop application and, from the outset in 2012, it was designed to be vendor-neutral and able to handle High Resolution Accurate Mass (HRAM) data without data compression, binning or loss of dynamic range, and to serve as an effective tool for building customized solutions **for** the MSI community **with** input from the MSI community. In 2021, MSI Software Solutions, LLC was formed and received the exclusive license from NC State University to transform the platform to better support the entire MSI community. In 2023, a BioPharma mode was added to support high throughput screening (HTS) and high content (phenotypic) screening (HCS). In 2024, we added the Freeware mode to share tools written by others but shared via the MSiReader GUI. These tools are absolutely free.

MSiReader (except Freeware mode) is a subscription-based software program for MSI and BioPharma and is designed with an architecture which focuses on the following key attributes: vendor neutrality, intuitive GUI's, computational efficiency, generation of publication-quality images, while also supporting all MSI ionization methods and HRAM data. The overarching goal of MSiReader is to provide a research-based productivity suite of tools and functions which are constantly evolving to support the needs of the MSI community. Moreover, the new BioPharma mode is offered with a slightly higher cost paid subscription (BioPharma mode includes the MSI Mode) and was written to support HTS/HCS experiments. Additionally, MSiReader is written with an intuitive graphical user interface (GUI) which enables researchers to read and process their MSI and HTS/HCS data in a computationally efficient and effective manner. MSiReader v3.03 is compatible with the common MS file sharing formats and offers a wide range of data analysis features. These features are being improved and expanded upon and will be included in future releases. MSiReader also outputs high-resolution publication-quality figures. A small dataset is provided in the installation folder to check for proper installation. This dataset is of little value for any other purpose. Please do not hesitate to contact us through our website www.msireader.com or by email at support@msireader.com for any suggestions to improve the software, user manual or to report any issues.

1.2 EULA for MSI Software Solutions, LLC

The end-user license agreement (EULA) can be found on our website at (https://msireader.com/docs/MSiReader_License.pdf). The EULA must be agreed to prior to installing the software on your computer.

1.3 Copyright

1.4 Citing MSiReader v3.03

If you use MSiReader in your publications, please cite the software in your experimental section as: MSiReader v3.03 (MSI Software Solutions, LLC, Raleigh, NC). MSiReader v3.03 is based on the peer-reviewed and copyrighted MSiReader platform with the following two key citations.[1,2] For a complete list of publications reporting different functions and/or applications of MSiReader, as well as other authors who have cited our work, please see our website:

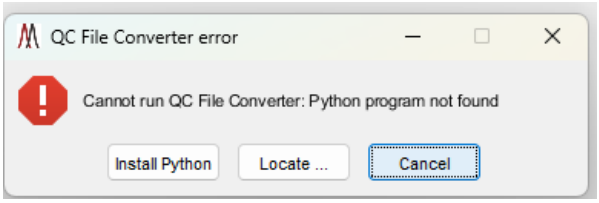
https://msireader.com/publications

1.5 Freeware Version

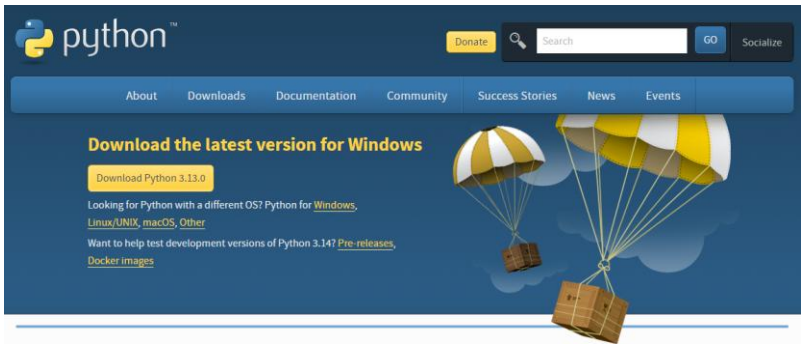
We are now sharing new tools, which are quite advanced, for free using MSiReader as a launch point. These tools are written in a variety of programming languages including MATLAB, Python, and R and are launched from the MSiReader main GUI. The end-user does not need to know how to use these different programming languages, the tools are embedded into the code. However, as we add new tools to the Freeware version, the end-user will need to download and add programs to their computer (*e.g.*, Python) – this is accomplished by MSiReader making it very easy to use these free tools.

The first two tools that are Freeware are the AutoQC file convertor (to get data structure in the proper format) and Supervised Learning for Instrument Classification and Evaluation for Mass Spectrometry Imaging (SLICE-MSI). These two tools require that the end user has Python downloaded onto their computer. The external tools depend on the specific Python version that is installed on the user's computer, so if users see a run-time error related to Python when they run such a tool, then installing the latest Python version might resolve the problem. MSiReader, upon launching either of these two tools, will check to determine if the end-user has Python on their computer, if not, it will provide directions to download this to your computer (as shown below). Importantly, as other Python based tools are added to MSiReader, this process will only require a quick locate step but not require downloading Python.

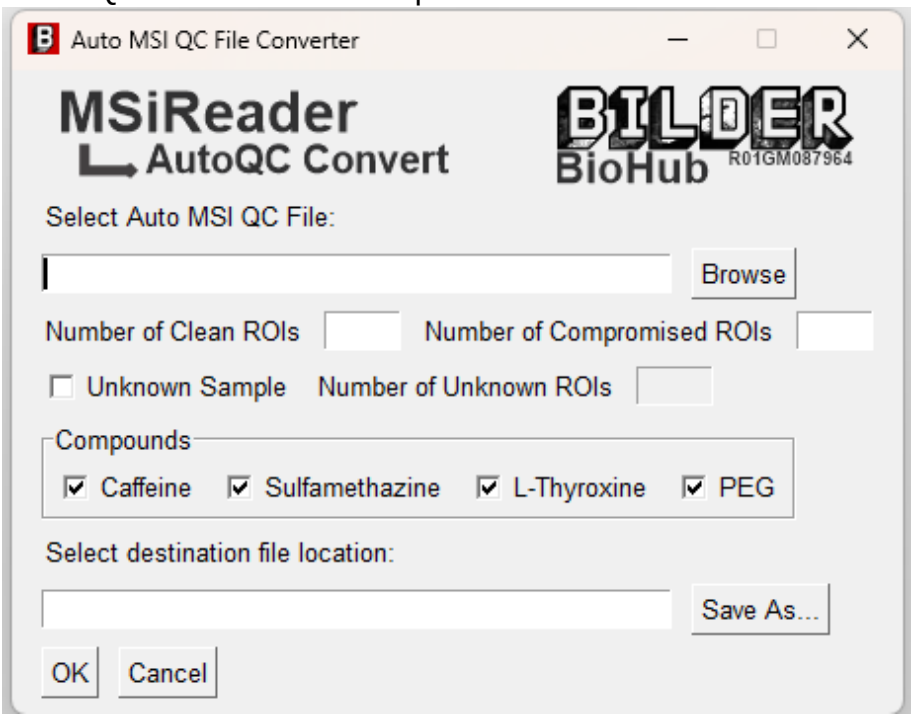
When the Trial version expires these tools will remain available in the Freeware version. Of course, they are also active in MSiReader Pro and BioPharma as well but they do not require a license. The end-user decides that they would like to use these tools, they click on the Main Menu QA/QC and the select either the AutoQC file convertor or the SLICE-



MSI tools. Let's assume that the end user does not have Python on their computer. Please select the AutoQC file converter from the QA/QC menu – the warning subGUI shown above will be displayed. Upon selecting **Install Python**, the user will be taken to the download Python page as shown below.

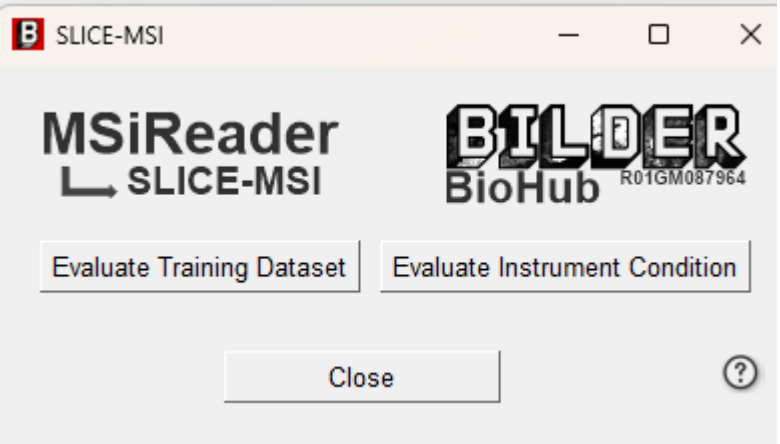


Click on Download Python and then run the executable to install Python on your computer. It should indicate that the setup was successful. Next, go back to MSiReader and re-launch the AutoQC file converter. The subGUI below should be displayed just as before but this time, choose locate. It will open a file explorer box and you will then choose Python(version number) subfolder, then select the Python.exe file then OPEN. The subGUI for the AutoQC file converter will open and is shown below.



The help file for this tool is being written and will be available shortly as a link from the GUI shown above. Note, subsequent uses of this freeware program will not require these steps – it will automatically run.

Next, launch the SLICE MSI tool. MSiReader identifies that Python is on the computer and where it is located. The subGUI should just appear as shown below.



The help file for SLICE-MSI is already available – just click on the **?** and the user manual will open.

We will be adding more Freeware tools in the future and MSiReader will serve as a convenient portal to disseminate them.

2 Computer Requirements and Installation of MSiReader

2.1 Minimum Computer Requirements

MSiReader v3.03 was developed and extensively tested on a WIN 10 and WIN 11 64-bit operating systems. It is highly recommended to have at least 32 GB of RAM but ≥ 64 GB is preferred. We are working on a version of MSiReader that is fully tested on Mac computers.

2.2 Installation and Start-Up

For a **video tutorial** on how to properly install MSiReader, click **HERE**.

To begin the installation process, go to https://msireader.com/compare-plans and scroll down the page to see the 15-day free trial, Professional or BioPharma version. Click the Download button which will then add that product to your cart (15-day free trial will indicate $0 and Professional License since all functions will work during a trial period). **Figure 1** shows what is displayed.

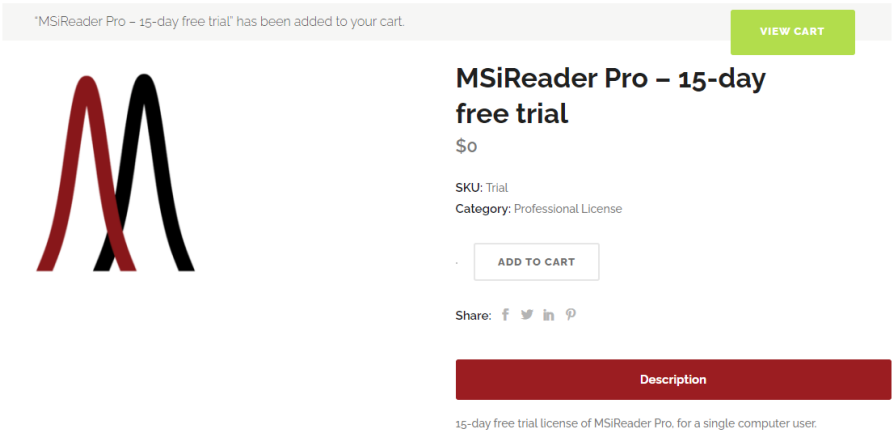


Figure 1: Display after 15-day free trial is selected. The SKU is Trial and the Category is Professional License since all functionality is provided for the trial period. At the end of the trial period, the license automatically expires.

Click the Green Button “View Cart” and verify details on the new page that is displayed. Next, click “Proceed to Checkout” which will require the end-user to enter in a valid name, address and email. The end-user will also have to agree to website terms and conditions. Next click “Place Order”. This will provide you with details about your license, when it expires, etc.

If you are beyond the trial period and are purchasing a license (paid subscription for Pro or BioPharma mode), the process is the same except you will need to choose the product and term that you wish to purchase and add each item after adjusting the quantity to your cart. When you Proceed to Checkout, there will now be a place to pay by credit card, direct wire transfer or pay by check.

In either case outlined above, once you land on the checkout page, look under the downloads section and click on Program Installer which will download the MSiReader installer to your local hard drive. The installer automatically downloads and installs the Matlab Compiler Runtime engine (MCR, approximately 3 GB); this could take ~10 minutes depending on your computer and network speed. This MCR installation is a one-time download because it will rarely need updating. Future installations and upgrades of new MSiReader versions will be much faster and is done by simply clicking “Update this program” under the Home menu.

Double click the MSiReader_install.exe file. A popup box will appear and ask if you allow this program to make changes to your device. MSiReader is a verified publisher from MSI Software Solutions, LLC as noted. Then select Yes and then Next and Next once again. The installer GUI will indicate that the installation is complete – please close this GUI.

IMPORTANT: Install the program in a folder of your choosing but **DO NOT** choose a location that is not local (*e.g.,* DropBox) and **DO NOT** place under the WinOS Programs Folder. The installer will default to creating a MSiReader folder and install the program in that directory. We highly recommend following this suggestion.

After you install the software, go to the folder which contains the program files. Double click on the post-install.bat file to complete the installation process. Once it is complete, this file will be removed from this folder.

After you install the software, go to the folder which contains the program files. Select the MSiReader.exe application file and right click and then select “show more options” and then select “pin to taskbar”. Start the program by clicking once on the ICON that is now in your taskbar. You can also pin the app to the taskbar from the start menu in Win11.

When running the program for the first time, the MCR loads in the background, which can take up to a full minute without any visual feedback so please be patient and don't try to double-click the executable again. All console output of the program is sent to an MSiReader.log file (one log file is created each day) so any errors/messages will automatically be recorded. Subsequent launches of MSiReader will be much faster since the MCR has been initialized.

The end-user can edit the MSiReaderPrefs.INI file which allows the end-user to customize the user interface; default settings are provided in the MSiReaderPrefs.INI file. See **§5** of the user manual for details.

Minimally, you will probably want to change the default Pathname for file Open and Save dialogs. Again, **DO NOT** choose a location that is not local (such as a networked or server location) as this will significantly decrease the performance of MSiReader.

A folder of data files that are used in this manual can be found on our website (www.msireader.com/support). This folder is called *MSiReader Test Data* and contains 18 subfolders for different functions in MSiReader. In each subfolder, a file called *README.txt* is present and contains specific information that you will help you get started with that specific dataset.

Successful Installation Check: "Example file.imzML" is located in the application subfolder called Example File. This is a computer-generated file used only for testing installation. Click '*Load Data*' in the MSi Data Attributes pane, select the Example file.imzML file by double-clicking on its filename, and then type in *m/z* = 369.3516 with ± 2.5 ppm in the MS Navigation pane and slightly adjust the minimal abundance using slider bar (please note that the *m/z* and tolerance fields should already be populated with these values). The default colormap is cividisblack which is color vision deficiency (CVD)-compliant[3,4]; the "black" is used for data below a user-defined threshold. The threshold is (max abundance - min abundance) / N, where N is the number of colors in the colormap. These

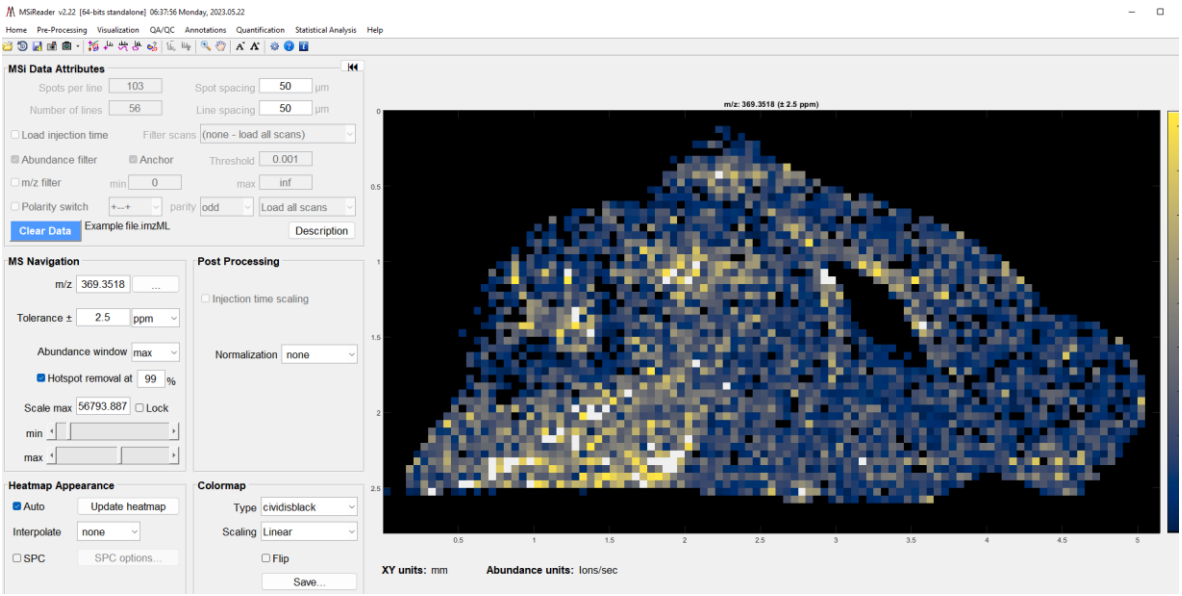


Figure 2: Loading of test data "Example.imzML" upon installation of MSiReader.

abundances are global for all scans loaded. Upon following these steps, you should see the GUI and heatmap as shown in **Figure 2** which indicates your installation was successful.

2.3 Status Bar and STOP Button

The MSiReader v3.03 main GUI includes a status bar for reporting the progress of loading datasets and a STOP button for early termination on the lower middle. **Figure 3** shows the MSiReader status bar updating while a file is being loaded. The status bar is always from 0% to 100% with a text overlay in the center which provides further information about the current operation. In this example, spectrum 6796 out of 17632 spectra in this data set have been being loaded. Note that for HTS/HCS experiments from well-plate data, file loading is near instantaneous and thus, the end user will likely only see the status/progress bar flash on the screen and then disappear.

Figure 3: Status/progress bar when loading data into MSiReader.

Clicking the **STOP** button will launch a confirmation dialog box as shown in **Figure 4**. In this case, the user can cancel loading the data file or continue.

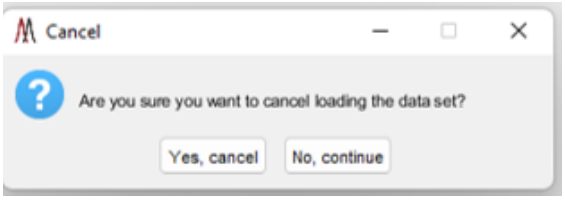


Figure 4: After clicking the red STOP button when loading data (see **Figure 3**) the user will be prompted with the dialog box shown here. The user can either select Yes to Cancel loading the data or No to continue loading the data.

2.4 Processing Data with MSiReader and Suggestions for Reducing File Sizes

Note: The following sections are directly related to the MSI Mode of MSiReader as HTS/HCS file sizes are typically very small; thus, to date, the BioPharma Mode has not experienced any issues with exceeding available RAM.

Predicting the maximum size data set MSiReader can process is not trivial since it depends on the amount of available RAM, which depends on the number of applications running on the computer. There needs to be enough RAM to store the data set, and sufficient memory left for additional processing operations. Some of these may require the creation and temporary storage of large variables. An example memory usage calculation follows.

Using a Windows 11 Pro computer with 32 GB of RAM, MATLAB R2022a consumes about 700 MB. Windows OS uses about 2 GB and MSiReader requires approximately 150 MB before a data set is loaded. Thus, it follows that ~29 GB is therefore available to MSiReader for loading, storing and analyzing MSI data. A single data point in a scan (*m/z* value and abundance) requires 16 bytes of memory, so for example, if a region of interest (ROI) of 100 lines and 100 columns is scanned and each generated spectrum contains an average of 10,000 *m/z* data points, the memory requirement can be calculated as shown in **Figure 5**.

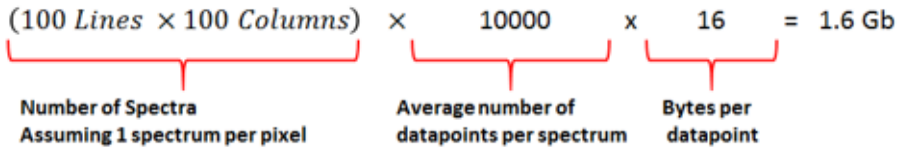


Figure 5: Basic calculation of total file size based on ROI, number of datapoints per spectrum and bytes per datapoint.

MSi PeakFinder (**§7.7.1**) over a large ROI (*e.g.,* 2500 scans) uses about 1 additional GB. Total memory used is then 2.6 GB. With 29 GB of RAM available to MSiReader, a data set approximately 10 times as large as this example can be processed. It is common for

MSiReader to work with files that contain more than 10,000 spectra and where the number of data points varies from 20,000 to 100,000 for each scan (processed raw files converted to imzML format). For example, the file used to demonstrate the MSi Quantification tool (**§7.7.2**) contains 7426 scans and almost 250M non-zero data points. The total amount of memory used by Windows, MATLAB, and MSiReader after loading this data set is 7.1 GB. Note that a desktop PC with 32 GB of RAM may be purchased for less than $1000. We highly recommend purchasing a desktop PC with at least 64 GB of RAM with the ability to add more in case your applications evolve and processing of larger and/or multiple files is demanded.

This example also highlights the advantage of using processed data where zero abundance data values have been eliminated. See the discussion about imzML files in **§3.1**. 10,000 data points/spectrum corresponds to an *m/z* data step size of 0.1 Th (Thompson units) over an *m/z* range of 1000 Th in the case of continuous (evenly spaced) data points on the *m/z* scale. If data is processed and zeros have been eliminated, 10^{-6} Th step sizes can be stored with 10,000 data points/spectrum over the same *m/z* range.

In case the data file is too large to be handled by your computer, here are some potential solutions:

2.4.1 File Conversion

Use a converter like the MSConvert tool from Proteowizard[5,6] to convert your original instrument raw file or another file sharing format (mzXML, mzML) into multiple files with smaller *m/z* ranges (*e.g.,* instead of having one single file for 200-2000 *m/z*, have 4 files covering 450 *m/z* each). Also see **§2.4.1** for more information about file conversion. Although this approach will not allow visualizing all the data at once, it will not impact the mass resolution of each image file.

2.4.2 m/z range filter

Use the *m/z* range filter to reduce the spectrum saved for each scan as it's read into MSiReader's internal data set storage. See **§6.2.1** for how to use the *m/z* range filter. This only applies to the MSI mode.

2.4.3 ROI location file

A region of interest (ROI) location file can be used to read a portion of a large file. See **§3.1** and **§6.2.1** for details. Note that this is more effective for reducing RAM requirements for imzML files than for mzXML files; therefore, it is advantageous to convert your files to imzML format.

2.4.4 Abundance threshold filter

Use the abundance threshold filter to reduce the number of low abundance peaks that are saved when the file is being loaded. Files from some instrument vendors may contain an extremely large number of zeros or nearly zero abundances (up to 95% of the total data points!). The abundance filter should always be used with an appropriate threshold to eliminate these zero values. In the MSi Data Attributes (MSI mode) or the Plate Attributes pane (BioPharma Mode), **BEFORE** you load your file, check the box for Abundance Filter and then enter in a value that is appropriate for your data in the Threshold Box as shown in **Figure 6** for the MSI Mode.

Checking “Anchor” within the MSI Data Attributes (or Plate Attributes Pane) dialog box causes consecutive runs of abundance values below the filter threshold to be replaced by two zero values, one at each end of the *m/z* range being filtered out, instead of being eliminated. This forces spectral line plots to be “anchored” to the *m/z* axis between peaks. However, note that using this filter increases loading time for most data sets.

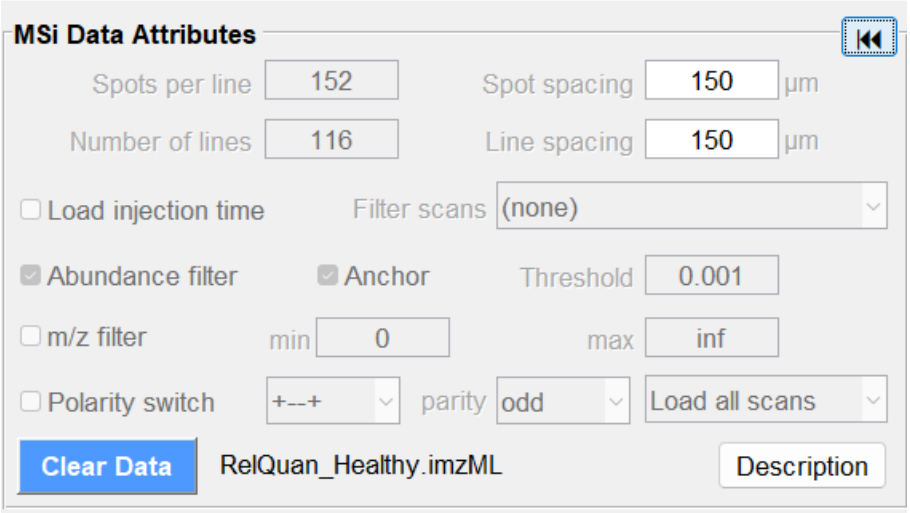


Figure 6: Thresholding data can be done to avoid loading zeros (or nearly zero values) in your dataset to reduce RAM requirements and improve data analysis performance in MSiReader.

The option to anchor filtered peaks to the baseline is unnecessary when loading centroid data sets. To improve performance, this option is ignored if any of the scans have been centroided. This only applies to imzML data files and can be disabled by setting the preferences INI file variable, *NoFilteredPeakAnchorCentroid*, to false (§5).

With all file formats except *.mss, the mass spectra may contain many very low abundance data points. These *m/z* (x) abundance (y) pairs can be filtered out of the data as the file is loaded to conserve memory space and to enhance performance. The threshold value is the smallest abundance value that is retained for any data pair in a scan. The default value is 0.001 and this can be changed with the *MinAbundanceThreshold* value in the preferences INI file (§5). You can change it to any value between zero and 1e6. Threshold filtering is most effective for continuous data (IMG

format), imzML files and native *.raw files. Due to the structure of the parser, it is less useful at reducing memory usage for mzXML files; however, after a data set is loaded, processing and heatmap updates will be faster.

2.4.5 Centroiding Data

Centroid your data during conversion to imzML files or collect your data natively in centroid mode.[7] These are mentioned here as alternatives to collecting your data natively or carrying this out during file conversion using second party algorithms.

During conversion of your profile data, you can significantly reduce the file size by centroiding your data during conversion. Alternatively, you could collect your MSI data in centroid mode on your mass spectrometry imaging platform. In either case, MSI Software Solutions highly recommends that you carefully check either route to make sure that this does not “compromise” your data in a way that is unacceptable to you. For example, collect datasets natively in both profile and centroid mode on the same sample and then check to make sure that they provide you with the same results including such attributes as mass measurement accuracy, spectral accuracy, # of annotations.[7] These are your best options to reduce file size and maximize performance of MSiReader for any given computer configuration.

As an example, a dataset was collected using the same ROI area in both profile mode and then in centroid mode on a MSI platform. The physical memory in the laptop was 32 GB of which 25 GB was available before opening up MSiReader; opening MSiReader reduced the RAM available to 24 GB. Importantly, loading the profile data (*.ibd file ~ 3 GB) and the natively collected centroided data (*.ibd file ~ 250 MB), the amount of available RAM increased from 20 GB (profile) to 23 GB (centroid) as expected. Conversely, for HTS/HCS 384 well-plate data, the file size is < 100 MB.

MSiReader can take your profile data and centroid it for you – this will reduce the file size and therefore reduce the amount of RAM required. This feature was added to enable some tools to be used in MSiReader that require centroided data but can also be used to reduce file size to the data in memory or using the batch mode. Under the Main Menu item “Pre-Processing” select “Centroid Data” function as shown in **Figure 7**. You can select from three different centroid algorithms (if you are using this tool on data that is already centroided, you must choose “Local Maximum” as the Centroid algorithm), set an abundance threshold, turn on or off the peak exclusion filter (peak exclusion must be carried out on centroided data – in this case, it will centroid your data and then apply the exclusion filter) and set your peak tolerance in this panel. Then select OK. If you check

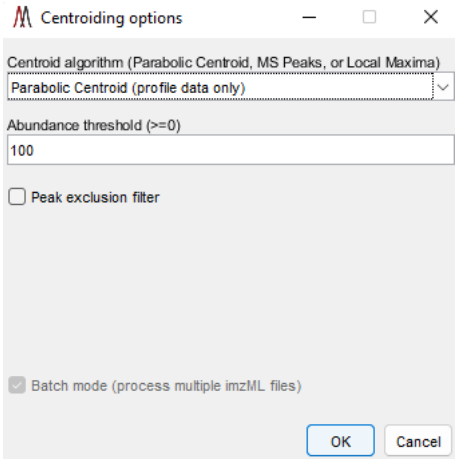


Figure 7: Centroid Data Options Panel in MSiReader

the peak exclusion filter, you will be prompted for a list of *m/z* values (.txt file). If the clipboard contains a positive number within the *m/z* range of the loaded data set you will be asked if you want to use those values as the exclusion list. If it contains other content or you decline you will be prompted to select a .txt or .xlsx file with the *m/z* values that you wish to exclude. In the case of selecting a .xlsx file with more than one worksheet you will be prompted to select a specific worksheet. For both file types, the first column of values will be used.

The exclusion list could be background ions that are present in high abundance in every spectrum that will be removed from the spectra and heatmap or a series of MALDI matrix peaks for MSI data files. In MSI Mode, this could also be a list of peaks that were

determined to be background ions versus on tissue. In BioPharma Mode, this exclusion list could be excess drug as one example that a user would like to remove prior to using MSiExport and use of multivariate analysis. If you set peak exclusion as “false” (by not having the check box checked), it will centroid your data using the other parameters you have selected in the options panel (**Figure 7**).

Upon centroiding your data in memory, you will be prompted to save a new imzML file in the same folder – MSiReader will add the extension _centroided to the original filename but the user can enter in any filename they choose prior to saving. For batch mode centroiding of data, it will aways add _centroided to each filename automatically and save them in the same folder as the original data.

IMPORTANT: Centroiding data may produce unexpected results if the input file is not an actual mass spectrum but a peak list (preprocessed centroid data). All data preprocessing steps (whether MSiReader or other software) should be validated in your workflow prior to applying them to your data to ensure artefacts are not introduced.

If you check Batch mode **Figure 7** and then OK, a file explorer box will open and then you can chose a folder and then select one, several or all .imzML files that the user wants to centroid. This process is carried out in the background. As an example, the .ibd file size for the profile data (in Mass Correction Folder) was ~1.2 GB but after the centroid algorithm was applied, the file size dropped to ~89 MB.

You can simultaneously do abundance thresholding which will further reduce the RAM required (**§2.4.4**). Moreover, prior to (or after using Scan scrubber tool) centroiding your entire dataset that has been loaded, you can use the ROI selection icon for a polygon and after you select the data of interest, then go to Pre-Processing and then Centroid Data and it will prompt you to select “ROI Scans” or “All Scans”. Once you draw your polygon for the data of interest, if you want the polygon to be a square, right click on the heatmap and select “Make ROI a Rectangle”. You can click on the square and move it around to position it over the ROI of your choice.

There are 3 options to centroiding as shown in **Figure 8** which include Local Maxima, Parabolic Centroid and MS Peaks (wavelet transform – not shown). Only use Local Maxima for data that was previously centroided data in the case where you wish to apply a threshold and/or peak exclusion filter. This is recommended because Local Maximum, when applied to profile data using most software, will likely compromise your mass measurement accuracy and ion abundance. This is of course the fastest of the three centroiding algorithm; however, be cautious centroiding data using this approach.

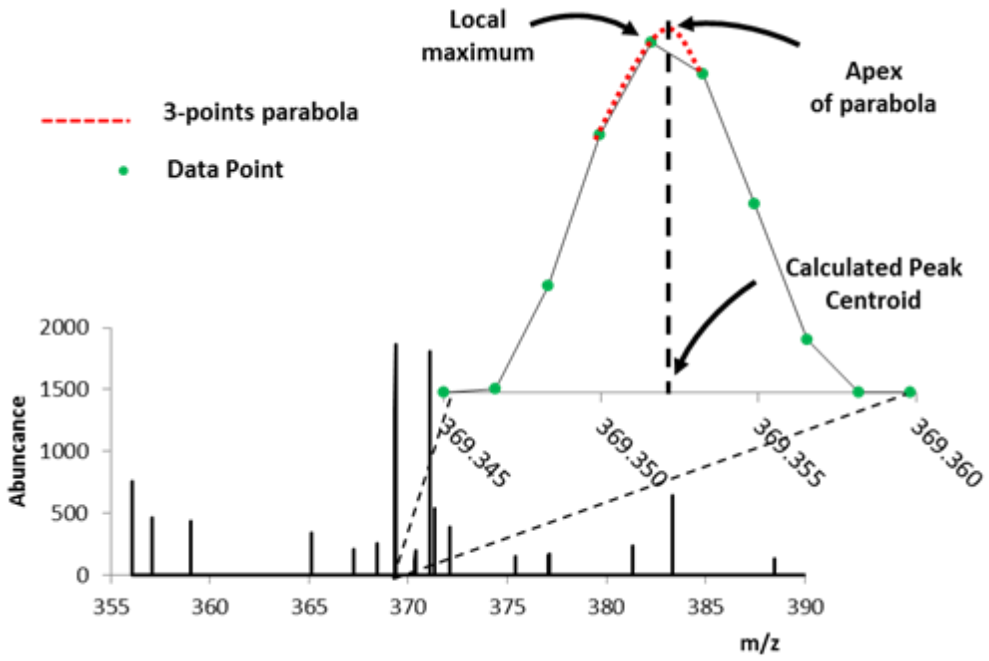


Figure 8: Illustration of centroid algorithms / calculations showing local maximum and parabolic functions.

MS Peaks uses a wavelet transform and filter to find peaks and is similar to the CWT algorithm for peak picking in MSConvert[5]. This algorithm finds peaks in a noisy signal by smoothing the data using a wavelet transform (Daubechies filter banks), putative peak locations are determined and then post-filtering to reduce over segmented and noisy peaks. This approach to centroiding will likely increase computational time significantly.

Important Note: Once the user has centroided their data in MSiReader, the modified imzML file can only be read by MSiReader due to our proprietary parsing algorithm and padding to reduce file size. Regardless, the .ibh, .imzML and .ibd files must all be present to open the file for analysis.

2.4.6 MSiReader imzML Parsing Tool

Use the MSiReader imzML parsing tool. Load each of the imzML files for your project into MSiReader and then clear them and then re-load them. The **first** time you load an imzML file into MSiReader, it makes use of our own parsing tool to retrieve necessary metadata for MSiReader from the file. During this process MSiReader will create an .ibh file. The **second and all subsequent times** you load the imzML for a given dataset, MSiReader looks for the .ibh file and the .ibd (mass spectrometry binary data file) and loads those files. Please do retain the imzML file in the same folder. This process is done automatically, and you will notice when you load the same datafile, that load times will be significantly faster. For example, the mouse bladder MSI data shown on the front cover of the user manual, the imzML file is 55 MB while the .ibh file is only 500 kB. **Note:** this is done automatically for all imzML datasets loaded into MSiReader. If you don't want to use our parsing tool, set the preferences file variable UseFastimzMLFileLoader to false and reload preferences or restart MSiReader (**§5**). No .ibh file will be created and file loading will be slower. Please report any errors to the developers by sending a screen capture of the error dialog and/or the MSiReader log file (in application folder, logs subfolder) to support@msireader.com. You can also validate the format of your imzML file using the imzMLValidator tool. This tool may be able to identify and correct an error in the file due to errors in the conversion process.

3 Supported Data File Formats

Several different data formats can be read with MSiReader: an imzML file (single file and multi-select), an mzXML file (single file and multi-select), an IMG (Analyze7.5) file, a folder of ASCII files (one per scan), a MSI saved session file (*.mss) which is a previously saved MSiReader session and a *.mim file custom format to MSiReader. Finally, loading *.mss, *.mim, or loading native instrument files is now available and steps to load Thermo *.raw files directly can be found in **§7.1** and **§8.2** (these all require a paid subscription). Steps to load each type are described below.

3.1 imzML File Format

This file format was developed by the EU project Computis[8] to efficiently store and share mass spectrometry imaging data. It is the preferred format for MSiReader since it contains the required meta-information, it is compact, and it loads faster than the other formats. An imzML data set consists of two files, one XML file (imzML extension) that contains the metadata such as the instrument parameters and the scan pattern information and a binary file (.ibd extension) that contains the mass spectrometry data. There are multiple converters that can translate directly from a vendor's format to imzML. Users can also use MSConvert from Proteowizard to convert vendor's format to HUPO-PSI's mzML and then use imzMLConverter to convert the mzML file to imzML format. Recently a user-friendly validation and editing tool, the imzMLValidator[9,10] has become available to ensure that converted files conform to the standard. Please use the imzMLValidator if you have problems loading your data. It includes tools to repair incorrect metadata and to add missing metadata. imzML files can be generated in processed (one *m/z* array per scan) or continuous mode (one *m/z* array per file). Both continuous and processed formats can be loaded into MSiReader, but the user must be careful to choose the appropriate format for the type of instrument used when converting a file to imzML.

For the imzML format, datasets typically include the *Spots per Line*, *Number of Lines*, *Spot Spacing* and *Line Spacing* parameters. *Spots per Line* and *Number of Lines* must be present. If *Spot Spacing* or *Line Spacing* are not in the file and cannot be determined from other imzML accession values, the current values in the MSi Data Attributes pane

are used. After the file is read, these can be modified by the user manually to obtain the correct plot scaling and aspect ratio.

Note 1: Alternate regular expression based imzML file parser. If there are errors reading an imzML file, you can try an alternative regular expression-based parser included with MSiReader to search for the required header information. But first, you should use the imzMLValidator discussed above to repair any problems with your file. To enable the alternate expression parser, set the MSiReader preferences INI file (§5) variable *imzMLRegexpParams* value to *true*. If you still have problems loading it with MSiReader please send an error report with a screenshot of any error dialog messages to support@msireader.com.

For an ROI of any shape (polygonal, rectangular or an arbitrary arrangement of scans), a text file (.txt extension) describing the position of each scan can be used to load an imzML file. The .txt location file must be in the same folder as the imzML file. This feature is enabled under *Filter Scans* pull down menu in the MSi Data Attributes pane (**Figure 9**).

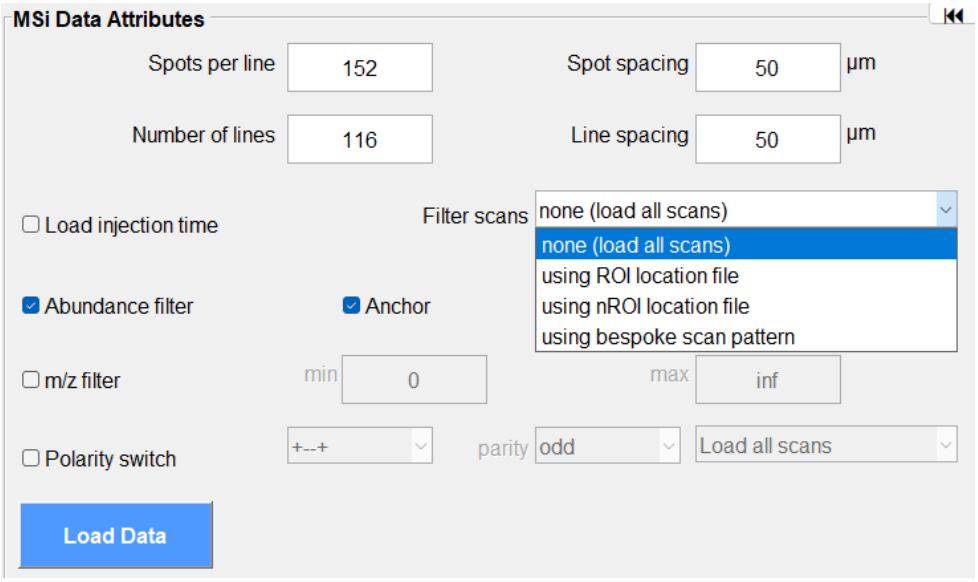


Figure 9: How to load a dataset using a locator file. For example, data collected using an arbitrary ROI. This feature filters the scans of interest using the .txt locator file that must be present in the same folder as the imzML file.

Select under *Filter Scans* “using ROI location file” or “using nROI location file”. More information is discussed below in Notes 3-6. This feature is also useful for reading a portion of a very large file that might otherwise exhaust the memory of your computer.

Note 2: Alternate text based imzML parser in MSiReader v3.03; this is a text-based imzML parser that can be used instead of the Java parser distributed with the imzMLConverter. The first time a file is loaded the parser reads and parses the imzML data set and creates a small file that it saves in the same folder as the imzML and .ibd files, using the same name. The new file has the extension .ibh. When the data set is loaded again, MSiReader loads the .ibh file and then the scan data without reading or parsing the imzML file. The new parser can be significantly faster than the imzMLConverter: about 20% faster for the first load and as much as 70% faster for files with a large number of scans after the .ibh file is created. It is enabled by setting the preferences INI file (**§5**) variable, *UseFastimzMLFileLoader* to “true”. The imzMLConverter can be used if you prefer or if you have problems loading a data set.

Note 3: Location files for an ROI of any shape. The format of an ROI location file is similar to the one used by the imzMLConverter and is presented in **Figure 10** in its simplest form. Each line of the file contains the X, Y location of a scan in the image file. By default, the smallest enclosing rectangular image (5 × 5 for this example) will be generated with an

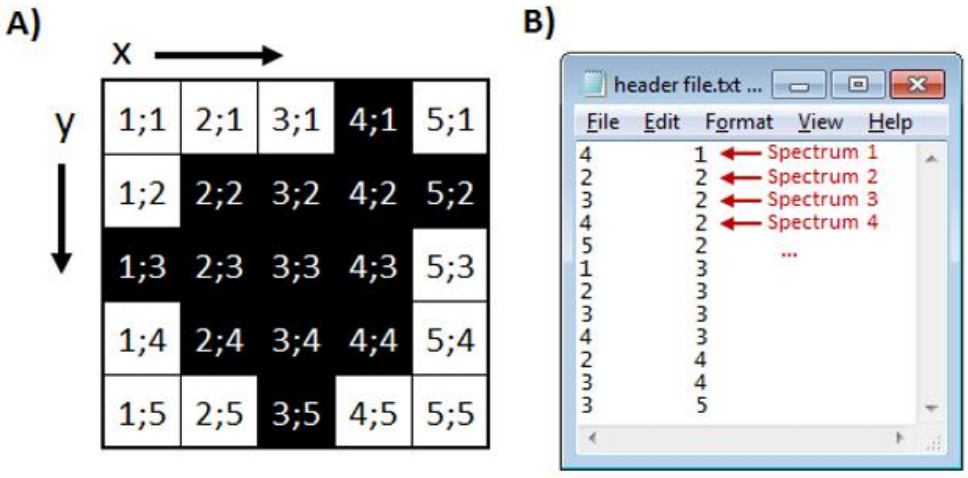


Figure 10: Location file format for non-rectangular ROI. A) Grid showing x and y of location where spectra were collected (black) and B) location file associating the location where each spectrum is located on the grid.

empty scan placeholder for unloaded scans. The preferences INI file (§5) variable *SqueezeROIEmptyScans* is a two-element logical vector that is used to expand the column and row dimensions to their original size in the image and fill with empty scans or to squeeze completely empty columns and rows from the image. For the example, setting *SqueezeROIEmptyScans* to "true false" will reduce the number of columns to 5, but will not reduce the number of rows in the loaded image. Empty columns and rows are removed according to these parameters regardless of how they became empty (ROI filter, *m/z* range filter, abundance filter, bespoke user scan filter, or polarity filter). If you prefer to only remove completely empty columns and rows from the border, preserving the spacing in the interior of the ROI, set the preferences INI file (§5) variable, *SqueezeROIBorderScansOnly* to "true". Your image will be cropped to the smallest enclosing rectangle.

Prior to loading an imzML file, if the "using ROI location file" was selected, the user will be prompted (see **Figure 11**) to select a text file containing the ROI information for that dataset.

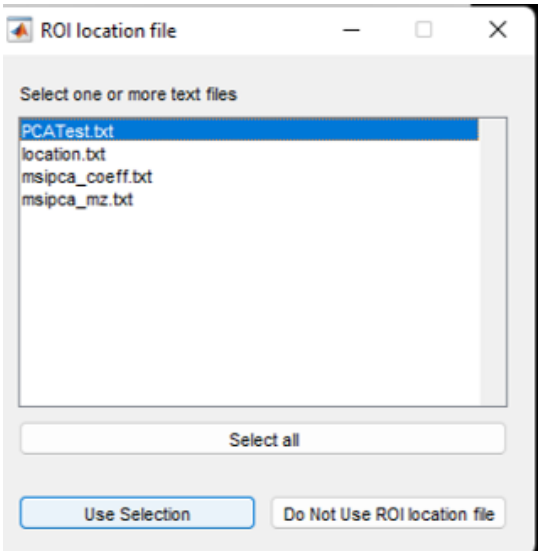


Figure 11: ROI location file selection dialog.

Note 4: Alternate location file formats. If the location file contains 3 or more columns of numeric values, the first column is taken to be the scan number and columns 2 and 3 are the x and y locations of that scan. If a 4th column is present, it is read and ignored unless its value is less than or equal to zero in which case that scan is skipped. This can be

useful for loading a portion of an ROI or selectively excluding individual scans without creating a new location text file. If a file containing a single column of values is selected then those values are interpreted as scan numbers and the x and y locations are calculated using the Spots per line value (*i.e.,* number of columns) as

$$x = \left[\frac{scan\ number}{Spots\ per\ line}\right] * Spots\ per\ line + 1$$

$$y = \left[\frac{scan\ number}{Spots\ per\ line}\right] + 1$$

Note 5: Loading an image ROI saved with the MSiReader Export ROI Location info. The file format generated by the Export ROI tool can be used as a location file as described above in Note 3. Load your imzML file into MSiReader and then using the pixel, line or polygon ROI tool using the appropriate ICON (**§4.2**), select the ROI. Next, right click to access the context menu (**Figure 12**) in MSiReader and select "*Export ROI Location Info*"

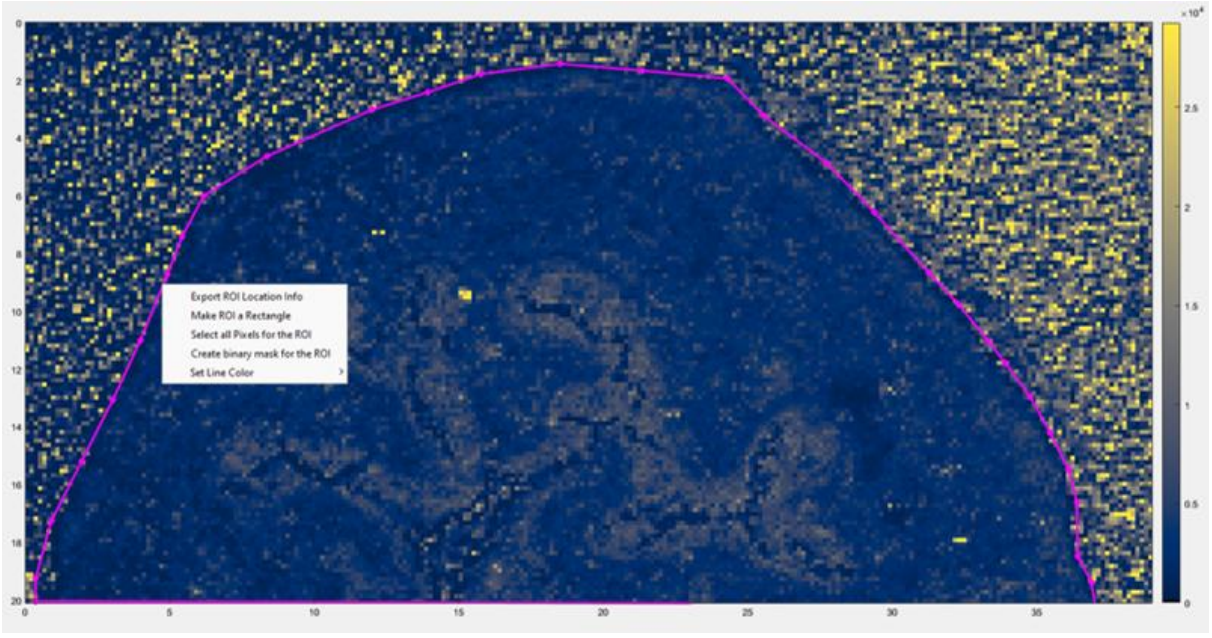


Figure 12: How to access the Export ROI Location Tool in MSiReader. It is a context menu (right click on the heatmap) after you have selected an ROI using the polygon selection tool.

(this is for the polygon tool as shown with the purple line). You will be prompted to enter a filename, and it should default to the same folder where your imzML data is stored.

The structure of the ROI location file you just created is shown in **Figure 13**. The first column is the scan number, the second and third columns are the x, y locations and the fourth column is read and ignored unless its value is less than or equal to zero in which case that scan is skipped.

Figure 13: Table showing the structure of the location file generated using the Export ROI Location Tool. Notice that in the fourth column, the values are all -1 which means those scans will be skipped when loading the data with the ROI location file. Inspection of the ROI in **Figure 12** will show that starting in the top lefthand corner, the ROI that has been drawn (purple line) excludes these data.

You can now reload the original dataset but use the location file to only include data in the ROI that you just created (shown in **Figure 14**). To do this, you select *"*using ROI location file*"* in the MSi Data Attributes pane using the pull-down menu "filter scans" and then load the imzML file. It will prompt the user to select the location file (**Figure 11**). Select the location file and the click "Use Selection" and it will load only data that was chosen in the location file. In the folder with the data is a text file "ROI Location.txt" that can be used to reproduce the data in **Figure 14**.

Note 6: nROI is a feature to visualize data that was collected with two different spatial resolutions. This will take the lower resolution and fill in a 2 by 2 pixel with one pixel using the same data and retain the XY dimensions of your data while the higher spatial resolution data will be displayed natively. This feature is used when you have a large ROI and want to image the entire tissue but with some regions within the ROI (called the nested ROI or nROI) collected at a high spatial resolution.

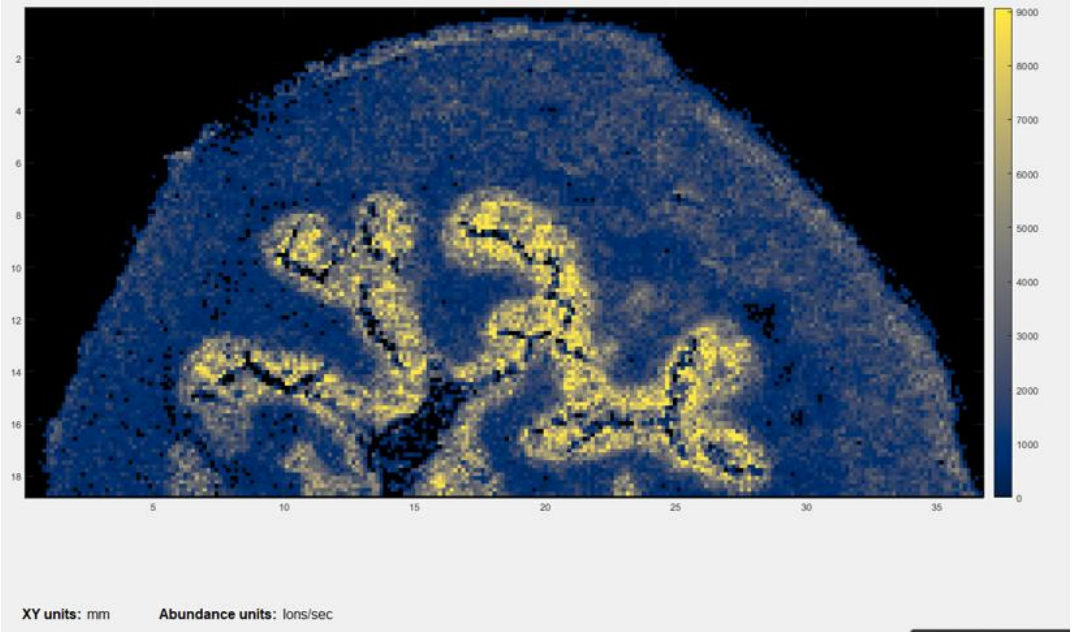


Figure 14: Reloading data shown in **Figure 12** using the location file shown in **Figure 13**.

3.2 mzXML File Format

mzXML is an open representation for MS data introduced by Pedrioli *et al*.[6] An mzXML file can be obtained from RAW data using the MSConvert tool from Proteowizard.[5] The mzXML format does contain most of the information about the instrument parameters but does not contain the spatial information necessary to associate each spectrum to a scan location on the image. If you are using the mzXML format for a rectangular ROI in fly back mode (left to right, top-down) image, you need to manually enter the number of spots per

line and the number of lines per row **BEFORE** loading the file. An alternative is to use the text format described in Note 7 below to automatically load those parameters.

For an ROI of any shape, a text file describing the position of each scan can be used to load mzXML files. This feature is enabled prior to loading the mzXML file type in the MSi Data Attributes pane by selecting using ROI location file in the filter scans pull down menu. If this option is checked, the user will be asked to select a file as described in Notes 3, 4 and 5 above.

Note 7: Header file for Rectangular ROI in Fly back mode. If you have developed your own imaging source control software that automatically saves a text file that contains the four spatial parameters mentioned above in a format similar to the MSiReader preferences INI file (§5), MSiReader will look for this file when loading an mzXML file (or folder of mzXML files). Instead of manually entering the values, you can create a text file containing the following four lines in the same folder as the mzXML file(s) you are reading. See example in **Figure 15**. If there is more than one text file present in that folder, you will be asked to choose the correct one (see example that was shown in **Figure 11**). If no suitable .txt file is selected, the values entered on the main GUI of MSiReader will be used (which you can see after loading the mzXML file).

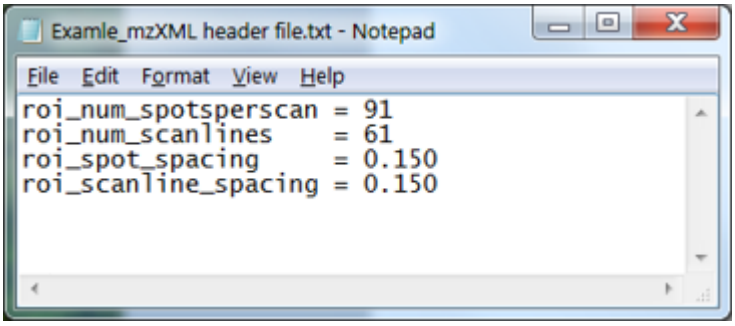


Figure 15: Example of optional header file for rectangular ROI (fly back mode). Spot spacing and scan line spacing parameter units are in mm.

If the mzXML file does not contain the correct number of scans (*i.e.,* Spots per Line * Number of Lines), the user is given the option to read the file anyway discarding extra scans or padding the image with empty scans or to select any of the possible rectangular dimensions that will load all of the scans from a list. See **Figure 16** and **Figure 17** below.

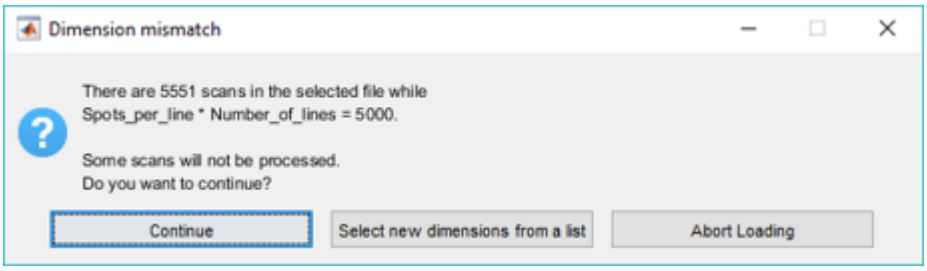


Figure 16: Dimension mismatch dialog box for an mzXML data set.

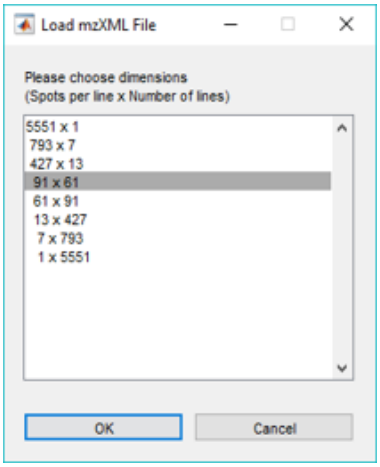


Figure 17: Dimension selection dialog for an mzXML data set. Based on the total number of scans, this dialog box suggests logical data structures.

3.3 IMG File Format

IMG or Analyze 7.5 is a file format originally developed by the Mayo Clinic to share MRI imaging data. It consists of 2 files: a binary file (.img) that contains the actual data and a header file (.hdr) with content details about the image such as the size and position of scans (spot spacing, line spacing, etc.). A third file (.t2m) containing the corresponding *m/z* array was added by Stoeckli et al. to adapt the Analyze 7.5 format for MSI data when developing Biomap. All three of these files are required by MSiReader. If the images are inverted compared to what you expect, you can change the *IMGFlipRows* value to *false*

in the preferences INI file (See §5). As with the imzML and mzXML formats you can also change the Spot Spacing and Line Spacing parameters after the file is loaded, using a negative value to flip the image along the corresponding dimension.

Note 8: Filtering low abundance data points. IMG format files are continuous (same *m/z* values for every scan) and may contain many spectral pairs with zero or very low abundance. If your data set is large, you should enable the abundance threshold filter to reduce the memory requirements and speed up data processing (§2.4.4).

Note 9: File buffering. The performance of scan loading is greatly improved if a group of scans is read at once into a buffer and then each scan is extracted, processed and saved. However, for some files memory may be exhausted during file loading by a large number of low abundance values that will eventually be filtered out. Two preferences INI variables (§5) are used to control buffered IMG file loading. *IMGLoadBufferSize* is the amount of memory in megabytes reserved for the buffer. The default is 40. For files with a large number of *m/z* values, a buffer of this size may only hold a few scans. In this case, it is more efficient to revert to reading one scan at a time. The *IMGBufferScanThreshold* variable is the number of scans per buffer below which scan-by-scan loading is used. The default value is 20.

Note 10: ROI Location Files. If an ROI location file is being used to load a portion of an IMG data set, the ROI must have been saved with the same *IMGFlipRows* setting that is being used to load the data. Otherwise, the wrong scans will be loaded.

3.4 imzML Folder

A collection of imzML files can also be loaded into MSiReader (you can also use the CTRL button to select a subset of the files in the folder); all files must be placed in a single folder and named sequentially as anyname1.imzML, anyname2.imzML, anyname3.imzML, etc. If the file names do not end with a numeric suffix they will be read in alphabetic order. The arrangement of the files into a tiled image mosaic is row-major order (*i.e.,* row 1 is filled left to right, then row 2, etc.).

The user is prompted to enter the number of columns and rows of tiles in the mosaic with a dialog as shown in **Figure 18**. The default values are for the smallest rectangular arrangement that can include all the files. If there are more files in the folder than the product of the columns and rows in the tiling pattern some of the files will not be read. If there are fewer files, the last row will be filled with empty tiles. Extra, completely empty rows are not created. There is also a checkbox called *inter-tile border*. Selecting this option will insert a white border around each loaded dataset (see **Figure 74** for an example) which is greatly preferred when viewing multiple datasets and when generating publication quality figures.

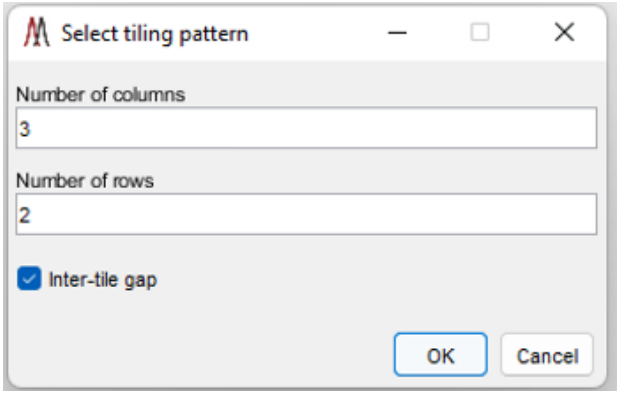


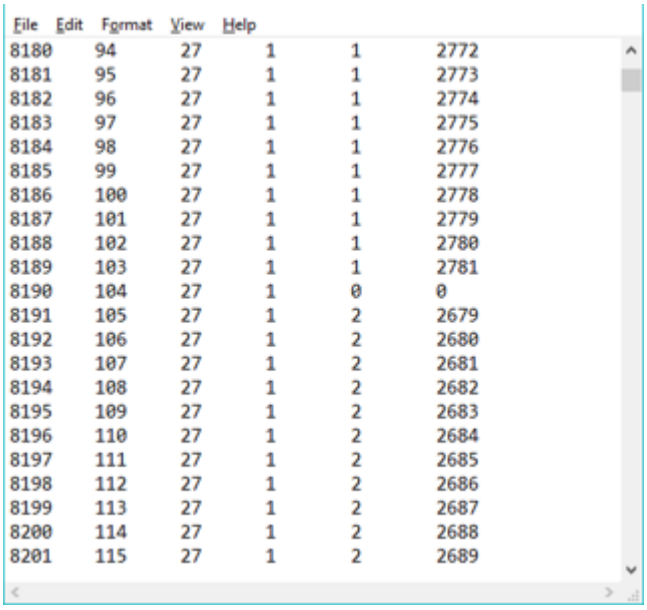
Figure 18: Tiling pattern dialog for a folder of imzML images (or multi-select).

It is convenient if a user labels their files in order of which sample type they are derived from. For example, if a user has 5 cancer tissue and 5 healthy tissues, it would be best to label the 5 cancer tissues as filename1-filename5 and the 5 healthy tissues filename 6-filename10. Then, when loading, the user should choose 5 columns and 2 rows. This

will make ROI selection and data export much easier to remember and visually easier to understand differences between the two sample types. However, this is not a necessary step.

The imzML files need not have the same ROI. Each file is centered in its tile with an optional empty scan separator between tiles along with as many empty columns and rows as needed to make the mosaic pattern plaid. That is, each tile in a row has the same number of columns and each tile in a column has the same number of rows. The preferences INI file (See **§5**) variable *TileScanSeparator* can be set to *false* if you do not want an empty scan line and empty scan column between tiles.

Location files as described in Notes 3, 4, and 5 above can be used with imzML folders; however, there is an important difference. The scan number and X, Y values in the location file refer to the position of a scan in the mosaic. Additional columns are needed to specify the original location of the scan in its file. All that is needed are the file number and the local scan number. This information is created automatically when the ROI export tool described in **§2.4.3** is used to create the file. A portion of such a file is shown in **Figure 19**. Saved location files are thus dependent on the tiling pattern and line and spot spacing when they were created.



File Edit Format View Help

8180	94	27	1	1	2772
8181	95	27	1	1	2773
8182	96	27	1	1	2774
8183	97	27	1	1	2775
8184	98	27	1	1	2776
8185	99	27	1	1	2777
8186	100	27	1	1	2778
8187	101	27	1	1	2779
8188	102	27	1	1	2780
8189	103	27	1	1	2781
8190	104	27	1	0	0
8191	105	27	1	2	2679
8192	106	27	1	2	2680
8193	107	27	1	2	2681
8194	108	27	1	2	2682
8195	109	27	1	2	2683
8196	110	27	1	2	2684
8197	111	27	1	2	2685
8198	112	27	1	2	2686
8199	113	27	1	2	2687
8200	114	27	1	2	2688
8201	115	27	1	2	2689

Figure 19: Example of an ROI location file for a folder of images. The first four columns are the scan number and x,y,z location in the image mosaic. The last two columns are the source file number and scan number. File number 0 and scan 0 are used for the empty scan separator.

3.5 mzXML Folder

For cases where the imaging data originates from multiple instrument files, it is possible to load multiple mzXML files into MSiReader. To do so, they must be placed in the same folder and named sequentially as anyname1.mzXML, anyname2.mzXML, anyname3.mzXML, etc. If the file names do not end with a numeric suffix they will be read in alphabetic order. Since the mzXML files are a sequence of scans and do not contain spatial information, there are two ways that MSiReader can arrange the scans into an image based on external information.

Stream mode. Scans from the files are read sequentially and distributed across columns and down rows to form a single image. Either the *Spots per Line* and *Number of Lines* values from the MSi Data Attributes panel are used to set the image size or an mzXML

info text file is selected as described in **Note 5** above. Dimension mismatches are handled in the same manner as for a single mzXML file.

Tiled mode. Scans from each file are read as a single image and the images from the files are arranged into tiles in the same manner as for an imzML folder. In this case the number of spots per line and number of lines in each file must be specified as well as the number of columns and rows of tiles in the image mosaic. This can be done by **1)** entering a vector of numbers for the Spots per Line and Number of Lines values in the MSi Data Attributes pane or **2)** the spotsperscan and scanlines values in the descriptive text info file can be expanded into a list of integers, one value for each mzXML file. The tiling

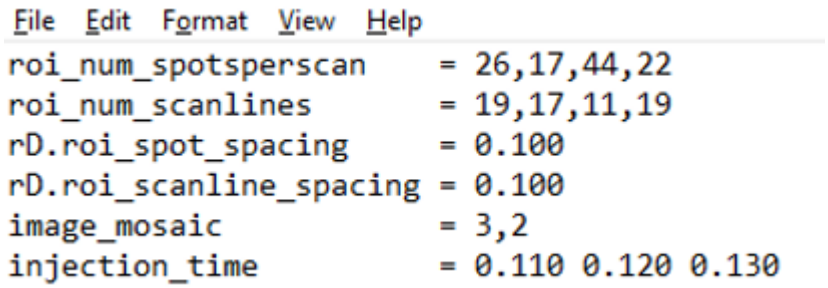


Figure 20: Example of an mzXML header text file for a folder of images. In this example there are four files with dimensions 26 × 17, 17 × 17, 44 × 11, and 22 × 19. They will be arranged into a tiled image with 3 columns and 2 rows.

pattern may also be specified in this file using an additional parameter, *image_mosaic*, whose value is two integers: the number of image columns and the number of image rows. See **Figure 20**. If *image_mosaic* is omitted the images are tiled into a single column or a single row. The choice is made automatically in favor of the smallest aspect ratio for the resulting image mosaic.

3.6 ASCII Folder

The ASCII format consists of multiple files grouped in a folder where each file contains the mass spectrometric data for a single scan and the name of the file identifies the location of the scan in the image (see **Figure 21** for more details about the file format). This was one of the output options for previous generations of Bruker instruments (converted using their CompassXport software). Although it is not a very common file sharing format it is supported by MSiReader and can be easily implemented by MSI groups developing an imaging source with in-house built instruments or those wanting to visualize post-processed data. Note that this format supports non-rectangular and unconnected regions of interest. If data is only available over a region of interest that is not rectangular or sparse, empty scans will be created to convert the region into the smallest rectangle enclosing all the scans in the ROI. Four preferences INI file (**§5**) variables are used to specify the format of the file as shown in **Table 1**. The default values are correct for the Bruker ASCII files.

Table 1: ASCII data file format variables

ASCIIFolderHeaderLines	Number of header or label lines in each file	1
ASCIIFolderNCols	Number of data columns	5
ASCIIFolderMZCol	Column containing *m/z* values	2
ASCIIFolderAbundanceCol	Column containing abundance values	3

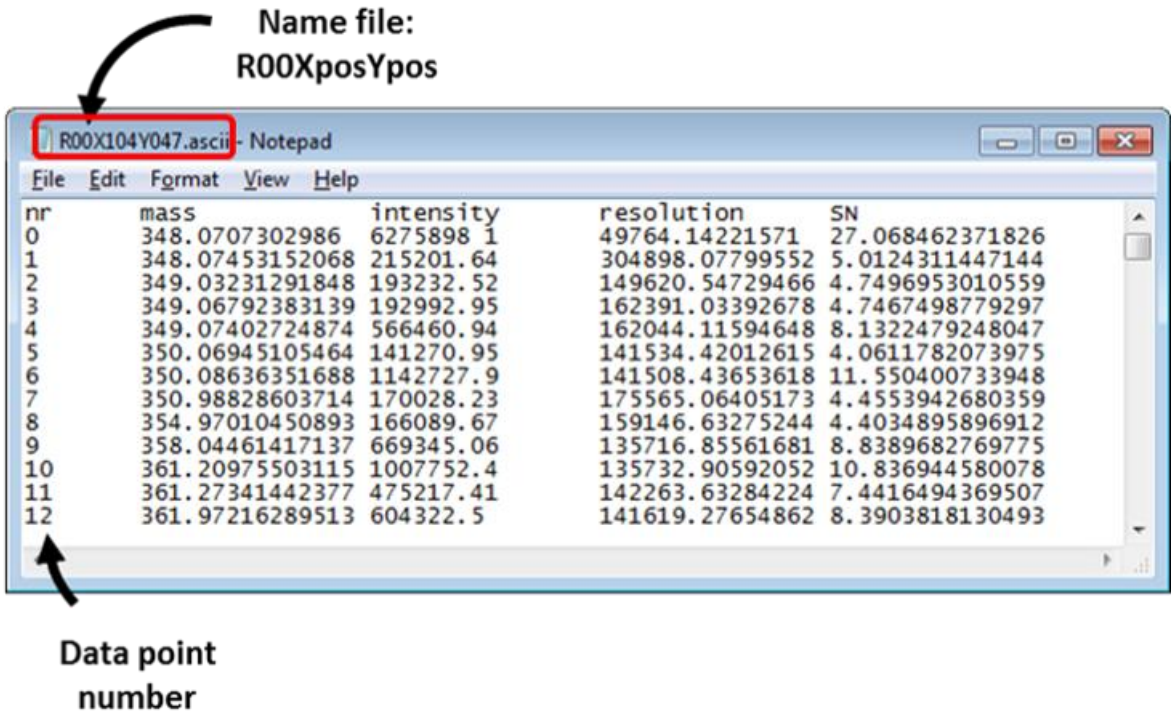


Figure 21: ASCII file format. The folder should contain one file per image scan and each file is named after the location of the scan. The file names are arbitrary, provided that each contains the letter 'X' followed by a sequence of numerals and the letter 'Y' followed by another sequence of numerals. The names are not case sensitive. MSiReader will read the two columns of the file that contain the *m/z* value and the abundance. In this case (**Figure 19**) columns 2 and 3. The header label line and the other columns are ignored. The columns are delimited with spaces, tabs, or commas.

*3.7 Loading Native (Thermo *.raw), *.mss and *.mim File Formats*

Once a data file of any type (Pro or BioPharma) has been loaded into MSiReader, the user is given the option to save the active session in a binary *.mss or *.mim file format for later use. Please see **§7.1** and **§8.2** for more information.

4 Pull-Down Menus, Toolbar Icons, Context Menus and Resizing Windows

There are a lot of options on the main GUI for Pro and BioPharma modes which will be described in this user manual. Below you will find general information on the pull-down menus, toolbar icons and what they do, and a detailed list of context menus embedded throughout the program.

A **video tutorial** on navigating the main GUI of MSiReader for Pro (MSI) can be found **HERE**.

A **video tutorial** on navigating the main GUI of MSiReader for BioPharma can be found **HERE**.

4.1 Pull-Down-Menus

The user-interface for MSiReader v3.03 has been completely re-designed to enhance the user-experience. The menu items and ICONS across the top are shown in **Figure 22** with the different tools and functions that are available. When you first load MSiReader, the only functions allowed from the menu are load data and the information under the help menu. Once you load a file, the allowable functions will become active, which also depends on your license.

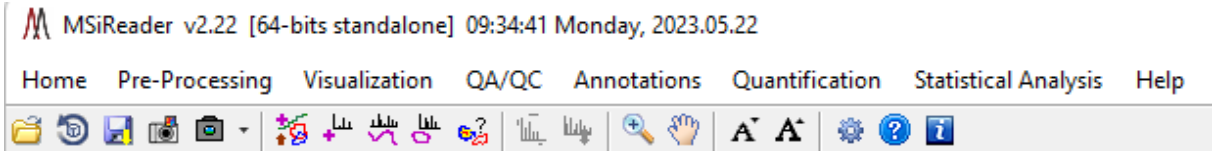


Figure 22: The top-level menu items and ICONS in MSiReader for both Pro and BioPharma modes (in BioPharma mode, Quantification is absent in the top menu). To access functions and tools under a given menu item, click on the menu item of interest and then again on the sub-menu. To view which license you have (Pro or BioPharma), click on Help and then About MSiReader.

After opening MSiReader, you can adjust the size of the fonts for the user interface to match your display by simply clicking on the “A” ICONS in the taskbar (increase or decrease).

4.2 Toolbar Icons

There are a total of 19 toolbar icons that are provided to greatly enhance the user experience given that each is common to many different tools in MSiReader. Descriptions of each icon are provided below. You can also mouse over each icon when using MSiReader to get a tooltip of the functionality it provides. Icons are enabled and disabled as needed/allowable by the program.

 Load Data

The “load data” icon provides the same functionality as the drop-down menu under HOME “load data” or when clicking the “load data” button in the MSi Data/Plate Attributes pane. Prior to loading your data, make sure that you decide if you will do abundance thresholding, *m/z* filtering, want to load injection times, etc. – all of these choices must be set in the MSi Data/Plate Attributes pane **prior to** loading your file.

This “re-load” icon allows a user to have loaded a datafile(s) and carried out some operations. The user then decides that they want to re-load the current dataset forgetting any operations that have been carried out. Thus, this icon clears the data from memory including operations that have been carried out and re-loads the data.

 Save the current session (*.mss and *.mim formats). See **§3.7**.

This file format saves all selections you have made with the currently loaded dataset so when you return to MSiReader, you can simply load this file and continue your data analysis from precisely the same point.

 Save the Current Heatmap Image (file formats available from pull-down menu).

Click here to save the current heatmap to several different file formats (14 different formats are available). Once you click on this icon, it will display the figure and labels and then from there, using the toolbar in the new window, you can do the following: 1) save the figure; 2) print the figure; or 3) toggle to show heatmap with and without the colorbar – if you want to save the figure without the colorbar, do this first prior to saving.

 Export this figure to a .pdf or image file format.

 Restore a previously saved ROI.

 Select a single scan ROI using the cursor tool.

Click this icon to select a single pixel (or well) in the heatmap. The user can drag this around using the mouse. Once a selected pixel of interest is decided upon, the user can right click and a sub-menu comes up allowing one to export the coordinates, change the color of the cursor, or plot the mass spectrum for the collected (or filtered) *m/z* range. When viewing the mass spectrum, there are new icons in the toolbar at the top of the figure. These allow one to save the plot as a MATLAB .fig file, several different graphics file formats (*e.g.,* *.png) or print the spectrum. The next band of icons in the toolbar allows you to select a peak in the mass spectrum and update the heatmap for that specific *m/z*. You can also copy details to the clipboard for the selected peak and enable the interactive peak browser.

 Select an ROI using the segmented line drawing tool.

Click this icon to select a line of any length or direction through the heatmap (well plate). Upon doing so, the length of the line will be shown on the top left of the heatmap. If you right click on the line in the image, you will be able to: 1) export the line ROI details; 2) set line color; 3) plot the ion abundance as a function of distance along the ROI line; and 4) select plot type (stem, stairs, or line).

 Select an ROI using the polygon drawing tool.

After an ROI is drawn the MSiExport, MSiSpectrum, and MSiQuantification tools are enabled as appropriate for each mode of MSiReader.

 Select interrogated and reference ROIs using the polygon tool.

After two ROI's are drawn, the MSiPeakfinder and other tools are enabled.

 Export and view mass spectrum data for selected pixels.

Launches the MSiSpectrum tool.

 Launch the MSiExport tool for a ROI.

Zoom in on the heatmap using the mouse clicks or the wheel.

To zoom in on a region within the displayed heatmap, click the zoom icon and then use mouse clicks or the mouse wheel and zoom in (out by moving mouse wheel in opposite direction) to more closely inspect a region of the image you are interested in.

Pan the image with the mouse.

The user can pan over the heat map image to inspect different regions. This can be useful after zooming into the image.

Decrease or increase the font size to match your display characteristics.

Rather than use the OS display settings (which affect all your programs), we have added a feature to allow the user to increase or decrease the font size to match their display resolution and/or personal preference. This works for the main GUI put also the sub-GUI's.

Reload the MSiReaderPrefs.INI File.

Information about this function is detailed in §5 of this manual.

 Launch the default PDF viewer on your system and load the user manual.

Opens the comprehensive user manual complete with table of contents. The manual is divided into four sections: **Part I: Getting Started**; **Part II: User Guide; Basic Functions**; and **Part III: User Guide: Advanced Functions; and Part IV: User Guide: BioPharma**. Moreover, there are links to internal and external content embedded within the manual.

 Display information about MSiReader including copyright and license information.

This provides information regarding the copyrights owned by NC State University and MSI Software Solutions, LLC.

4.3 Context Menus

Several important features are accessed with right-click context menus on buttons, panels, checkboxes, dialog boxes and plots. These are available on the main GUI of MSiReader as well as for numerous tools in MSiReader. This was done as these menus apply to a specific tool or function that you are currently working with. An example is shown below; however, it is important that you right click on new tools that you are using to make the most efficient and effective use of MSiReader. These context menus will be highlighted in later sections. Since these context menus are self-explanatory, an exhaustive list with screenshots is not provided here. It is important to note that in previous versions of MSiReader, numerous functions were “hidden” in these context menus and difficult to find. We have moved all primary functions to the toolbar (menu items or icons) to avoid users having to “right click” to discover these functions and tools. What remains in the context menus allows users to be more productive while keeping the interface simple.

Figure 23 shows a screenshot of the MS Navigation pane after right clicking on the *m/z* field. This menu allows the user to use a variety of methods to re-load previously used *m/z* values, append new ones to the clipboard, clear the clipboard, etc. This context menu avoids the need to re-type accurate *m/z* values for analytes of interest. A detailed description of each menu item can be found in **§6.2.3**.

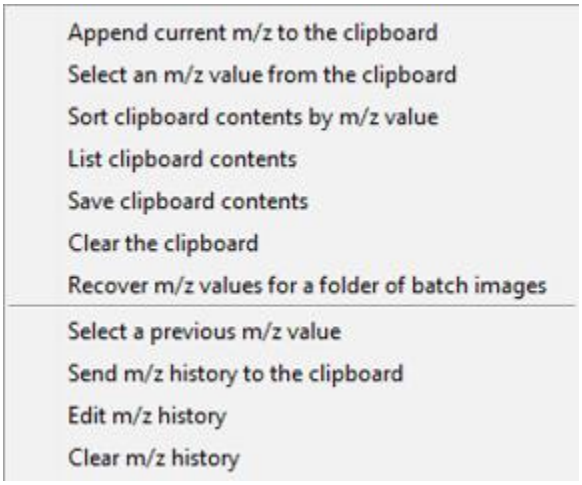


Figure 23: Context menu for clipboard and *m/z* history functions.

4.4 Resizing MSiReader Windows

The main MSiReader figure window and all sub-GUI windows are resizable. This allows MSiReader to be more effectively used with a variety of monitor resolutions, aspect ratios and dpi (dots-per-inch) settings. When started, the main window is sized to nearly fill the main screen of the computer. After startup, it can be made larger or smaller or moved to another monitor. For small, high-resolution screens, it may be necessary to adjust the dpi setting to greater than 100% text size. For WIN10/11 this is done from the display control panel. Changing the window size and the heatmap aspect ratio is also very useful if images from multiple data sets are stacked for simultaneous processing.

5 Customizing MSiReader Pro – the Preferences .INI file

A **video tutorial** can be found **HERE** on how to customize your MSiReader experience.

Note: Some of these preferences apply only to MSI data streams while others apply to both MSI and HTS/HCS data streams.

A plain text preferences file is read by MSiReader on startup that contains most of the default settings and function parameters. Editing it allows the user to easily change settings to customize the UX in MSiReader, so the default edit box values and checkbox options are the ones that you commonly use. Recall that you can edit the preferences and reload them without restarting MSiReader or affecting the loaded data set. This file is in the installation folder. It is called *MSiReaderPrefs.INI* and contains the fields in **Table 2** below in the same order that they are listed in the table. The file is divided into sections for different GUI's in MSiReader. Non-numeric values (*e.g.,* color names, interpolation types) should not be surrounded by quotation marks unless otherwise noted. The alternate colormap selector and the alternate spectrum plot are used regardless of the settings in the preferences file. Finally, the user can always export data to text files (.txt) instead of Excel workbooks by setting '*ExportToExcel'* to *false*.

The end user can change numerous settings and save them as the default parameters. When MSiReader is launched it will use these settings to avoid the user having to change these parameters each and every time the program loads. As the user becomes more

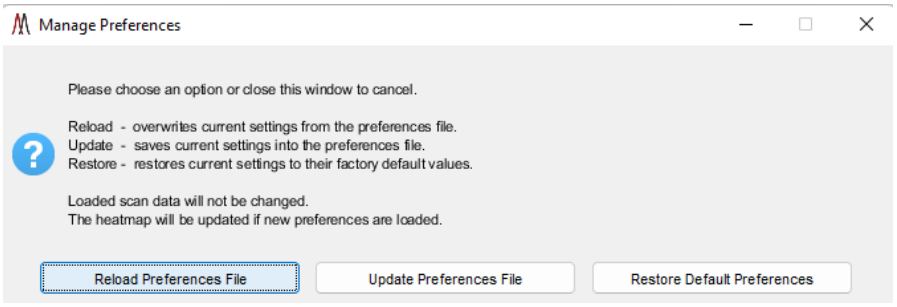


Figure 24: Manage preferences file while using MSiReader. Descriptions for each choice are provided in the dialog box which include Reload, Update, and Restore.

knowledgeable about the parameters described below and how they relate to their research, they can revisit the .INI file in the MSiReader folder (.txt file) to make changes, save them and then reload MSiReader. If MSiReader is already in operation, you can make changes to the .INI file and then reload .INI file using the following steps including restoring the default parameters. Click on the  icon in the toolbar. Upon doing so, the dialog box shown in **Figure 24** is displayed. Note that restoring to the default values does not change your preferences INI file, rather it changes the current in-memory settings by loading a copy of the default values that is saved in the *msiresources* folder. If you are processing data and have all your settings that you prefer for current and future sessions using MSiReader, simply click the "Update Preferences File" in the dialog box (**Figure 24**) and these will save the settings into the .INI file.

If you have a request to add additional attributes to the .INI preferences file, please email us at support@msireader.com.

Table 2 below shows the list of parameters that can be customized to improve the end-user experience. The entries in this table match the order of the entries in the .INI file stored in the installation folder and the values in the table are the values in MSiReaderPrefs.INI after installation.

This page intentionally left blank.

Table 2: MSiReader v3.03 Preferences .INI File

Parameter Name	Default	Description
[MSiPrefs]		
MSiReader main window		
Loadtype	imzML File (single or multiple files)	Default file type. Other options are *mzXML File, IMG File, ASCII Folder*, and M*SS File*.
NCol	100	Default value for the number of columns in the data set.
NRow	100	Default value for the number of rows in the data set.
XSpacing	150	Spacing in µm between scans across the columns.
YSpacing	150	Spacing in µm between scan down the rows.
mzCenter	369.3516	Default value for *m/z* when a file is loaded. If it is out of range, the median *m/z* value of the loaded data will be used instead.
mzCenterHistoryLength	100	Number of previous *m/z* values to save in the navigation history list. The value must be between 1 and 1000.
mzStep	0.01	Default value for *m/z* step.
mzWindow	2.5	Default *m/z* window size value.
mzWindowUnits	ppm	Default *m/z* window units. Specify either *Th* for Thomson or *ppm* for parts-per-million. Default is *ppm*.

mzWindowStyle	tolerance	Meaning of the window size value. *Window* means it is the full window width, *tolerance* means it is one-half the window width.
Pixelation	window max	Method used to calculate abundance: *window sum, window mean* or *window max*.
MaxColorScale	100000	Default Scale Override value. The value is changed to the actual maximum abundance when a file is loaded.
NormalizeOption	none	Default normalization option. Specify either *none*, *Ref Peak*, *TIC, Max, Mean, Median, Midpoint* or *Custom*. The *Custom* option is only available after custom heatmap data has been loaded.
NormScale	1	After normalization, all abundance values are multiplied by this value.
NormCutoff	1	Minimum abundance value for normalization peak in any given spectrum for normalization to be applied.
NormPeak	519.14	Default value for reference peak normalization.
LocalTICNormBoundaries	200 600 1000	Default *m/z* ranges for local TIC normalization
BaselineThreshold	10	Minimum number of data points for baseline correction.
mzResampleOption	roi scans	Default option used for resampling mass spectra when averaging is performed. Values are,

		roi scans : Resample to all *m/z* values from the scans in the ROI. *all scans* : Resample to all *m/z* values from all scans in the image. *uniform* : Resample uniformly over the full *m/z* range with a “match” tolerance.
mzResampleMaxPts	1e+7	Maximum number of resampled *m/z* values allowed.
ItType	linear	Default heatmap interpolation type. Other options are *spline* and *cubic.*
ItOrder	0	Default heatmap interpolation order, 0 to 5.
ColorMap	cividisblack	Default colormap. Other options are discussed in **§6.2.6**.
LogColorScaleBase	10	Base used for logarithmic color scale mapping, one of *10*, *2,* or *e*.
MaxLogColorBarLabels	15	Maximum number of logarithmic colorbar labels.
MMAColormap	cividishi	Colormap for mass measurement accuracy plots.
MMAHistogramDirection	vertical	Direction of the bars in a mass measurement accuracy histogram. *Vertical* or *horizontal*.
MMAHistogramBinMethod	auto	Binning method used to create a histogram. Possible values are: *auto*, *scott*, *fd*, *integers*, *sturges*, or *sqrt*.
MMAHistogramMargin	2	Margin for mass measurement accuracy histogram plots. Units are *ppm*.

DefaultFontSize	9	Default Font Size for MSiReader GUI's
File loading		
ResetLoadOptions	false	Set to *true* to reset the options for data set loading (including the data set type, ROI location file prompt, polarity switching and filtering, etc.) to their default setting when a loaded data set is cleared.
AbundanceFilterEnable	true	Default setting for the abundance filter. See **§2.4.4**.
MinAbundanceThreshold	0.001	Abundance threshold for a data point to be created in the MSi structure when loading data. Smaller values are removed as each scan is loaded. See
FilteredPeakAnchor	true	Default setting for "smart" abundance filtering, whereby a consecutive sequence of low abundance values is replaced by two zero values, one at each end of the sequence. See **§6.2.1**.
NoFilteredPeakAnchorCentroid	true	Default setting when centroiding data to anchor peaks to baseline.
mzFilterEnable	false	Default setting for the *m/z* range filter. See **§6.2.1**.
mzFilterMin	0	Minimum *m/z* value allowed in the filtered data stream for each scan.
mzFilterMax	Inf	Maximum *m/z* value allowed in the filtered data stream for each scan.

SqueezeROIEmptyScans	true true	Removes empty columns and rows, respectively, from the image.
SqueezeROIBorderScansOnly	false	Remove only the empty columns and rows around the border of an image.
UseFastimzMLFileLoader	true	Enables fast text-based parsing of imzML files. Set to false to use the JAVA based imzMLConverter instead.
EnableUUIDimzMLWarning	true	Warn the user if imzML and ibd universally unique identifiers do not match. Set to *false* to suppress the warning dialog.
ImzMLRegexpParams	false	Use an alternate, regular expression-based method of reading imzML header information.
ImzMLDimensionWarning	true	The user is warned if an imzML file does not contain sufficient information to set the X and Y pixel sizes.
LoadInjectionTime	false	Initial state of the load injection time main GUI checkbox.
LoadScanStartTime	false	
IMGLoadBufferSize	40	Size of IMG file load buffer in megabytes.
IMGBufferScanThreshold	20	Minimum number of scans per buffer to use buffered loading instead of reading one scan at a time.
IMGFlipRows	true	Flip IMG file by rows after it is loaded.
ASCIIFolderHeaderLines	1	Number of header and labels lines in an ASCII format data file.

ASCIIFolderNCols	5	Number of columns of data in an ASCII format data file.
ASCIIFolderMZCols	2	Column containing the *m/z* values in ASCII format data file.
ASCIIFolderAbundanceCol	3	Column containing the abundance values in an ASCII format data file.
TileScanSeparator	true	Insert an empty column and an empty row between the tiles in an image mosaic.
MissingTICWarning	true	Display a message “Total Ion Current was not found in the original file and will be calculated instead” if TIC values were not found in the file. Set to *false* to disable display of the warning dialog.
Peak picking and ROI spectrum export		
IntPeakCutoff	80	Default selection criteria for peak picking. The minimum percentage of scans in which an *m/z* must occur in the interrogated ROI to be a candidate when peak picking. See **§7.7.1.**
RefPeakCutoff	20	Default selection criteria for peak picking. The percentage of scans in which an *m/z* can occur in the reference ROI to be a candidate when peak picking. See **§7.7.1.**

MinPeakRatio	2.0	Default selection criteria for peak picking. See **§7.7.1.** Abundance ratio threshold for a peak that doesn't meet the RefPeakCutoff criteria.
AllROIScans	true	
PeakThreshold	100	Default selection criteria for spectrum export and peak picking. See **§7.6.6.** Minimum abundance for an *m/z* to be considered when peak picking.
PeakAlgorithm	Parabolic Centroid	Algorithm for spectrum export and peak picking. See **§7.6.6**. Choices are: *Parabolic Centroid*, *MS Peaks*, or *Local Maxima*.
ViewPeaks	true	Default value for spectrum export and peak picking. See **§7.6.6.** Plot the average spectrum for the ROI.
IncludePeakMarkers	true	Default value for spectrum export and peak picking. See **§7.6.6.** Markers for the peaks are included in the spectrum plot.
ExportPeaks	true	Default value for spectrum export and peak picking. See **§7.6.6.**
ExportPixelPeaks	false	Default value for spectrum export and peak picking. See **§7.6.6.**
SendPeakstoClipboard	true	Default value for spectrum export and peak picking. See **§7.6.6.** The *m/z* values for the peaks found are sent to the clipboard.

AppendtoClipboard	true	Default value for spectrum export and peak picking. See **§7.6.6.** Peaks are appended to the current clipboard contents. Set to *false* to replace the clipboard contents instead.
IncludeAbundance	false	Default value for spectrum export and peak picking. See **§7.6.6.** The exported data file includes abundance values for each *m/z* in the peaks list.
SpectrumAverageAllScans	false	Set to false to average only the scans with abundance greater than the threshold.
SpectrumAverageThreshold	100	Abundance threshold for the spectrum averaging.
ExportToExcel	true	Default value for spectrum export and peak picking. See **§7.6.6.** If set to *false*, data will be exported to a text (.txt) file.
ExportPixelsToText	false	Set to *true* to always export full scan values for ROI scans into a text file, regardless of the number of scans or the number of *m/z*, abundance values.
UseXLSTemplate	true	Default value for spectrum export and peak picking. See **§7.6.6.** A template file is used for exporting peak picking and spectrum ROI data to Excel workbooks.
ExportSpectrum	false	Default value for spectrum export and peak picking. See **§7.6.6.** The average *m/z*, abundance values for the scans in the ROI are included in the export file.

MaxMarkerstoView	50	Number of markers to trigger a confirmation dialog before calling *msviewer* when peak picking or exporting a spectrum.
CursorColor	magenta	Color of cursor for scan selection in spectrum extraction. Valid color values are *blue, green, red, cyan, magenta, yellow, black, white*.
SpectrumROIColor	magenta	Color of tool for ROI selection in spectrum extraction. Valid color values are *blue, green, red, cyan, magenta, yellow, black, white*.
InterrogatedROIColor	magenta	Color of tool for ROI selection in Peak Picking (interrogated zone). Valid color values are *blue, green, red, cyan, magenta, yellow, black, white*.
ReferenceROIColor	green	Color of tool for ROI selection in Peak Picking (reference zone). Valid color values are *blue, green, red, cyan, magenta, yellow, black, white*.
ExportSpectrumChunk	1000	Target number of values to write to Excel at a time. A larger value will use more memory but may be faster. Note that progress bar updates will also be less frequent.
TextExportChunk	1000	Number of lines to write to a text file between progress bar updates.
ROISelectorStyle	polygon	Type of region of interest selector tool. Possible values are *polygon*, *rectangle* or *freehand*.

ROISaveLocationToMAT	true	Saves ROI location info into a MAT file reloading the ROI into MSiReader.
SpectrumPlotStyle	line	Plot style for spectrum plots. Possible values are *line*, *stem* or *stairs*.
SpectrumPlotMarkerStyle	point	Marker style for spectrum plots. Possible values are *point* or *line*.
Database Annotation		
mzMatchTolerance	5	Tolerance for matching target *m/z* values to key values in an annotation database. Units are ppm.
AnotatePeaks	false	Default peak annotation setting for MSiSpectrum (**§7.6.6)** and MSiPeakfinder (**§7.7.1)**.
LoadDatabasesOnStartup	false	Loads the positive and negative ionization database files when MSiReader starts. They will be loaded when needed by MSiDatabase (**§7.7.1**), MSiSpectrum (**§7.6.6**) or MSiPeakfinder (**§7.7.1**) if set to *false*.
IncludeNullMatchResults	false	When set to *true* only the peaks that match a key in the selected ionization database will be saved in results files. If *false*, peaks that match database keys will have additional columns of information from the database, other exported peaks will have empty annotation columns.

PositiveIonDatabaseFile	MSiReaderPositiveIons.xlsx	Name of the position mode ionization database workbook file. The full path and name must be given if the file is not in the MSiReader installation folder. The example included in the installation is *MSiReaderPositiveIons.xlsx.*
PositiveIonWorksheet	Lipid Ions	Name of the worksheet in the positive ionization workbook containing the annotation information and *m/z* key values.
PositiveIonKeyColumn	6	Column of the positive ionization *m/z* key values.
PositiveIonInfoColumn	1 2 3 4 5	List of columns to include as annotations for matching key values in an exported worksheet.
PositiveIonHeaderRow	1	Row in the positive ionization worksheet containing column labels for the annotation information.
PositiveIonFirstRow	2	First row containing data in the positive ionization worksheet.
NegativeIonDatabaseFile	MSiReaderNegativeIons.xlsx	Name of the negative mode ionization database workbook file. The full path and name must be given if the file is not in the MSiReader installation folder. The example included in the installation is *MSiReaderNegativeIons.xlsx.*
NegativeIonWorksheet	Lipid Ions	Name of the worksheet in the negative ionization workbook containing the annotation information and *m/z* key values.

NegativeIonKeyColumn	6	Column of the negative ionization *m/z* key values.
NegativeIonInfoColumn	1 2 3 4 5	List of columns to include as annotations for matching key values in an exported worksheet.
NegativeIonHeaderRow	1	Row in the negative ionization worksheet containing column labels for the annotation information.
NegativeIonFirstRow	2	First row containing data in the negative ionization worksheet.
MSiPeaksDatabaseFile		Name of the peaks database workbook Excel file. The full path and name must be given if the file is not in the MSiReader installation folder.
MSiPeaksWorksheet	Centroid data	Name of the worksheet in the peaks workbook containing the annotation information and *m/z* key values.
MSiPeaksmzKeyColumn	1	Column of the peaks *m/z* key values.
MSiPeaksInfoColumns	2	List of columns to include as annotations for matching key values in an exported worksheet.
MSiPeaksHeaderRow	1	Row in the peaks worksheet containing column labels for the annotation information.
MSiPeaksFirstRow	2	First row containing data in the peaks worksheet.
Image overlay		

ImgMoveResizeColor	magenta	Color of tool for moving and resizing an optical image in MSiImage. Valid color values are *blue, green, red, cyan, magenta, yellow, black, white*.
ImgCropColor	green	Color of tool for cropping an optical image in MSiImage. Valid color values are *blue, green, red, cyan, magenta, yellow, black, white*.
ImgAlpha	0.5	Default optical image overlay transparency. 0 is opaque and 1 is transparent.
ImgBorder	0.1	Size of the empty border region around the heatmap plot drawn by MSiImage. Expressed as a percentage of the heatmap size. Valid values are between 0 and 1.
ImgRotateMag	0.2	MSiImage rotate tool magnification factor. Smaller values move the object slower than the mouse. Valid values are between 0.01 and 1.
Figure export		
ExportFileType	png	Default exported graphics image type. Other options are *bmp, emf, eps, jpg, pdf, tif.*
ExportFigureSize	880 440	Width and height in pixels of an exported figure window as reported by MATLAB. The figures can be resized after they are created.

ExportFigTitleStyle	trim	Exported figure title style. Valid values are *full*, *trim*, *batch*, *metaspace*, *none* or a binary character mask. The default is *trim*. *Trim* contains the data set comment string (if any) and the *m/z* value and window size. *Full* also includes the file name, dimensions, interpolation option, abundance display option and normalization setting.
ExportFigAxes	true	Include or exclude axis scales, labels and tick marks from exported figures.
ExportFigAxesLabels	true	Include or exclude axis labels from exported figures.
ExportFigColorbarStyle	compact	Style of the colorbar for exported figures: one of *compact*, *fixed*, *exponent*, *general*, *auto* or *none*. *Compact* is the default.
ExportFigColorbarLabels	6	Number of labels on the colorbar of exported figures.
ExportFigColorbarPrecision	2	Number of digits to the right of the decimal point displayed for exported figure colorbar abundance labels.
ExportFigFontName	Arial	Font name for text in exported figures. Any installed font can be specified.
ExportFigFontSize	12	Font size for text in exported figures.
ExportToMAT	false	Save a MATLAB MAT file containing Xdata, Ydata and Zdata when the heatmap plot is exported as a graphics format file.

ExportToFIG	true	Save a MATLAB FIG file for exported heatmap figures in addition to the graphics format file.
Batch Processing		
BatchFileNameStyle	1	The naming convention for exported batch image files.
BatchFileType	png	Default batch graphics image type. Other options are *bmp, emf, eps, jpg, pdf, tif.*
BatchFigureSize	880 440	Width and height in pixels of a batch figure window as reported by MATLAB. The figures can be resized after they are created.
BatchFigTitleStyle	metaspace	Batch figure title style. Valid values are *full*, *trim*, *batch*, *metaspace, none* or a binary character mask. The default is *batch*. *Batch* contains only the *m/z* value and window size. *Full* also includes the data set comment string (if any), the file name, dimensions, interpolation option, abundance display option and normalization setting.
BatchFigAxes	true	Include or exclude axis scales, labels and tick marks from exported batch figures.
BatchFigAxesLabels	false	Include or exclude axis labels from exported batch figures.

BatchColorbarStyle	compact	Style of the colorbar for batch processing figures: one of *compact*, *fixed*, *exponent*, *general*, *auto* or *none*. *Compact* is the default.
BatchColorbarLabels	6	Number of labels on the colorbar of batch figures.
BatchColorbarPrecision	2	Number of digits to the right of the decimal point displayed for batch figure colorbar abundance labels.
BatchFontName	Arial	Font name for text in batch figures. Any installed font can be specified.
BatchFontSize	12	Font size for text in batch figures.
ExportToMATBatch	false	Save a MATLAB MAT file containing Xdata, Ydata and Zdata for each figure generated during batch processing.
ExportToFIGBatch	false	Save a MATLAB FIG file for batch heatmap figures in addition to the graphics format file.
ExportDuplicateMZBatch	false	If a peak list contains duplicate *m/z* values, generate duplicate figures during batch processing. A sequence number is appended to each file name.
BatchPeakHeatmapUpdate	false	Controls updating of the heatmap plot during batch processing of a file of peaks. Not updating the main MSiReader heatmap as each *m/z* value is processed improves performance.

BatchPeakHeatmapVisible	false	Set to *true* to make figures visible while batch processing. This will significantly increase the amount of time required to make a folder of images.
Numeric precision for displayed items		
mzExportPrecision	4	Number of digits to the right of the decimal point in *m/z* values for exported plot titles and file names produced with the batch processing feature. Internal file contents are not affected.
mzDisplayPrecision	4	Number of digits to the right of the decimal point for *m/z* values displayed in value and edit boxes.
mzDeltaPrecision	4	Number of digits to the right of the decimal point for delta *m/z* values (*e.g., m/z* Step) displayed in value and edit boxes.
AbundancePrecision	3	Number of digits to the right of the decimal point for abundance values (*e.g.,* Scale Override) displayed in value and edit boxes.
ConcentrationPrecision	4	Number of digits to the right of the decimal point for solution concentrations displayed in value and edit boxes.
SpatialPrecision	3	Number of digits to the right of the decimal point for spatial dimensions displayed in value and edit boxes.

Abundance Labeling		
RawAbundanceLabel	Ions/sec	Label to use on plots for raw abundance values.
IonAbundanceLabel	Ions	Label to use on plots when abundance is scaled by injection time.
NormSuffixLabels	'Peak normalization' 'TIC normalization' 'Local TIC normalization' 'Max normalization' 'Mean normalization' 'Sum normalization' 'Median normalization' 'Midpoint normalization' 'Custom normalization'	Character string labels for the nine normalization selections: *none*, *Ref peak*, *TIC*, *Max*, *Mean*, *Median*, *Midpoint*, and *Custom*. The labels must all be on one line in the INI file. Any label containing whitespace characters (space, comma or tab) must be enclosed in single quotes. The default values are ‘’ ‘Peak normalization’ ‘TIC normalization’ ‘Max normalization’ ‘Mean normalization’ ‘Median normalization’ ‘Midpoint normalization’ ‘Custom normalization’
Binned data export options		
mzExportBinUnits	ppm	Bin size units for exporting data. Set to *ppm* or *Th*.
mzExportBinWidth	10	The *m/z* width of the export bins.
mzExportBinThreshold	100	Minimum abundance value included in a bin.

mzExportBinFill	0	Value used to fill empty bins. Any character string including the not-a-number placeholder NaN is allowed.
mzExportBinOption	1	Method used to compute the bin value: 1=mean, 2=sum, 3=max of all abundance values within the bin window.
mzExportBinPlot	0	Optional plotting while computing bins. 0=none, 1=one plot updated in real-time, 2=a separate plot for each bin. Be careful with option 2, too many figure windows may be created!
MSiColocalization		
ImageDimensionMaxRatio	80	Maximum ratio of the smallest to largest image in either dimension allowed for colocalizing saved image files. The value is expressed as a percentage. See **§7.6.3.**
ColocalNormOption	data max	Normalization method used by MSiColocalization.
MSiSlicer		
SliceLinePlotStyle	stem	Default type of 2D slice abundance plot in MSiSlicer. One of stem, stairs or line. The default is *stem*. See **§7.4.6.**

SlicerExportPlotStyle	stem3	Type of 3D plot exported from the MSiSlicer tool. Either *stem3* or *surface*. The default is *stem3*.
Polarity switching		
PolaritySwitchPattern	+--+	Default polarity switching pattern. Possible values are +--+, -++-, +- and -+.
PolaritySwitchMode	odd	Default polarity switching mode. Specify *odd* or *even*.
MSiQuantification		
QuantifyMinSpots	3	Minimum number of spots to enable regression in the MSiQuantification tool. The minimum value is 3 and the max is 10. See **§7.7.2**.
QuantifyDensity	1	Default MSiQuantification tissue density (g/cm^3).
QuantifyThickness	10	Default MSiQuantification tissue thickness (µm).
QuantifyCalibrationMZ	251.04739	Default MSiQuantification calibration ion *m/z* value.
QuantifyLiquidVolume	0.1	Default volume of MSiQuantification calibration spots (µL).
QuantifyResponseTreatment	Abundance mean	Default treatment of response abundances for the MSiQuantification tool. One of *Abundance mean*, *Abundance sum*, or *Abundance max*.
QuantifyRiskFactor	0.05	MSiQuantification risk factor = 1 – confidence level. Any value between 0.68 and 0.001 is acceptable.
QuantifyPlotSpotLabels	true	Add spot labels on MSiQuantification regression plots.

MSiCorrelation		
CorrelateEmptyImages	false	Enables the correlation of empty images. The abundance is zero for all scans for a candidate *m/z* value. See **§7.6.5.**
AllowBinaryReferenceImage	true	Enables the optional treatment of a reference image ROI as all *ones* or all *zeros*.
CorrelationScoresToSave	110	Number of correlation scores to save. Any positive value including *0*, *none*, *All* and *Inf*.
CorrelationImagesToSave	100	Number of top ranked images to save. Any positive value including *0*, *none*, *All* and *Inf*.
CorrelationAbundanceThreshold	0	Abundance threshold used by MSiCorrelation. A candidate image is considered empty if all scan abundances are below this value or they are all equal to zero.
CorrelationReference	none	The default correlation method. One of *none*, *Reference m/z*, and *External data*.
CorrelationMetric	Structural Similarity Index	The default algorithm selection for image correlation. One of *Absolute Difference*, *Mean Squared Error*, or *Structural Similarity Index*. The abbreviations *AbsDiff*, *MSE*, and *SSIM* are recognized.
InvertReferenceImage	false	The default state of the *Invert reference* checkbox for MSiCorrelation.

NormalizeExternalRefData	true	The default state of the *Normalize external reference data* context menu item for MSiCorrelation.
MSIPCA		
MassRangeSelectionThreshold	25	Threshold number of mass channels for switching from a list selection dialog to a range selection dialog. Any value greater than 2 can be entered.
External Calibration		
ExternalMasses	371.1012 391.28428 413.26623	Rank ordered *m/z* values for performing mass correction of a dataset. These values can be changed when the tool is launched.
ExternalMassWindow	30	Default mass tolerance to consider a calibrant peak as being correct; units are in ppm
CommandWindowProgressLog	false	Print progress message for each scan in the Matlab command window. This value can be changed when the tool is launched.
RealtimeMassWindowPlot	true	Plot the progress of calibration for each scan. This value can be changed when the tool is launched.
Adduct Calculator		
AdductCalculatorFile	MSiAdducts.xlsx	
AdductCalculatorWorksheet	Common Adducts	

AdductCalculatorLegend	Legend	These are how the data are stored in different fields in the output file specified above.
AdductCalculatorHeaderRow	1	
AdductCalculatorFirstRow	2	
AdductNameColumn	1	
AdductModeColumn	2	
AdductFormulaColumn	3	
AdductChargeColumn	4	
AdductMultiplierColumn	5	
AdductDeltaMassColumn	6	
AdductDefaultSelection	0	
METASPACE (Annotation File Structure)		
MetaspaceHeaderRow	3	Row containing column heading character strings.
MetaspaceFirstRow	4	The first row containing data values.
MetaspaceMassSelectionColumns	6	Columns used to create the *m/z* pull-down menu.
MetaspaceRankSelectionColumns	4 5 6 7 8 9 10 11	Columns used to create the *Image Order* pull-down menu.
MetaspaceMassSelectionDefault	6	Default *m/z* selection.
MetaspaceRankSelectionDefault	6	Default *Image order* selection.
MetaspaceMoleFormColumn	4	Column containing molecular formulas.
MetaspaceAdductColumn	5	Column containing adduct formulas.

Hotspot Removal		
HotSpotRemoval	true	Default is to have Hotspot removal ON
HotSpotPercentile	99	Default is set parameter to the 99th percentile
Peak Browser		
PeakThresholdIncrements	*500*	Increment for raising and lowering the peak browser abundance threshold with the arrow keys.
Sequential Paired Covariance		
SPCEnable	False	Default is that SPC tool is not enabled via checkbox on main GUI.
SPCThreshold	1	Default threshold abundance to include in the calculation.
SPCLogbase	2.718281828459046	Default scaling of abundance after SPC is applied.
SPCFilter	product	Default is to take the product for SPC algorithm. Other choices are sum, median and midpoint.
Miscellaneous		
OpenglOption	auto	OpenGL rendering method. One of *auto*, *software*, *hardware* or *hardwarebasic*. The default is *auto*.
MaxExcelRows	1048576	Maximum number of rows in an Excel worksheet. This value can be decreased to limit the size of an exported .xlsx file, but it cannot be increased.

MaxExcelCols	16384	Maximum number of columns in an Excel worksheet. This value can be decreased to limit the size of an exported .xlsx file, but it cannot be increased.
MaxExcelCells	1e7	Maximum number of cells allowed in an exported Excel worksheet.
XLSTemplateFilename		Data processing export template file name. The default file is *MassExcessTemplate.xlsx*. It is stored in the installation directory.
Pathname		Default path when selecting a file, C:\MSiReader\Application\Example File

Part II: User Guide - Basic Functions

6 Basic Functions and Tools in MSiReader

There are a number of basic functions and tools in MSiReader that are commonly used by the MSI community; these include loading single, multiple or a folder of files, abundance threshold filtering, injection time scaling, single pixel selection tool with view and export mass spectrum, customizable user preferences, and export of high quality heatmaps for publication in different file formats. These functions are highly integrated into the main GUI along with functions that are useful for the MSI Mode and the BioPharma mode.

6.1 The Home Menu

The Home pull down menu has 9 options to choose from, 5 of which are also have an ICON to carry out the same function as shown in **Figure 25**. When you choose “load data”, you will be prompted to a pull-down menu with several options for filetypes. After your data is loaded, the menu item and the “LOAD” button in MSi Data attributes will change to “CLEAR”.

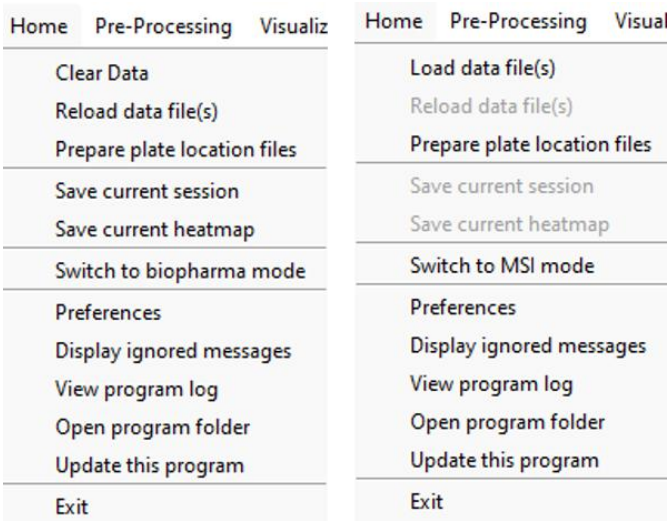


Figure 25: A screenshot of the twelve different functions a user can carry out from the Home menu which include: **1)** Clear Data (**left**) …Load Data (**right**) ( ); **2)** Reload MSI data ( ); **3)** Prepare plate locations (only for Biopharma mode); **4)** Save current session as a *.mss file ( ); **5)** Save current heatmap ( ); **6)** Switch to BioPharma mode …. Switch to MSI mode; **7)** Preferences ( ); **8)** Display ignored messages. If a user checks a box “don’t show me this message again”, they can choose to display these messages here. **9)** View program log (useful for troubleshooting errors with support team at MSI Software Solutions); **10)** Open program folder; **11)** Update this program – when new releases are available, simply chose this option to update your program. **12)** Exit (no icon but the x in top right corner carries out the same function).

“Save current session” (*.mss file format) not only saves all your data but also the precise tools and settings you were using when working up your data. When you choose “Save current heatmap” you will have 14 different file formats to choose from in the pull-down menu. This will generate the current heatmap into a new figure that you can customize. Using the menu in the pop-up GUI, you can change font type and size, change the title (default is the *m/z* value and *m/z* tolerance that was used to generate the heatmap), axes

labels, etc. Once you make your edits to your figure, using the pull-down menu, the end user can save the graphic in several different file formats.

The “Preferences” opens up a dialog box as discussed in §5 which allows a user to customize MSiReader with default values that pertain to your research projects. Using the items in the Home menu (or their respective ICONS when available) in conjunction with the main GUI of MSiReader, a user can efficiently explore their dataset and begin the process of discovery, generate figures for documenting findings in notebooks or for publications, and save your session (*.mss) for re-visiting at a later date with precisely the same parameters when you exited MSiReader.

6.2 The MSiReader Main Graphical User Interface (GUI) for MSI Mode

A **video tutorial** on navigating the main GUI of MSiReader for MSI mode can be found **HERE**. Note this is an overview of the entire GUI for MSI mode.

The main GUI in MSiReader MSI Mode contains 6 panes and includes: **1)** MSi Data Attributes; **2)** Post-Processing; **3)** MS Navigation; **4)** Heatmap Mode; **5)** Heatmap Appearance; and **6)** Colormap. Collectively, these serve as a simple and effective interface to efficiently begin to look at your MSI data with a large heatmap display on the right. Below, each of these panes will be discussed. Please note that when you load MSiReader for a given session, only the MSi Data Attributes pane is shown until a data file is loaded. The overarching GUI with menus, sub-menus, and context menus were discussed in §4 and the description and function of the MSiReaderPrefs.INI were presented in §5. In this section, details will be provided to guide you through the process.

Recall that you should adjust your font sizes to match your display resolution as described in §4.2 using the “A” ICONS in the taskbar for an improved user experience. This can also be done by clicking on the “Visualization” tab and going down and clicking on

"Increase font size" or "Decrease font size" repeatedly until the GUI is appropriated sized for your display. You can also set this in the preferences .INI file (§5) default value = 9.

6.2.1 MSi Data Attributes Pane

Figure 26 shows the MSi Data Attributes Pane; when you first load MSiReader for a session, this is the only pane that is displayed until a file is loaded. However, prior to loading a specific data set, you can still make changes to these default checkboxes and values.

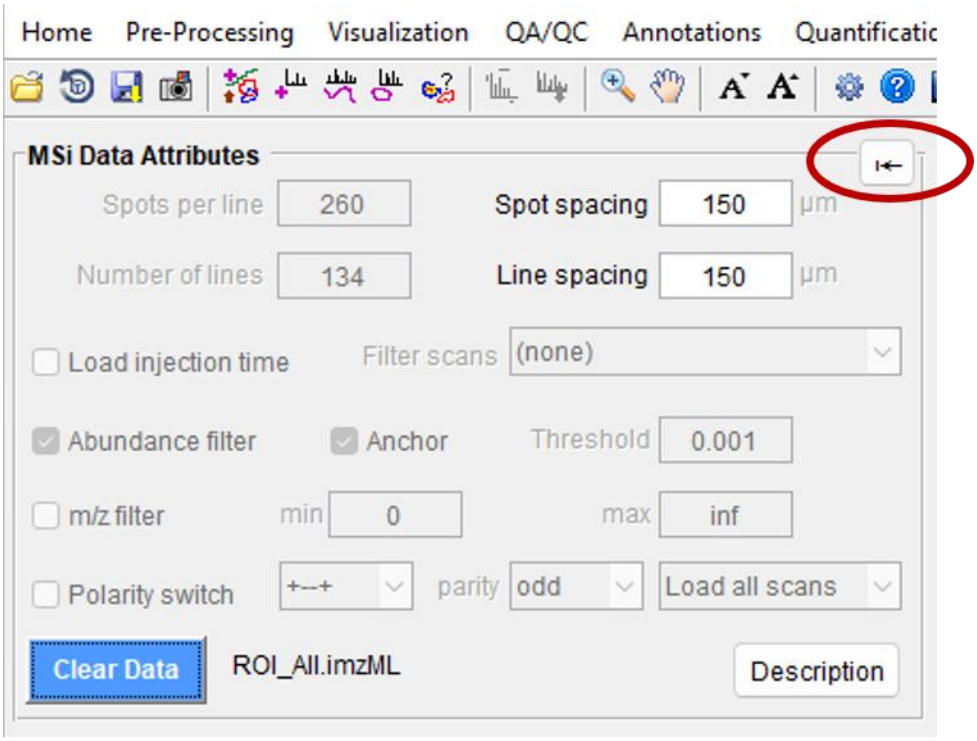


Figure 26: The MSi Data Attributes pane displays the choices that you have set as a default in the MSiReaderPrefs.INI file as well as some values that were imported from the imzML file.

6.2.1.1 Characteristics of the MSI data

The spots per line, number of lines, spot spacing and line spacing are initially set to the default values in the MSiReaderPrefs.INI file (§5). Spots per Line and Number of Lines entries will be filled in automatically when you load a file unless the mzXML format (single

file, multiple files or an entire folder) is selected. The Spot Spacing and Line Spacing fields, relating to the horizontal and vertical spacing, respectively, are loaded automatically from imzML and IMG format files and will affect the dimensions and aspect ratio of the heatmap plots. The Spot Spacing and Line Spacing fields can be changed at any time after the file has been loaded by typing new values into their edit boxes; these manually entered dimensions will be applied immediately. After the file is read, these values can also be modified to change the heatmap plot X and Y axis scaling and the aspect ratio. If set to a negative value the corresponding axis direction is reversed, that is, the heatmap is flipped left to right or turned upside down, respectively. If a value of zero is entered, one unit per pixel scaling is used. Default values can be given in the preferences .INI file in **§5**.

6.2.1.2 Injection time scaling

Heatmap abundance can be loaded and subsequently scaled by injection time with a checkbox in the MSi Data Attributes pane “Load injection time”. Injection times will either be read from the data file directly or, if not found in the file, the user will be prompted to enter a value during the load process. When the load injection time box is checked, the injection time is read into (or an injection time is manually entered in the dialog box). All of the scans in an image do not have to have the same injection time. For example, an imzML file that is a “stitched” composite of multiple data sets or a folder of imzML or mzXML files. How to use the injection time values will be discussed in **§6.2.2**.

6.2.1.3 Filtering data

Data sets can be filtered during loading in a number of ways including abundance threshold (**§2.4.4**).

6.2.2.1 Injection time scaling

If you loaded the injection time when you loaded your data file(s) or manually entered a value via the dialog box, in this pane there is a toggle to either use the injection time (checked) or not use the injection times (unchecked – default value). When using the injection time(s), the ion flux (ions/sec) is multiplied by the scan injection time and the heatmap is updated immediately as well as the abundance units. For example, changing from “ions/sec” (ion flux) to “ions” (total number of ions) if you go from not using ion injection time to using injection time information. If you then uncheck the box for injection time scaling, the abundance data is restored to its previous state by simply dividing by the injection time. The heatmap plot(s) is(are) immediately updated. The default labels (*e.g.,* ions, ions/sec) can be changed in the MSiReaderPrefs.INI file (**§5**) to match the output of your specific mass spectrometry platform.

6.2.3 MS Navigation

The MS Navigation is shown in **Figure 27** to navigate your data using different analytical figures of merit.

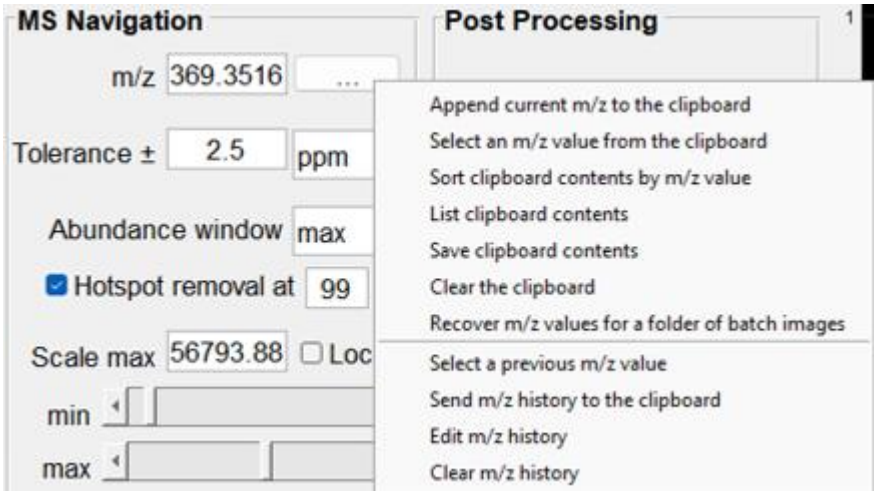


Figure 27: The MS Navigation Pane which includes data entry fields of *m/z*, tolerance, abundance determination, hotspot removal, scale max. scale lock and min and max slider bars to scale the heatmap. If you right click on the "…" next to the *m/z* field, there is a context sensitive menu that provides options for moving *m/z* values to the clipboard as shown.

Once an image is loaded in MSiReader, the user can manually enter values in the *m/z* field. Below are descriptions of the options available in the MS Navigation pane.

6.2.3.1 *m/z*

Location on the *m/z* scale where the *m/z* window is centered. Note that it is possible to append *m/z* values to the clipboard by accessing the right-click context menu of the *m/z* edit box. A peak list can therefore be easily generated while navigating the data set and then used with the correlation and batch processing tools (**§7.6.5**) or pasted in Excel for later use. The right-click context menu for the *m/z* edit box (see **Figure 28**), contains items to access clipboard and history features that aid image navigation and make it easier to build lists of *m/z* values for batch processing and for saving in a document or spreadsheet.

Whenever the heatmap plot is updated the *m/z* value is automatically added to the history. The clipboard is the windows system clipboard, so it is not necessarily empty when MSiReader is launched and anything added to it is available after exiting MSiReader. For example, the *m/z* values can be pasted into a column of an Excel worksheet while MSiReader is active or after exiting. Both the clipboard and the history are preserved when the loaded data set is cleared and new data is loaded. The *m/z* history is lost when the MSiReader session terminates.

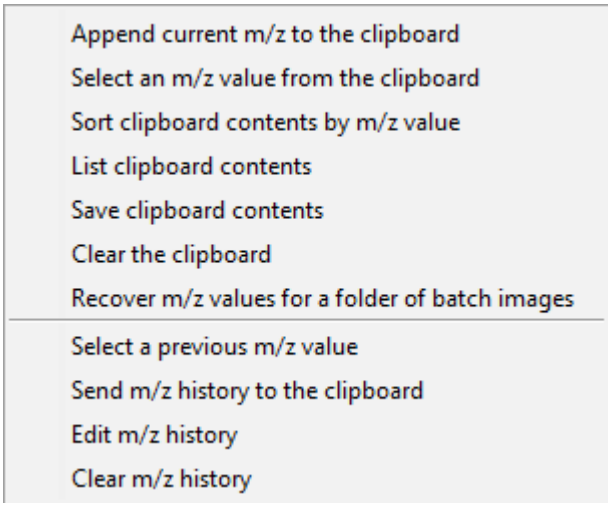


Figure 28: Context menu for clipboard and *m/z* history functions.

Selecting the last item in the top section of the context menu, *Recover m/z values for a folder of batch images*, will prompt the user to select a folder and then attempt to build an *m/z* list from the names of the graphics files (*bmp, emf, eps, jpg, pdf, png, tif,* or *fig*) in the folder. For example, MSiReader's correlation and batch processing tools (**§7.6.5**) and figure export (**§7.6.6**) tools create file names containing *mmm_zzzzz.ext*, where

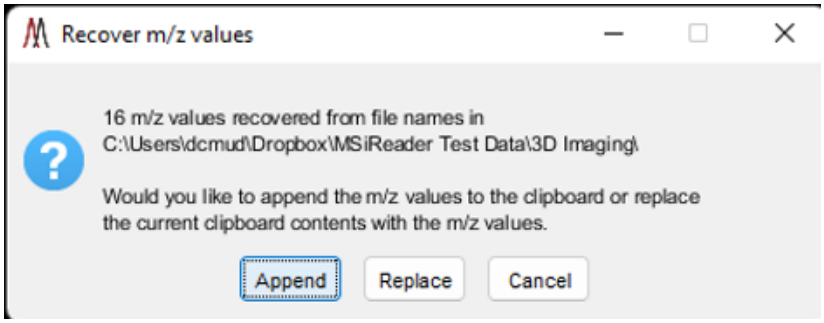


Figure 29: *m/z* recovery clipboard dialog.

mmm.zzzzz is an *m/z* value and *ext* is one of the graphics file type extensions. This can be particularly useful when the contents of a folder have changed. For example, curating a folder of putative peaks with a viewing application. If any *m/z* values are recovered from the file names in the folder the user is prompted to either append them to the clipboard or replace the contents of the clipboard with the list as shown in **Figure 29**.

6.2.3.2 Tolerance

Size of the window considered for the calculation of the abundance of the *m/z* peaks. The user can choose to have a fixed *m/z* window in Thomson (Th) or a relative window in parts-per-million (ppm). Note that the *m/z* window size units selected will also be used by the MSiPeakfinder tool (**§7.7.1**).

6.2.3.3 Abundance

MSiReader offers three different methods to map abundance to a color displayed on the heatmap: **1)** the maximum abundance value in the *m/z* window (window max); **2)** the sum of the abundance values in the *m/z* window (window sum); or **3)** the mean of the abundance values in the *m/z* window (window sum). The meaning of these three options is shown in **Figure 30**. Note that the reported abundance and *m/z* value are not necessarily the values at the center of the window.

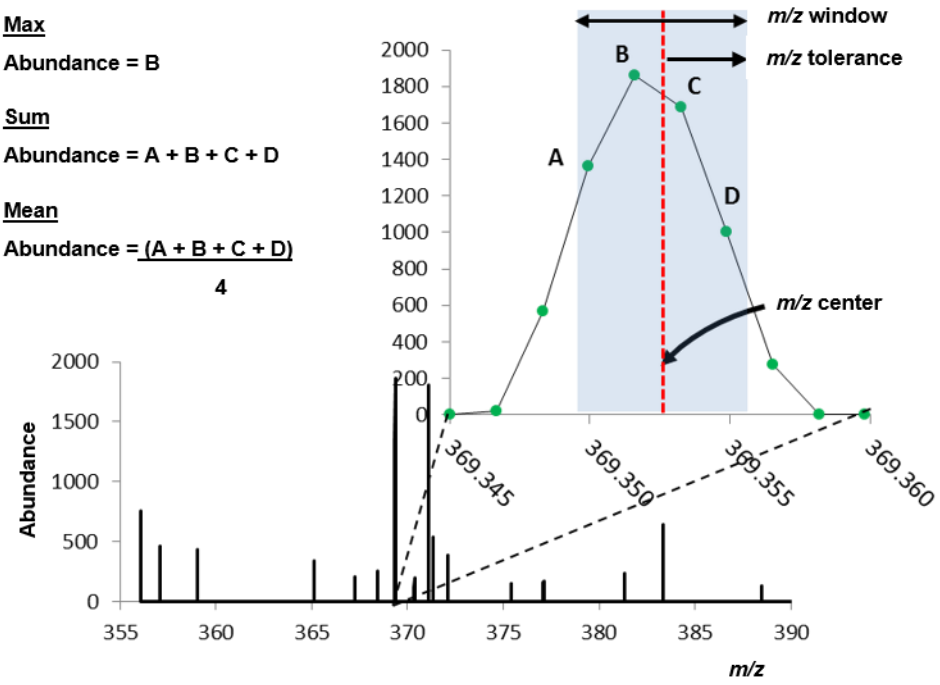


Figure 30: Definition of *m/z* window, *m/z* tolerance, *m/z* center and the three methods used by MSiReader to report ion abundance (max, sum, and mean).

6.2.3.4 Hotspot Removal Tool

Described elsewhere in manual.

6.2.3.5 Min/max slider bars

Values for the minimum and the maximum abundance values represented by the color scale. All scans with an abundance outside of this range will be displayed with the most and least color intensities.

6.2.3.6 Scale Max

The default value for the abundance max slider bar is the maximum abundance of all the scans. For finer adjustment of the color abundance scale, simply change this value. There is also a *Lock colorscale* checkbox context menu item on the scale override field. This is useful for comparing images visually by forcing identical color bar scales regardless of the maximum abundance value. It applies to normalized data set in the MS Navigation pane as well as batch processing (**§7.6.5**).

Note: all default values can be modified in the preferences INI file (**§5**).

6.2.4 Heatmap Mode

Described elsewhere in manual.

6.2.5 Heatmap Appearance

Described elsewhere in manual.

6.2.6 Colormap

The default colormap is *cividisblack* which is color vision deficiency compliant[3,4] and presents a heatmap that is representative of the data. It is a perceptually linear colormap instead of a “rainbow” style colormap like the previous default, *jet,* which has long been considered misleading for the presentation of scientific data[3], especially when converted to grayscale and printed.

The scaling is a simple a way to better display large dynamic range data in the heatmap when you have an analyte that varies over orders of magnitude in abundance within your image. The user can choose from linear, log base 10, log base 2, and log base e. If you wish to “flip” which color is most abundant and which is least abundant, check the “flip” checkbox. Other colormaps are available as well.

6.3 Pre-Processing Menu

Note: Mass Correction, Centroid Data and Scan Scrubber are more advanced functions and are described in **§7.3**.

6.4 The Visualization Menu

Note: All the tools under this menu item including MSi Slicer, Image Overlay, RGB colocalization and 3D plotting are advanced functions and described in **§7.4**. Refer to respective sub-sections to learn more about those functions/tools.

6.4.1 Increase font size / Decrease font size

To increase the font size of MSiReader, you can use the icons “A” and “A” icons in the toolbar or access them through this drop-down menu. It will allow users to set MSiReader font sizes without having to do this through their global OS which will change all of their program displays.

6.5 QA/QC Menu

Note: Mass measurement accuracy, spectral accuracy and Auto MSI QC are advanced options and described in **§7.5**. Refer to respective sub-sections to learn more about those functions/tools.

6.6 The Annotations Menu

Note: Database, Molecular Formula Adduct Search, SSIM colocalization, MSi Spectrum (generate mass spectrum) and MSi Export (export abundance data) are advanced options in MSiReader Professional and BioPharma versions (**§7.6**). Refer to respective sub-sections to learn more about those functions/tools.

6.6.1 METASPACE Annotations

METASPACE Annotation Format. The MSiReaderPrefs.INI file contains variables that define the meaning of the columns in a METASPACE annotation file. This provides a means to accommodate future changes to the format and for the user to enhance the format with additional columns of information or create an entirely new format of annotation file. These variables are defined in **Table 3**.

Table 3: METASPACE variables in the MSiReader preferences INI file.

Variable	Default Value	Meaning
MetaspaceHeaderRow	3	Row containing column heading character strings.
MetaspaceFirstRow	4	The first row containing data values.
MetaspaceMassSelectionColumns	6	Columns used to create the *m/z* pull-down menu.
MetaspaceRankSelectionColumns	4 5 6 7 8 9 10 11	Columns used to create the *Image order* pull-down menu.
MetaspaceMassSelectionDefault	6	Default *m/z* selection column.
MetaspaceRankSelectionDefault	6	Default *Image order* selection column.
MetaspaceMoleFormColumn	4	Column containing molecular formulas.
MetaspaceAdductColumn	5	Column containing adduct names.

After loading a METASPACE annotation file the *Image order* pull-down menu appears as shown in **Figure 31**. Note that an additional item, *Sequence number*, is always added to the end of the list. If it is selected the *m/z* values are not sorted and the images are assigned sequence numbers according to their order in the CSV file. *Sequence number* is also added to the menu when an Excel or text file is selected.

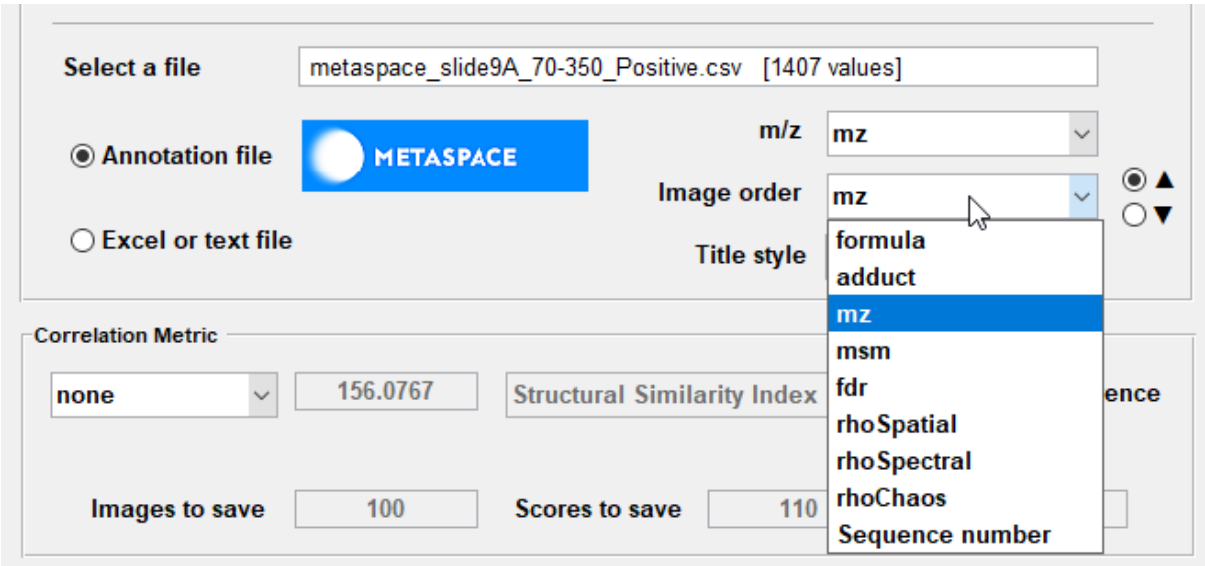


Figure 31: Image order selection for a METASPACE annotation file.

The right-click context menu for the *Annotation file* and *Excel or text file* radio buttons is used to clear loaded data so a new file can be selected as the *m/z* peak source as shown in **Figure 32**.

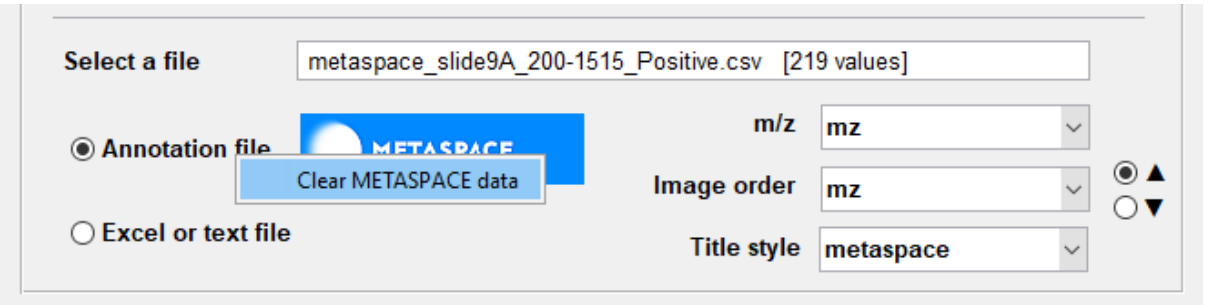


Figure 32: Clearing a loaded METASPACE annotation file.

6.6.2 ROI Functions

Select a single scan ROI using the cursor tool. Use this to select a single pixel (or voxel) in the heatmap. The user can drag this around using the mouse. Once a selected pixel of interest is decided upon, the user can right click and a sub-menu comes up allowing one to export the coordinates, change the color of the cursor, or plot the mass spectrum for the collected (or filtered) *m/z* range. When viewing the mass spectrum, there is a new toolbar at the top. These allow one to save the plot as a MATLAB .fig file, or print the spectrum. The next band of icons in the toolbar allows you to select a peak in the mass spectrum and update the heatmap for that specific *m/z*. You can also copy details to the clipboard for the selected peak. Finally, once a user selects a single pixel (voxel) and displays the mass spectrum, moving the cursor on the heatmap automatically updates the mass spectrum.

Other ROI functions are more advanced and described *vide infra.*

6.7 Quantification Menu

Note: Relative – MSi PeakFinder, Absolute – QMSI Spatial Calibration Curve, and fMSI PIE tool are advanced functions and described in **§7.7**. Refer to respective sub-sections to learn more about those functions/tools. **Note:** There is no quantification menu item for the BioPharma mode in MSiReader.

6.8 Statistical Analysis Menu

Note: PCA (**§7.8.2**) and t-SNE (**§7.8.3**) are advanced functions in MSiReader Professional version. Refer to respective section to learn more about this function.

6.9 The Help Menu

This section contains links to the following: 1) user manual; 2) end-user license agreement (EULA); 3) link to publications; 4) information about MSiReader; and 5) a link to contact us via email. The user manual can also be accessed by the icon  and information about MSiReader can also be accessed by the icon  .

Part III: User Guide – Advanced Functions

7 Advanced Functions and Tools in MSiReader (MSI and BioPharma Modes)

There are a significant number of advanced functions and tools for MSI mode as well as the BioPharma mode (§8) in MSiReader. MSI Software Solutions, LLC is constantly evolving the software to add new features, tools and functions to MSiReader. If you have a suggestion for improvement of existing tools, a request for a new tool, or if you need a customized solution for your research, please email us at support@msireader.com.

Make sure to go to §6 for the basic functions and tools in MSiReader. The description of these is not repeated here.

7.1 Loading Native File Formats from Different Vendors in MSI Mode

The imzML has been very useful in standardizing the field allowing users to convert their data in a two-step process into this format. However, for large projects, experiments with unique data structures, etc. it can be tedious to convert all the files. To this end, in the paid version of MSiReader, we are adding the ability to load native file formats. Once loaded and a user carries out operations on the data, it can then be saved as a *.mss or a *.mim file format that can be used solely in MSiReader and it will load much faster than the original native data file. Please see §3.7 for more information.

There are many different types of data collection in mass spectrometry imaging including, but not limited to, rectangular ROI in flyback or meander mode and arbitrary ROI in flyback or meander mode which requires a location file. The MSI Mode includes information regarding reading native file formats in MSiReader (other formats were described in detail in §3). Moreover, there are many different experimental workflows for HTS/HCS and thus, we have included these different ways to read in these types of data (§8.2) for both imzML and *.raw files. Each scenario has been implemented in MSiReader.

*7.1.1 Loading Thermo Fisher Scientific *.raw files in MSI Mode*

The simplest data collection of mass spectrometry imaging data is a rectangular ROI where the user must know the following information: spots per line, number of lines (the product of these two numbers is the total number of scans), spot spacing and line spacing, and whether they were collected in meander or flyback mode. There is a test dataset online under Thermo RAW and then subfolder Rectangular ROI for testing this feature. More advanced data collection strategies such as Arbitrary ROI, requires a location file. For the Arbitrary ROI, test data are provided with a location file in Arbitrary ROI subfolder. Both of these scenarios are described in the next two sections.

7.1.1.1 Loading a *.raw file with Rectangular ROI (no location file)

First, **BEFORE** you LOAD the Thermo *.raw data file, enter in the spots per line, number of lines, line spacing and spot spacing to match your experiment. The test data has values of spots per line = 112 and number of lines = 58 for a total number of scans = 6496 with a spatial resolution of 100 microns. A location file dialog box will appear, select “Do Not use ROI location file”. Another dialog box will appear prompting the user to indicate the scan type as meander, flyback or cancel. These test data were collected in meander mode so click meander. Then it will tell the user the total number of scans based on the

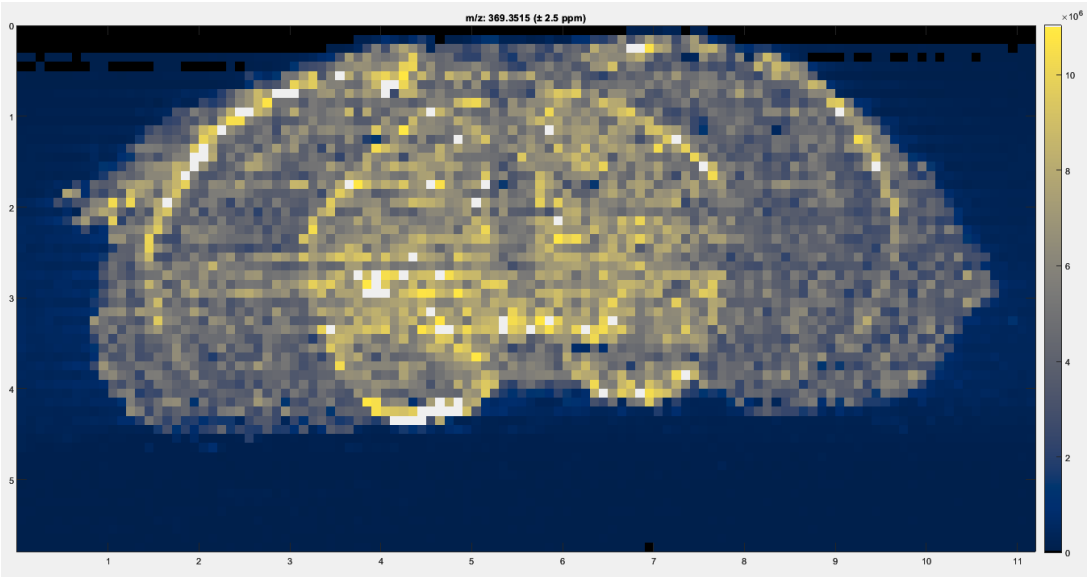


Figure 33: Loading of Thermo *.raw file from data collected with a rectangular ROI in meander mode.

information entered into the GUI. If it is not correct (number of scans in file is not equal to the product of spots per line and the number of lines), it will give the user an error message. If it is correct, no error is present, just click OK and it will load the data. For this test data set using 369.3516 as the *m/z*, the resulting image should appear as shown in **Figure 33**.

7.1.1.2 Loading a Thermo *.raw file with Arbitrary ROI (location file required)

To LOAD a dataset that was collected with an arbitrary ROI, simply click load data. The user will be prompted to select a location file. In the test case, the file is ZF_80*.raw and the location file is ZF_80_location.txt. Once the location file is selected, a message will appear informing the user that 13343 scans are reported in the RAW file but 36140 scans are declared in the location file. This is due to the fact that when using ArbROI, the data is collected for only specific pixels/voxels within the rectangle as define by the location file; hence, these are removed during the loading process. Click OK and the data will load and should look like that shown in **Figure 34**.

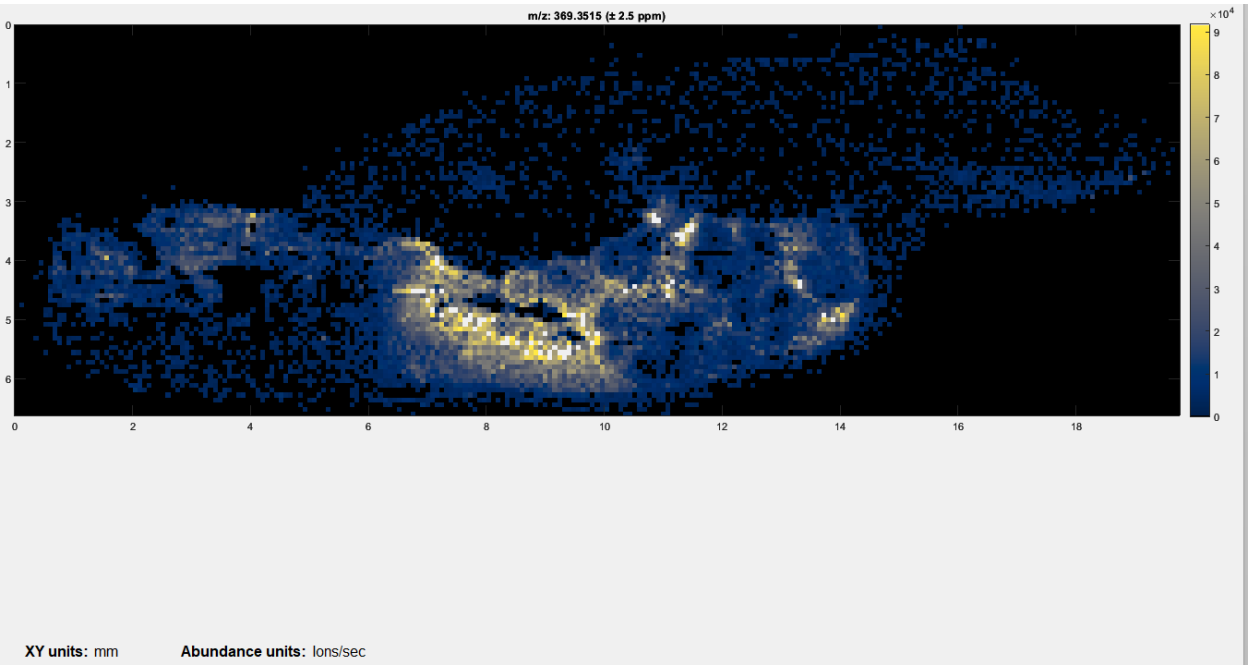


Figure 34: Loading of a mass spectrometry image that was collected with an arbitrary ROI. This is an image of a whole-body zebrafish at 80 micron spatial resolution.

7.1.1.3 The *.mss and *.mim file formats

Once a user has loaded a *.raw file and wishes to save it, it cannot be saved as a Thermo *.raw file (that is a proprietary file format). MSiReader will save this file as a *.mss or *.mim file which then can be directly loaded into MSiReader when further working up a dataset. Please see **§3.7** for more information. For example, if a user takes the data shown in **Figure 34** and tries to centroid it, MSiReader already knows that it was collected in centroid mode and will automatically chose local maxima for peak picking. If a user applies an abundance threshold of 0.001 (data was loaded with threshold = 0.001 so no change is actually being made), it will prompt the user to save the file as a *.mim file. This file is in the test data folder. Inspection of the file size will reveal the *.mim file is about 1/3 the size of the original *.raw file and loads over 100× faster.

It is important to note that while these data were collected with an ArbROI, the *.mim file format now includes all the information for the data and thus, when loading the *.mim file, the user will not be prompted for a location file.

Once a data file of any type (MSI or BioPharma mode) has been loaded into MSiReader, the user is given the option to save the active session in a binary *.mss or *.mim file format for later use. **All data, settings and processing (including colormap and slider positions) will be saved.** The session file (*.mss) and the *.mim format also have the advantage of loading up to 10 times faster than the original file, depending on the original format.

When the dialogue box shows up to save the file that is loaded into memory via the main menu (Home/Save session) or the main toolbar icon, the user can select *.mss or *.mim. Here are the details of what will be saved.

- If a *.mss file is selected (default), then all the GUI parameters/configs are saved, alongside the data.
- If a *.mim file is selected, then on the loaded data is saved, without the GUI parameters/configs.

For both of these file formats, the user can save a loaded dataset without making any modifications to them.

If you wish to save imzML-modified data as a *.mim, a user can now do this.

- If the original data came from a single imzML file, then the *.imzML output is default but users can also select *.mim.
- If the original data came from multiple files or not from an imzML, then only *.mim output is selectable.

7.2 The MSiReader Main Graphical User Interface (GUI) for MSI Mode

A **video tutorial** on navigating the main GUI of MSiReader for MSI mode can be found **HERE**. Note this is an overview of the entire GUI for MSI mode – some of the functions were already described in the user manual.

The main GUI in MSiReader MSI Mode contains 6 panes and includes: **1)** MSi Data Attributes; **2)** Post-Processing; **3)** MS Navigation; **4)** Heatmap Mode; **5)** Heatmap Appearance; and **6)** Colormap. Collectively, these serve as a simple and effective interface to efficiently begin to look at your MSI data with a large heatmap display on the right. Below, each of these panes will be discussed. Please note that when you load MSiReader for a given session, only the MSi Data Attributes pane is shown until a data file is loaded. The overarching GUI with menus, sub-menus, and context menus were discussed in §4 and the description and function of the MSiReaderPrefs.INI were presented in §5. In this section, details will be provided to guide you through the process.

Recall that you should adjust your font sizes to match your display resolution as described in §4.2 using the “A” ICONS in the taskbar for an improved user experience. This can

also be done by clicking on the “Visualization” tab and going down and clicking on “Increase font size” or “Decrease font size” repeatedly until the GUI is appropriated sized for your display. You can also set this in the preferences .INI file (§5) default value = 9.

7.2.1 MSi Data Attributes Pane

Figure 26 shows the MSi Data Attributes Pane; when you first load MSiReader for a session, this is the only pane that is displayed until a file is loaded. However, prior to loading a specific data set, you can still make changes to these default checkboxes and values. For example, if you used an Arbitrary ROI to collect your data and you have a

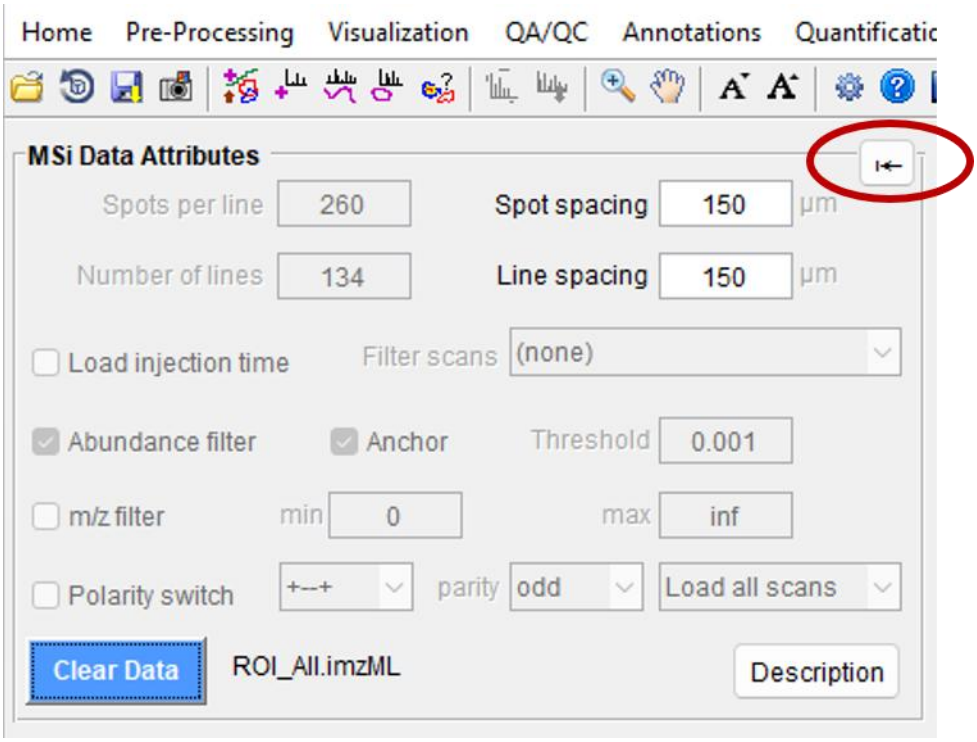


Figure 35: The MSi Data Attributes pane displays the choices that you have set as a default in the MSiReaderPrefs.INI file as well as some values that were imported from the imzML file.

location file, you can use the pull-down menus in Filter scans to select “using ROI location file” prior to loading your data.

7.2.1.1 Characteristics of the MSI data

The spots per line, number of lines, spot spacing and line spacing are initially set to the default values in the MSiReaderPrefs.INI file (§5). Spots per Line and Number of Lines entries will be filled in automatically when you load a file unless the mzXML format (single file, multiple files or an entire folder) is selected. The Spot Spacing and Line Spacing fields, relating to the horizontal and vertical spacing, respectively, are loaded automatically from imzML and IMG format files and will affect the dimensions and aspect ratio of the heatmap plots. The Spot Spacing and Line Spacing fields can be changed at any time after the file has been loaded by typing new values into their edit boxes; these manually entered dimensions will be applied immediately. After the file is read, these values can also be modified to change the heatmap plot X and Y axis scaling and the aspect ratio. If set to a negative value the corresponding axis direction is reversed, that is, the heatmap is flipped left to right or turned upside down, respectively. If a value of zero is entered, one unit per pixel scaling is used. Default values can be given in the preferences .INI file in §5.

7.2.1.2 Injection time scaling

Heatmap abundance can be loaded and subsequently scaled by injection time with a checkbox in the MSi Data Attributes pane “Load injection time”. Injection times will either be read from the data file directly or, if not found in the file, the user will be prompted to enter a value during the load process. When the load injection time box is checked, the injection time is read into (or an injection time is manually entered in the dialog box). All of the scans in an image do not have to have the same injection time. For example, an imzML file that is a “stitched” composite of multiple data sets or a folder of imzML or mzXML files. How to use the injection time values will be discussed in §6.2.2.

7.2.1.3 Filtering data

Data sets can be filtered during loading in a number of ways including **1)** using an ROI location file (§2.4.3) or a bespoke scan pattern; **2)** abundance threshold (§2.4.4); **3)** *m/z* range; and **4)** polarity switching. Abundance filtering is the most commonly used but each one of these filtering approaches will be described here individually; some can be carried out simultaneously (*e.g.,* abundance threshold and *m/z* range) while others are mutually exclusive (*e.g.,* ROI location and bespoke scan pattern). Using an ROI location file when loading data was described in §3.1 and setting an abundance threshold (including the meaning of the anchor checkbox) was described in §2.4.4 and thus, these will not be discussed here.

Unwanted scans that follow a regular pattern can be filtered from a data set as it is read with a bespoke scan filter by selecting “using bespoke scan pattern” in the pull-down menu to the right of filter scans. When the load button is clicked the user will be prompted with the dialog shown in Error! Reference source not found. to describe the scan pattern. The pattern specifies the scans to keep from each pattern replication across rows of the image. If the pattern length is not an integer multiple of the number of columns in the image, the last pattern replicate can be trimmed and the pattern will start again on the

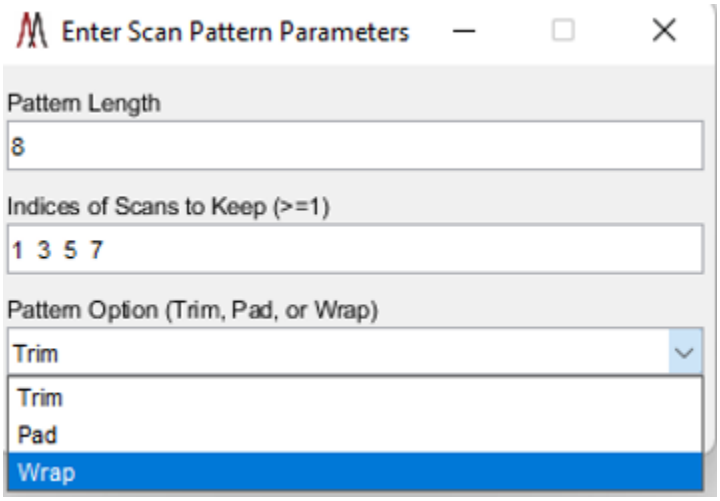


Figure 36: Bespoke scan filter dialog box.

next row (*Trim*), rows of the image can be padded with empty scans to fulfill the pattern (*Pad*), or the pattern can be wrapped around to the next row (*Wrap*).

While the file is being loaded scans that are filtered from the image are set to empty. For the parameters shown in **Figure 26** the odd numbered scans in each row would be read and saved while the even numbered scans would be skipped. After loading is finished, rows and columns that are completely empty will be removed from the image if the preferences INI file variable *SqueezeROIEmptyScans* is *true* (§5). Only those rows and columns outside of a bounding box around the non-empty scans are removed if the variable *SqueezeROIBorderScansOnly* is also *true* (§5).

m/z range filtering can be carried out for all file formats except *.mss; the scans are filtered by *m/z* value as they are read. As shown in **Figure 26**, one can check the *m/z* filter which will allow the user to set the minimum and maximum values allowed which are zero and infinity, respectively. Data pairs (*m/z*, abundance) outside of this range will not be saved in the loaded image. The default values for the filter (0 and infinity) can be changed in the INI preferences file (§5). This filtering can be done to break a large file into several smaller ones or perhaps, a user collected 100 images from *m/z* = 200 to *m/z* = 2500 and upon inspection of all the data, it is observed that there are no analyte peaks between *m/z* 1000 and 2500. In either scenario, *m/z* range filtering will reduce the demand for physical memory.

Data collected natively using polarity switching can be filtered in MSiReader. MSiReader supports analysis of data sets with mixed polarity scans in two ways: polarity filtering and polarity switching. Options for filtering and switching are accessed by right-clicking on the file type pull-down menu in the MSi Data Attributes pane (**Figure 26**) before a file is read. Polarity filtering and switching are only implemented for the *imzML file*, *mzXML file, imzML folder* and *mzXML folder* data type selections. Also note that polarity information for all scans must be stored in the data set for this feature to be meaningful.

Any data set that contains both positive and negative scans can be filtered by polarity as it is read, retaining only the positive (+) image, the negative (-) image or both (load all scans). The distribution of polarity is arbitrary and an empty scan will be inserted in the image in place of each filtered-out scan. The type of polarity filter is selected before the file is read from the context menu to conserve memory. The default is to load all scans.

In the case where all scans are kept, the MSiSpectrum (**§7.6.6**) and MSiPeakfinder (**§7.7.1**) tools have a button to use the positive scans, negative scans or all scans for processing selected ROI(s) if there were both (+) and (-) polarity scans in the data set.

MSiReader supports files that contain four polarity patterns replicated across the rows of the image matrix: [+ - - +], [- + + -], [+ -] and [- +] along with two scan retention options: *keep odd* and *keep even*. These choices are accessed by the pull-down menu which is activated if the Polarity switch checkbox is selected. Polarity switching options are selected before the data set is loaded. The defaults are the [+ - - +] pattern and the *keep odd* option. This can be changed in the preferences INI file (**§5**).

For the 4-tuple patterns [+ - - +] and [- + + -] either the odd (1,3) or the even (2,4) scans are equilibrium scans with no advancement of the sample raster stage and these scans are not loaded. Since the equilibrium scans are all in the same column of each raster scan line, that column can be “squeezed out” of the resulting image. The other scans, (2,4) and (1,3) respectively, are loaded and MSiReader will then have a positive image and a negative image interleaved by column. As with polarity filtering, the MSiSpectrum **(§7.6.6)** and MSi Peakfinder (**§7.7.1**) tools have a button group to use the positive image, negative image or both for processing selected ROI(s). The polarity filter can be used in conjunction with equilibrium scan switching so that only the positive or the negative image is loaded and the unwanted columns are eliminated from the image. For the 2-tuple patterns [+-] and [-+] the *keep odd* or *keep even* options are used to specify which image polarity to load (positive or negative) and the polarity filter is disabled (*i.e.*, set to *all scans*).

If polarity switching is enabled, the first four scans of the file are read and their polarities are compared with the selected pattern. An error is displayed if there is a mismatch. The remainder of the file is not checked for fidelity to the selected pattern. Rows of the data matrix are padded with empty scans, if necessary, so that the number of spots per line is an integer multiple of the selected pattern length (2 or 4). When a scan is selected with the cursor tool  , the polarity and abundance for the scan under the marker is displayed above the heatmap plot as the tool is moved on the screen. Similarly, if an *m/z* spectrum plot is enabled, the title of the plot includes the polarity of the scan.

If you need more working space for the heatmaps, you can click on the arrow as shown in **Figure 26** in the red oval. This will collapse the MSi Attributes pane and the Post Processing pane. Clicking “settings” will recover those two panes.

7.2.2 Post Processing

7.2.2.1 Injection time scaling

If you loaded the injection time when you loaded your data file(s) or manually entered a value via the dialog box, in this pane there is a toggle to either use the injection time (checked) or not use the injection times (unchecked – default value). When using the injection time(s), the ion flux (ions/sec) is multiplied by the scan injection time and the heatmap is updated immediately as well as the abundance units. For example, changing from “ions/sec” (ion flux) to “ions” (total number of ions) if you go from not using ion injection time to using injection time information. If you then uncheck the box for injection time scaling, the abundance data is restored to its previous state by simply dividing by the injection time. The heatmap plot(s) is(are) immediately updated. The default labels (*e.g.,* ions, ions/sec) can be changed in the MSiReaderPrefs.INI file (**§5**) to match the output of your specific mass spectrometry platform.

7.2.2.2 Peak Normalization

Numerous methods of peak normalization are implemented by MSiReader, including normalization by any arbitrary matrix that has the same number of scans as the loaded data set. In the normalization pane, select the type of normalization using the pull-down list. The default selection is “none”. A label is added after abundance units to reflect that normalization have been carried out. The character strings used for each type of normalization can be changed in the preferences INI file (**§5**).

Normalization using a single reference peak. The following function is used to normalize the abundance values in the image with the abundance of a specific *m/z* value. When you select Ref Peaks in the pull-down menu, another box will appear to type in the *m/z* value of the reference peak (single peak *m/z*). The tolerance of the reference *m/z* value is based on the tolerance set in the MS Navigation pane (**§6.2.3**).

$$A_{mz,x,y} = \frac{\mathrm{Max}\left(A_{range(mz),x,y}\right)}{\mathrm{Max}\left(A_{range(ReferencePeak),x,y}\right)} \times \mathrm{NormScale}$$

For each spectrum, normalization will only be performed if the abundance of the reference peak is above the user defined threshold, *NormCutoff*. The normalized abundance is scaled by the *NormScale* value. Default values for *NormCutoff*, *NormScale* and the *ReferencePeak* can be changed in the preferences INI file as described in **§5**.

A quick check to make sure proper function is to enter in the same *m/z* value in Ref Peak data entry field as did for the *m/z* of interest you entered in the MS Navigation pane. In this instance, you should observe two things: **1)** the scale should be one with a heatmap that has a scale of unity and a single color. This is because you have normalized the peak of interest to itself; **2)** after the abundance units (below the heatmap), you should see “peak normalization” so you will easily recall that these data were normalized.

Normalization using multiple reference peaks. The end user can also use multiple peaks to normalize the data based on the *m/z* value (range) of the data. To access this feature, in the main GUI in the post-processing pane, select “Ref Peaks” for normalization and

then check “multiple refs” and then click on *m/z* bounds. A table will appear that has “From *m/z*”, “To *m/z*” and “Reference *m/z*”. For example, if the user has collected data from *m/z* 200 – 800 and wishes to normalize the data from 200 to 500 and 500 – 800 using different reference *m/z* values, simply enter in 200 and 500 and the *m/z* value of reference peak (ref1) in first row and 500 and 800 and the *m/z* value of the reference peak (ref2) in the second column. Click OK to normalize the data using this approach. In this example, all of the data between *m/z* = 200 – 500 will be normalized to the abundance of ref1 and all of the data between *m/z* = 500 – 800 will be normalized to the abundance of ref2.

Normalization using the total ion current (TIC).

When normalization with TIC is selected, every scan is normalized with its total ion current value. If the TIC value for each scan was not provided with the original imaging data file, the user is given the option to use the sum of all abundance values in the spectra as the TIC. TIC normalization will therefore be calculated as follows,

$$A_{mz,x,y} = \frac{\mathrm{Max}\left(A_{\mathrm{range(mz)},x,y}\right)}{\mathrm{Sum}\left(A_{\mathrm{all\_mz},x,y}\right)} \times \mathrm{NormScale}$$

Note 1: For both Ref Peak and TIC normalization, if *sum of window* or *mean of window* is selected instead of *max of window*, *Sum* or *Mean* is used instead of *Max* in the numerator and denominator of Equation (2) and in the numerator of Equation (3).

Normalization using a **local** total ion current (local TIC).

In the pull-down menu, select local TIC and cutoff and scale will show up (default values are 1) and TIC Bounds. Click on TIC Bounds and a dialog box will pop up. Enter in the bounds for the local TIC (default values are 200 600 1000) which means that a local TIC from < 200, 200-600, 600-1000 and > 1000 will be calculated. More local TIC ranges can be input into the dialog box to define more local TIC regions. For example, 200 400 600 800 1000 will generate 4 local TIC ranges that will be used for normalization.

The *m/z* intervals are defined by their boundaries with the *LocalTICNormBoundaries* variable in the preferences INI file (§5). The global *m/z* range of the data set defines the lowest and highest *m/z* ranges. For example, setting *LocalTICNormBoundaries* to "200 600 1000" defines four local TIC intervals:

m/z < 200, 200 >= *m/z* < 600, 600 >= *m/z* < 1000, and *m/z* >= 1000.

If the user wishes to change the local TIC intervals after starting MSiReader, select TIC bounds to enter new values. Local TIC data is immediately calculated for the new intervals if a data set is currently loaded. The smallest value that can be entered is zero and the largest is Inf. The *m/z* values are sorted and duplicates are removed.

To obtain heatmaps of the TIC (and local TIC's if selected) go under visualization menu and then select TIC; TIC and local TIC plots are displayed for any *m/z* intervals that are not empty along with the global TIC plot. Note that the TIC values read from an imzML file are not necessarily the same as the sum of abundances for each scan. Thus, the sum of the local TIC over all *m/z* ranges may not be equal to the global TIC.

Normalizing to the maximum abundance. First the windowing options are applied. That is, the *Sum*, *Mean* or *Max* within the *m/z* window is found for each scan. Those values are divided by their maximum and the result is multiplied by the *NormScale* value. The maximum heatmap abundance will be *NormScale* and the minimum will most likely be zero.

Normalizing to the mean abundance. First the windowing options are applied. That is, the *Sum*, *Mean* or *Max* within the *m/z* window is found for each scan. Those values are divided by their average and the result is multiplied by the *NormScale* value.

Normalizing to the median abundance. First the windowing options are applied. That is, the *Sum*, *Mean* or *Max* within the *m/z* window is found for each scan. Those values are

divided by their median and the result is multiplied by the *NormScale* value. Note that a non-empty scan can have zero abundance after median normalization.

Normalizing to the midpoint of the abundance range. First the windowing options are applied. That is, the *Sum*, *Mean* or *Max* within the *m/z* window is found for each scan. Those values are divided by their midpoint and the result is multiplied by the *NormScale* value.

Normalizing with a custom heatmap. Custom heatmaps are typically created by combining or post processing abundance data in Excel. See **§7.2.6** below for more details about custom maps. Normalization with custom heatmaps enables the user to apply an arbitrary normalization scheme to a data set. As an example, if you want to normalize imaging data to the sum of the abundance of other analytes (*e.g.,* drugs and metabolites) you can use the image summation feature to create a file suitable for loading as a custom heatmap. See **§7.2.6** for more details. In this case, the exported data is a single column of values in a text (.txt extension) file. You can also create Excel files containing the results of combining exported image data in other ways. When selecting an Excel worksheet as a custom heatmap the data selected must have the same number of values as the number of scans in the loaded data set but they do not have to be arranged into the same number of columns and rows. The custom normalization data can also be a single value. It will be expanded to a matrix with dimensions that match the loaded data set.

7.2.3 MS Navigation

The MS Navigation is shown in **Figure 27** to navigate your data using different analytical figures of merit.

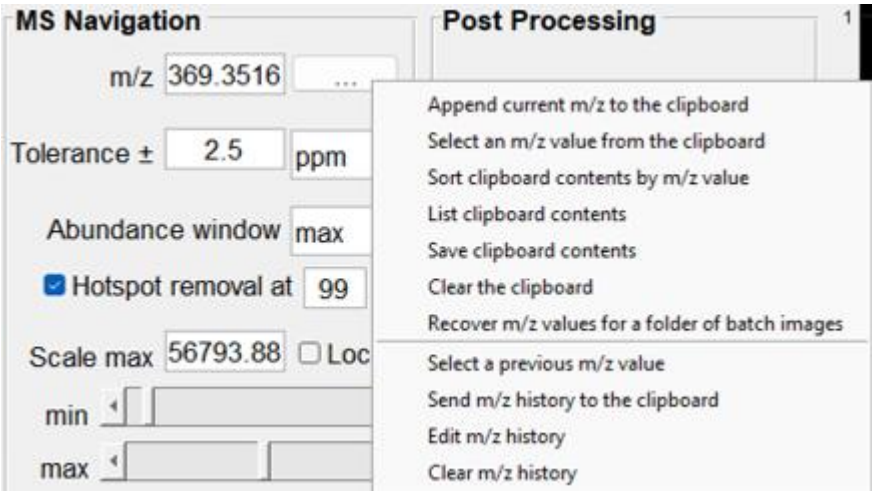


Figure 37: The MS Navigation Pane which includes data entry fields of *m/z*, tolerance, abundance determination, hotspot removal, scale max. scale lock and min and max slider bars to scale the heatmap. If you right click on the "…" next to the *m/z* field, there is a context sensitive menu that provides options for moving *m/z* values to the clipboard as shown.

Once an image is loaded in MSiReader, the user can manually enter va;ues in the *m/z* field. Below are descriptions of the options available in the MS Navigation pane.

7.2.3.1 *m/z*

Location on the *m/z* scale where the *m/z* window is centered. Note that it is possible to append *m/z* values to the clipboard by accessing the right-click context menu of the *m/z* edit box. A peak list can therefore be easily generated while navigating the data set and then used with the correlation and batch processing tools (**§7.6.5**) or pasted in Excel for later use. The right-click context menu for the *m/z* edit box (see **Figure 28**), contains items to access clipboard and history features that aid image navigation and make it easier to build lists of *m/z* values for batch processing and for saving in a document or spreadsheet.

Whenever the heatmap plot is updated the *m/z* value is automatically added to the history. The clipboard is the windows system clipboard, so it is not necessarily empty when MSiReader is launched and anything added to it is available after exiting MSiReader. For example, the *m/z* values can be pasted into a column of an Excel worksheet while MSiReader is active or after exiting. Both the clipboard and the history are preserved when the loaded data set is cleared and new data is loaded. The *m/z* history is lost when the MSiReader session terminates.

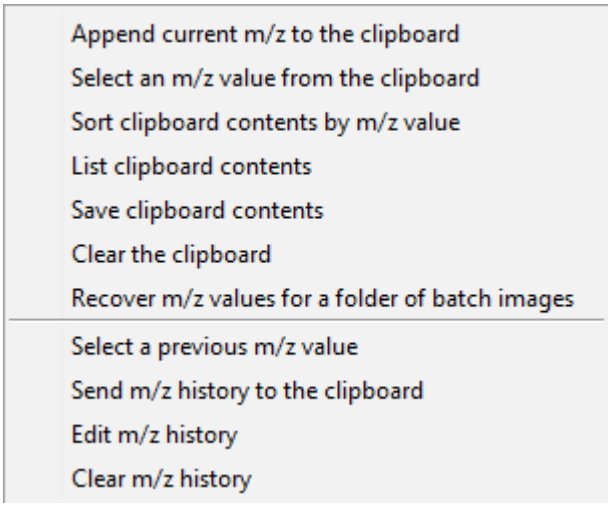


Figure 38: Context menu for clipboard and *m/z* history functions.

Selecting the last item in the top section of the context menu, *Recover m/z values for a folder of batch images*, will prompt the user to select a folder and then attempt to build an *m/z* list from the names of the graphics files (*bmp, emf, eps, jpg, pdf, png, tif,* or *fig*) in the folder. For example, MSiReader's correlation and batch processing tools (**§7.6.5**) and figure export (**§7.6.6**) tools create file names containing *mmm_zzzzz.ext*, where

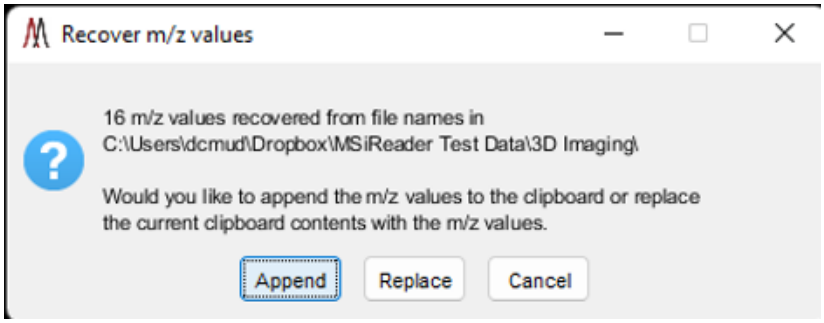


Figure 39: *m/z* recovery clipboard dialog.

mmm.zzzzz is an *m/z* value and *ext* is one of the graphics file type extensions. This can be particularly useful when the contents of a folder have changed. For example, curating a folder of putative peaks with a viewing application. If any *m/z* values are recovered from the file names in the folder the user is prompted to either append them to the clipboard or replace the contents of the clipboard with the list as shown in **Figure 29**.

7.2.3.2 Tolerance

Size of the window considered for the calculation of the abundance of the *m/z* peaks. The user can choose to have a fixed *m/z* window in Thomson (Th) or a relative window in parts-per-million (ppm). Note that the *m/z* window size units selected will also be used by the MSiPeakfinder tool (**§7.7.1**).

7.2.3.3 Abundance

MSiReader offers three different methods to map abundance to a color displayed on the heatmap: **1)** the maximum abundance value in the *m/z* window (window max); **2)** the sum of the abundance values in the *m/z* window (window sum); or **3)** the mean of the abundance values in the *m/z* window (window sum). The meaning of these three options is shown in **Figure 30**. Note that the reported abundance and *m/z* value are not necessarily the values at the center of the window.

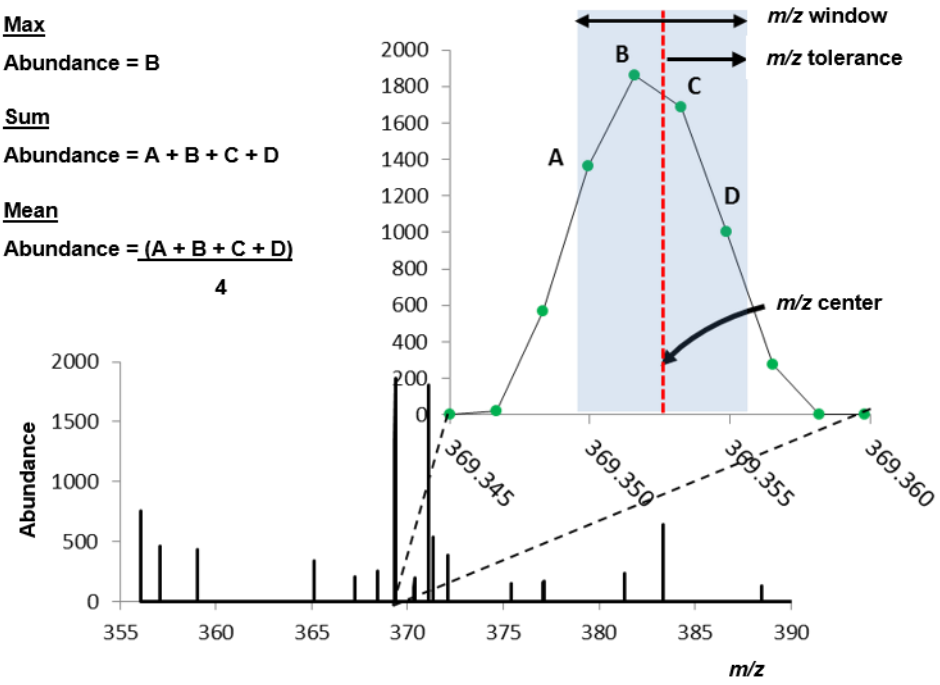


Figure 40: Definition of *m/z* window, *m/z* tolerance, *m/z* center and the three methods used by MSiReader to report ion abundance (max, sum, and mean).

7.2.3.4 Hotspot Removal Tool

The appearance of a heatmap image is occasionally dominated by a small number of pixels whose abundance is much greater than the rest of the image. To apply the hotspot removal tool, simply check the box and then set the percentile level; the default is to have it enabled with a percentile of 99%. Enabling hotspot removal dramatically improves the

appearance of the heatmap image by saturating pixels above the selected percentile level. This is achieved by automatically adjusting the max color scale slider bar to the abundance value corresponding to the hotspot percentile level. This algorithm is identical to the one used by METASPACE[11].

7.2.3.5 Min/max slider bars

Values for the minimum and the maximum abundance values represented by the color scale. All scans with an abundance outside of this range will be displayed with the most and least color intensities.

7.2.3.6 Scale Max

The default value for the abundance max slider bar is the maximum abundance of all the scans. For finer adjustment of the color abundance scale, simply change this value. There is also a *Lock colorscale* checkbox context menu item on the scale override field. This is useful for comparing images visually by forcing identical color bar scales regardless of the maximum abundance value. It applies to normalized data set in the MS Navigation pane as well as batch processing (**§7.6.5**).

Note: all default values can be modified in the preferences INI file (**§5**).

7.2.4 Heatmap Mode

The user can choose to generate heatmaps using the MSi data or to load custom abundance data from a file (Excel or text). These 2 modes are described below.

7.2.4.1 MSi File Data

When this mode is selected, heatmaps are generated using the data from the MSi file. In this mode, all data processing and toolbar tools are enabled. This is the default mode of operation. This can be selected under Visualization > Heatmap Normalization > Use Loaded File Data. In this case, the user can select different normalization in the Post-Processing pane of the main GUI in MSI and BioPharma Mode.

7.2.4.2 Custom Heatmap

When this mode is enabled, the user can use custom data to generate a heatmap. Data can be loaded from an Excel spreadsheet or a *.txt file. If the Excel file contains multiple worksheets the user is prompted to select one of them. MSiReader expects N abundance values where N is the number of scans in the image (number of rows times number of columns) or a single value. The first value is the scan in the top left corner and the last

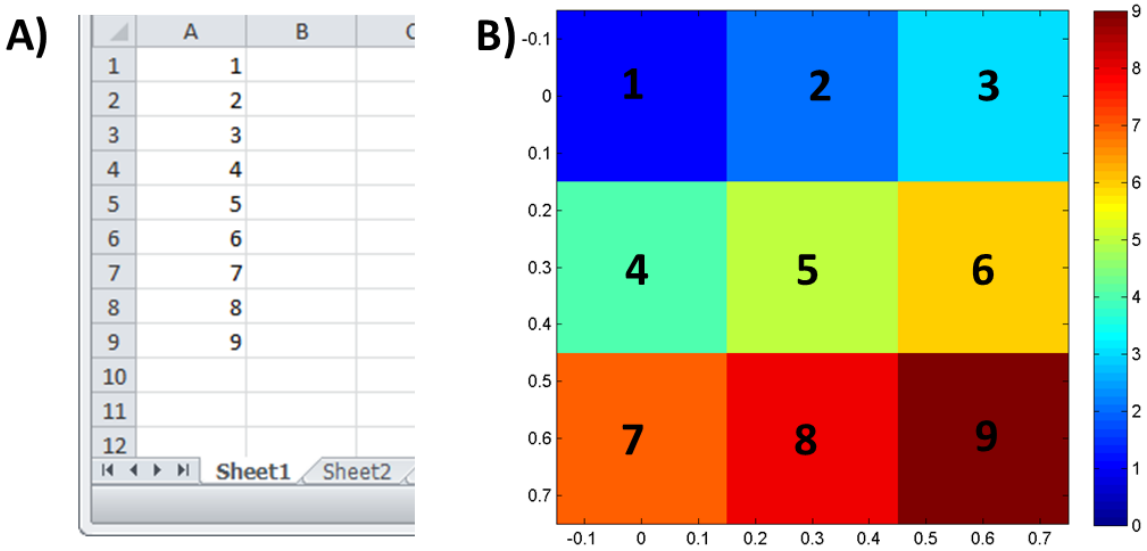


Figure 41: Simple example of custom heatmap loaded for a 3×3 image where abundance increases from 1 to 9. **A)** Excel spreadsheet containing the abundance values and **B)** resulting heatmap.

value is the abundance at the lower right corner (every line from left to right) as shown in Error! Reference source not found.. This input format (order of abundance data point)

was chosen since it is the same format as the output format for the abundance extraction tool. If the input worksheet (or text file) contains a matrix with the correct number of elements, then it will be used as the custom heatmap. If not, then if the first column contains N values it will be used. The user can therefore extract data points using the abundance extraction tool (after selecting all scans in the image as a ROI) and perform any processing of that data in Excel before reloading the results as a custom abundance heatmap. The MSiSlicer tool can also export a 2D cross-section of data and the entire abundance heatmap (**§7.4.6**). The format of this exported data is appropriate for input as a custom abundance heatmap.

For example, if a user would like to make a custom normalization by summing up the abundance of specific *m/z* values in the data, here are the steps.

1. Under the visualization menu, select “summed *m/z* abundance”. Enter in the *m/z* values that will be summed. It will automatically display a heatmap of the summed abundance and then prompt the end-user to enter in a *.txt filename for this custom heatmap. Save this file.
2. Next, under Visualization > Heatmap Normalization, select “use custom abundance data”. Because this custom abundance data was created with the data, it will have the same ROI dimensions. It will prompt the user to load the *.txt file that was just created. Notice then when it is applied, it automatically updates the heatmap. If a user wants to undo the application of the custom heatmap, simply return to the Visualization > Heatmap Normalization and select “use loaded file data”.
3. Finally, in creating the abundance data, another *.txt file was also automatically created with the same filename with added extension _mzlist so the user knows which *m/z* values were summed for the custom heatmap.

7.2.5 Heatmap Appearance

7.2.5.1 Interpolation

The pixel interpolation scheme can be changed in the Heatmap Appearance menu. Three types of interpolation are available (linear, spline and cubic) and each type can be applied

up to the 5th order. For each type, selecting zero order will revert to non-interpolated data (*i.e.*, none). Applying an interpolation scheme does not change the stored data since it is only an image processing step. Default interpolation is linear of order zero (*i.e.*, no interpolation). This can be modified in the preferences INI file (**§5**). When an ROI drawing tool is enabled, the interpolation order is temporarily changed to zero so that scan boundaries are clearly visible.

7.2.5.2 Sequential Paired Covariance (SPC)

A sequential paired covariance (SPC) visualization feature has been added to the Heatmap Appearance panel. SPC reduces the effect of variable noise peaks in an image. It is a visualization tool and does not modify the underlying spectral data.

SPC is a way of visualizing data with large dynamic range as well as defining changes (*e.g.,* tumor margin) which may otherwise not be apparent. This algorithm was recently published for mass spectrometry imaging[12] and was based on previous work with liquid separations coupled to mass spectrometry[13,14]. First, the user selects the checkbox and that allows the user to then choose the SPC Options. The first is the threshold with a default value of 1. The second is the log base you wish to use. The third entry is the filter function which can be product, sum, median or midpoint. Given that the default for the heatmap update is checked, when you chose these different options, using the *m/z* value entered in the MS Navigation pane, you will observe the SPC heatmap.

SPC is calculated for each pixel in a heatmap as the logarithm of the product of that pixel's abundance with the abundance of the adjacent pixels. The corner pixels have only three neighbors, the other pixels on the first and last row and column have five neighbors, and all interior pixels in the image have eight neighbors. SPC is enabled with a checkbox and has three options that are accessed from a context menu on the checkbox label: an abundance threshold, the base of the logarithm, and the filter function. Abundances below the threshold are excluded from the calculation and the default threshold is 1. This prevents zero or very low abundance values from propagating in the image. The default

logarithm base is *e* (2.7183). Setting the base to any value less than or equal to 1 disables the logarithm step after the dot product is formed. The filter function default value is *product*. Three other choices are *sum, median* and *midpoint*.

The colorscale slider bars and colorscale override can be used to reduce the upper and lower abundance assigned to the most and least intense colors respectively. Increasing the minimum value can be helpful for reducing the influence of the background on the image.

SPC can be used with any of the abundance treatments (window mean, window max, window sum), normalization options, hotspot removal, interpolation, and log color scales. If enabled when the MSiCorrelation and batch processing tool (**§7.6.5**) is launched it with be applied to batch images as they are generated. Three variables were added to preferences INI (**§5**) file to set the default values for the SPC options.

Variable	**Value**
SPCEnable	false
SPCThreshold	1
SPCLogbase	2.7183
SPCFilter	product

7.2.6 Colormap

The default colormap is *cividisblack* which is color vision deficiency compliant[3,4] and presents a heatmap that is representative of the data. It is a perceptually linear colormap instead of a “rainbow” style colormap like the previous default, *jet,* which has long been considered misleading for the presentation of scientific data[3], especially when converted to grayscale and printed.

The scaling is a simple a way to better display large dynamic range data in the heatmap when you have an analyte that varies over orders of magnitude in abundance within your image. The user can choose from linear, log base 10, log base 2, and log base e. If you

wish to “flip” which color is most abundant and which is least abundant, check the “flip” checkbox.

7.3 Pre-Processing Menu

7.3.1 Mass Correction

A **video tutorial** on how to use the mass correction tool can be found **HERE**.

You can check your MSI data quality using the QA/QC tools described in **§7.5**. In the event that upon plotting out your data using the mass measurement accuracy heatmap and/or histogram tool is not within the specification of your instrument, you can use this tool to do a single-point mass correction (this is not a full mass re-calibration routine).

The MSiReader external mass correction tool can be used to improve the MMA for a given data set. The calibrated results are displayed as a heatmap showing the ppm shift for each pixel in the image and are optionally saved into an Excel workbook. The loaded data can also be updated in MSiReader with the new *m/z* values for each scan. Finally, the corrected data may be saved as a new .imzML and .ibd file for permanent storage. Alternatively, the user can save the mass corrected data as a *.mim file which is about 1/3 the file size and loads significantly faster.

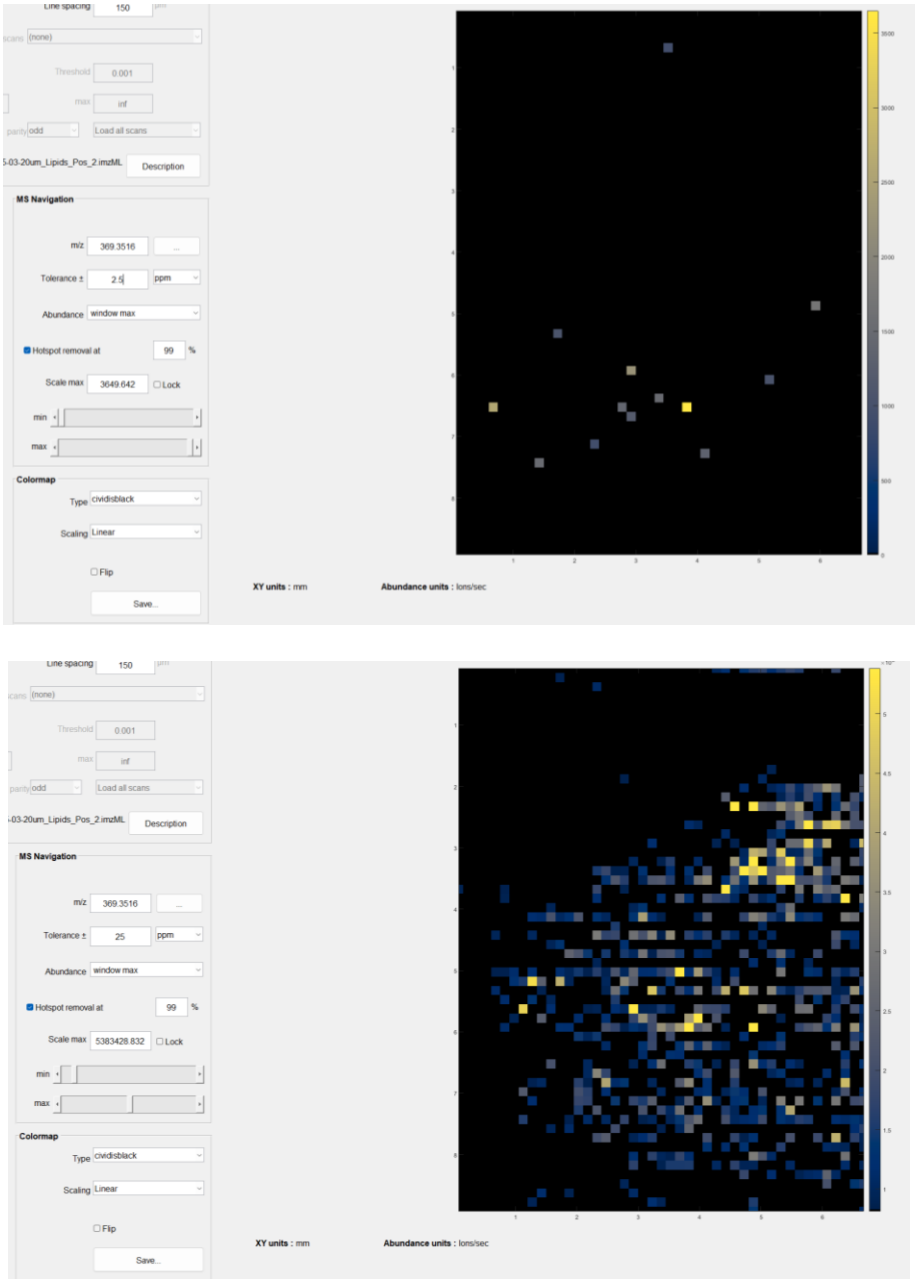


Figure 42: Heatmap of cholesterol in mouse placenta with ± 2.5 ppm (**top**) and ± 25 ppm (**bottom**) tolerance windows. This indicates that the mass measurement accuracy is not within specification and thus, a mass correction of the data is required.

For example, for the mouse placenta tissue in **Figure 42** (top) cholesterol (*m/z* 369.3516) with a tolerance of ± 2.5 ppm should be highly abundant across the sample and in fact it is when the tolerance window is increased to ± 25 ppm as shown in **Figure 42** (bottom).

It is also apparent from the mass spectrum shown in **Figure 43** that a lock mass (m/z_{theo} 391.2843) typically used on this instrument platform also has poor MMA (19.934 ppm; m/z_{obs} = 391.2921). Using these or other known ions, a mass correction can be determined for each scan and applied.

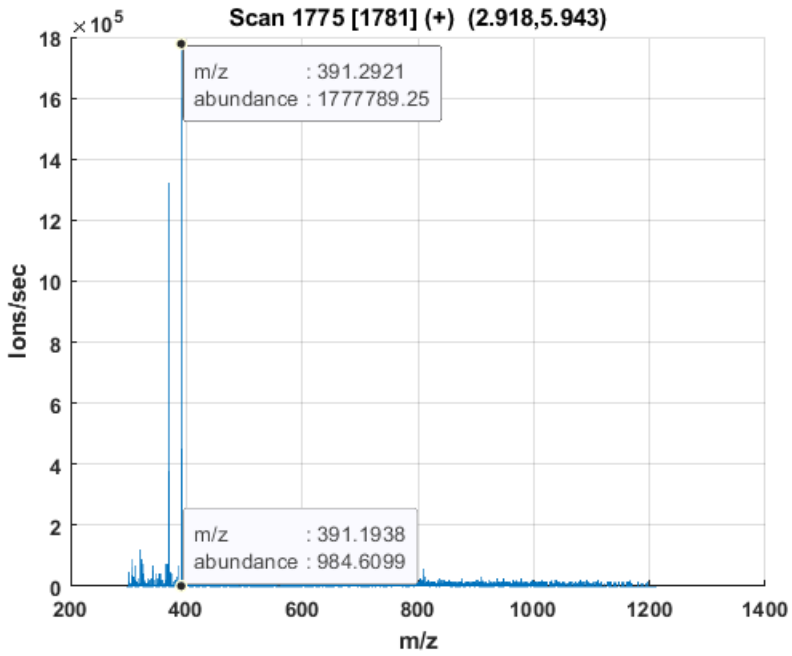


Figure 43: The lock mass is expected at *m/z* 391.2843 but occurs at *m/z* 391.2921.

Selecting “Mass correction” under the “Pre-processing” pull down menu will launch the single-point mass correction tool for the currently loaded data set. Note that this feature is only available for imzML data sets and if the user wishes to save the calibrated data into a new imzML data set all input filters (ROI location file, bespoke scan pattern, *m/z* range, and polarity) **must be disabled** when the data is loaded; by default, only the abundance threshold filter is enabled which is allowable.

After selecting the “Mass correction” tool, the dialog box shown in **Figure 44** is displayed. The default values for these settings can be changed in the preferences INI file (§5). Any number of *m/z* values can be entered and the search window ppm value can be specified as a vector with a value for each *m/z*. For each scan, the most abundant peak within the mass window for each calibration value will be found and the most abundant of those

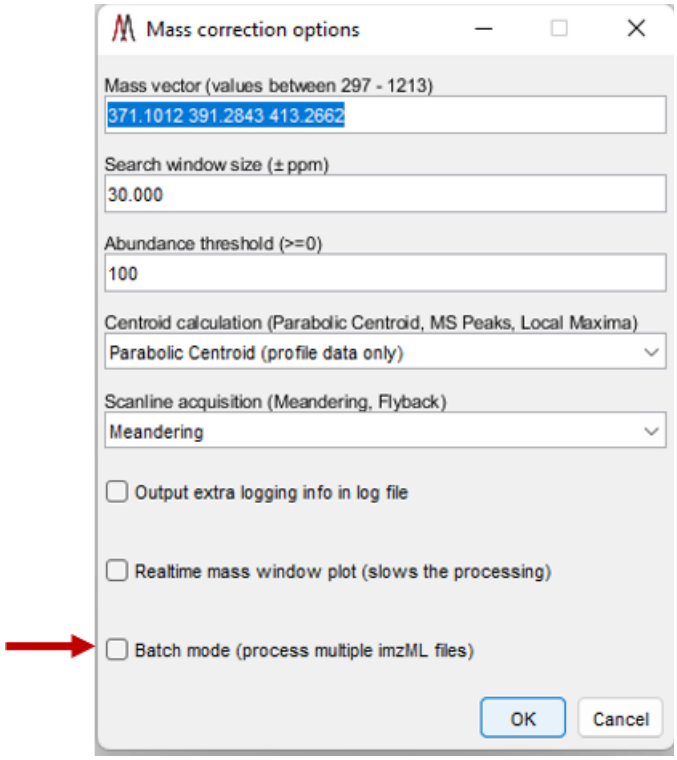


Figure 44: External mass calibration user options dialog box. If the user inputs values that are not allowed, when OK is selected, the dialog box will remain present on the screen until the user fixes the error. For example, notice that the input for mass vector for the calibrant ions, an *m/z* range is noted that is allowed for the loaded dataset.

entered will be used for calibration. Peaks are found using one of the three centroid algorithms implemented by MSiReader. If you are mass correcting profile data, you **must** use either Parabolic Centroid OR MS Peaks. If the user is mass correcting centroided data, one **must** use Local Maxima. The scanline acquisition parameter will default to the value read from the metadata by MSiReader. However, it should be noted that this is not a required parameter in the file so the value in the dialog should be confirmed by the user or the results cannot be correctly saved to a new imzML dataset. The batch mode (red arrow in **Figure 44**) for mass correction is discussed below at the end of this section.

The real-time mass window plot option shows the tolerance window around each calibration mass as the scans are processed in the same plot as shown in **Figure 45**. The plot can be closed at any time and it will not be recreated. Note that the real-time plot degrades performance substantially.

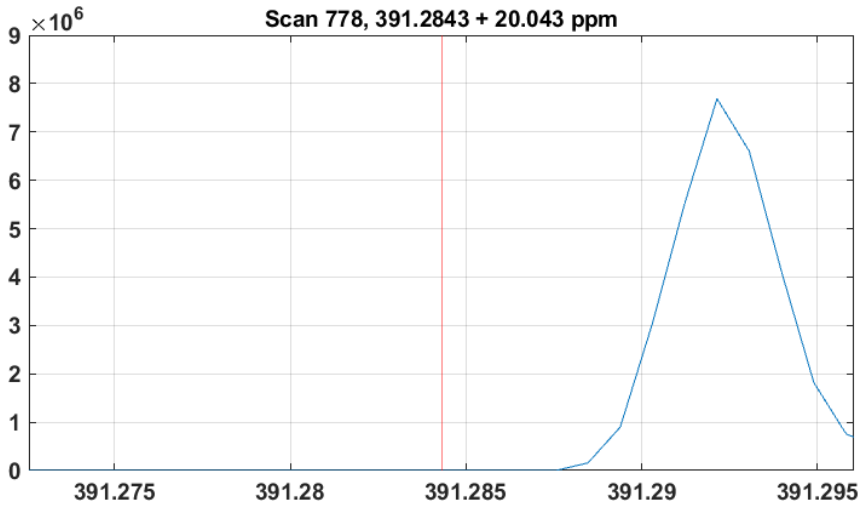
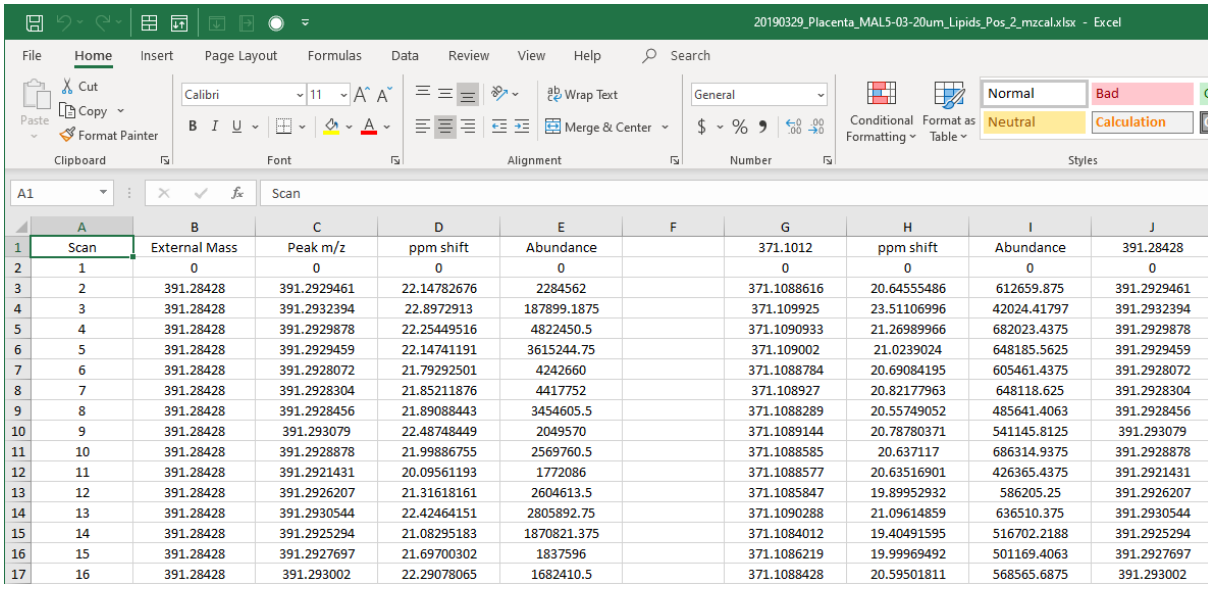


Figure 45: Real-time plot is updated while searching calibrant *m/z* values.

Upon completion the user is asked to select a place to save a report summarizing the results in an Excel workbook. An example is shown in **Figure 46**. The report includes the *m/z*, abundance and ppm shift for all masses and tolerance windows. The ppm shift



20190329_Placenta_MAL5-03-20um_Lipids_Pos_2_mzcal.xlsx - Excel

File Home Insert Page Layout Formulas Data Review View Help Search

Cut Copy Format Painter Paste Clipboard | Calibri 11 Font | Wrap Text Merge & Center Alignment | General Number | Conditional Formatting Format as Table | Normal Bad Neutral Calculation Styles

A1 Scan

	A	B	C	D	E	F	G	H	I	J
1	Scan	External Mass	Peak m/z	ppm shift	Abundance		371.1012	ppm shift	Abundance	391.28428
2	1	0	0	0	0		0	0	0	0
3	2	391.28428	391.2929461	22.14782676	2284562		371.1088616	20.64555486	612659.875	391.2929461
4	3	391.28428	391.2932394	22.8972913	187899.1875		371.109925	23.51106996	42024.41797	391.2932394
5	4	391.28428	391.2929878	22.25449516	4822450.5		371.1090933	21.26989966	682023.4375	391.2929878
6	5	391.28428	391.2929459	22.14741191	3615244.75		371.109002	21.0239024	648185.5625	391.2929459
7	6	391.28428	391.2928072	21.79292501	4242660		371.1088784	20.69084195	605461.4375	391.2928072
8	7	391.28428	391.2928304	21.85211876	4417752		371.108927	20.82177963	648118.625	391.2928304
9	8	391.28428	391.2928456	21.89088443	3454605.5		371.1088289	20.55749052	485641.4063	391.2928456
10	9	391.28428	391.293079	22.48748449	2049570		371.1089144	20.78780371	541145.8125	391.293079
11	10	391.28428	391.2928878	21.99886755	2569760.5		371.1088585	20.637117	686314.9375	391.2928878
12	11	391.28428	391.2921431	20.09561193	1772086		371.1088577	20.63516901	426365.4375	391.2921431
13	12	391.28428	391.2926207	21.31618161	2604613.5		371.1085847	19.89952932	586205.25	391.2926207
14	13	391.28428	391.2930544	22.42464151	2805892.75		371.1090288	21.09614859	636510.375	391.2930544
15	14	391.28428	391.2925294	21.08295183	1870821.375		371.1084012	19.40491595	516702.2188	391.2925294
16	15	391.28428	391.2927697	21.69700302	1837596		371.1086219	19.99969492	501169.4063	391.2927697
17	16	391.28428	391.293002	22.29078065	1682410.5		371.1088428	20.59501811	568565.6875	391.293002

Figure 46: The observed *m/z*, ppm shift and ion abundance for each of the external calibration masses that were entered into the dialog box (**Figure 45**).

heatmap shown in **Figure 47** summarizes the results graphically. This is optional and does not have to be saved.

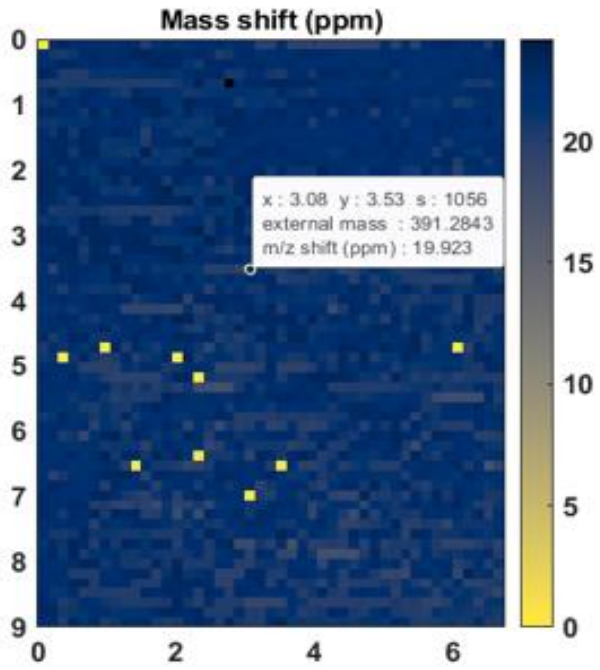


Figure 47: External mass calibration ppm shift heatmap. The data cursor shows the selected peak *m/z* and MMA (ppm) for the queried scan.

The mass shift plot toolbar icons allow the user to see a before-and-after spectrum plot for the scan under the cursor (**Figure 48**), update the *m/z* values in MSiReader (**Figure 49**), and save the results into a new imzML data file or a *.mim file format.

When saving the calibration results into a new imzML data set, the .imzML file is copied unchanged and the .ibd file copied and then new *m/z* vectors are written for each scan. Note that if one or more ROIs are active when the external tool is launched, the user will be prompted to select either ROI scans or all scans. Only the selected scans are processed, plotted and modified.

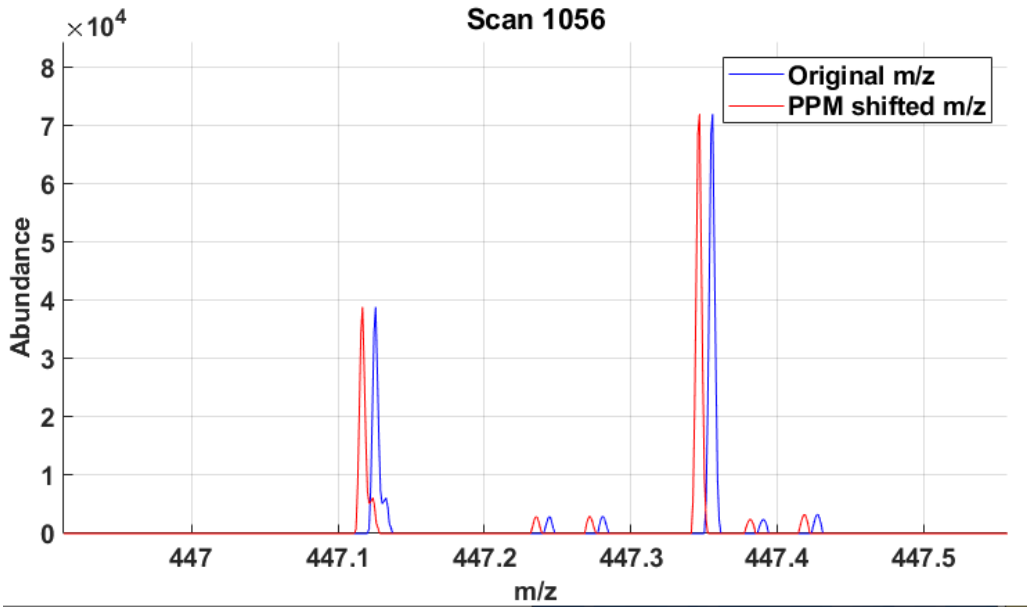


Figure 48: The original (blue) and *m/z* corrected red) spectrum for scan 1056.

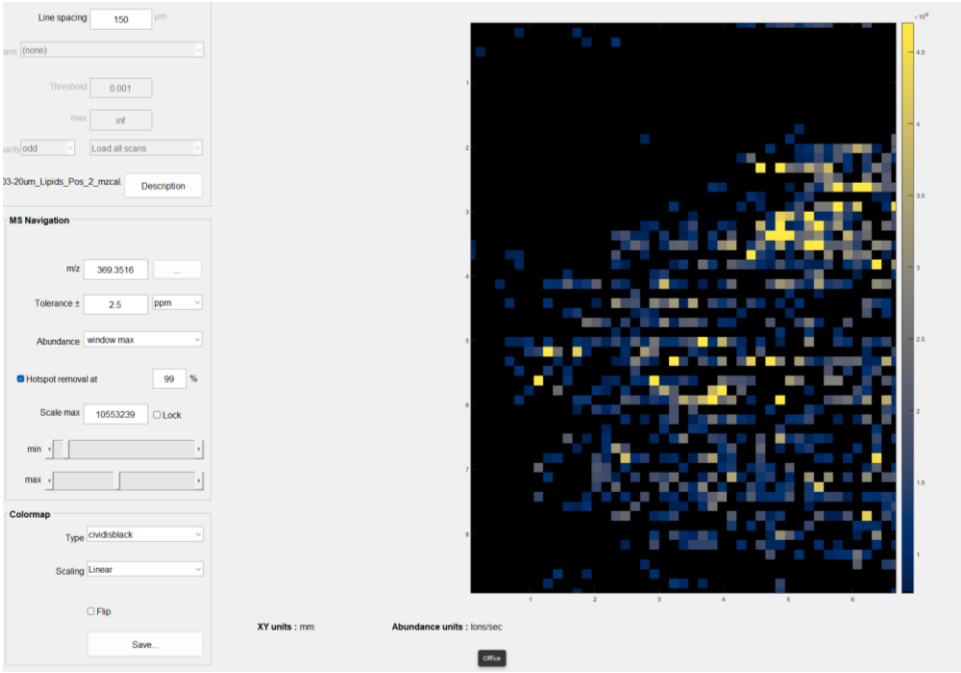


Figure 49: Heatmap for cholesterol with +/- 2.5 ppm tolerance after loading the mass corrected data.

The batch mode for mass correction is to allow the end-user to carry out a single-point mass correction on multiple imzML files without having to load them, save the calibrated data and then save the new imzML file. This function can be accessed in two different ways. The batch mode is not yet functional for *.raw files.

First, you can load an imzML file as before and then launch the mass correction tool which will show you dialog box as shown in **Figure 44**. Make sure the values in this dialog box are suitable for your dataset. Next, check the box that says *Batch Mode* and then *OK*. This will open a folder for the user to select one, two or an entire folder of imzML files. After selecting the files and then OPEN, MSiReader will automatically do a single-point mass correction on every imzML file that was selected and then write the corrected data to a new imzML file with an extension to each filename _mzcal. Since this is batch mode processing, the user does not have to load a dataset – the user can access *Mass correction* as before directly from the *Pre-Processing Menu.* In this approach, since no data is loaded into memory, the Batch Mode is automatically checked (and cannot be unchecked). After your parameters are set, select *OK* and then the file explorer will open as before, select which imzML file(s) you want to mass correct and then *OPEN*.

MSiReader will do as before and recalibrate the data and add an _mzcal to each file that was selected for mass correction.

7.3.2 Centroid Data and Peak Exclusion Filter

A **video tutorial** on centroiding data and peak exclusion filter can be found **HERE**.

Note: This section is partially repeated from **Part I: Getting Started**

MSiReader can take your profile data and centroid it for you – this will reduce the file size and therefore reduce the amount of RAM required. This feature was added to enable some tools to be used in MSiReader that require centroided data but can also be used to reduce file size to the data in memory or using the batch mode. Under the Main Menu item “Pre-Processing” select “Centroid Data” function as shown in **Figure 50**. You can select from three different centroid algorithms (if you are using this tool on data that is already centroided, you must choose “Local Maximum” as the Centroid algorithm), set an abundance threshold, turn on or off the peak exclusion filter (peak exclusion can only be carried out using centroided data – in this case, it will centroid your data and then apply the exclusion filter) and set your peak tolerance in this panel. Then select OK. If you check the peak exclusion filter, you will be prompted for a list of *m/z* values. If the clipboard contains a positive number within the *m/z* range of the loaded data set you will

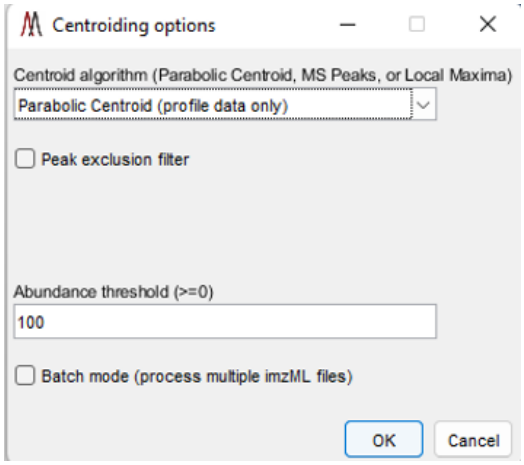


Figure 50: Centroid Data Options Panel in MSiReader

be asked if you want to use those values as the exclusion list. If it contains other content or you decline you will be prompted to select a .txt or .xlsx file with the *m/z* values that you wish to exclude. In the case of selecting a .xlsx file with more than one worksheet you will be prompted to select one. For both types of files, the first column of values will be used.

The exclusion list could be background ions that are present in high abundance in every spectrum that will be removed from the spectra and heatmap. They could also be MALDI matrix ions. If the user does not check the peak exclusion filter, it will centroid your data using the other parameters you have selected in the options panel (**Figure 50**).

Upon centroiding your data in memory, you will be prompted to save a new imzML file in the same folder – MSiReader will add the extension _centroided to the original filename but the user can enter in any filename they choose prior to saving. For batch mode centroiding of data, it will aways add _centroided to each filename automatically and save them in the same folder as the original data. The user can also opt to save these as a *.mim file format.

IMPORTANT: Centroiding data may produce unexpected results if the input file is not an actual mass spectrum but a peak list (preprocessed centroid data). All data preprocessing steps (whether MSiReader or other software) should be validated in your workflow prior to applying them to your data to ensure artefacts are not introduced.

If you check Batch mode and then OK, a file explorer box will open and then you can chose a folder and then select one, several or all .imzML files that the user wants to centroid. This process is carried out in the background. As an example, the .ibd file size for the profile data (in Mass Correction Folder) was ~1.2 GB but after the centroid algorithm was applied, the file size dropped to ~89 MB.

You can simultaneously do abundance thresholding which will further reduce the RAM required (**§2.4.4**). Moreover, prior to (or after using Scan scrubber tool) centroiding your

entire dataset that has been loaded, you can use the ROI selection icon for a polygon and after you select the data of interest, then go to Pre-Processing and then Centroid Data and it will prompt you to select “ROI Scans” or “All Scans”. Once you draw your polygon for the data of interest, if you want the polygon to be a square, right click on the heatmap and select “Make ROI a Rectangle”. You can click on the square and move it around to position it over the ROI of your choice.

There are 3 options to centroiding as shown in **Figure 51** which include Local Maxima, Parabolic Centroid and MS Peaks (wavelet transform – not shown). Only use Local Maxima for data for previously centroided data in the case where you wish to apply a threshold and/or peak exclusion filter. This is recommended because Local Maximum, when applied to profile data using most software, will likely compromise your mass measurement accuracy and ion abundance. This is of course the fastest of the three centroiding algorithm; however, be cautious centroiding data using this approach.

.

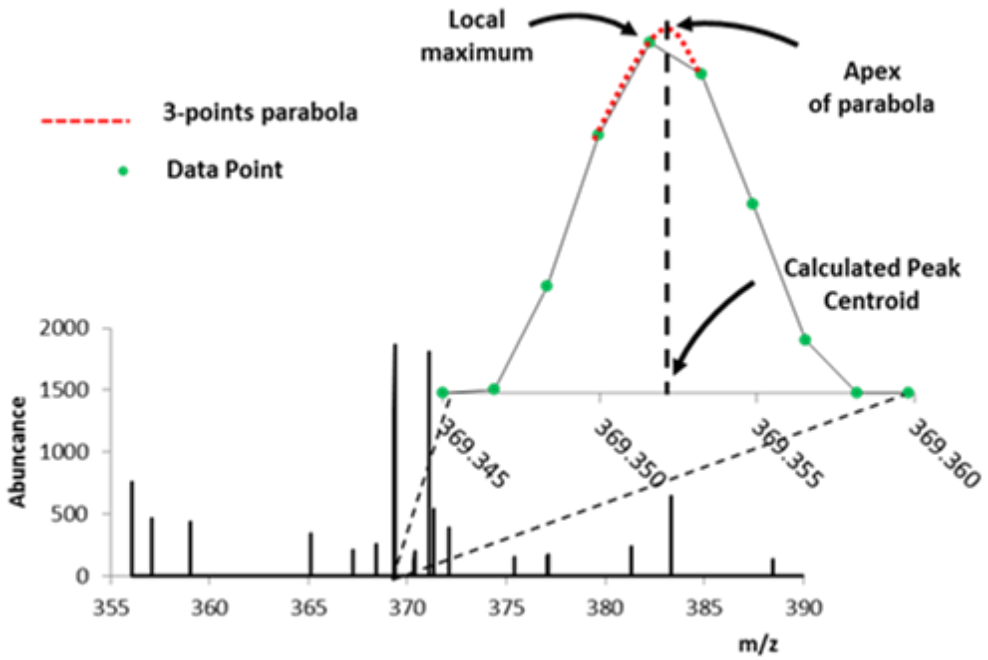


Figure 51: Illustration of centroid algorithms / calculations showing local maximum and parabolic functions.

MS Peaks uses a wavelet transform and filter to find peaks and is similar to the CWT algorithm for peak picking in MSConvert[5]. This algorithm finds peaks in a noisy signal by smoothing the data using a wavelet transform (Daubechies filter banks), putative peak locations are determined and then post-filtering to reduce over segmented and noisy peaks. This approach to centroiding will likely increase computational time significantly.

These could be background ions that are present in high abundance in every spectrum that will be removed from the spectra and heatmap. If you set peak exclusion as “false”, it will centroid your data using the other parameters you have selected in the options panel (**Figure 50**).

Important Note: Once the user has centroided their data in MSiReader, the modified imzML file can only be read by MSiReader due to our proprietary parsing algorithm and

padding to reduce file size. Regardless, the .ibh, .imzML and .ibd files must all be present to open the file for analysis.

7.3.3 Scan Scrubber

For a **video tutorial** on how to use Scan Scrubber, click **HERE**.

The scan scrubber allows a user to load a file, select a single pixel, line or polygon (using the ROI selection tools) and then remove the data either inside the ROI or outside of the ROI. After this process is carried out, the user can then update the heatmap and save these new data. For example, if a user has a file with a lot of noise in the off-tissue pixels or a very abundant pixel that is skewing downstream statistical analysis, one can select an off-tissue polygon ROI or single pixel ROI and then clear them and then save as a new imzML file or *.mim file. It will prompt the user to enter in a filename; however, MSiReader will automatically add "_scrubout" or "_scrubin" to the end of the original filename.

7.3.4 Ion Classification Tool

In mass spectrometry imaging of tissues, it is important to objectively determine whether an ion is tissue-related (on-tissue) or is a background ion. A tool based on object image analysis was recently reported and is now part of MSiReader.[15] Below are the steps in order to use this new tool properly. It is important to note that the published algorithm was modified to significant enhance computational speed; the version in MSiReader v3.03 is over 100 times faster than the published algorithm.

The first step is to create to a list of *m/z* values from the imaging data prior to running the ICT algorithm. This can be done using two different methods:

Method 1 uses the polygon ROI tool to draw an ROI across the entire image. Right click on the image and "select all pixels for the ROI". Next, under Annotations > Data Export, launch MSiSpectrum. Under "Algorithm for Peak Centroid Calculation" – make the appropriate selection for the data. If the data is already centroided, the end user must select "Local Maxima". If the data is profile data, the end user must select "Parabolic Centroid" or "MS Peaks". The user can also apply an abundance filter at this step as well. Next, click "Browse" and enter a filename for these data. This will generate a *.xlsx file that will be used in Step 2.

Method 2 can also be used by exporting the annotation file from METASPACE that can be used in Step 2.

1. Launch the ICT algorithm which is found under the menu item pre-processing and sub-GUI will be displayed as shown in **Figure 52**.

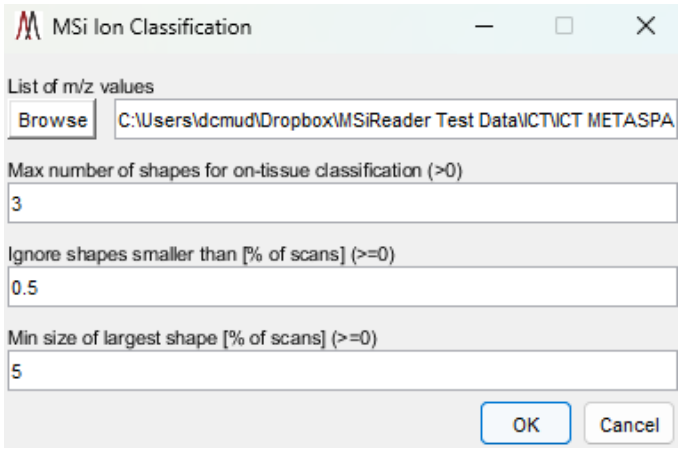


Figure 52: ICT sub-GUI where the list of *m/z* values (generated by method 1 or 2 above) is selected as well as user-defined variables for the algorithm.

2. The default values for the ICT are based on experience; however, if you are imaging a multi-organ system (*e.g.*, zebrafish), it is normal for a molecular to be distributed in specific organs only and thus, the maximum number of shapes should be increased to allow for that heterogeneity. The higher the value, the more conservative the ICT is to not calling a species a background ion.

3. Click OK to run ICT. After it is done, it will prompt the user to give a filename for the output Excel file which will have 4 worksheets: 1) unified results: 2) On tissue; 3) Background; 4) mz out of range; and 5) not detected. These are provided so that the end-user can look at the results and perhaps make changes to the variables that go into the ICT specific to their study.

   The logic for determining which classification is chosen, is based on object-based image analysis (OBIA) is as follows:

   If no shapes are detected, the ion is classified as **not detected** (essentially low detection frequency)**;**

   If the largest shape consists < 5% of MSI scans, the ion is classified as **background** regardless of the number of detected shapes;

   Given an entry for the number of shapes = 3, if 4 or more shapes are detected, the ion is classified as **background**; if 1-3 shapes are detected, the ion is classified as **on-tissue**.

4. This Excel file can then be used to filter the dataset using the peak exclusion filter (**§7.3.2)** to remove these background ions from the data set and then re-writing a new imzML file. In this case, when the user checks the peak exclusion filter, they will be prompted to choose which worksheet contains the data they wish to exclude. This is an important pre-processing step prior to doing downstream statistical analysis. For example, in DESI and ESI post-ionization methods, removing significant numbers of ambient ions from the data is critical to ensure that an end-user is not, for example, using PCA, separating out a cancer versus healthy tissue sample based on these ambient ions. In MALDI, matrix ions should be removed from the data prior to further processing as these can also drive incorrect conclusions which have nothing to do with disease versus healthy but variability in the MALDI matrix ion signals.

7.4 The Visualization Menu

7.4.1 Abundance Rank

A rank plot is a quick tool to plot *m/z* spatial distribution as a function of rank in abundance. Using the drop-down menu, choose abundance rank and a dialog box will pop up. If the user selects 1, the *m/z* value for the most abundant peak (base peak) at

each scan will be plotted on the heatmap. By selecting 2, the *m/z* distribution of the second most abundant peak will be shown for every scan. Although the usefulness of these plots is limited to higher abundance peaks, it is a quick way to extract some features. An example of this type of plot is shown in **Figure 53**. Two icons have been added to the rank plot toolbar. Clicking on the  icon updates the MSiReader heatmap with a data cursor *m/z* value. If there are multiple data cursors the user is prompted to select one. The selected *m/z* value is also added to the *m/z* history list when the heatmap is updated. The  icon appends all of the data cursor *m/z* values to the clipboard.

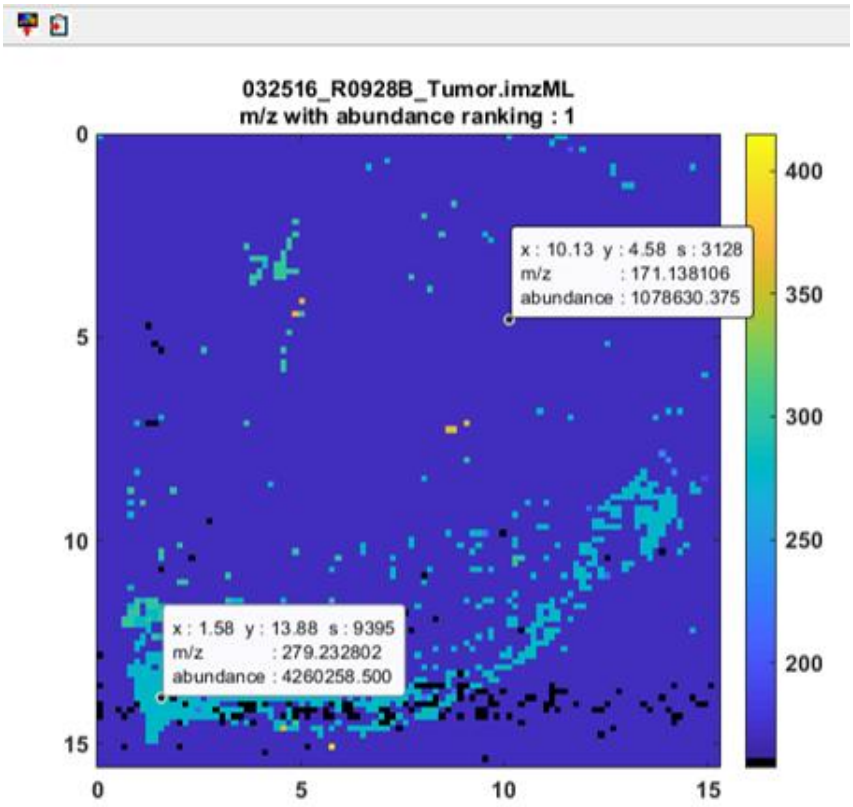


Figure 53: Example of an abundance rank plot. The data cursors show the X,Y location, scan number and *m/z* values for the most abundant ion in scan 9395 (*m/z* = 279.2328) and for a widely distributed ion at scan 3128 (m/z = 171.1381).

7.4.2 Total Ion Current (TIC)

The total ion current (TIC) for each scan is plotted in a heatmap by as shown in **Figure 54**. This is simply a tool to visual the TIC at each pixel across the heatmap.

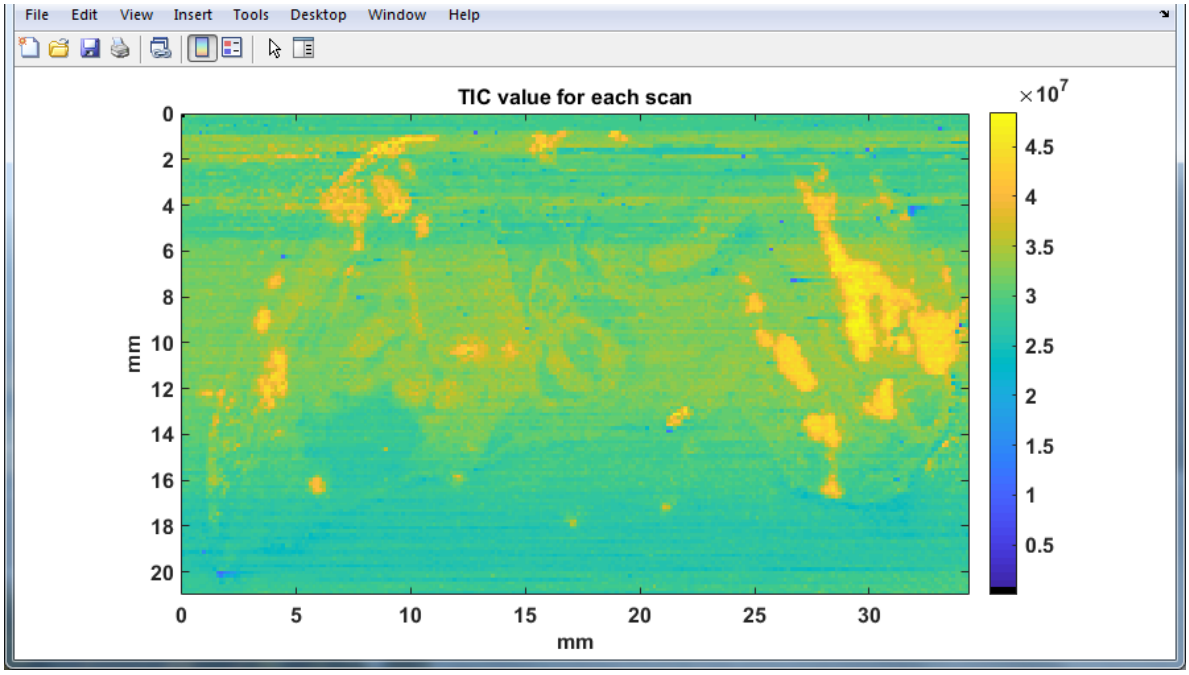


Figure 54: The total ion current for each scan in the image. Notice that in this TIC heatmap there are distinct regions that have higher abundances than other regions. This could be biological in origin based on the structure and or tissue type *(e.g.*, cancerous or healthy) or could be related to the variability of the analytical platform.

7.4.3 Number of analytes

The density of *m/z* values (analytes) across an image can quickly be viewed by clicking on “number of analytes” from the drop-down menu under visualization. A heatmap whose color is proportional to the number of *m/z* values in each scan is plotted in a new figure. An example is shown below in **Figure 55.**

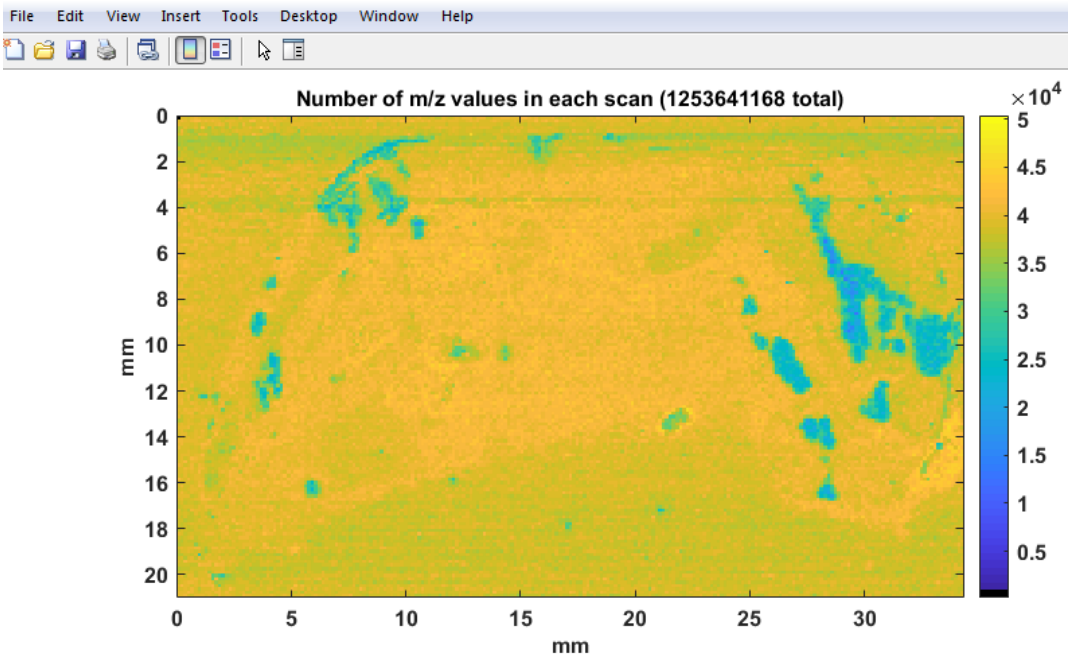


Figure 55: The number of *m/z* values for each scan.

7.4.4 Summed m/z abundance

Images for a list of *m/z* values can be summed by choosing the “summed *m/z* abundance” in the drop-down menu under visualization. The user is prompted to enter a list of *m/z*

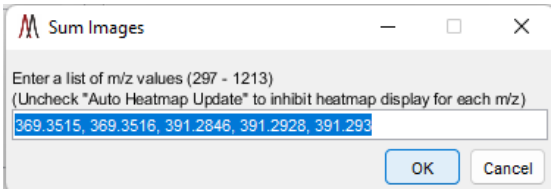


Figure 56: Image summation dialog.

values separated by commas or spaces as shown in **Figure 56**. The default entry is the most recent five values in the *m/z* history list.

A heatmap showing the total ion abundance of the *m/z* values chosen by the user is exported and an example is shown in **Figure 57** and the user is prompted to save the summation matrix into a text file. The text file can be loaded as a custom heatmap (§7.2.4) and used to normalize the loaded data set (§7.2.5). A second text file is also saved containing the *m/z* list. Note that the summation is for the normalized and windowed *m/z* as displayed using the criteria in the MS Navigation pane, not the abundances of the RAW scan data.

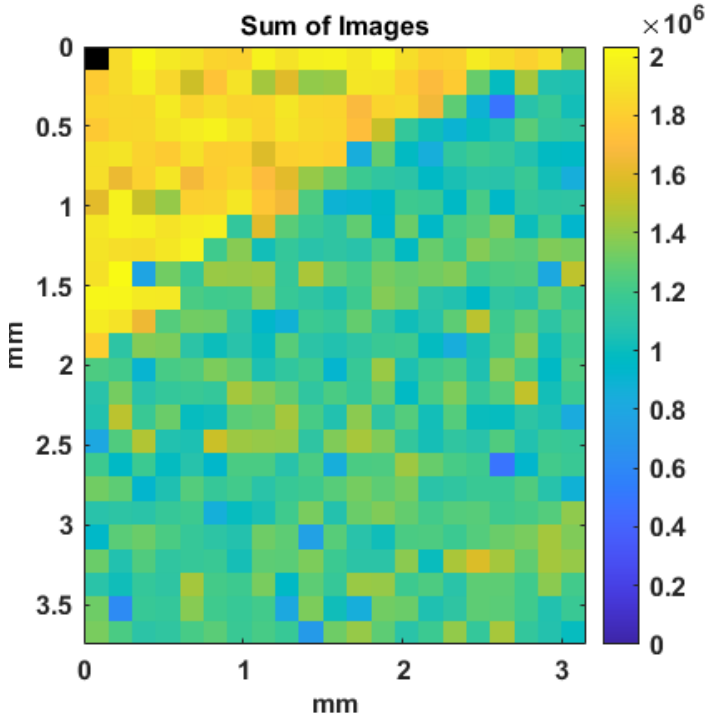


Figure 57: Heatmap of the summation of the user selected *m/z* values.

7.4.5 Spectra above and below a threshold

The heatmap of the distribution of scans above and below a threshold for the current *m/z* chosen in the MS Navigation pane is carried out using this tool. The user is prompted to enter an abundance threshold and an abundance tolerance with the dialog box shown in

Figure 58. The default values are the median abundance and 1/100th of that value for the abundance tolerance for the current *m/z.* A plot similar to the one in **Figure 59** is displayed showing the distribution of scans whose abundance is within the tolerance range in white, below *(threshold – tolerance)* in blue and above *(threshold + tolerance)* in red. Scans not in any selected ROI are shown in black. The plot colorbar has been customized to show the number of scans in each of these four categories.

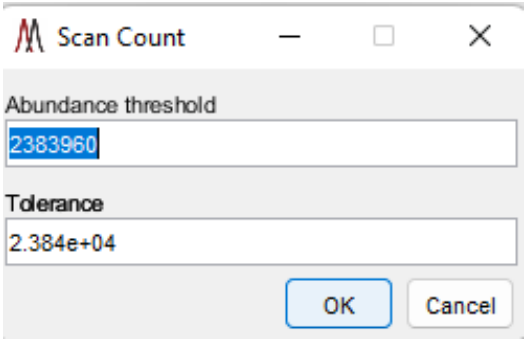


Figure 58: Threshold and tolerance dialog for the abundance distribution plot.

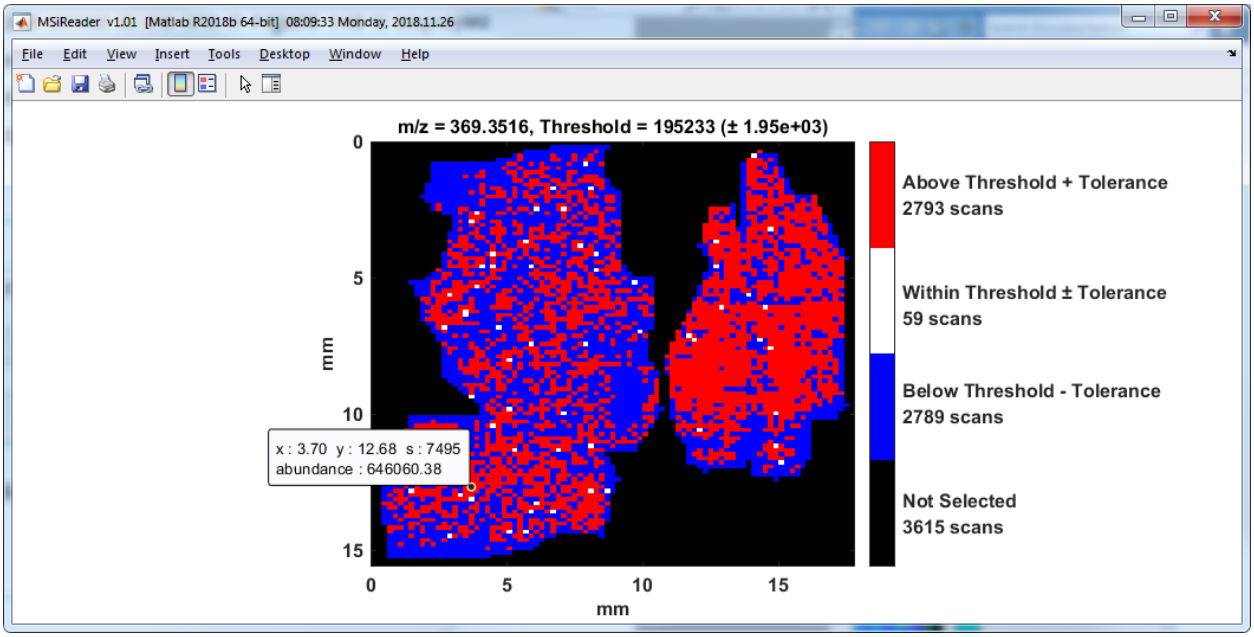


Figure 59: Abundance distribution relative to user defined threshold.

7.4.6 MSiSlicer

Selecting MSiSlicer in the pull-down menu, the current heatmap is loaded into a new GUI called MSiSlicer. The cursor immediately changes to a +, and the user can draw a segmented line ROI across the image. As shown in **Figure 60**, MSiSlicer then displays the ion abundance (bottom) along a segment line in the lower plot window. Black x's mark the positions of the connecting points in the ROI. As the line is moved or the points edited the ion abundance plot is automatically updated. Three plot styles: line, stem and stairs, can be selected from a pull-down list. A checkbox locks the vertical axis of the plot and prevents the axis from automatically updating to accommodate the changing abundance range as the line is moved. The left panel of the GUI also contains information about the applied slice and tools to invert the heatmap colors, refresh the plots and redraw the ROI. The colormap can be edited by right-clicking on the colorbar to the right of the heatmap.

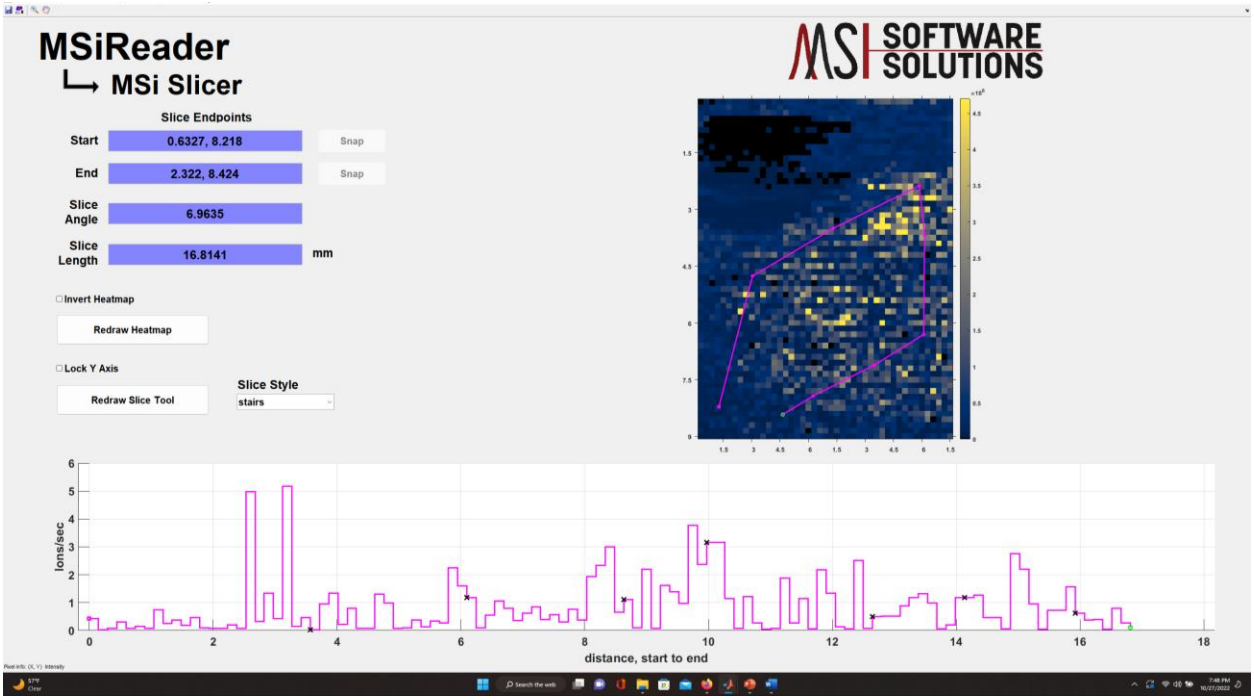


Figure 60: MSiSlicer GUI showing a "stairs plot" of the abundance of cholesterol across this tissue.

By pressing the icon in the MSiSlicer GUI, the heatmap will be extracted as a 3D figure where abundance is simultaneously represented as a heatmap and as an elevation on the z axis as shown in **Figure 61**. The plot shown is a 3D stem plot. Either *stem3* or *surface* can be selected in the preferences INI file (**§5).** The view of the 3D heatmap can

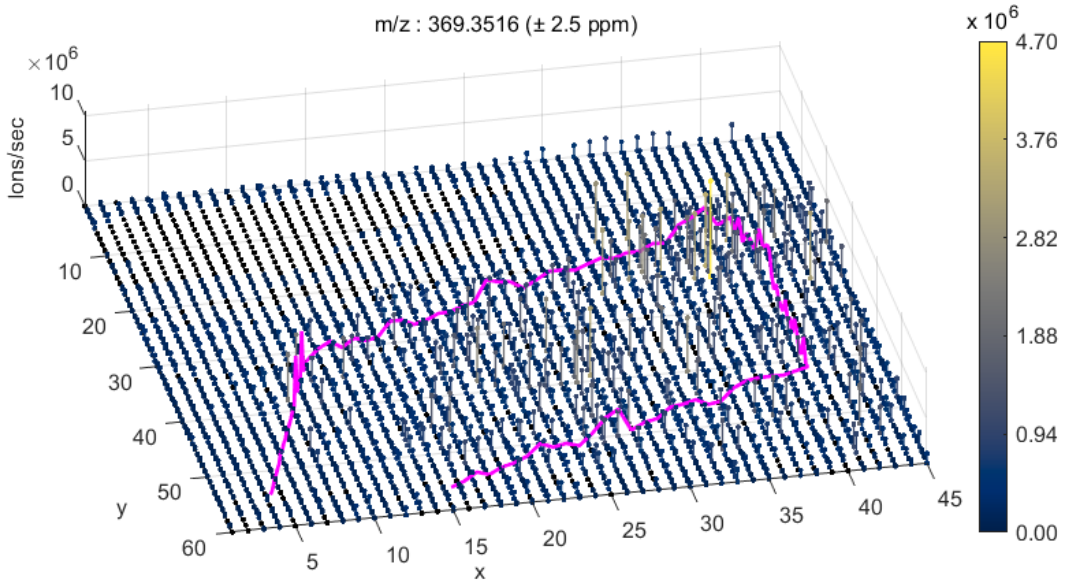


Figure 61: 3D Heatmap extracted from MSiSlicer.

be rotated by clicking on the icon and then dragging the pointer over the figure. In addition to the 3D heatmap, the ion abundance plot is also extracted into a new window. Both figures can be saved to another format (*e.g.* .jpg, .png) from the *File/Save as* menu or saved as Matlab .fig files.

The data used to generate the 3D heatmap and the graph can also be extracted into an Excel workbook by clicking the icon on MSiSlicer's toolbar. In addition to information about the data set, the workbook will contain the raw heatmap ion abundance data as a matrix and the interpolated data used to generate the ion abundance plot (scan location and abundance vs distance along the segmented line) in separate worksheets. The exported heatmap data matrix can be read back into MSiReader as a custom abundance heatmap and used for normalization. Using this approach, you can normalize an image to a reference peak from another image, provided the images are the same size.

7.4.7 Image Overlay

The MSiImage tool can be used to combine another image with the heatmap, for example, an optical image of the same tissue. We recommend using third party tools to prepare your optical image prior to importing it into MSiReader; our tool works on editing images but it not overly sophisticated. The image can be in any graphics file format that Matlab can read (*e.g.*, png, jpg, tiff) and any image can be used as the overlay including an exported heatmap plot for a different *m/z* value or even another tissue sample.

To use the tool, click on *Image Overlay* in the drop-down *Visualization* menu. This will open the MSiImage interface containing the current molecular image in the main MSiReader GUI as shown in **Figure 62**. After pressing the  icon, the user is asked to select an optical image file which will be resized to fit within the axes and displayed on top of the heatmap as shown in **Figure 63**.

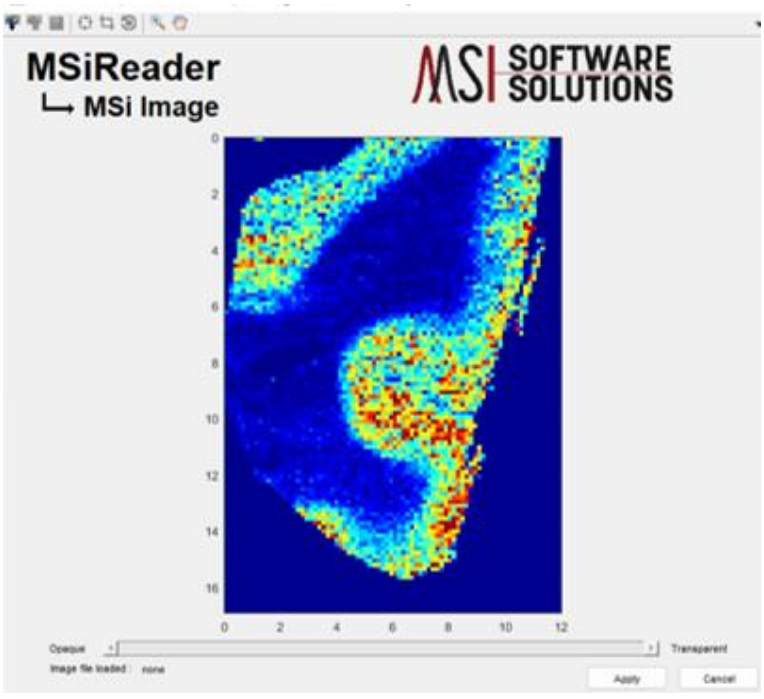


Figure 62: MSiImage loaded with current heatmap (*m/z* = 329.2475).

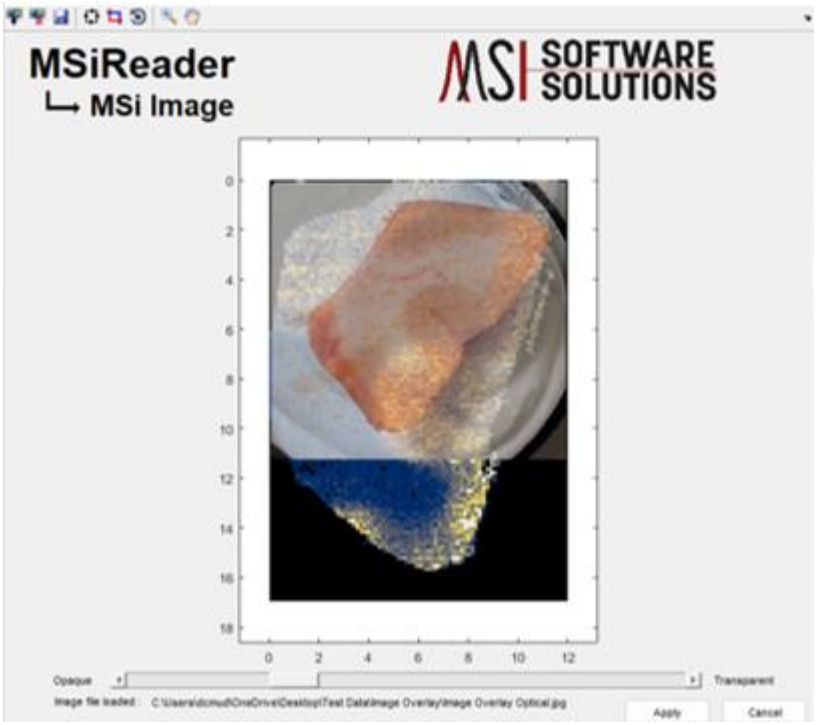


Figure 63: MSiImage after inserting an optical image (no alignment as been done at this stage). The toolbar contains icons to resize, crop and rotate the optical image overlay. Transparency of the optical image can be adjusted with the slider bar at the bottom.

The overlay image can be aligned with the underlying heatmap using the adjustment icons on the MSiImage toolbar. They are move/resize , crop and rotate . It is recommended to crop your image prior to loading or do that first using the tool in MSiReader. The image aspect ratio is 1:1 and locked by default but it can be unlocked by right-clicking on the heatmap after selecting the move/resize tool. The zoom and pan tools remain functional while adjusting the overlay. The rotate tool rotates the image about its center using mouse motion as input. A motion magnification factor, *ImgRotateMag*, can be set in the preferences INI file to speed up or slow down the rotation (**§5).**

Transparency of the optical image can be adjusted at any time using the slider bar at the bottom. After you are satisfied with the alignment and any resizing or cropping you have done the resulting image overlay can be saved for future use by clicking on the save icon

 as a .png file. The .png file contains the image as seen in MSiImage. The image overlay can be removed by clicking the  icon.

Upon clicking the *Apply* button, the MSiImage tool will close and the optical image combined with the MSI data will appear in MSiReader as shown in **Figure 64**. A transparency slider bar is added to the MSiReader main window under the *m/z* slider bar.

All MSiReader tools are fully functional with the overlaid optical image (browsing, data extraction, MSiPeakfinder, batch processing of images, etc.). Transparency of the optical image can be readjusted at any time using the bottom slider bar in the main MSiReader interface. Hint: You can make the optical image temporarily disappear by making it 100% transparent. Alternatively, the user can click on Remove overlay in the main MSiReader GUI on the bottom right-hand corner.

At any time, the user can press the MSiImage button again to realign the optical image, erase it, save it, or load a new one.

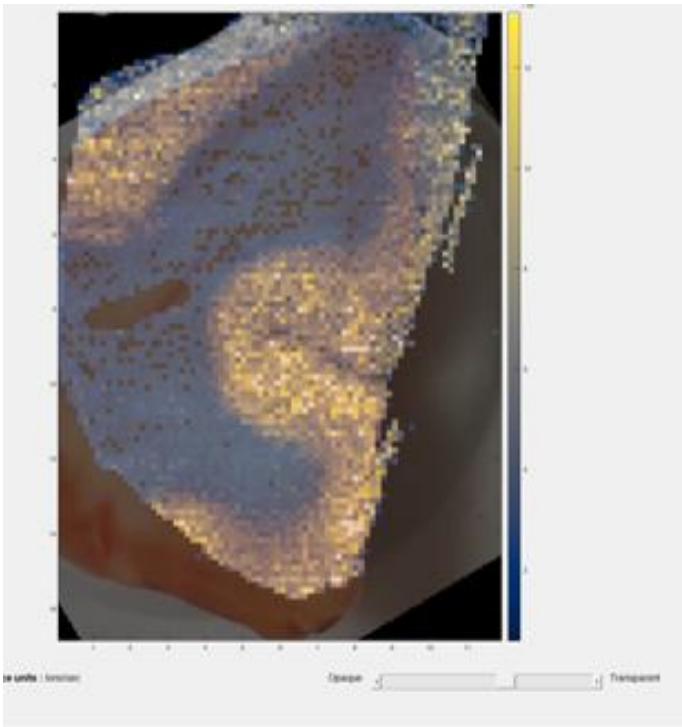


Figure 64: Overlaid optical picture and molecular ion map. The overlay transparency level is adjusted using the slider bar. To remove the image overlay, click on remove overlay on the bottom right-hand corner. This is NOT a toggle, once removed the user will have to go back into the image overlay tool to recover it. However, to temporarily hide the optical image, move the slider bar to 100% transparent.

7.4.8 8-Color Colocalization Plots

A **video tutorial** on generating colocalization color plots can be found **HERE**.

The user can overlap up to eight heatmaps using the color channels in MSiReader. Spatial overlap is often used to perform *qualitative* comparison of the distribution of specific molecules over the sample surface. To create a colocalization image, save figure (.fig) files for up to 8 individual heatmaps that you want to overlap. Any interpolation scheme, hotspot removal, tolerance, etc. can be used provided that these are the same for all the figures. When all the images are saved, select *Colocalization Plot* in the drop-down visualization menu to launch the colocalization interface. Using the interface, select a color to apply and browse to choose the corresponding figure file (See **Figure 65**); the

red color channel is automatically selected and the remaining 7 channels can be enabled by checking the box next to them. If the user only selects the red, green and blue color, the default setting is to normalize channels separately and produce a blended plot of the channels that the user has chosen. However, the end-user can also change the normalization to max of all channels at which time a slider bar to change the gain for each of the three colors. Moreover, the plot mode can also be changed from blended to dominant mode; a dominant plot means that whatever *m/z* value is dominant in that pixel, that color will be displayed. If the end-user adds additional colors beyond red, green and blue, the plot is fixed and will produce a dominant plot of the color channels that are normalized separately.

Note 1: Separate .fig files are used as the input so users can integrate complex normalized heatmaps or custom heatmaps into colocalization plots.

Clicking the  icon will save the colocalization plot as a .fig file with relevant information in the title.

Another image can be overlaid on the colocalization heatmap using the toolbar icons for loading, deleting, moving, resizing, cropping and rotating. The slider bar at the bottom of the MSiColocalization GUI can be used to adjust the opacity of the overlay once an image is loaded. See **§7.4.7** for details on using the image overlay tools. The slider bar is only displayed if an image file is loaded to map to the mass spectrometry imaging data.

The data for each color channel is divided by a normalization scaling factor and multiplied by a gain. The gains are initially set to one and the slider bars for relative color intensity range logarithmically from 0.00001 to 100000. The channel normalization factors can be selected with the right-click context menu in the *Gain* panel in MSiColocalization as shown in **Figure 66 (**right-click on the word Gain to access this menu). None means that the data

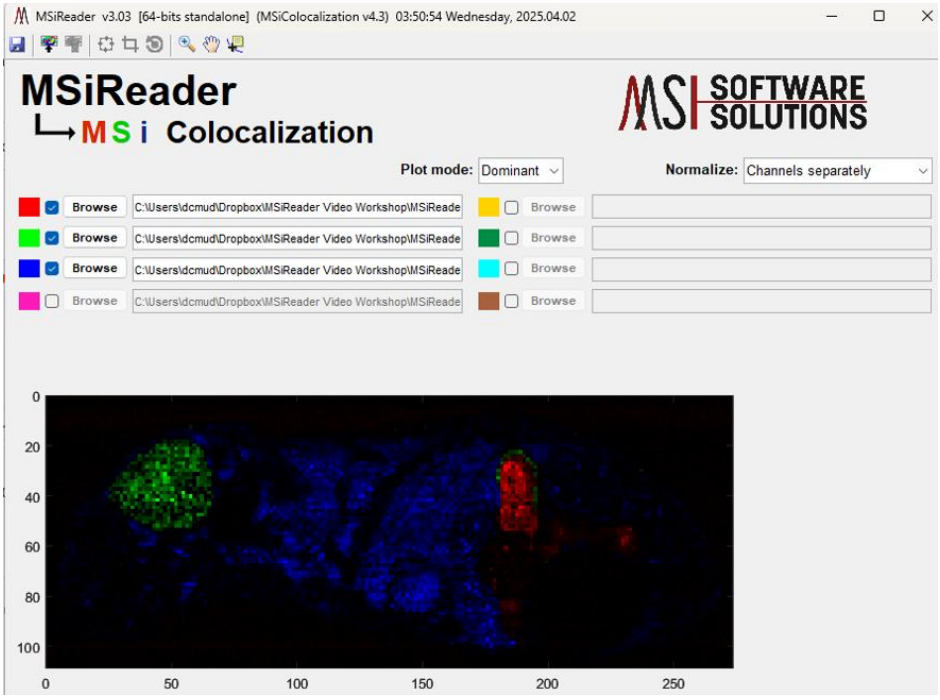


Figure 65: MSiColocalization interface showing three m/z values on a whole mouse tissue imaging dataset (red channel is *m/z* = 303.2531, green channel is *m/z* = 367.3301 and the blue channel is *m/z* = 617.1808. The default is to normalize each channel separately in blended mode and this is what is shown.

is not normalized (*i.e.*, the scaling factor is unity). The other two choices normalize each channel to its maximum abundance value or globally to the maximum abundance in any of the data sets. When the normalization method is changed the colocalization plot is immediately updated. The default method can be set in the preference INI file with the *ColocalNormOption* value (**§5).**

Note 2: The figure files do not have to be the same size. The smaller figures will be resized to match the largest one. The maximum allowed size difference in either the column or row dimension is 80%. This value can be changed in the preference INI file (**§5)**.

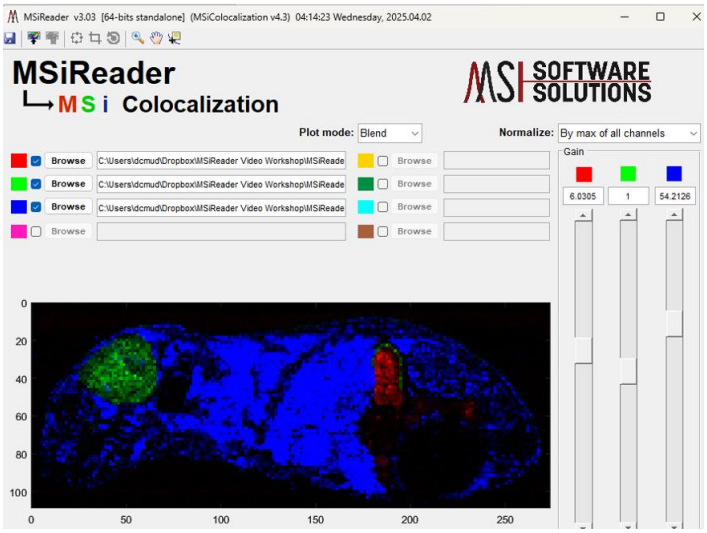


Figure 66: Color channel normalization menu. Click on “Normalize” and choose from None, Channels Separately, or Max Abundance in all channels.

Note 3: DO NOT use .fig files that contain drawn ROIs or image overlays. The user can add image overlays after the co-localization plot is made using the icons in the toolbar in the co-localization GUI.

7.4.9 3D Plotting

For a **video tutorial** on how to generate 3D Heatmaps, click **HERE**.

Three 3D plots are available by selecting 3D Plotting from the drop-down menu. The choices are mass spectra plot, an image stack and a 3D colocalization plot. The spectral plot is either a waterfall line plot or stem plot for a selection of previously exported centroid or average spectra. The image stack plot is either a stack of spatial heatmaps, one each for a list of *m/z* values or a stack of spatial heatmaps, one each for the files in an image mosaic. 3D colocalization plots are a stack of image layers, one each for a set of previously saved .fig files created by the MSiColocalization tool.

7.5 QA/QC Menu

7.5.1 Mass Measurement Accuracy

A **video tutorial** on the use of the MMA QA/QC tool can be found **HERE**.

MSiReader provides tools for the calculation and plotting of mass measurement accuracy (MMA) for any *m/z* in an ROI or for the entire image. Access this tool by selecting *Mass Measurement Accuracy* under the QA/QC menu. For a given *m/z* and tolerance, MSiReader finds the most abundance peak, max_peak_k, in each scan that is within the tolerance window. It then calculates the MMA for each scan as,

$$MMA_k = \frac{max\_peak_k - m/z}{m/z} \times 10^6$$

After the MMA value is calculated for all scans in the ROI (or image), several types of plots can be produced or the MMA data can be saved into an Excel or text file. The submenus for the MSiReader heatmap axes have seven items for selecting these MMA functions as shown in **Figure 67**. Each is described below. Note that if an ROI is active when an MMA function is invoked the user is prompted to select either *all the scans* or *only the ROI scans* for processing.

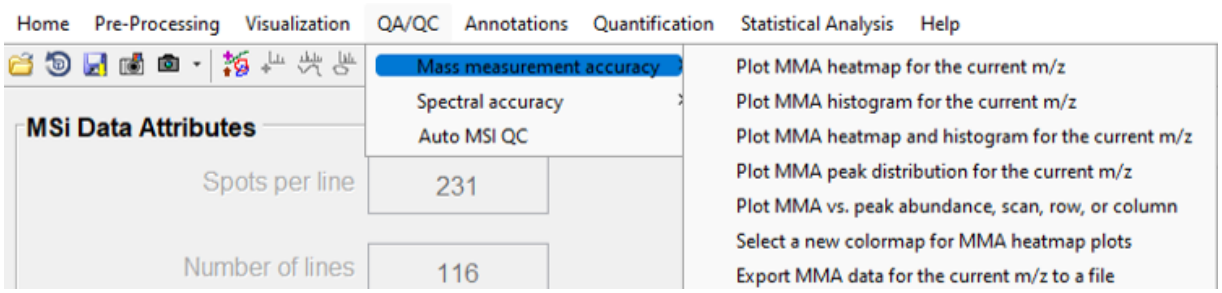


Figure 67: Mass measurement accuracy drop down-menu choices (described below).

7.5.1.1 Plot MMA heatmap for the current *m/z*

A mass measurement accuracy heatmap for the current *m/z* center value and tolerance is displayed in a new figure. The color of each pixel in the exported heatmap is proportional to the MMA value of the most abundance peak in the *m/z* window. The default colormap used for the plot is a balanced colormap with the most intense color in the center

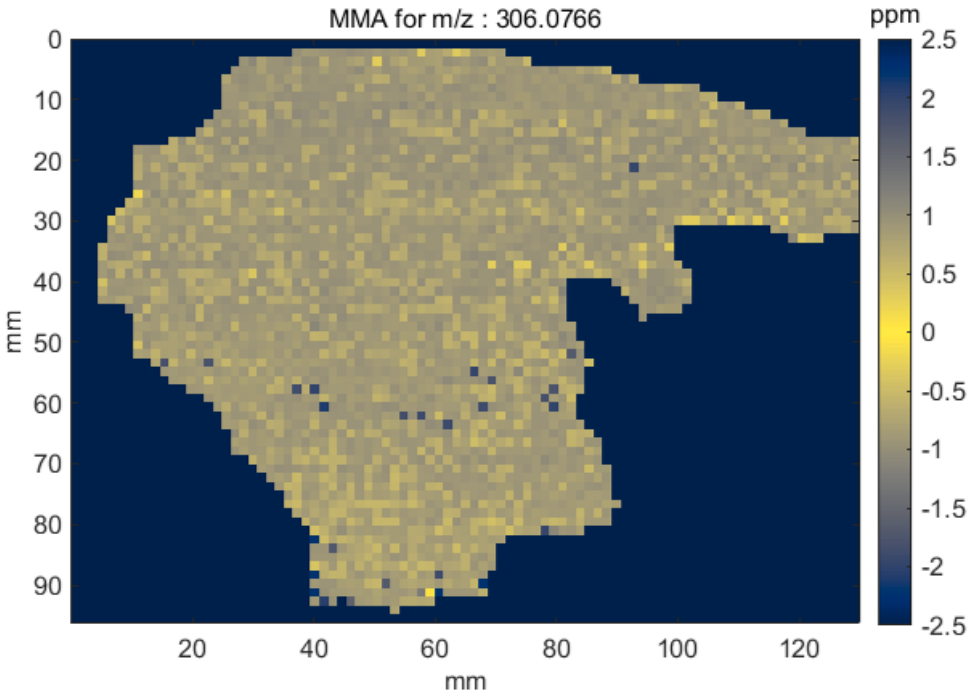


Figure 68: MMA heatmap for *m/z* = 306.0766

where the MMA equals zero and the least intense color at the plus and minus limits of the *m/z* tolerance (set these in the main GUI for MSiReader). The colormap used for the MMA heatmap is specified by a preferences INI file variable *MMAColorMap* (§5). The default colormap is *parulahi.mat*. An example of this plot is shown in **Figure 68**.

The MSiReader installation folder includes the default colormap as well as a parula based colormap with the highest intensity at the *m/z* tolerance limits and the lowest intensity at the center value. It is named p*arulalo.mat*. Balanced versions of six other colormaps are also in the colormap folder, *msicolormaps*.

The data tips tool (“transparent menu” above the heatmap, second from the left) can be used to query the spatial coordinates (X, Y, and scan number), the MMA value and the most abundant peak (*m/z* and abundance) that was used to calculate MMA for that scan.

7.5.1.2 Plot MMA histogram for the current *m/z*

Two histogram plots are produced. One shows the number of scans in each ppm bin and the other normalizes the bins counts such that the height of each bar is proportional to

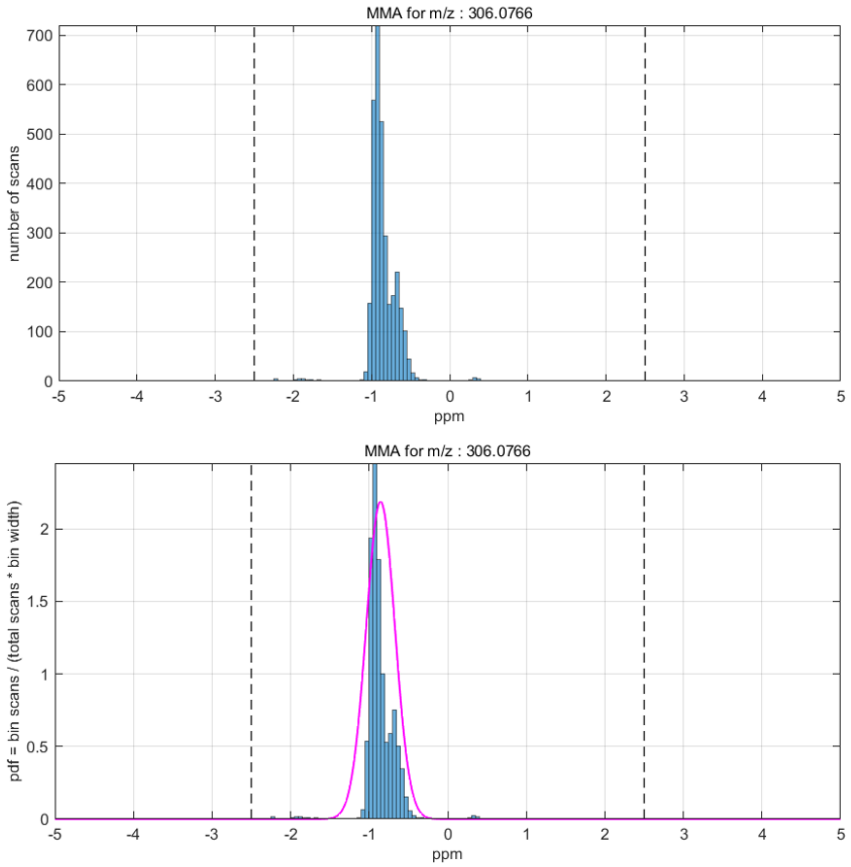


Figure 69: Mass measurement accuracy histogram and probability density function.

the number of scans in each bin divided by the product of the total number of scans and the bin width (*i.e.*, the probability density function or PDF). The area of each bar is the relative number of observations. A Gaussian normal curve with the mean and standard deviation of the binned data is also plotted on the probability density histogram. An example of these two plots in **Figure 69** shows a systematic mass shift of -0.9 ppm.

Two preferences INI variables (**§5)** control the histogram bar direction, M*MAHistogramDirection*, and the bin selection method, *MMAHistogramBinMethod*. The

default values are *vertical* and *auto*. The bar direction can also be horizontal as shown in **Figure 70**.

By default, the histogram has the same *m/z* window size as the heatmap. However, the *m/z* window for the histogram can be extended (by a multiplicative factor) to reveal any significant values outside of the tolerance window. This is done by setting the *MMAHistogramMargin* variable in the INI preferences (§5) file to rescale the *m/z* window for the histogram calculation. The default value is 1 ppm.

7.5.1.3 Plot MMA heatmap and histogram for the current *m/z*

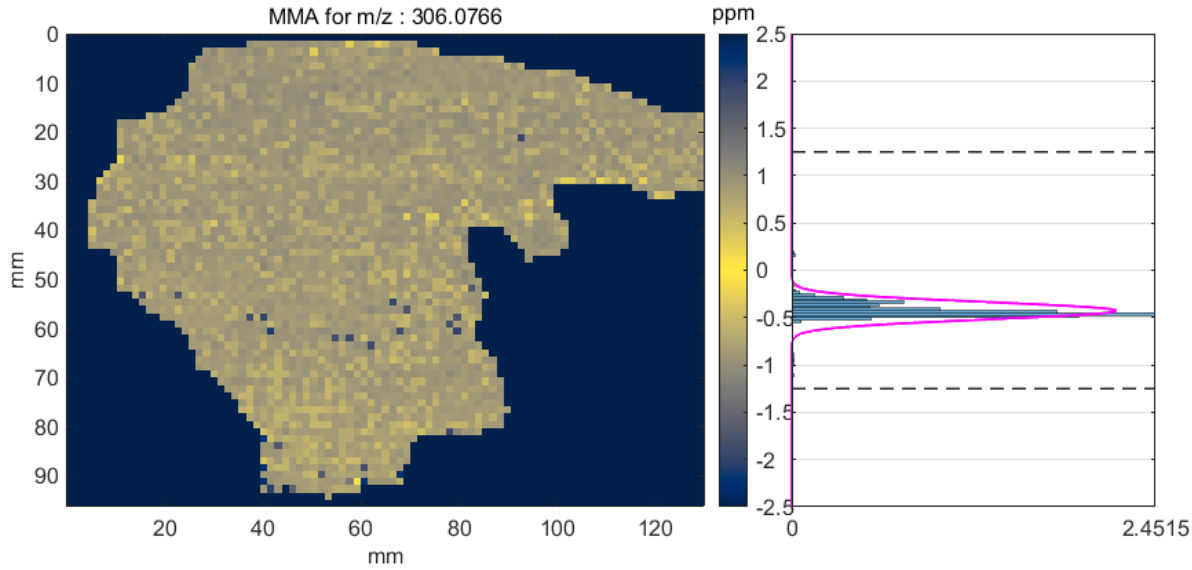


Figure 70: Mass measurement accuracy heatmap and histogram. Please note that the heatmap uses the scale from 2.5 ppm to -2.5 ppm while the histogram goes from 5 ppm to -5 ppm (the dashed lines are the ± 2.5 ppm limits) but the values are not listed on the dual plot as it becomes too crowded.

This plot is a combination of the first two plots; a mass measurement accuracy heatmap and a PDF histogram. The histogram is oriented in the horizontal direction. Elements of this figure can be moved, resized or deleted using the figure toolbar edit tool (left leaning arrow). If the MMAHistogramMargin Preferences INI (§5) value is larger than one, the histogram window is extended as described above. The dotted lines in the figure delimit

the *m/z* window used to calculate the MMA and the limits of the plot show the additional margin value described above.

7.5.1.4 Plot MMA peak distribution for the current *m/z*

An animated scatter plot is drawn showing the distribution of MMA vs ion abundance for all peaks in the *m/z* window. The animation is by scan number and the pixels for each scan have the same color. The data tips tool ("transparent menu" above the heatmap, second from the left) can be used to query the spatial coordinates scan number, MMA and abundance.

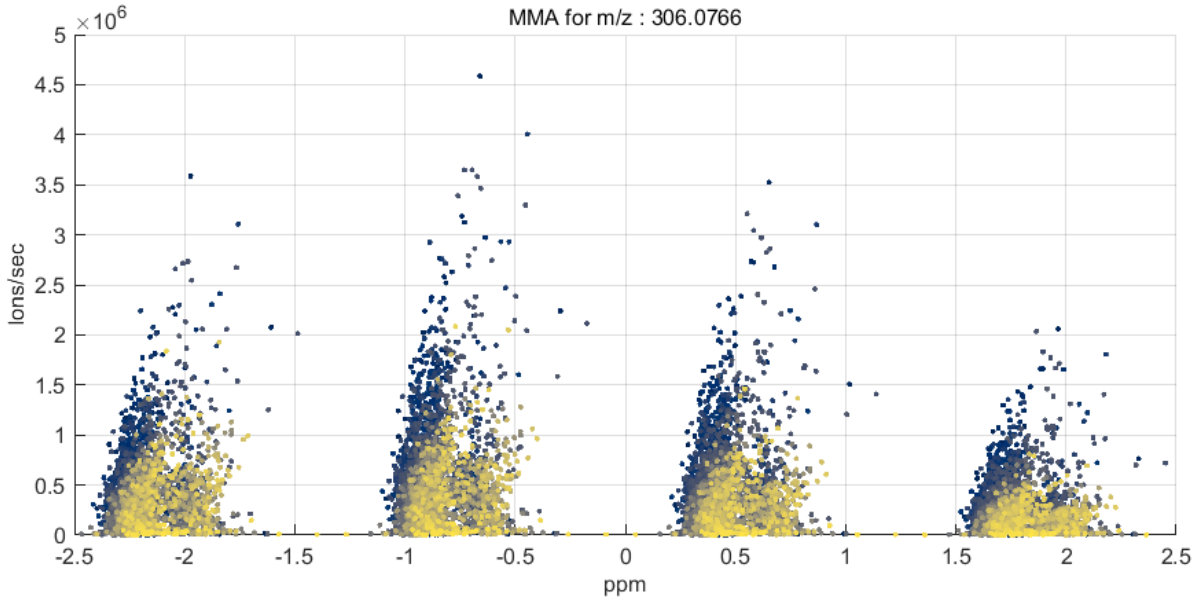


Figure 71: Mass measurement accuracy peak distribution.

Note 2: The *mmaabundance* and *mmadistribution* plots are 3D plots viewed from above the XY plane. The Z direction is the scan number. This can be seen by using the figure toolbar rotate icon  to change the viewing angle. After enabling this tool, the right-click context menu can be used to quickly view the data relative to the XY, XZ or YZ planes. For example, when **Figure 71** is rotated as shown in **Figure 72**, not only is the MMA clustering visible, but it can be seen that the MMA shifted (-) and then (+) slightly as the sample was scanned.

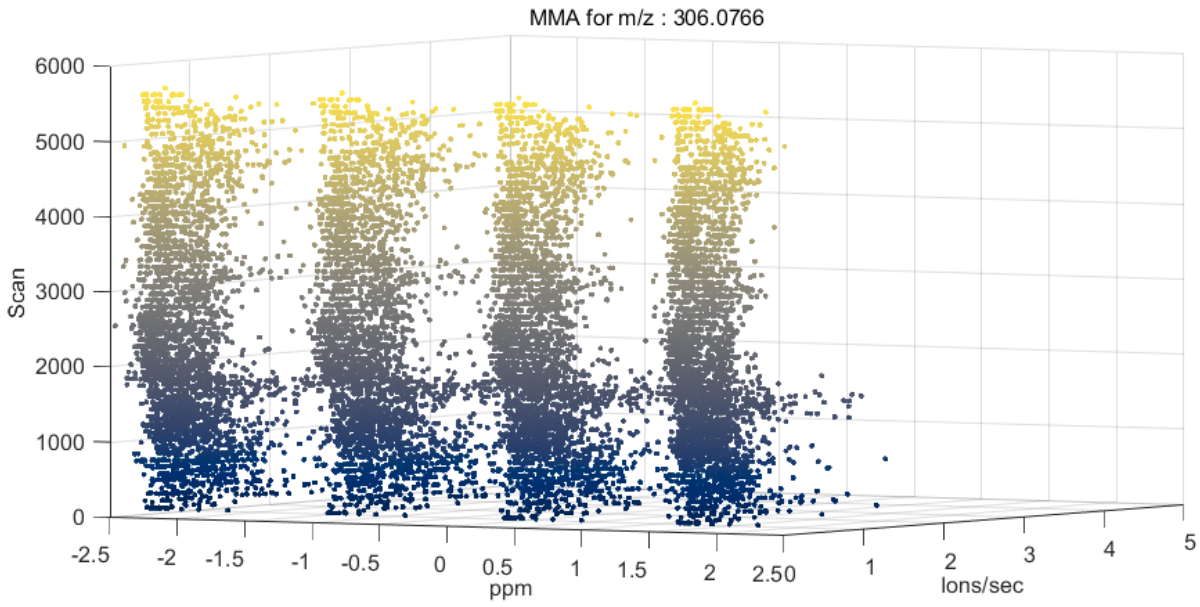


Figure 72: Mass measurement accuracy peak distribution by scan.

7.5.1.5 Plot MMA vs. peak abundance, scan, row, or column

A scatter plot is produced with a point for each scan showing the abundance vs MMA values used to produce the MMA histogram. This is the raw data that is used to make the histogram plots. The data cursor tool can be used to reveal the scan number, MMA value and abundance for any point. There is a drop-down menu on the bottom left of **Figure 73** for which the end-user can select ions/sec (abundance), scan number (time), row or column. This sets the x-axis accordingly. Moreover, if the user wishes to make a plot where the y-axis is absolute MMA (the direction of the MMA does not matter), click the "Use ppm" checkbox in **Figure 73**.

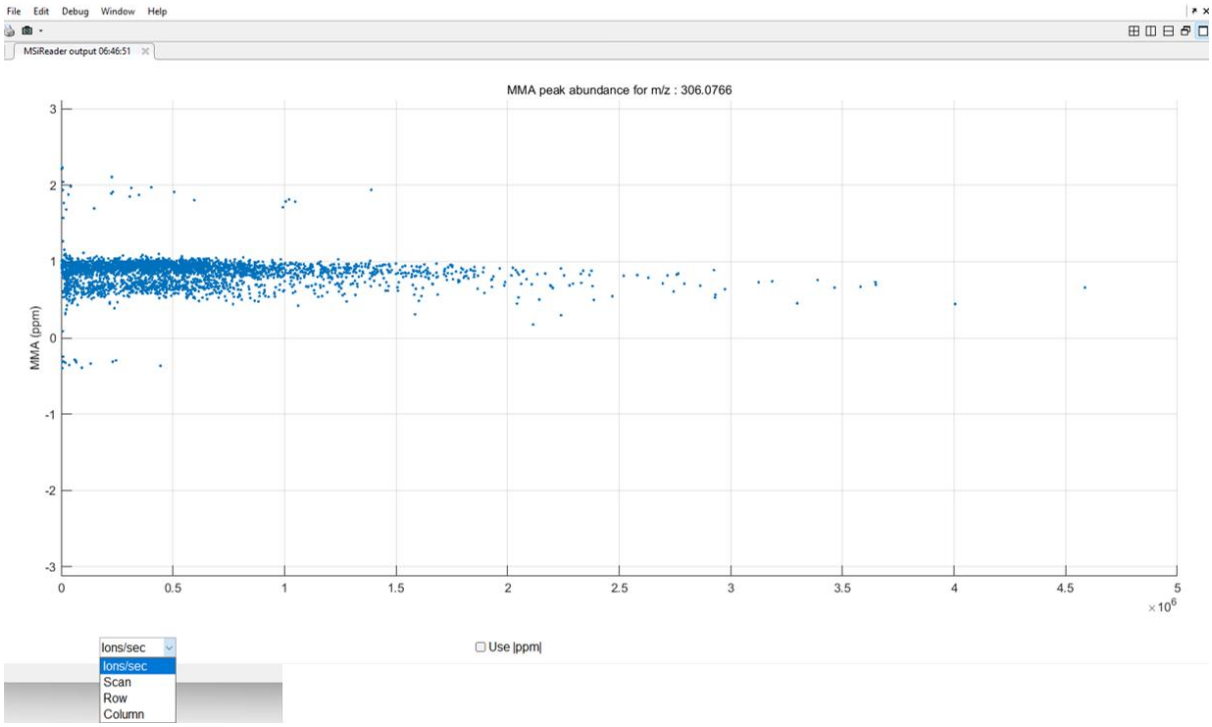


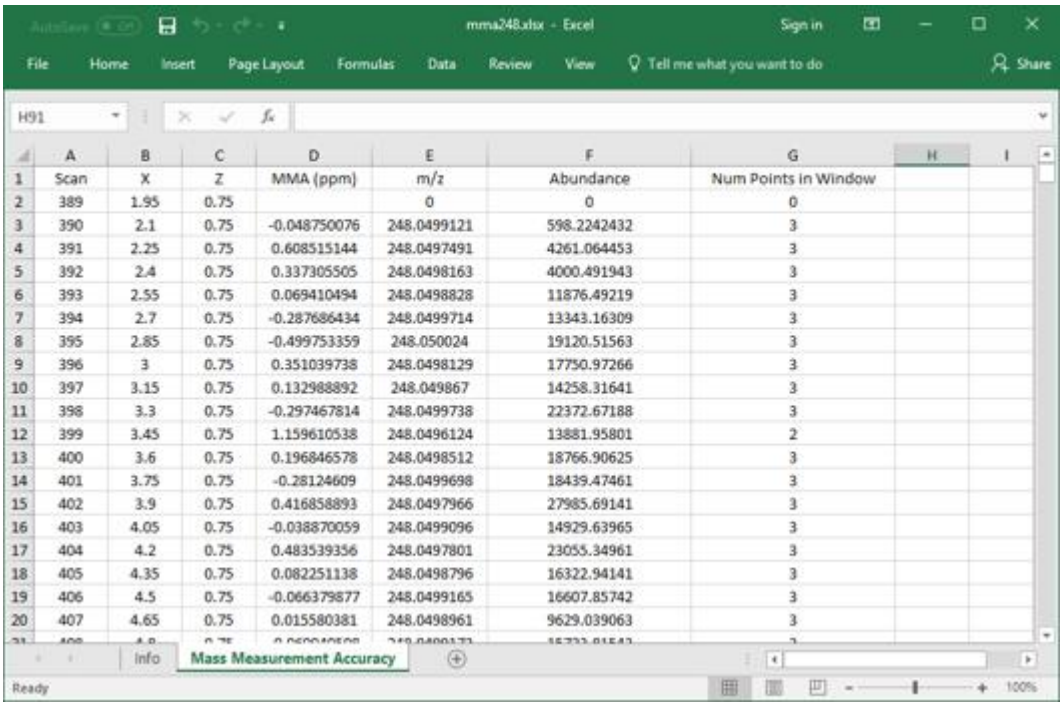
Figure 73: Mass measurement accuracy as a function of the ion abundance of a specific *m/z* value.

7.5.1.6 Select a new colormap for MMA heatmap plots

Allows the user to select a MMA heatmap colormap of their choice. Most of these will not be bi-directional. In other words, going to + MMA values will be a different color than going to – MMA values.

7.5.1.7 Export MMA data for the current *m/z* to a file

The MMA heatmap data is calculated and exported to an Excel workbook or a text file if Excel is not available. Included for each scan are its X and Y location, the MMA value in ppm, the *m/z* and abundance of the maximum peak in the *m/z* window and the number of data points in the window. An example of exported MMA data is shown in **Figure 74**.



	A	B	C	D	E	F	G	H	I
1	Scan	X	Z	MMA (ppm)	m/z	Abundance	Num Points in Window		
2	389	1.95	0.75		0	0	0		
3	390	2.1	0.75	-0.048750076	248.0499121	598.2242432	3		
4	391	2.25	0.75	0.608515144	248.0497491	4261.064453	3		
5	392	2.4	0.75	0.337305505	248.0498163	4000.491943	3		
6	393	2.55	0.75	0.069410494	248.0498828	11876.49219	3		
7	394	2.7	0.75	-0.287686434	248.0499714	13343.16309	3		
8	395	2.85	0.75	-0.499753359	248.050024	19120.51563	3		
9	396	3	0.75	0.351039738	248.0498129	17750.97266	3		
10	397	3.15	0.75	0.132988892	248.049867	14258.31641	3		
11	398	3.3	0.75	-0.297467814	248.0499738	22372.67188	3		
12	399	3.45	0.75	1.159610538	248.0496124	13881.95801	2		
13	400	3.6	0.75	0.196846578	248.0498512	18766.90625	3		
14	401	3.75	0.75	-0.28124609	248.0499698	18439.47461	3		
15	402	3.9	0.75	0.416858893	248.0497966	27985.69141	3		
16	403	4.05	0.75	-0.038870059	248.0499096	14929.63965	3		
17	404	4.2	0.75	0.483539356	248.0497801	23055.34961	3		
18	405	4.35	0.75	0.082251138	248.0498796	16322.94141	3		
19	406	4.5	0.75	-0.066379877	248.0499165	16607.85742	3		
20	407	4.65	0.75	0.015580381	248.0498961	9629.039063	3		

Figure 74: Mass measurement accuracy data exported to an Excel file.

7.5.2 Spectral Accuracy

A **video tutorial** on how to use the spectral accuracy QA/QC tool can be found **HERE**.

Access the spectral accuracy (atom counting tool) by selecting Spectral Accuracy in the drop-down menu under QA/QC. This will give the user two different pieces of data. The first, *Plot ion count heatmap*, produces an ion count plot for an *m/z*, which is the heatmap data scaled by injection time. Access this function by selecting “Spectral Accuracy” in the pull-down menu and then “Plot ion count heatmap”. If ROIs are active, the user is prompted to select either *ROI scans* or *ALL Scans*. The user is then prompted to enter an *m/z* value (peak of interest), the injection time, and a colorbar label (**Figure 75**). The plot is the same if a user exported a heatmap in the main MSiReader GUI if injection time was loaded along with the data set. An example is shown in **Figure 76**. This output shows absolute number of ions in a heatmap as it has been shown that spectral accuracy is related to the abundance[16]. This can be used in conjunction with the plot isotope count heatmap (see below).

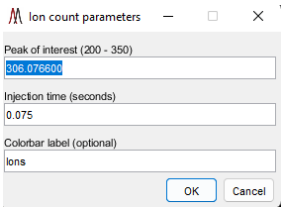


Figure 75: Isotope ratio parameters dialog.

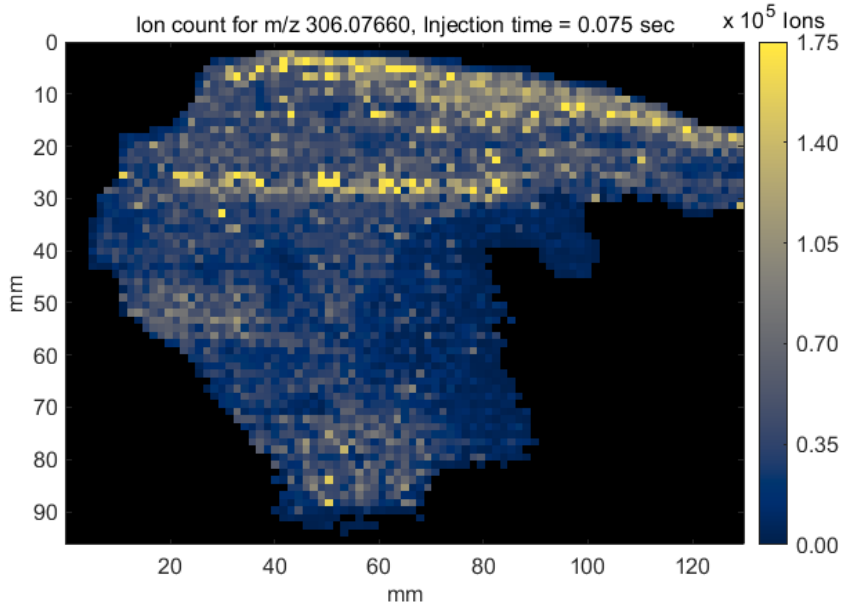


Figure 76: Exported ion count heatmap.

The second option in the pull-down menu is *Plot isotope count heatmap*, which compares the ratio of abundance for two *m/z* values (*e.g.*, the monoisotopic *m/z* (M) and the M+1 peaks of an atom) against the known ratio of those two isotopes. The dialog shown in **Figure 77** is launched for the user to enter two isotopes, the abundance of the heavier

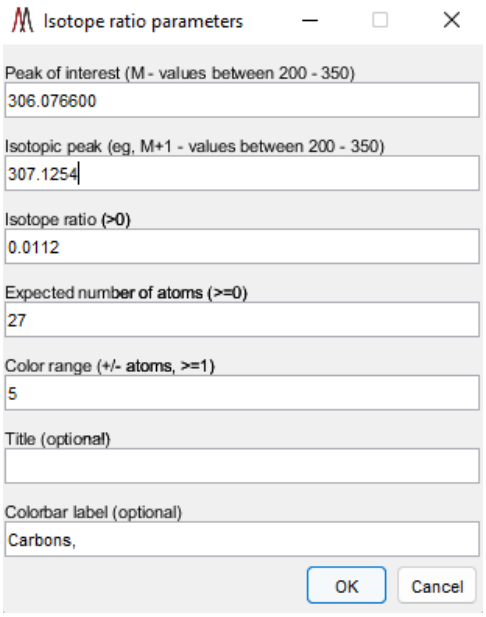


Figure 77: Dialog box for determining the accuracy of counting atoms.

isotope, the expected atom count, a tolerance range and optional labels. Once all user input is entered, select OK and then a heatmap will be displayed (which can be saved) and the user will be prompted to enter a filename (abcd.xlsx) to save the metadata as well as the spectral accuracy data in Excel format in two different worksheets.

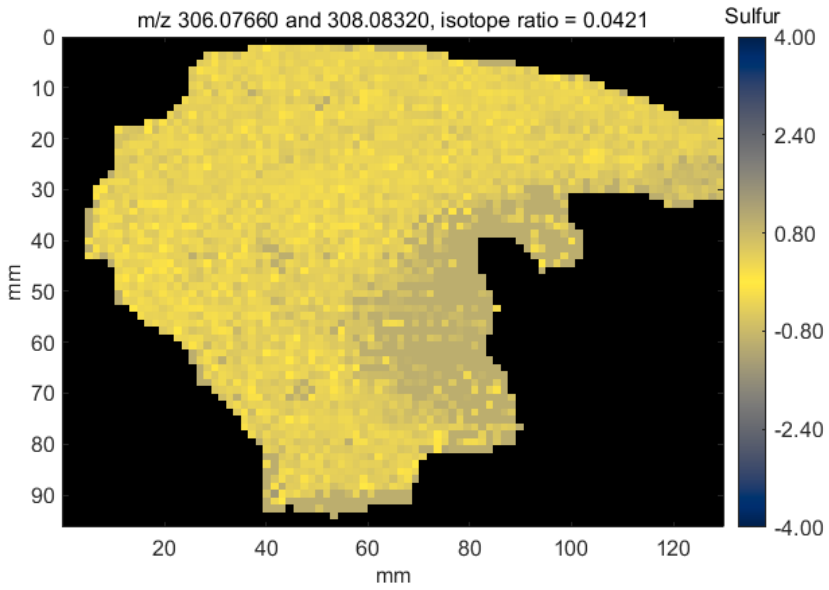


Figure 78: Heatmap showing the deviation from the expected number of sulfur atoms and that supported by the data. In this example, the expected number of sulfur atoms is 1.

A heatmap plot showing the deviation from the expected number of atoms at each scan is displayed (**Figure 78**). Note that the spectral accuracy is, in general, very good (yellow means that the deviation from the expected number of atoms is zero); however, the region on the right center of this tissue does not do as well as the rest of the ROI – this is attributed to higher ion abundance of this molecule on the left-hand side of this tissue than on the right (see **Figure 76**).

7.5.3 AutoQC MSI

Summary statistics for all tiles in a multifile image mosaic for each *m/z* value in a list. The results are saved into an Excel workbook with the following 16 worksheets.

Info	Folder, date, version, settings, and options
File	File names in the selected folder
Tile Sum	Abundance sum
Tile Max	Maximum abundance
Tile Mean	Mean abundance of scans above threshold
Tile Std Dev	Standard deviation of scans above threshold
Tile RSD	RSD of scans above threshold
Tile Mean Zero	Mean including low abundance scans
Tile Std Dev Zero	Standard deviation including low abundance scans
Tile RSD Zero	RSD including low abundance scans
Tile MMA	Mass measurement accuracy
Tile Isotope Count	Spectral accuracy for the isotope peaks
Tile Detection Frequency	Ratio of scans above threshold to total number of scans
Tile Iso Detection Frequency	Detection frequency for the isotope peaks
Tile Num Scans	Number of scans above threshold
Tile Iso Num Scans	Number of scans above threshold for isotope peaks

The current MSiReader *Navigation* panel values are used to generate abundance matrices for each *m/z* and isotope peak. The Info worksheet contains the current options, *m/z* values, isotope data and thresholds used to create the results. Each peak of interest and each isotope can have a different abundance threshold. The results worksheets contain a value for each *m/z* and each file. They can be arranged with file data down the columns or across the rows.

AutoQC MSI is an algorithm to compile a significant number of statistics on a folder of data for which the expected answer is known. This is evolving into a system suitability testing approach and will be in a future release. First, this function only works for a folder of files so the first step is for the user to load 2 or more files in MSiReader. Next, from the pull-down menu, select AutoQC MSI. The dialog box shown in **Figure 79** will pop-up. In this example, the molecular formula of the molecule is $C_{13}H_{12}F_2N_6O$.

Here the user will enter in the peak(s) of interest, isotope abundance of the heavier isotope (^{13}C = 0.0112), isotope ratio that is expected, number of atoms for each peak, abundance threshold, isotope abundance threshold and how the results should be output (tile or *m/z*). The user will then be prompted to enter a filename for the result will be in Excel format. The output gives a lot of basic statistics for each file that was loaded (for

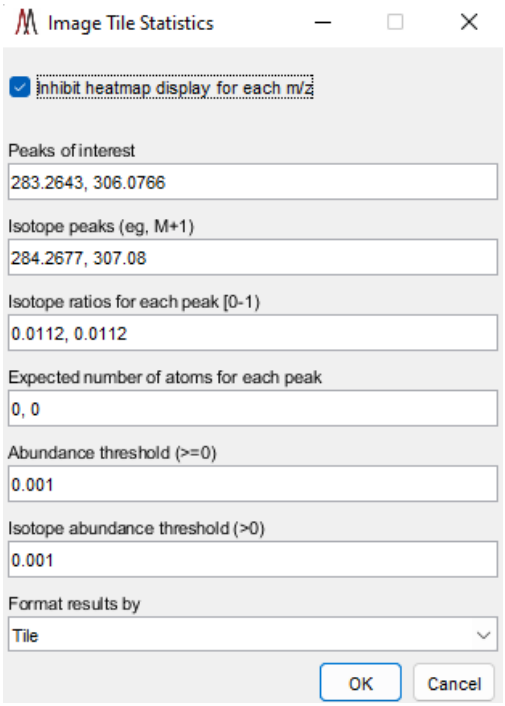


Figure 79: AutoQC MSI dialog box.

each peak of interest and its isotope) including the mean, RSD, isotope count, detection frequency, and number of scans the peak was found in (must be above the user set abundance threshold for both the peak of interest and its isotope threshold abundance (*e.g.*, ^{13}C). These can be useful QC statistics (even in raw format) to make sure that your MSI platform is operating properly.

For example, six ROI's were collected where each image was 20 × 20 pixels (voxels) with a QC standard homogenously sprayed on a glass slide (data shown in **Figure 80**). These

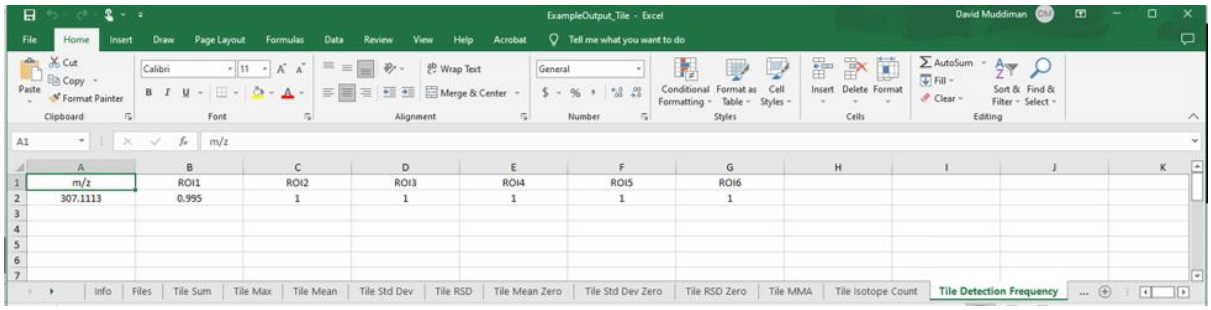


Figure 81: Example of data (summary statistics) generated from the Auto QC MSI tool. This is an example spreadsheet that shows the data for detection frequency for each ROI, each containing 400 voxels, that were measured. In this case, 5 out of the 6 ROI's detected the compound in every single voxel while in ROI1, the data indicate it was detected in 398 out of 400 voxels which equates to a detection frequency of 0.995.

were loaded into MSiReader and processed as outlined above using *tile* as the output setting (enter "1" in the dialog box shown in **Figure 79**). **Figure 81** shows an example of the data output to an Excel spreadsheet when using this tool.

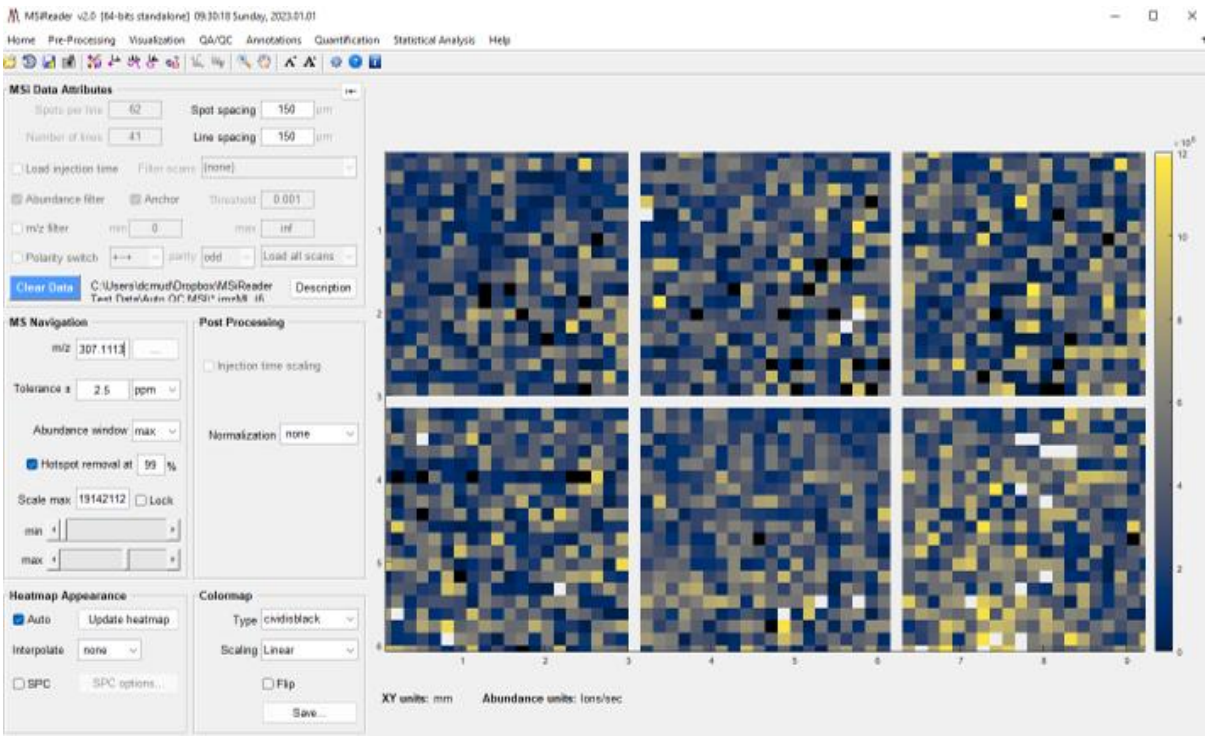


Figure 80: Loading of 6 QC ROI's into MSiReader. The peak of interest is *m/z* = 307.1113 with a M+1 peak (*m/z* = 308.1147).

There are several ways to annotate data in MSiReader which include the *MSiSpectrum* tool, the *MSiPeakfinder* tool, and the Search *Custom DB* tool. The first two use the current dataset loaded into MSiReader to search while the *Custom DB* tool can annotate a file from any source provided it is in the correct format. It is important to note that the default databases (MSiReaderPositiveIons and MSiReaderNegativeIons) are NOT loaded when MSI Reader starts. The user can change the default settings in the preferences .INI file (§5) to load them when MSiReader is started. Alternatively, the user can right click on the *MSiReaderPositiveIons* or *MSiReaderNegativeIons* and click on *Reload*. This is true for all three ways to annotate a dataset. For MSiSpectrum and MSiPeakfinder, the user has to select a ROI (two for *MSiPeakfinder*) to generate the .xlsx file while the search *Custom DB* tool does not require data to be loaded as this is annotating a file that was previously exported.

7.6.1 Database Searching

A tool for searching a results file in Excel format to find matches for a list of putative *m/z* can be selected by choosing *Database* under the *Annotations* menu and then *Search custom database* in the sub-menu. This will launch the MSiDatabase tool shown in **Figure 82**. Note that you must first use MSiExport, MSiSpectrum or MSiPeakfinder to generate the results file from your imzML file. It allows the user to annotate a single peak (the default for this field is the current *m/z* value in the main MSiReader GUI but it can be edited), the peaks in the clipboard, or the user can select a file. The file could be a

previously exported results file from a peak picking session or any Excel worksheet containing a list of *m/z* values in the first column.

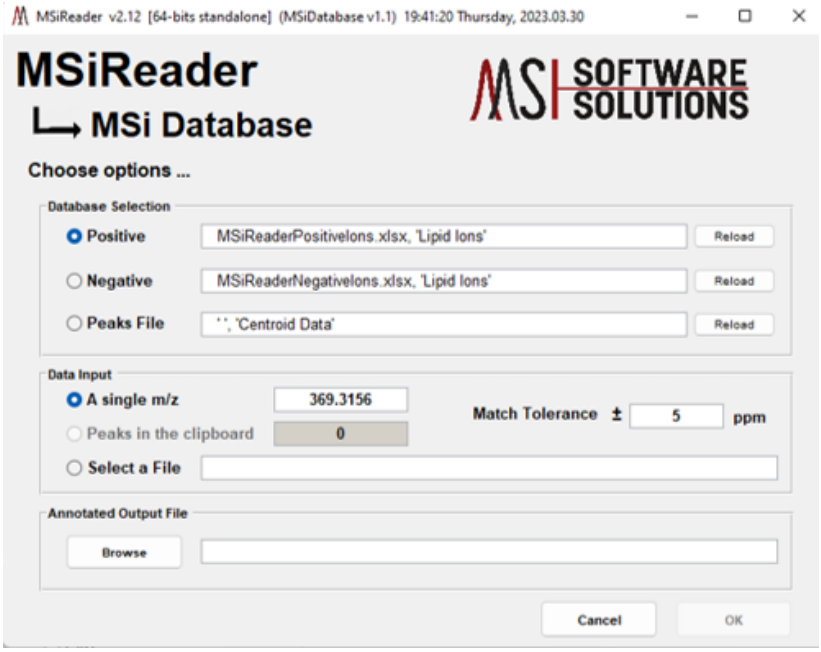


Figure 82: MSiDatabase GUI. Selection is for a file that was generated in this example using MSiSpectrum – in this example the filename is Mouse ROI.xlsx.

To annotate a single *m/z* value that maybe present in your data, enter the *m/z* value in the *Mass Selection* pane as shown in **Figure 73** along with a *Match Tolerance* in ppm. In this way, the dataset that you have exported into an Excel spreadsheet will be annotated.

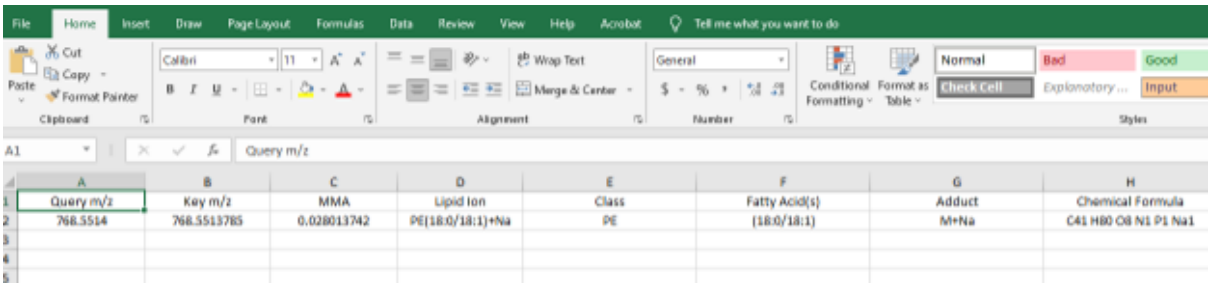


Figure 83: Output using the Custom DB annotation tool with a query for a single *m/z* value. Note that the only the information about the single *m/z* is included in the output.

For the mouse dataset, enter 768.5514 and load the *File to Annotate* (Mouse ROI.xlsx). The output from the single *m/z* query is shown in **Figure 83**.

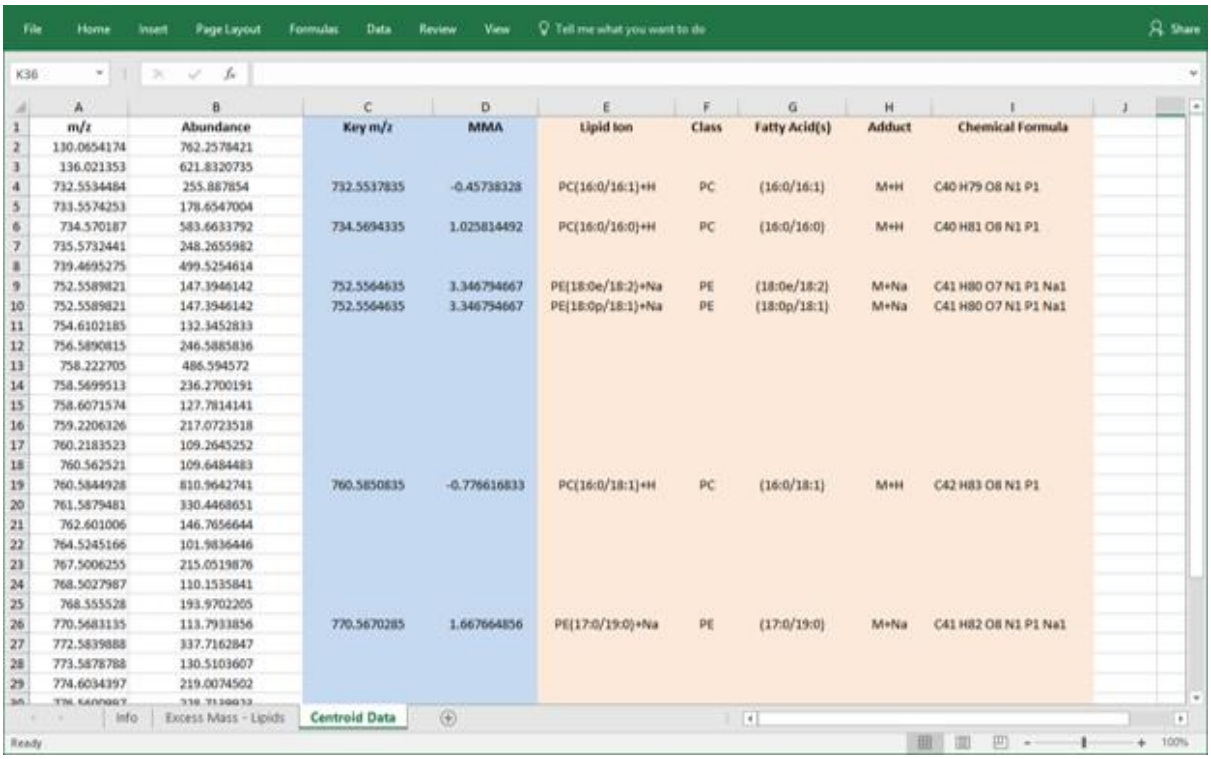
The user can select the *Select a File* radio button (**Figure 82**) and then a dialog is launched for the user to choose an Excel file. If the selected file contains multiple worksheets, a prompt is issued to pick the correct one. In each scenario, the user can also choose a *m/z* match tolerance and a database that is for either positive or negative ionization mode.

Importantly, an alternate database file can be loaded at any time by right-clicking on the positive or negative mode file name and a dialog will be launched that says “reload positive ion database” or “reload negative ion database”. This can be custom made by the end-user specific to their project. The default databases can be set in the Preferences .INI File (**§5**). Or simply click “Reload” next to the relevant database.

For the example shown, a positive-ion mode data file has been selected as the *m/z* source which contains a large number of *m/z* values. In this case, the output location and filename has defaulted to be the same filename. The user can click the Browse button to select some other output file (including a new empty file). In either case, information from the database is always added to output file in columns to the right of any existing columns. This allows the user to annotate a previously annotated file without overwriting information. For example, after a database has been updated, an old file could be processed by MSiDatabase again to search for new matches.

Figure 84 shows an annotation of the *Centroid Data* worksheet from a previous export of the MSiSpectrum tool. The first two columns are the original output results. The next two columns with light blue background contain the *m/z* keys from a positive ionization database that match within 5 ppm and the MMA calculation for each match. The five

columns with light tan background were copied from the database file for the matching keys.



	A	B	C	D	E	F	G	H	I
1	m/z	Abundance	Key m/z	MMA	Lipid Ion	Class	Fatty Acid(s)	Adduct	Chemical Formula
2	130.0654174	762.2578421							
3	136.021353	621.8320735							
4	732.5534484	255.887854	732.5537835	-0.45738328	PC(16:0/16:1)+H	PC	(16:0/16:1)	M+H	C40 H79 O8 N1 P1
5	733.5574253	178.6547004							
6	734.570187	583.6633792	734.5694335	1.025834492	PC(16:0/16:0)+H	PC	(16:0/16:0)	M+H	C40 H81 O8 N1 P1
7	735.5732441	248.2655982							
8	739.4695275	499.5254614							
9	752.5589821	147.3946142	752.5564635	3.346794667	PE(18:0e/18:2)+Na	PE	(18:0e/18:2)	M+Na	C41 H80 O7 N1 P1 Na1
10	752.5589821	147.3946142	752.5564635	3.346794667	PE(18:0p/18:1)+Na	PE	(18:0p/18:1)	M+Na	C41 H80 O7 N1 P1 Na1
11	754.6102185	132.3452833							
12	756.5890815	246.5885836							
13	758.222705	486.594572							
14	758.5699513	236.2700191							
15	758.6071574	127.7814141							
16	759.2206326	217.0723518							
17	760.2183523	109.2645252							
18	760.562521	109.6484483							
19	760.5844928	810.9642741	760.5850835	-0.776616833	PC(16:0/18:1)+H	PC	(16:0/18:1)	M+H	C42 H83 O8 N1 P1
20	761.5879481	330.4468651							
21	762.601006	146.7656644							
22	764.5245166	101.9836446							
23	767.5006255	215.0519876							
24	768.5027987	110.1535841							
25	768.555528	193.9702205							
26	770.5683135	113.7933856	770.5670285	1.667664856	PE(17:0/19:0)+Na	PE	(17:0/19:0)	M+Na	C41 H82 O8 N1 P1 Na1
27	772.5839888	337.7162847							
28	773.5878788	130.5103607							
29	774.6034197	219.0074502							

Figure 84: An example of the annotated file using a data export from MSiSpectrum.

If the mass source was a single *m/z* or a list of peaks in the clipboard, then the output file will contain the query *m/z* values in the first column. Any matching database keys, MMA and annotation information would follow in columns two through eight.

The MSiSpectrum and MSiPeakfinder tools have checkboxes in the *Peak Export (centroided data)* panel that enable the database annotation and tolerance selection in addition to the other options for exporting results. In this case, the *Centroid Data* worksheet is annotated as soon as it is created without the need to run the MSiDatabase tool separately. **Figure 85** shows the annotation functionality in the MSiSpectrum tool. The user should check *Use Database to Annotate Peaks* to annotate the data during the export.

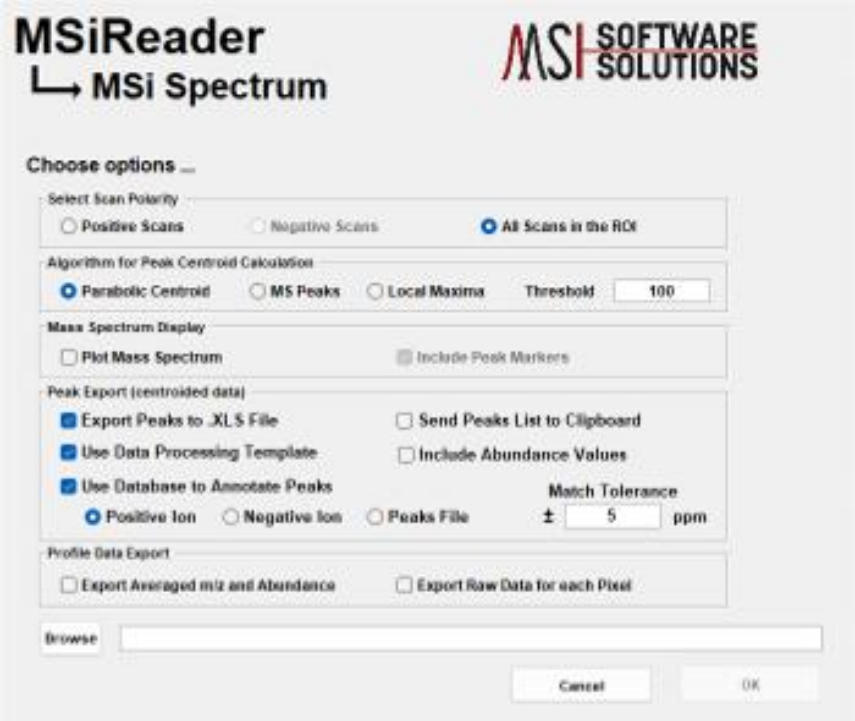


Figure 85: The annotation tool as deployed in the MSiSpectrum tool.

The preferences INI file (§5) has variables for the name and format of each database Excel file. The format of each is specified with six values: a worksheet name, the column containing *m/z* key values, a list of the information column numbers to be added to an annotated worksheet, the row number containing labels for the information columns, and the first row that contains data.

Example databases are provided with the MSiReader distribution. There are files for positive and negative mode ionization from a shotgun lipidomics study which can be used as an example of how a user might build their own database with ions of interest; one for positive ion mode and one for negative ion mode.

One can make additions to an existing file simply by clicking Database in the pull-down menu and then selecting “add record to database” from the Annotations menu. Upon doing so, an ionization selection dialog box shown in **Figure 86** and then opens the

current database file for the selected mode in Excel. A user can make edits / additions and then have an option to save or not save those changes.

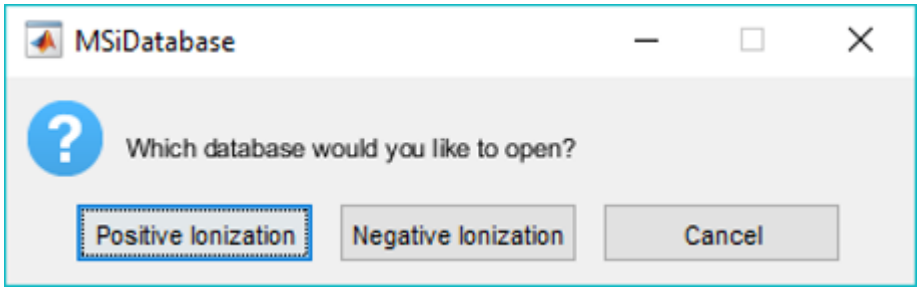


Figure 86: Ionization database selection dialog to open an existing database and make additions to it.

7.6.2 Molecular formula adduct search

A **video tutorial** on using the molecular formula tool can be found **HERE**.

An isotopic distribution and adduct calculator, MSiFormula, can be launched from the drop-down menu by selecting molecular formula adduct search. The MSiFormula GUI is shown in **Figure 87**. A formula can be entered using the standard names of the first 92 elements of the periodic table and the theoretical monoisotopic mass and isotopic distribution will be automatically calculated using the NIST standard table of stable isotopes. Inclusion of heavy isotopes can be entered explicitly (for example, $^{13}C_2$ or $^{15}N_4$) and the exact mass those isotopes will be used when calculating the isotopic distribution. The mass of any occurrence of an element not preceded by an isotope number will be found using the stable isotope distribution table. A context menu on the molecular formula entry box allows an immediate change of the ^{13}C or ^{15}N ratio for this session of MSiFormula. The ratios of ^{12}C and ^{14}N are automatically adjusted so the ratios sum to one for each element.

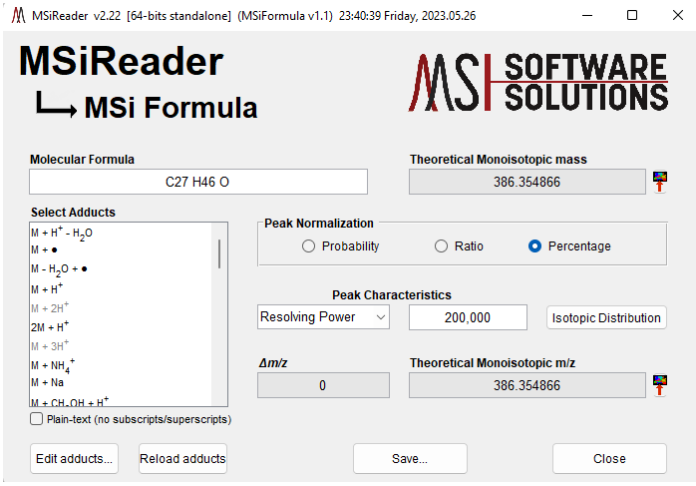


Figure 87: MSiFormula GUI for finding adducts of a given neutral molecular formula in a loaded dataset.

Elements may be entered in any order and may occur multiple times. An element name not followed by a numeral is assumed to occur once. The elemental parts of a formula must be separated with whitespace when an isotope is used. For example, the formula "$^{13}C_{15}N^{18}O$" without whitespace could be interpreted as "$^{13}C_{15}$ N_{18} O", "$^{13}C_{15}$ N ^{18}O", "^{13}C $^{15}N_{18}$ O", or "^{13}C $^{15}N_{18}$ O".

The peaks of the isotopic distribution can be normalized so their heights sum to one (probability), so the most abundant peak is one (ratio), or so the most abundant peak is 100 (percentage); the default is percentage. The theoretical isotopic distribution can be plotted with the peak shape specified by a single numeric value, either full width half max (FWHM) or mass resolving power – simply click the *isotopic distribution* button. The isotopic distribution is carried out using the Fourier transform methodology[17]. Either a line or stem plot can be drawn, with or without peak markers and a legend with menu options on the plot button. The theoretical monoisotopic mass can be sent to the MSiReader navigation panel and the heatmap will automatically be updated.

A pull-down list of common adducts is available to modify the isotopic mass of the formula. If multiple adducts are selected, multiple results will be displayed as expected monoisotopic masses. Any of these results can be sent to the MSiReader navigation panel to immediately update the heatmap.

The adducts are stored in an Excel workbook named, *MSiAdducts.xlsx*, in the MSiReader installation folder. It contains 35 positive mode adducts and 15 negative mode adducts. The file can be edited to add, remove, or change the order of the adducts. The names of the adducts are arbitrary and will appear in a list selection in MSiFormula. The columns of the *Common Adducts* worksheet needed by MSiFormula are described by a set of preferences INI file variables (**§5**). Other columns in the worksheet are ignored.

Variable	Value
AdductCalculatorFile	MSiAdducts.xlsx
AdductCalculatorWorksheet	Common Adducts
AdductCalculatorLegend	Legend
AdductCalculatorHeaderRow	1
AdductCalculatorFirstRow	2
AdductNameColumn	1
AdductModeColumn	2
AdductFormulaColumn	3
AdductChargeColumn	4
AdductMultiplierColumn	5
AdductDeltaMassColumn	6
AdductDefaultSelection	0

The results for a molecular formula (masses, ratios, and isotopic distribution) and all selected adducts can be saved into an Excel workbook. Multiple instances of MSiFormula can be run simultaneously and data does not have to be loaded into MSiReader prior to launching the tool. Multiple default adducts can be specified with a vector of row numbers in the *Common Adducts* worksheet for the value of *AdductDefaultSelection*. Zero means that no adducts are initially selected.

If you want to add to the adducts, click *Edit Adducts* and the Excel spreadsheet will open. Make edits and save the spreadsheet using the same file name. Then click *Reload Adducts* so it re-reads the file. Those options will now appear in the window.

To determine how a molecule ionizes (*e.g.*, $M+H^+$ or $M+Na^+$), enter in the molecular formula and select the adducts you want to be considered. This will automatically populate the fields $\Delta m/z$ and *Theoretical Monoisotopic m/z*. Next, click the ICON to the right of *Theoretical Monoisotopic m/z* – this will pop up a window with the *m/z* values of the different adducts under consideration. If you only selected one adduct, it will automatically update heatmap with that single *m/z* value. Repeat this process for each *m/z* (adduct) being considered and make note of the abundance of each type of adduct. This tool can be very effective at determining how a molecule ionizes.

7.6.3 SSIM Colocalization Tool and Batch Processing

7.6.3.1 Background Information

Navigating large data sets and generating heatmaps for more than a few *m/z* values can be tedious. MSiReader has implemented a batch processing feature to automatically generate and save a heatmap plot for all of the *m/z* values in a list. After selecting a source of peaks and a destination folder, images are generated (default is a .png file) and saved for each peak. Thousands of image files can be created in less than an hour and rapidly sorted by the user later using any graphical image viewer software. This is particularly useful to visualize the output from the automatic peak picking function. It is also very handy for users who want to quickly extract images for a list of target *m/z* values and for peaks that are found and saved while navigating the data set. The source of the *m/z* values can be the contents of the clipboard, an Excel worksheet, a text file, or a sequence of values uniformly spaced over an *m/z* range. If an Excel workbook is selected that has more than one worksheet, the user is prompted to select the correct one. The *m/z* values in the first column of the selected worksheet will be used as the peak list.

Batch image processing has been incorporated into the image correlation tool which is launched by selecting SSIM co-localization tool under the *Annotation* menu. The MSiCorrelation GUI is shown in **Figure 88**.

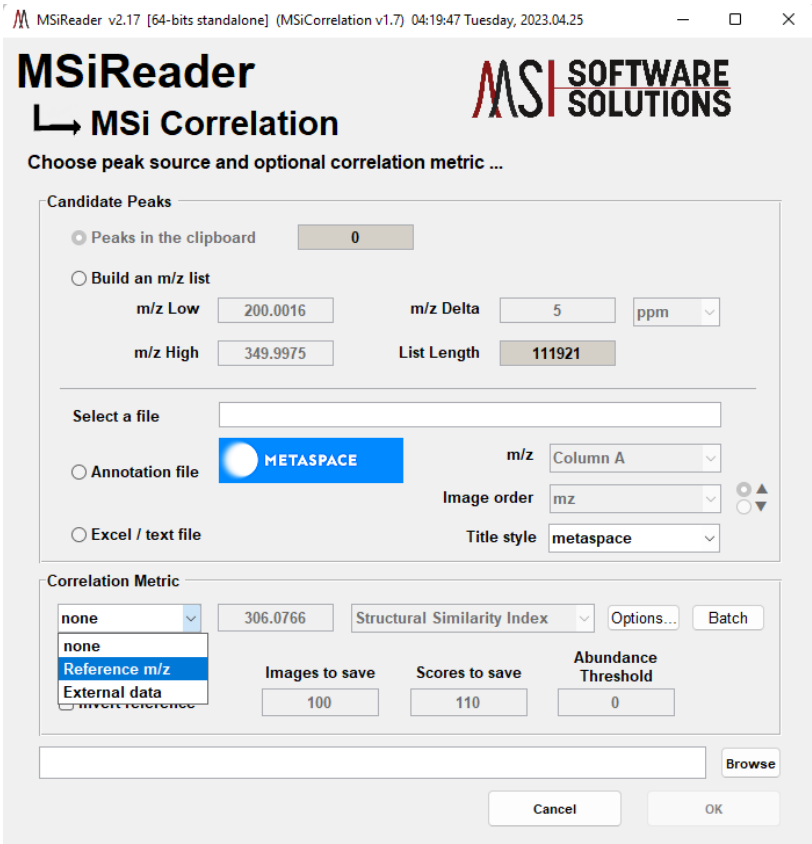


Figure 88: The correlation and batch processing user interface in MSiReader.

MSiCorrelation generates a folder of images in exactly the same way as the batch processing tool if the *Correlation Metric* pane pull-down selection is *none*. Images in the format described below in §7.6.5 will be saved in a folder selected by the user. The files will be named after the *m/z* value of the image (*e.g.*, the image for *m/z* 369.3516 will be saved as 369_3516.png) for easy sorting. The title information, figure size and colorbar characteristics can be customized with the preferences .INI variables (§5) shown in **Table**

4. Their meaning is identical to that described above for exported figures. However, note that default values are different for batch and figure export.

Table 4. Batch figure formatting options.

BatchFileType	*png*
BatchFigureSize	*system dependent*
BatchFigTitleStyle	*batch*
BatchFigAxes	*true*
BatchFigAxesLabels	*true*
BatchFigColorbarStyle	*compact*
BatchFigColorbarLabels	6
BatchFigColorbarPrecision	2
BatchFigFontName	Arial
BatchFigFontSize	12
ExportToMATBatch	*false*
ExportToFIGBatch	*false*

The additional settings for batch processing are set with preferences file INI variables (§5). They are: *ExportDuplicateMZBatch*, *BatchPeakHeatmapUpdate*, and *BatchHeatmapVisible*. The default value for all three is *false*. Setting them to *true* will cause duplicate *m/z* values to generate separate images, update the MSiReader heatmap as each value is processed, and make the separate figure window for each image visible as it is created. Enabling either of these last two settings will significantly slow down batch processing.

There is the option to generate a uniformly spaced list of values over any *m/z* range for batch processing and image correlation. Element spacing can be in absolute or ppm units. Current normalization and windowing selections from MSiReader are used when the images are generated. The *m/z Delta* value defaults to the current MSiReader window width. Increasing it will leave gaps between the windows and decreasing it will result in overlapping windows.

Other information relevant to the data set and options that were selected are saved into a short text file in the same folder as the images. An example is shown in **Figure 89**. A second text file is written into this folder containing the *m/z* peaks list. This later file can then be used as a peaks list for analyzing another data set. If the loaded data set was a folder of imzML or mzXML files a third file is saved containing the tiling pattern and a list of the file names. The default number of decimal places for the *m/z* value and the *m/z* window is 5 but can be modified in the INI file (§5) by changing the *mzExportPrecision* variable. If a photographic image overlay has been loaded, the overlay is included in each of the batch processing output files. The transparency slider can be used to temporarily hide this photo overlay without deleting it before launching MSiCorrelation. Recall that the color scale can be locked during batch processing so that the images all have the same abundance to color mapping. If the color scale is locked and a file containing an *m/z* peaks lists was selected that had a second column of values, then the user is given the option to use those values for the abundance to color mapping for each *m/z*. This is useful for generating images for an *m/z* target list that are visually comparable across multiple data sets with the same abundance range for all the images.

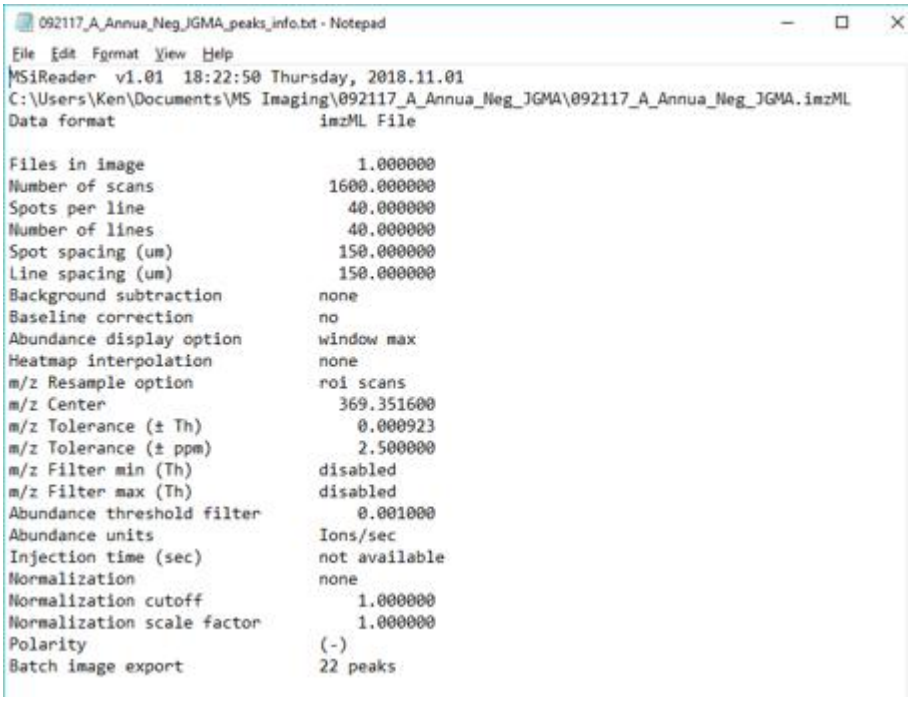


Figure 89: Example of the information saved as a text (.txt) file with a batch of images.

7.6.3.2 Image Correlation Algorithms

As shown in **Figure 88** there are two other choices for the *Correlation Metric* pane pull-down selection, *Reference m/z* and *External data*. Either of these will result in the list of candidate *m/z* values being compared and ranked by their similarity to a reference image. If *Reference m/z* is chosen, the image data for the value in the data entry box is used as the reference. If *External data* is chosen, the user will be asked to select a .mat file containing one or two matrices with the same dimensions as the loaded set. The first matrix becomes the reference data, the second optional matrix is a binary mask that can be used to exclude a portion of the reference matrix from consideration. For example, scans outside of a tissue ROI. A special case of an external reference is the down sampled version of an optical overlay imaged saved with the MSiImage tool as described in Section 7.3.2. In this case MSiCorrelation automatically chooses the correct matrices in the .mat file.

The current MSiReader normalization and window settings are used when generating candidate images for correlation with the reference. The value in the *Abundance Threshold* data entry box can be used to exclude very low abundance images. This can be particularly useful when using a very large list of *m/z* candidates. If all of the scans for an *m/z* have ion abundances below the threshold, that image is considered empty and excluded. Note that this evaluation is done after applying normalization, thus it may be difficult to select an appropriate threshold value. It may be helpful to normalize to the maximum abundance so that all of generated images and the reference have an abundance range between zero and *Normscale* (**§5**).

An external data set that is used as the reference image may have already been normalized. In this case, uncheck the *Normalize Data* in the *Correlation Metric* pane to disable normalization of the reference image so that it is not normalized twice.

Three algorithms are implemented for scoring candidate images with respect to the reference. In **Figure 90** the *Structural Similarity Index* has been selected from the pull-down list. The other three algorithms are *Absolute Difference*, *Mean Squared Error and 2D*.

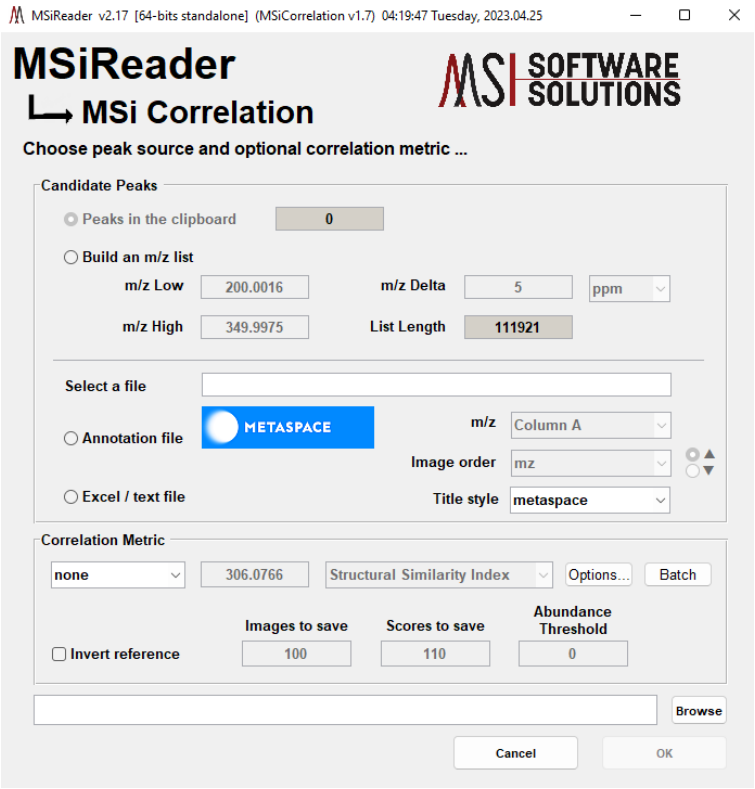


Figure 90: MSiCorrelation GUI showing selection of SSIM as the algorithm for image correlation.

The scoring algorithms for image correlation are:

Absolute Difference Subtract the reference image matrix from each candidate element-by-element and sum the absolute value of those differences. A score of zero means the images are identical.

Mean Squared Error The square root of the sum of the squared differences is divided by the number of elements. This is equivalent to the Euclidean 2-norm (*i.e.*, the vector length) divided by the number of elements. A score of zero means the images are identical.

2D This was added to the MSiCorrelation tool as an optional scoring algorithm selection.

Structural Similarity Index An index based on human perception of quality or image "likeness". The SSIM algorithm[18] was developed for digital image processing as a measure of the quality of images after compression, filtering, transmission and

reproduction. The similarity measure of a reference standard against candidate images is the weighted product of the luminance, contrast and structural components of the image. The SSIM default parameters can be accessed in MSiReader by clicking *Options* to the right of SSIM in the *correlation metric* pane; the input values are shown in **Figure 91**; these were determined in a previous study to be optimal for MSI[19].

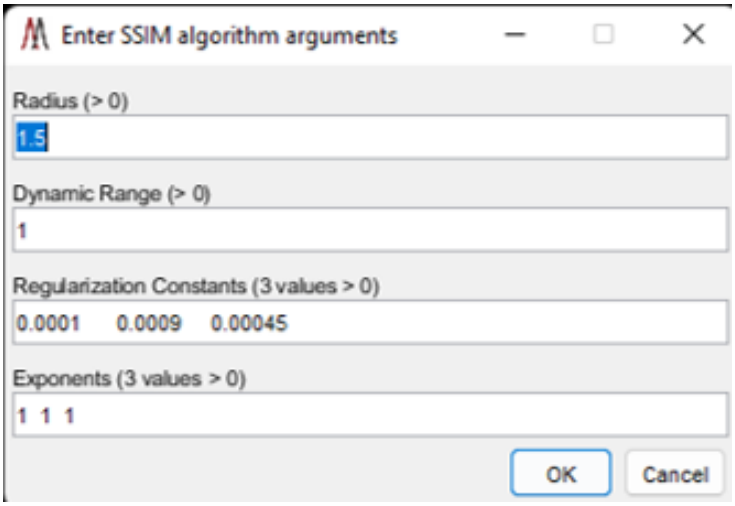
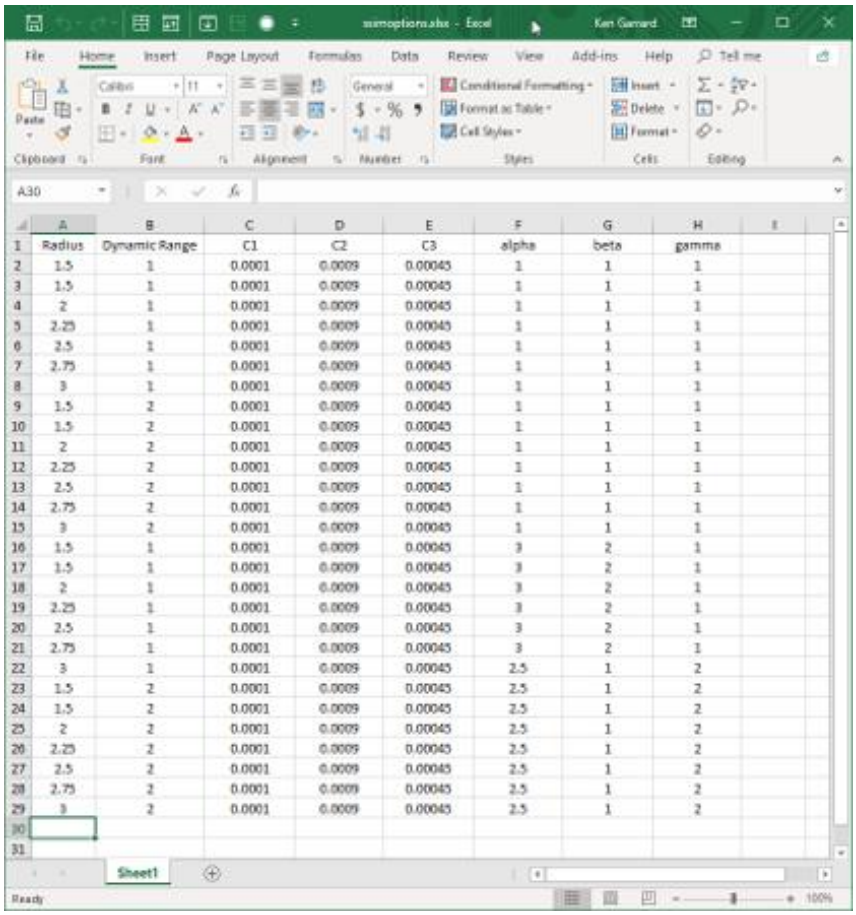


Figure 91: Parameters selection dialog for the SSIM algorithm.

The *Radius* parameter specifies the standard deviation of an isotropic Gaussian function used for weighting a window of pixels surrounding each pixel. The *Dynamic Range* parameter is the absolute range of values in the input images. The regularization constants are used to stabilize the luminance, contrast and structural components when the local mean or standard deviation in a small region tends toward zero. Lastly, three *Exponents* are used to adjust the relative importance of the luminance, contrast and structural components before combining them into a single score. While the default values work well, the user should refer to these references[18,19] for more details. SSIM scores range between -1 and +1, with one indicating a perfect match.

Choosing *Batch* to the right of SSIM in the *Correlation matrix* pane in **Figure 90** launches a dialog for the user to choose a file of SSIM parameters. The file can be Excel or text format with each row or line containing the eight parameter values. An example is shown in **Figure 92** with 28 sets of parameters varying the radius, dynamic range, and weighting exponents. MSiCorrelation will execute the SSIM algorithm using all combinations of parameter sets and candidate *m/z* values and rank them by their similarity score. Using

this option allows for great flexibility in testing a large number of parameters for your specific dataset.



	A	B	C	D	E	F	G	H
1	Radius	Dynamic Range	C1	C2	C3	alpha	beta	gamma
2	1.5	1	0.0001	0.0009	0.00045	1	1	1
3	1.5	1	0.0001	0.0009	0.00045	1	1	1
4	2	1	0.0001	0.0009	0.00045	1	1	1
5	2.25	1	0.0001	0.0009	0.00045	1	1	1
6	2.5	1	0.0001	0.0009	0.00045	1	1	1
7	2.75	1	0.0001	0.0009	0.00045	1	1	1
8	3	1	0.0001	0.0009	0.00045	1	1	1
9	1.5	2	0.0001	0.0009	0.00045	1	1	1
10	1.5	2	0.0001	0.0009	0.00045	1	1	1
11	2	2	0.0001	0.0009	0.00045	1	1	1
12	2.25	2	0.0001	0.0009	0.00045	1	1	1
13	2.5	2	0.0001	0.0009	0.00045	1	1	1
14	2.75	2	0.0001	0.0009	0.00045	1	1	1
15	3	2	0.0001	0.0009	0.00045	1	1	1
16	1.5	1	0.0001	0.0009	0.00045	3	2	1
17	1.5	1	0.0001	0.0009	0.00045	3	2	1
18	2	1	0.0001	0.0009	0.00045	3	2	1
19	2.25	1	0.0001	0.0009	0.00045	3	2	1
20	2.5	1	0.0001	0.0009	0.00045	3	2	1
21	2.75	1	0.0001	0.0009	0.00045	3	2	1
22	3	1	0.0001	0.0009	0.00045	2.5	1	2
23	1.5	2	0.0001	0.0009	0.00045	2.5	1	2
24	1.5	2	0.0001	0.0009	0.00045	2.5	1	2
25	2	2	0.0001	0.0009	0.00045	2.5	1	2
26	2.25	2	0.0001	0.0009	0.00045	2.5	1	2
27	2.5	2	0.0001	0.0009	0.00045	2.5	1	2
28	2.75	2	0.0001	0.0009	0.00045	2.5	1	2
29	3	2	0.0001	0.0009	0.00045	2.5	1	2

Figure 92: A file of parameters for SSIM image correlation. Each candidate image is evaluated with each set of parameters.

If ROIs are present in MSiReader when using MSiCorrelation, the user is prompted to select the scans of interest using one of the dialogs shown in **Figure 93**.

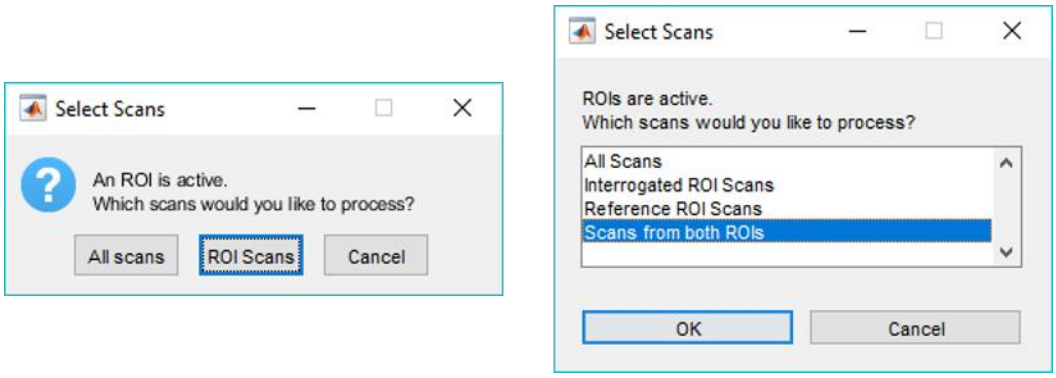


Figure 93: Active ROI dialogs for a single ROI (left) and two ROIs (right).

If scans from ROIs are selected a second dialog gives the user the option to make the reference image positive or negative, that is, 1's inside the ROI and 0's outside of it or the other way around. The dialog is not shown if *All scans* is selected. This dialog is shown in **Figure 94**.

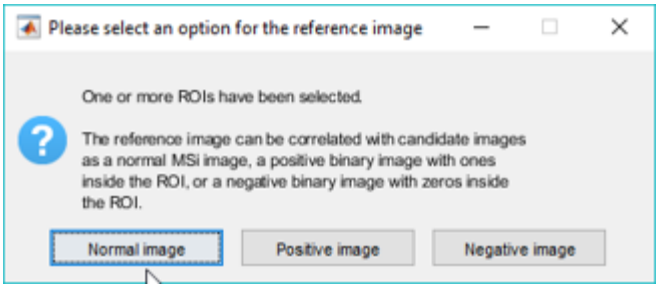


Figure 94: Normal or binary reference image dialog.

With the options selected in **Figure 90** and the parameters shown in either in **Figure 91** or **Figure 92**, "virtual" images would be generated in memory, correlated with the reference for each set of SSIM parameters and scored. The output folder would be filled with image files for the highest ranked results (user-defined – see **Figure 90**). The top 100 *m/z* values and scores would be saved in a text file in the same folder. If you want to save all of the scores, enter *Inf* or *All* in the *Scores to Save* data box (**Figure 90**). Before ranking over 280,000 images it is advisable to use peak picking to generate a smaller list of *m/z* candidates to see how long this process takes with your data. The SSIM algorithm is very fast, scoring a pair of images with 1 million elements each in less than 120

milliseconds on an ordinary laptop. However, generating the internal data for each candidate image will consume significantly more time.

A progress bar is displayed during the correlation process. If it is closed, the user is prompted to confirm terminating the image correlation or to continue. Continuing will **skip** the remaining candidate images and proceed with sorting the scores obtained so far and saving graphic image files. Another progress bar is displayed while the files are generated. You can open the folder where they are being saved to see a preview of your images. This process can also be stopped early by closing the progress bar again.

MSiCorrelation uses the current MSiReader parameters when generating images (*e.g.*, MMA, colorscale). You may lock the heatmap scale in the *MS Navigation* pane so that all

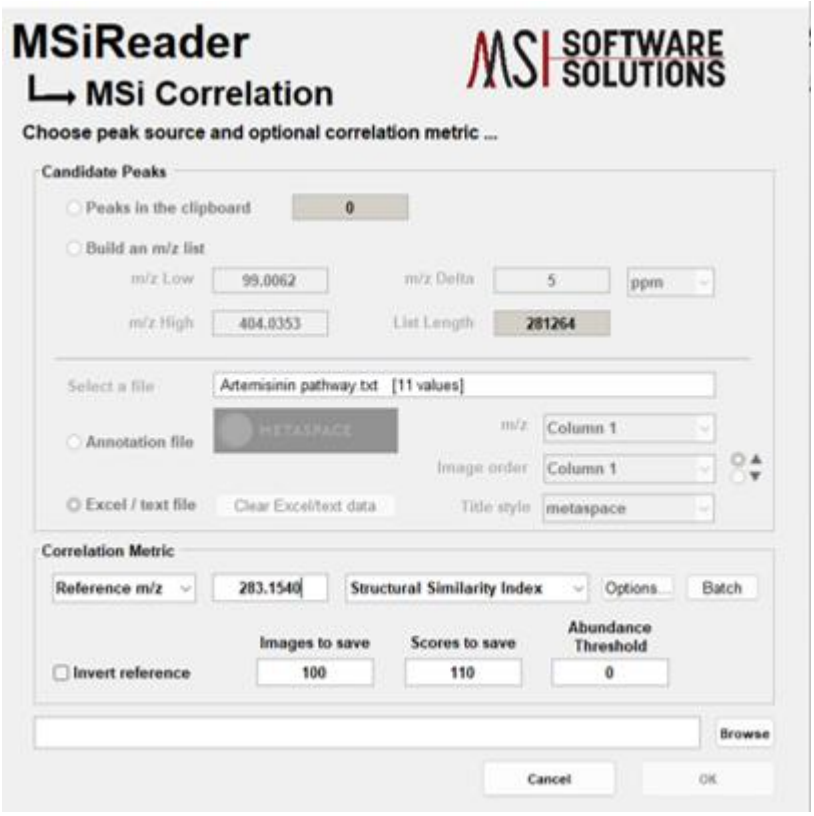


Figure 95: The MSiCorrelation tool using a .txt file as the input. Click browse to select a file for the output folder for the images to be generated.

images are generated on the same scale. Next, the user has 4 choices of candidate *m/z* values: clipboard contents; build an *m/z* list, choose an annotation file or select Excel or text file to upload a series of *m/z* values for the correlation analysis. In this example, a .txt file with a list of *m/z* values that the user would like to determine if they correlate with the reference *m/z* value. The reference *m/z* value that was chosen is 283.1540 as shown in **Figure 95**. If the user selects none – this is batch image generation; if the user selects reference *m/z*, this will be a MSI image of that *m/z* value and lastly, external data, this selection allows you to use an optical overlay for correlation.

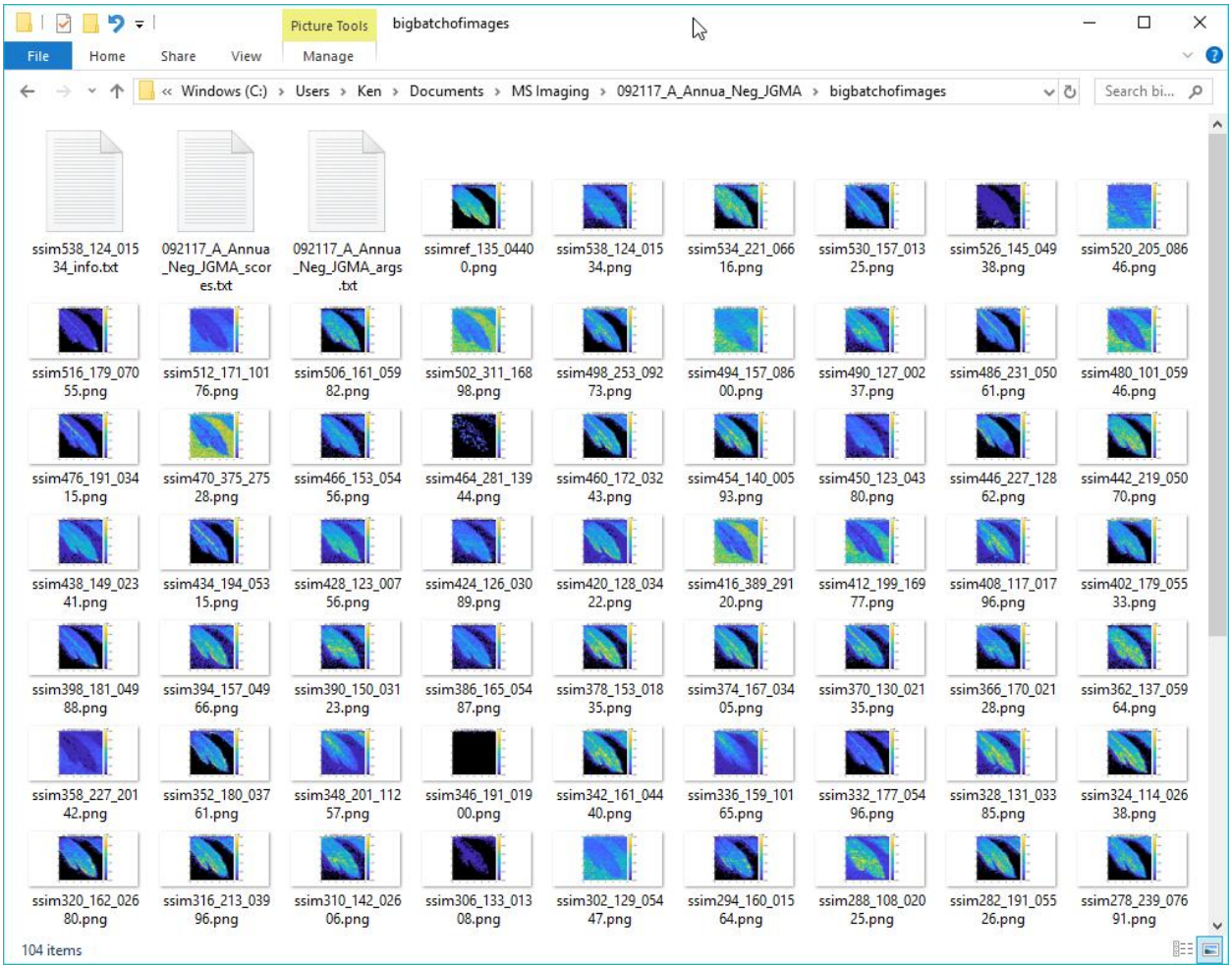


Figure 96: A folder of images created by MSiCorrelation.

An example of a folder of ranked images is shown in **Figure 96**. The image files are named using the correlation method, the rank and the *m/z* value for easy identification and sorting. The reference image is also saved in this folder.

Two or three text files are also saved using the data set name. A <name>_info.txt file containing current MSiReader and MSiCorrelation settings is shown in **Figure 97**. If SSIM was the selected method another file <name>_args.txt is written. Lastly, the scores are saved. An example is shown in **Figure 98** with the *m/z* values and their score. The third column is the index (*i.e.*, line number or row) into the SSIM parameter file used to obtain that score.

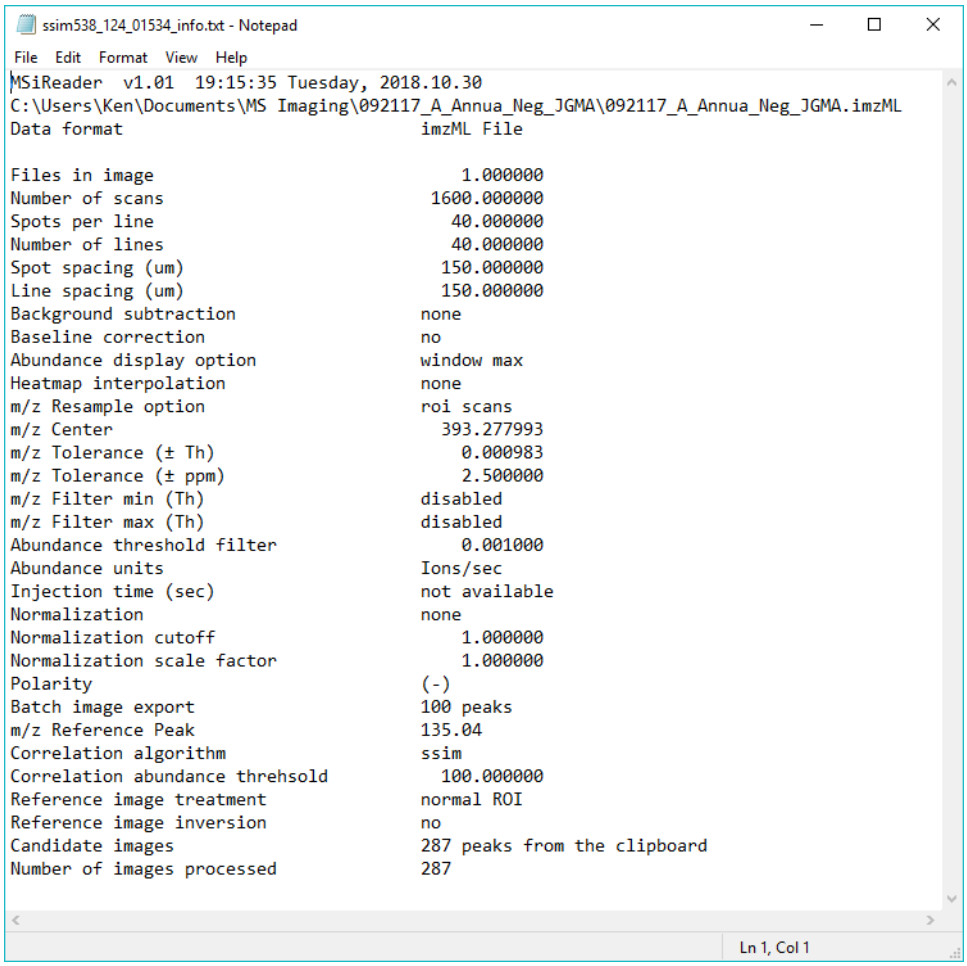


Figure 97: Text information file with data set and correlation parameters.

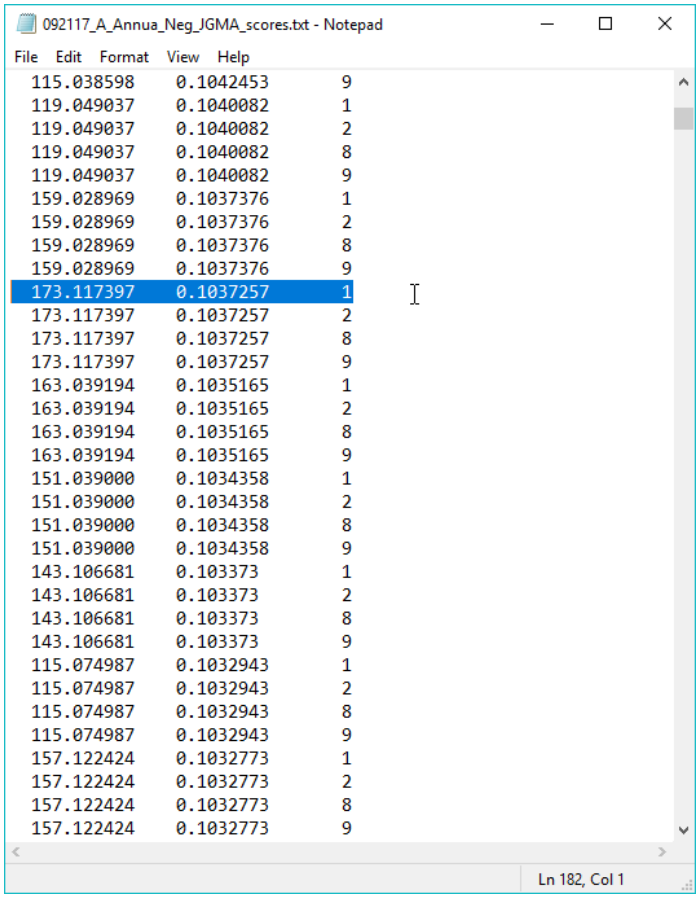


Figure 98: Correlation scores files showing *m/z* values, their score and the row of the corresponding parameter set that was used to obtain that ranking. The highlighted text is the 182nd highest ranked candidate on line 182 of the file.

7.6.4 ROI Functions

Restoring a previous ROI or defining an ROI (single pixel/voxel), line, or polygon can be done using these menu items or their associated ICONS in the toolbar. Restore a previously saved ROI. Select a single scan ROI using the cursor tool. Use this selection in the pull-down to select a single pixel (or voxel) in the heatmap. The user can drag this around using the mouse. Once a selected pixel of interest is decided upon, the user can right click and a sub-menu comes up allowing one to export the coordinates, change the color of the cursor, or plot the mass spectrum for the collected (or filtered) *m/z* range. When viewing the mass spectrum, there is a new toolbar at the top. These allow one to save the plot as a MATLAB .fig file, or print the spectrum. The next band of icons

in the toolbar allows you to select a peak in the mass spectrum and update the heatmap for that specific *m/z*. You can also copy details to the clipboard for the selected peak.

If the user wishes to select all pixels for data export, the user can load the data and go directly to MSiExport. Since no ROI was selected, it will export all the pixels/wells. This is particularly useful for the BioPharma mode so users do not have to select a polygon tool and then say "All Scans" when opening up the MSiExport tool.

 Select an ROI using the segmented line drawing tool. Click this selection in the pull-down to select a line of any length or direction through the heatmap. Upon doing so, the length of the line will be shown on the top left of the heatmap. If you right click on the line in the image, you will be able to: 1) export the line ROI details; 2) set line color; 3) plot the ion abundance as a function of distance along the ROI line; and 4) select plot type (stem, stairs, or line).

 Select an ROI using the polygon drawing tool. If you right click on the polygon in the image, you will be able to: **1)** export the line ROI details; **2)** make the ROI a rectangle; **3)** select all pixels for the ROI; **4)** create a binary mask for the ROI; and **5)** set line color. The user can generate a binary mask using the polygon drawing tool or the interrogated and reference regions; the binary mask is simply a matrix of 0's and 1's applied to a given dataset.

 Select interrogated and reference ROIs using the polygon tool. This is done to compare two regions of interest using, for example, MSiPeakfinder (**§7.7.1**).

The user can apply a binary mask that they created after selecting a ROI using the polygon tool or the interrogated and reference ROI tool. If the user selects from the Annotations menu then ROI then "Apply binary ROI mask to heatmap" but has not

selected one or two ROI's, the matrix applied is all 1's (no change in heat map). If one or two ROI's are defined, the user can right click on either ROI and a menu will allow you to create a binary mask and include interrogated ROI, reference ROI or all scan in both ROI's. If they are overlapping regions, you will have an additional option. To turn ON or OFF the binary mask, the user simply selects this option in the drop-down menu (toggle). One use of this function might be to focus on a cancer ROI and a healthy ROI and thus, visually remove the other regions of the tissue and the image saved.

7.6.5 MSiCorrelation and Batch Processing

The MSiCorrelation tool provides a convenient way to generate and order batch images for the *m/z* values in a METASPACE annotation file, an Excel file, the clipboard or for a uniformly spaced *m/z* list. The image files can be named so they appear in a desirable order in the output folder. This can be by *m/z* value, according to another column of the same input file, or by their correlation rank. For correlation ranking, the images are generated virtually and compared with the reference image using the selected algorithm. Then the specified number of top ranked images are saved in the output folder. The MSiCorrelation tool user interface is shown in **Figure 95**.

Peak Source and Image Order

The source of *m/z* values can be the clipboard contents, a METASPACE annotation file, an Excel worksheet, a text file, or a list with uniform spacing can be generated by the tool. In the case of an Excel workbook, the user is prompted to select a worksheet. The *m/z* and *Image order* pull-down menus in the *Candidate Peaks* panel are automatically populated with the column headings from the selected worksheet, a text file, or a METASPACE .csv format annotation file.

Only columns containing numeric data will be added to the *m/z* column list. The image order column can be numeric or text. If column headings are not present the strings "Column A", "Column B", etc. are used for Excel files and "Column 1", "Column 2", etc., for text files. Whenever a new column is selected the number of *m/z* values within the

range of the current data set is displayed next to the file name in square brackets. There must be at least one value for the *OK* button to be enabled.

The method for specifying exported and batch heatmap plot titles is defined as follows. There are six title styles and ten title elements. The preferences INI file (**§5)** variables *ExportFigTitleStyle* and *BatchFigTitleStyle* can be any of the style names or a custom style can be specified mnemonically. This allows the elements to be used in any order. The number of elements on each line of the title are controlled by using the semicolon character to indicate a line break.

The allowed title elements and their meaning are:

mz	*m/z* value	
tolerance	tolerance in ppm units	
mzwindow	tolerance in Th units	
formula	molecular formula	METASPACE annotation only
adduct	adduct formula	METASPACE annotation only
normalization	normalization option (*none, TIC, max, ...*)	
pixelation	*m/z* window treatment (*max, sum, mean*)	
dimensions	columns and rows in the data set	
interpolation	heatmap interpolation algorithm and order	
comment	user comment	
filename	data set name	

Using these elements, the predefined styles are:

none	no title
short	*mz*
batch	*mz tolerance*
trim	*comment*; *mz tolerance*
metaspace	*formula adduct; mz tolerance*

full mz tolerance mzwindow; formula adduct; normalization; pixelation; dimensions; interpolation; comment; filename

The file generated for each image is named using a prefix containing its sequence number in the sorted list followed by the *m/z* value of the image. For example, "seq032_369_3516.png", for the 32nd image with *m/z* 369.3516. This gives the user a way to order the images in the output folder other than by *m/z* value. The *up triangle* and *down triangle* radio buttons to the right of the *Image order* pull-down menu are used to select the sorting direction. If all of the values in the ordering column contain numbers, a numeric sort is used. Otherwise, sorting is alphabetically. For example, the molecular formula $C_{18}H_{36}O_2$ occurs before $C_6H_8O_7$ in ascending alphabetic order. In both cases, the sorting algorithm is stable, *i.e.*, the relative order of equal values is preserved.

An example of a folder of batch images is shown in **Figure 99**. The images are ordered according to decreasing MSM value for a METASPACE annotation file.

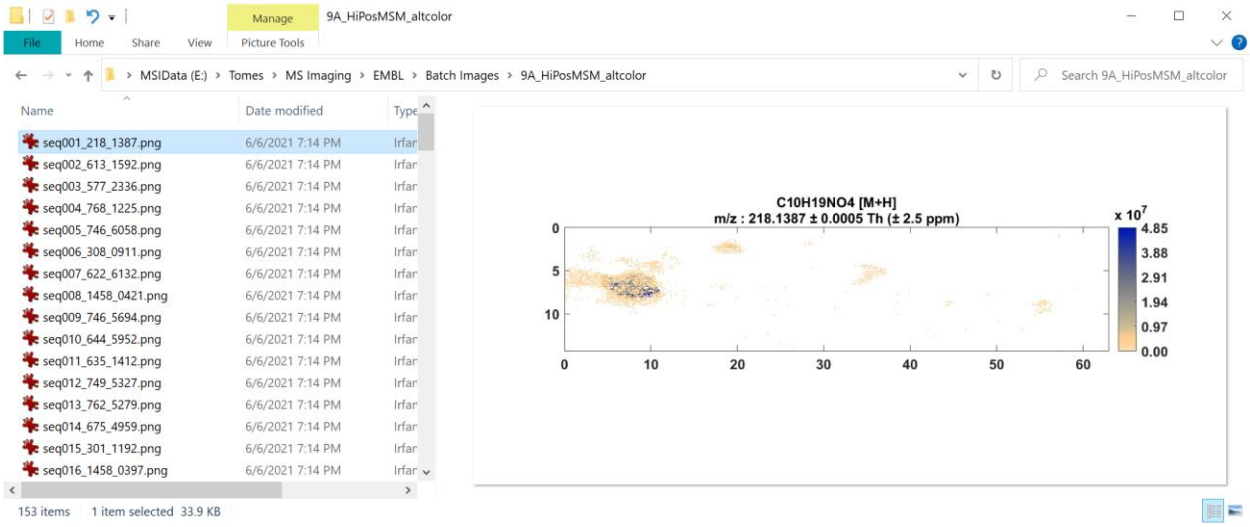
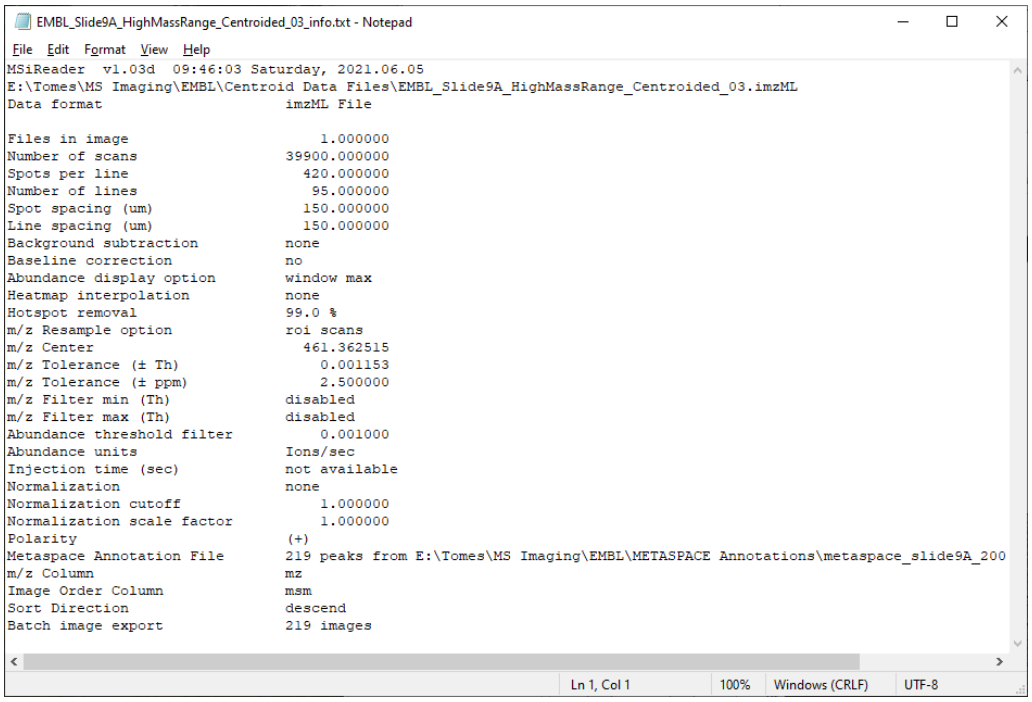


Figure 99: A folder of batch images ordered by MSM value.

In addition to the image files, two text files are saved in the same folder. **Figure 100** shows the contents of the "Info" file which contains details of the MSiReader settings and

MSiCorrelation selections (upper), and the list of *m/z* values and image order values for each image in the folder (lower). If batch image creation is stopped by clicking the MSiReader STOP button before completion, the peaks list file will have data only for the images that were generated.



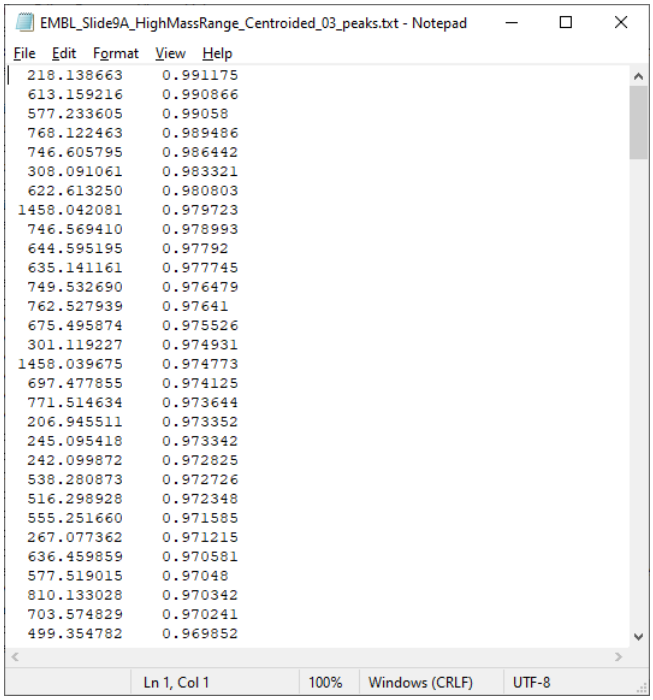


Figure 100: Information file with MSiReader and MSiCorrelation parameters (top) and peaks list file (bottom) for the images folder shown in **Figure 99**.

METASPACE Annotation Format

The MSiReaderPrefs.INI file (**§5)** contains variables that define the meaning of the columns in a METASPACE annotation file. This provides a means to accommodate future changes to the format and for the user to enhance the format with additional columns of information or create an entirely new format of annotation file. These variables are defined in **Table 5**.

Table 5. METASPACE variables in the MSiReader preferences INI file (§5).

Variable	Default Value	Meaning
MetaspaceHeaderRow	3	Row containing column heading character strings.
MetaspaceFirstRow	4	The first row containing data values.
MetaspaceMassSelectionColumns	6	Columns used to create the *m/z* pull-down menu.
MetaspaceRankSelectionColumns	4 5 6 7 8 9 10 11	Columns used to create the *Image order* pull-down menu.
MetaspaceMassSelectionDefault	6	Default *m/z* selection column.
MetaspaceRankSelectionDefault	6	Default *Image order* selection column.
MetaspaceMoleFormColumn	4	Column containing molecular formulas.
MetaspaceAdductColumn	5	Column containing adduct names.

After loading a METASPACE annotation file the *Image order* pull-down menu appears as shown in **Figure 101**. Note that an additional item, *Sequence number*, is always added to the end of the list. If it is selected the *m/z* values are not sorted and the images are assigned sequence numbers according to their order in the CSV file. *Sequence number* is also added to the menu when an Excel or text file is selected.

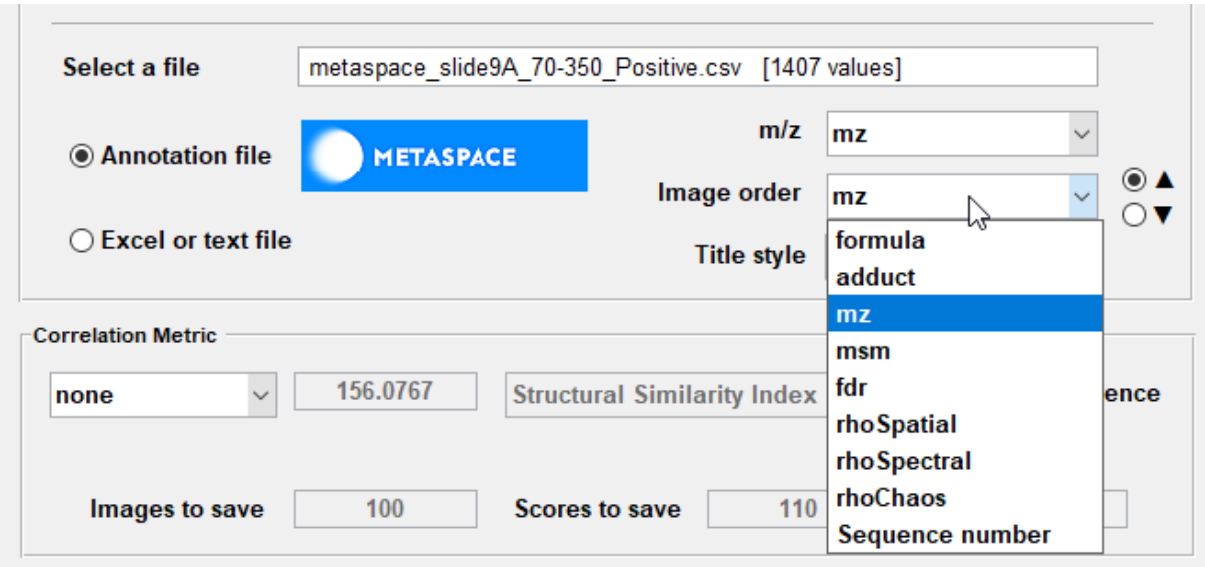


Figure 101: *Image order* selection for a METASPACE annotation file.

The right-click context menu for the *Annotation file* and *Excel or text file* radio buttons is used to clear loaded data so a new file can be selected as the *m/z* peak source as shown in **Figure 102**.

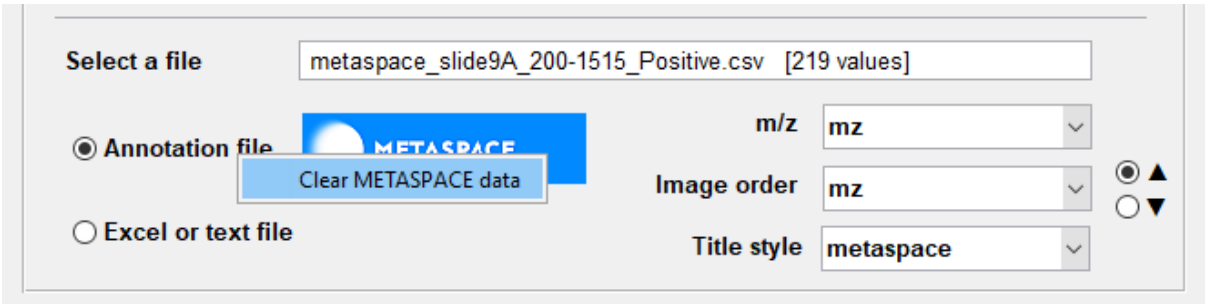


Figure 102: Clearing a loaded METASPACE annotation file (right click on Annotation file). If using Excel or a text file, right click on there will clear those annotation files.

Batch Image Titles

Batch image heatmap plot titles can be formed from the ten elements shown in **Table 6.**

Table 6. Batch image title elements.

Name	Meaning
mz	The *m/z* value for the heatmap.
tolerance	The tolerance window in Th and ppm units.
formula	The molecular formula.
adduct	The adduct formula.
normalization	The current normalization option: *TIC, max, mean, median*, etc.
pixelation	The abundance treatment for each pixel: *max, mean,* or *sum*.
dimensions	The image dimension in pixels.
interpolation	The heatmap interpolation algorithm and order.
comment	The user comment value.
filename	The name of the loaded data set.

There are six predefined formats for the titles of batch heatmap images: *none*, *short*, *batch*, *trim*, *metaspace*, and *full*. These styles are defined using these elements from **Table 7** and the semicolon character as shown in **Table 7**; a semicolon starts a new line in the title.

Table 7. Batch title styles created using the title elements.

Name	Title Style
none	*no title*
short	mz
trim	mz tolerance
batch	comment; mz tolerance
metaspace	formula adduct; mz tolerance
full	mz tolerance; formula adduct; normalization; pixelation; dimensions; interpolation; comment; filename

In the preferences INI file (§5) the *BatchFigTitleStyle* variable can be set to any of the style names or a custom style can be created using the element names from **Table 7** and the semicolon character. When each image is created the actual values are substituted into the title style string in the order specified. Blank lines are omitted from the plot title (*e.g.*, when no user comment is defined).

Figure 103 shows the title style pull-down menu. The additional style, *custom*, is always the last selection in the menu and refers to the style given for the *BatchFigTitleStyle* variable in the INI file (§5), whether or not it is the one of the six named styles in **Table 7**.

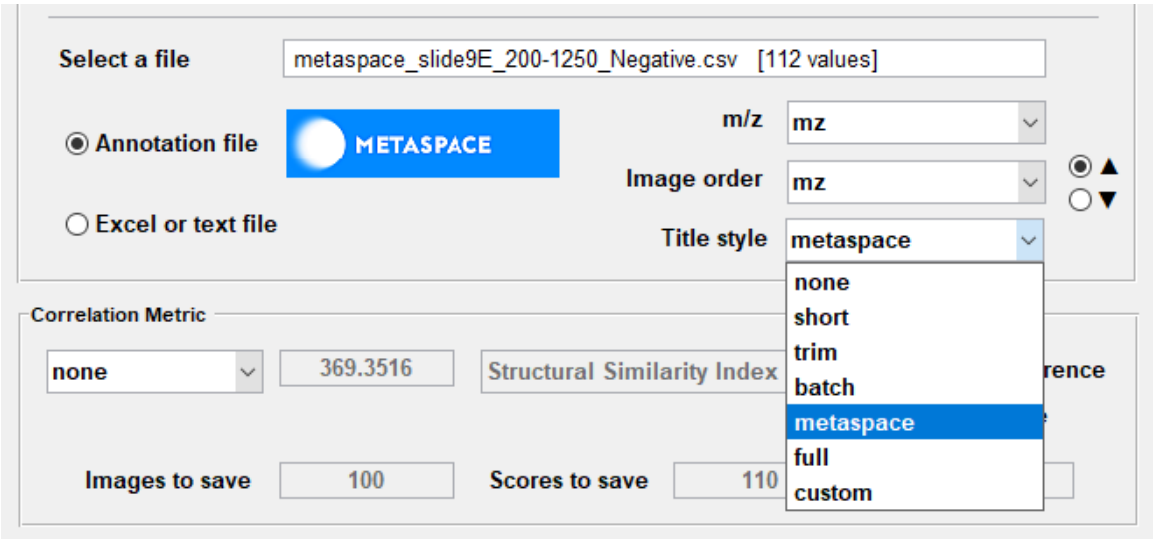


Figure 103: Title style selection.

For example, if the *MSiReaderPrefs.INI* file (§5) has,

BatchFigTitleStyle = mz tolerance; formula adduct; filename

and *custom* is selected from the pull-down menu, a three-line title will be created similar to the one shown in **Figure 104**.

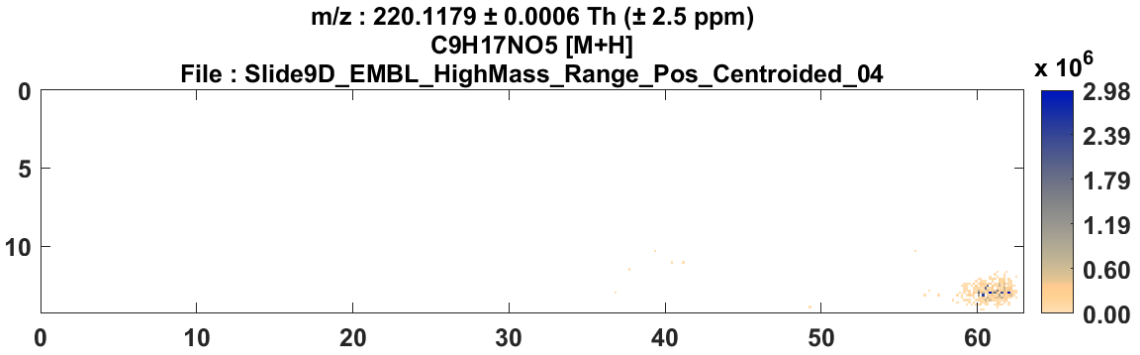


Figure 104: Example of the custom title style with *m/z*, tolerance, molecular formula, adduct, and file name.

However, if the *metaspace* style is selected from the pull-down menu, the same plot would have the title shown in **Figure 105**.

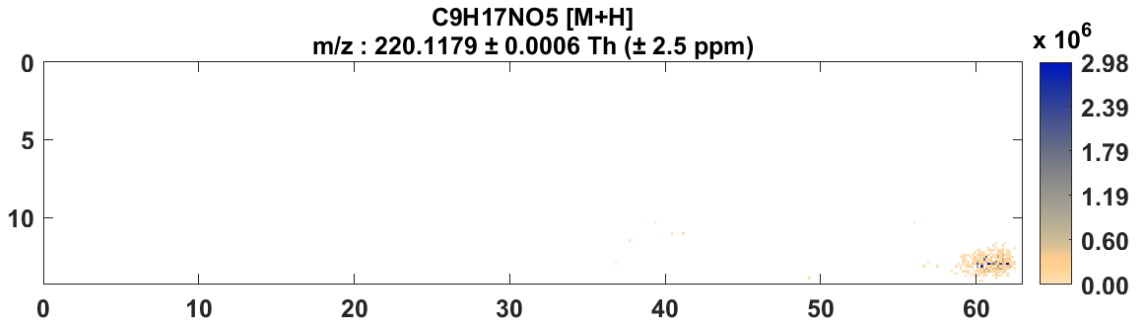


Figure 105: An example of the metaspace title style with molecular formula, adduct, *m/z*, and tolerance.

7.6.6 Data Export

7.6.6.1 MSi Spectrum (generate mass spectrum)

Using the above icons in the toolbar, the user can load a previously saved ROI, select a single scan, a region of interest or the entire data set and export useful information such as

- Averaged spectrum. See below.
- Raw individual mass spectra. See below.
- Abundance values for single or multiple *m/z* values.
- Spectral abundance binned over an *m/z* range.

The user must first select the scan(s) in the current image that are to be processed. The default tool to select a region is *polygon* where the region is chosen by selecting multiple points to form a closed area. Double click to connect the last point to the first and close the polygon. At any time, the user can right-click and force the region to become a rectangle. Points from the polygon region can be deleted, dragged around or new points added (hold down the “A” key and click on any line segment in the polygon) to modify the shape. A single scan region can be created by double clicking once without drawing a line. The default selection tool can be changed to *freehand* (region is drawn over tissue with mouse cursor) or *rectangle* in the preferences INI file (**§5**). *Freehand* ROIs are converted to polygons when the region is closed. The image zoom and pan tools are not reset during ROI selection so that you can zoom in to more accurately select a small region. Most of the other tools also remain enabled allowing you to change the *m/z* value, window units, normalization, interpolation, heatmap color scale and colormap.

 Load a previously saved ROI. Pressing this button will launch a file selection dialog for the user to choose a .mat file containing a previously saved ROI. Saving an ROI is described in **§2.4.3**. The user will be warned if the ROI was originally drawn on a different file or if the present dimensions and scaling do not match the saved information, but the

user will not be prevented from loading it. If the file contains two ROIs, the user will be prompted to select either one or both of them. The ROI(s) will be drawn at the saved X and Y locations on the current heatmap. The toolbar icons for MSiSpectrum, MSiPeakfinder, and MSiQuantification will be enabled if appropriate and the icon used to draw the original ROI will be "toggled".

Use cursor to select a single scan. A cursor will appear on the heatmap and can be dragged to any desired location. Note that any interpolation scheme will be automatically removed after user has pressed this button. A mass spectrum like the one shown in **Figure 106** can be enabled for the scan under the cursor by right-clicking the selected scan. The plot will update automatically as the cursor tool is moved. The abundance and polarity for the scan at the current *m/z* is displayed above the heatmap plot and updates automatically as either the cursor is moved, or the *m/z* value is changed.

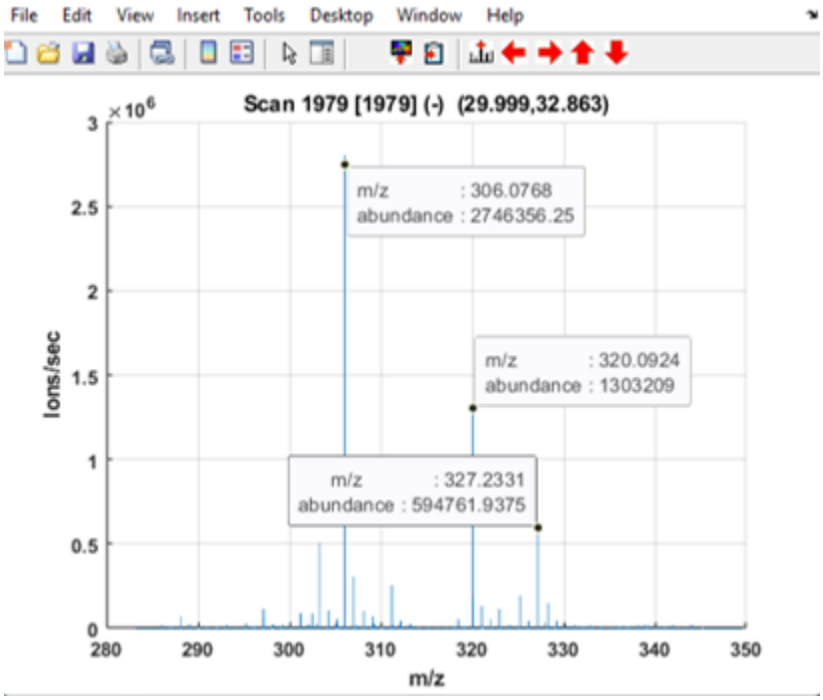


Figure 106: Spectrum plot for a single scan. The title contains the scan number, polarity and location in the image. The data cursor tool has been customized by MSiReader to show the *m/z* and abundance value for the selected peaks.

The plot also has a context menu with five items. The last one keeps the plot in the foreground and the others can be used to set or lock the horizontal and vertical axes (**Figure 107**). Two icons have been added to the spectrum plot toolbar. Clicking on the

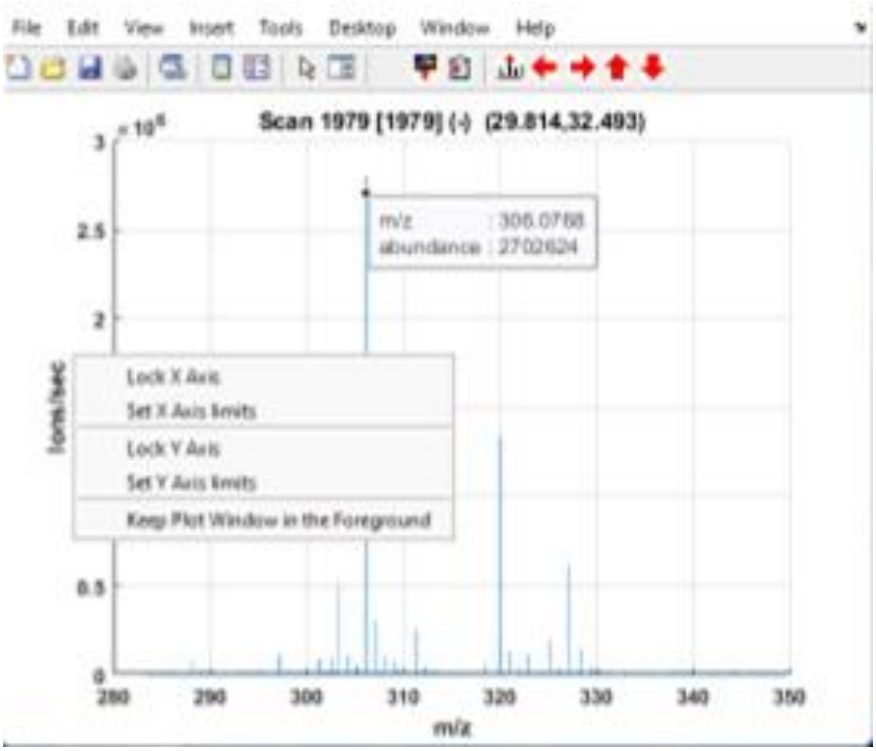


Figure 107: Context menu selections for spectrum plot options.

 icon updates the MSiReader heatmap with a data cursor *m/z* value. If there are multiple data cursors the user is prompted to select one. The selected *m/z* value is also added to the *m/z* history list. The  icon appends all of the data cursor *m/z* values to the clipboard.

 Use cursor to select the scans along a segmented line. After pressing this toggle button, the user can draw a segmented line on the heatmap to define the ROI. The scans that the line intersects are in the ROI. Note that any interpolation scheme will be automatically removed after the user has pressed this button. Like the polygon tool, data points on the line can be moved or added after the ROI is drawn. The line length, abundance mean, standard deviation, minimum and maximum values are displayed above the heatmap plot and update automatically as the ROI is moved or the *m/z* value

is changed. An abundance plot of the scans along this ROI like the one shown in **Figure 108** can be enabled by right-clicking on the line. As the line is moved or edited or the *m/z* value is changed the plot automatically updates.

The plot also has a context menu (right mouse click on the line segment) with five items: Export slice line location information, set color, delete vertex, plot slice line abundance, and select slice plot style (with sub-menu).

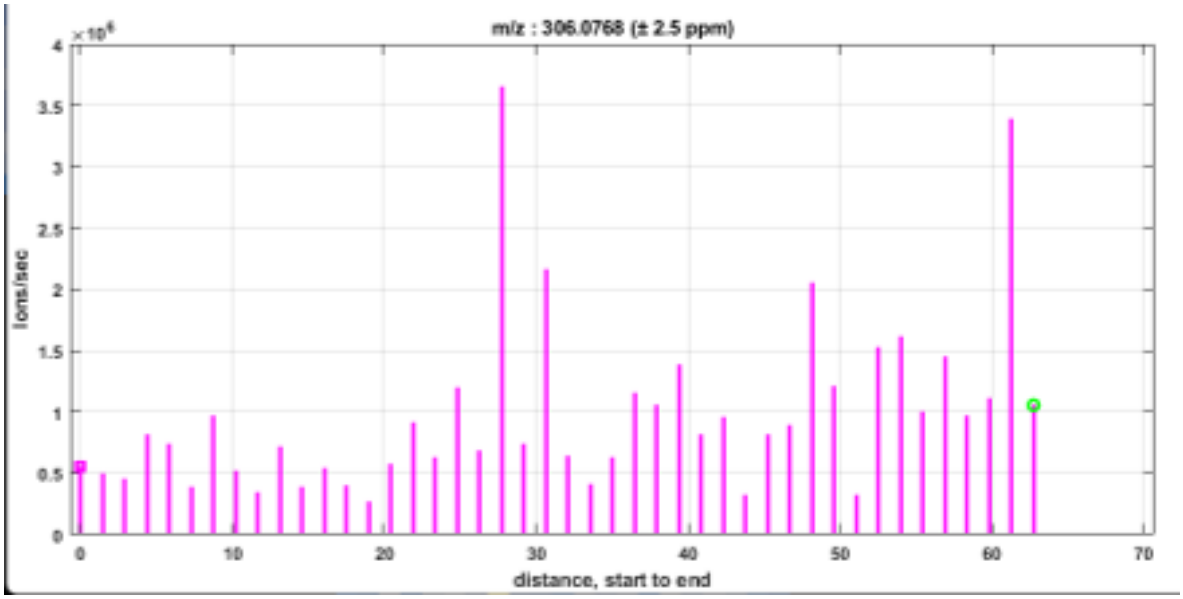


Figure 108: Abundance plot along a segmented line. Locations of the connecting points for the line segments are marked with a black x (if not just a single line), the magenta square and the green circle mark the beginning and end of the ROI, respectively.

Use ROI tool to select a region of interest containing multiple scans. After pressing this toggle button, the user can draw any shape on the heatmap to draw an ROI. Note that any interpolation scheme will be automatically removed after user has pressed this button. The area, and abundance mean, standard deviation, minimum and maximum values are displayed above the heatmap plot and update automatically as the ROI is moved or the *m/z* value is changed.

Select all scans from the image. After drawing the ROI (even a single scan ROI), the right-click context menu for the ROI object can be used to expand the ROI to include all scans in the image. This option is particularly useful for users who want to build heatmaps by combining the abundance value of multiple *m/z*'s (*e.g.*, sum, ratio). The user can export abundances for multiple *m/z*'s, process them in Excel and load the result as a custom heatmap. The area and abundance mean, standard deviation, minimum and maximum values are displayed above the heatmap plot and update automatically if the *m/z* value is changed.

Once the scan or region of interest has been successfully defined, the toolbar icons allowing the user to export spectrum data or heatmap abundances for those selected scans are enabled. These features are described here.

 Export and view mass spectrum data for a cursor or ROI. Upon pressing the icon (or drop down menu under *Annotations* then *Data export* then *Generate mass spectrum*) the MSiSpectrum sub-GUI is launched and several options for data processing and spectrum export will be offered to the user (see **Figure 109**). The user can also choose to extract the data and a centroided peak list to Excel. Options for centroid calculation can be selected in the GUI (minimum abundance threshold, centroid algorithm, etc.). When centroiding data for an ROI or building an average spectrum the *m/z* values from the scans must be resampled to a common set of values. If the data set was loaded with scans having both (+) and (-) polarity, a button group is enabled to select a polarity option, including both polarities.

Figure 109: MSiSpectrum GUI to export a mass spectrum.

Users can also export individual unprocessed profile data for each of the selected scans to an Excel spreadsheet. Note that this operation may take a long time for a large dataset. If the dataset is too large to be exported to Excel, or Excel is not installed an alternate text format is used. A preference .INI file entry (§5) is also available to select text instead of Excel export. Note that unprocessed scan export to a text file can be up to 300 times faster than to an Excel workbook.

The figure toolbar contains navigation tools (zoom, pan) as well as a data cursor tool  that shows the *m/z* and abundance at a selected point on the plot. As shown in the **Figure 110**, multiple data cursors can be added to customize a plot. Magenta marker lines or dots are added as well. The markers can be temporarily hidden by clicking on *Peak Markers* in the legend. The markers (dots or lines) will be hidden until *Peak Markers* is clicked again. Similarly, spectrum visibility can be toggled by clicking on *ROI Spectrum* in the legend. The preferences INI file (§5) variable *MaxMarkerstoView* sets the marker

dialog threshold and *SpectrumPlotStyle* and *SpectumPlotMarkerStyle* define the line style (*line*, *stem* or *stairs*) and the marker style (*point* or *line*).

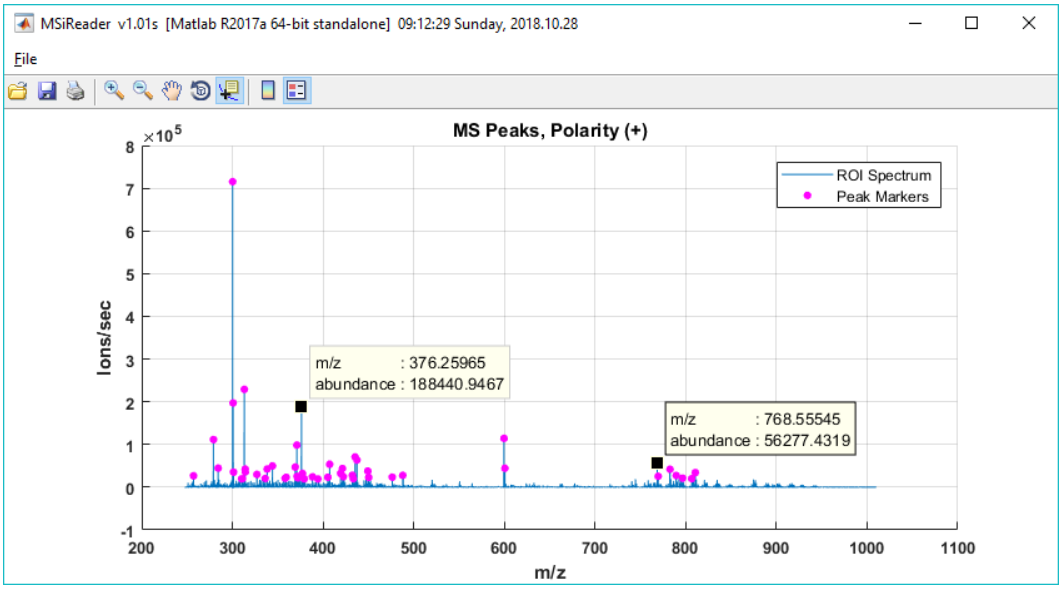


Figure 110: Plotting a mass spectrum in MSiReader.

Exporting ROI location information

After selecting an ROI, the location and scan number of all scans in the ROI can be exported into a text file simply by right-clicking on the ROI and selecting *Export ROI Location Info*. The text file generated will contain 4 columns:

Column 1: Scan number from the original file (assuming meandering, fly back raster pattern)
Column 2: X location on image (column number)
Column 3: Y location on image (row number)
Column 4: Z location on image. Note that if a non-square ROI is exported, the smallest enclosing rectangle will be exported and the scans that are were not in the ROI will be marked as –Z (*e.g.* -1).

Note that this file format can be used as a location file when loading mzXML and imzML files. Only the scans in the ROI will be loaded!

For a tiled image mosaic of a folder of imzML or mzXML files, the ROI location file will have two more columns containing the file number and original scan number of each scan in its file. The first three columns are always the scan number and X,Y location of the scan in the displayed image. An example is shown in **Figure 111**.

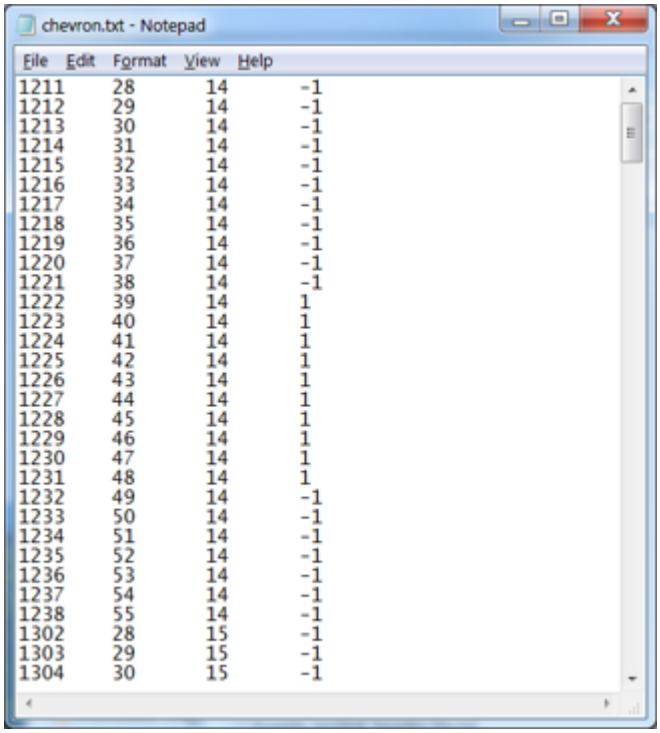


Figure 111: Example of output file generated by ROI export tool.

In addition to the text file described above, a .mat file using the same name is created containing the location information, the file name, and the number of columns and rows. This file can be used to load the ROI into a future MSiReader session in the same location where it was originally drawn.

7.6.6.2 MSi Export (export abundance data)

This icon launches the MSiExport tool which allows the user to export raw or normalized abundance values for each scan in an ROI or to bin the abundance values into a specified *m/z* range. First the user selects a scan, an ROI, or all of the image scans using the icons

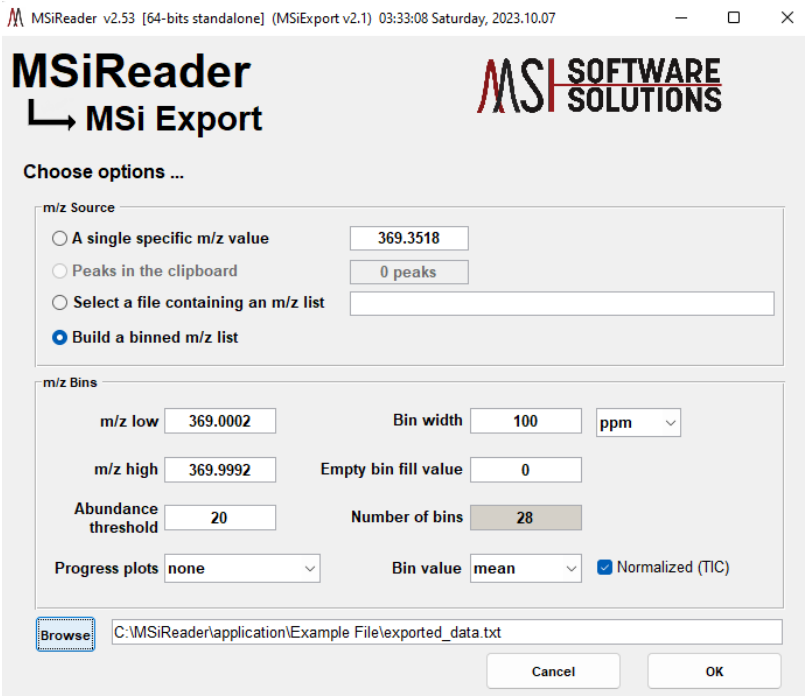


Figure 112: GUI for MSiExport tool.

in the toolbar. Then after selecting the icon in the toolbar or from drop-down menu, select “export abundance data” under *Annotations* and then *Data Export* - the MSiExport GUI shown in **Figure 112** is launched.

Four options are given for the *m/z* value(s) to export: a single value, the values in the clipboard, a list of *m/z* values in a file (*.xlsx, *.txt or *.csv), or a uniformly spaced list (by Th or ppm). For the first three options the raw scan data can be exported to the Excel format shown in **Figure 113**.

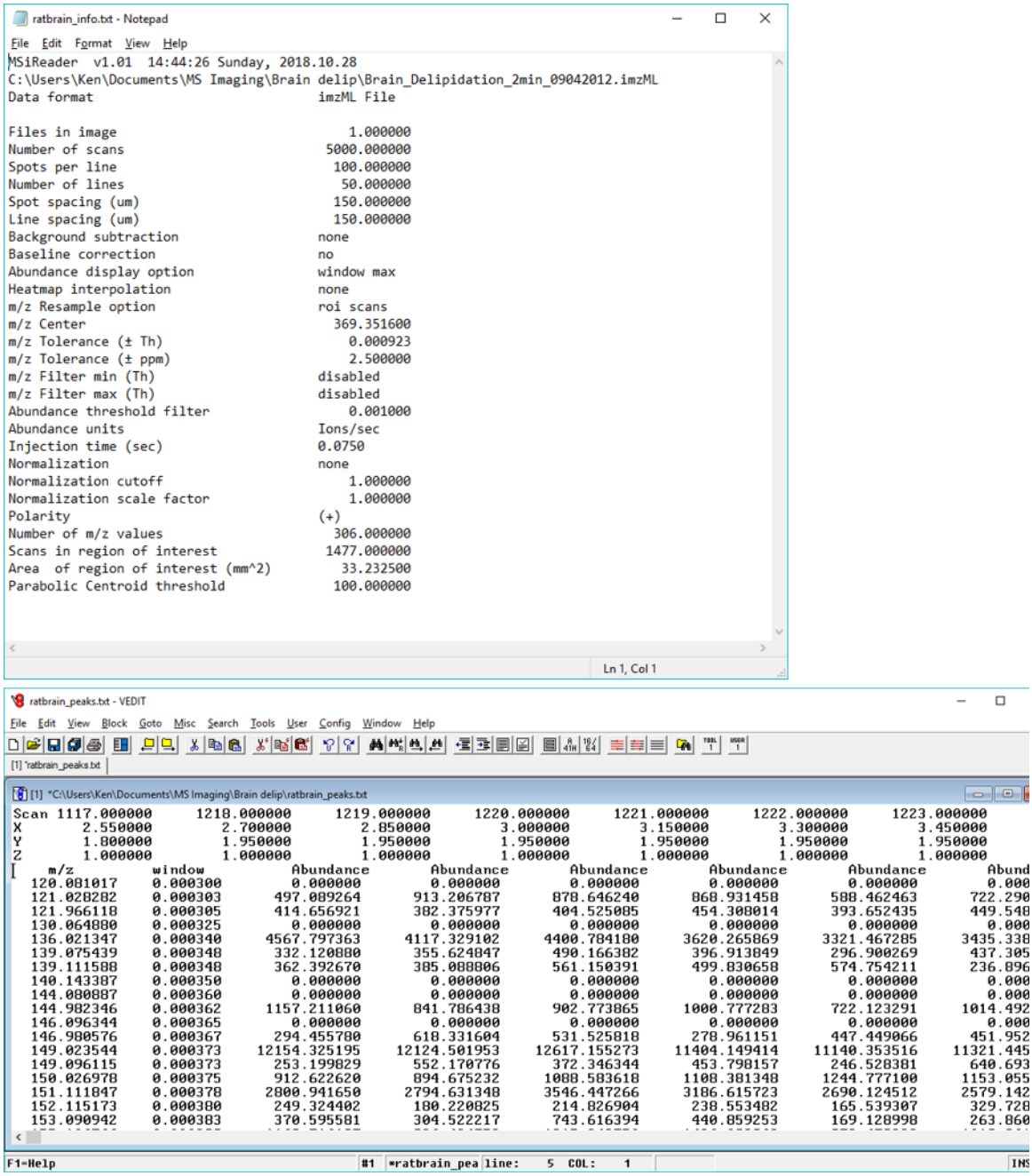
If Excel is not installed, the amount of data is too large, or the preferences INI (§5) variable *ExportPixelsToText* is set to *true*, two text files are created instead. One contains the MSiReader settings normally sent to the Info worksheet (*e.g.*, date and time, file name, number of scans, window and filtering options, etc.) and the other has the raw scan data. Examples of these files are shown in **Figure 114**. All abundance values are calculated using the current *m/z* window and abundance calculation method. Normalization is also applied if any method has been selected. If the loaded data is from a folder of imzML or mzXML files, two columns of data are added to the Excel file shown in **Figure 113**, the file index number and the local scan number from that file. For the text format, two

	A	B	C	D	E	F	G
1				m/z	462.3579	476.3736	548.5404
2				window	0.01	0.01	0.01
3	Scan	X	Y	Z	Intensity	Intensity	Intensity
4	1986	12.9	3	1	160.9126	267.9367	647.2319
5	1987	13.05	3	1	218.4212	186.1572	3362.562
6	2084	12.6	3.15	1	653.3719	590.6677	147.6358
7	2085	12.75	3.15	1	273.7264	297.6645	477.3943
8	2086	12.9	3.15	1	385.0843	641.5623	818.9492
9	2087	13.05	3.15	1	313.7498	110.313	5352.607
10	2088	13.2	3.15	1	667.158	376.2081	1325.003
11	2089	13.35	3.15	1	220.6659	94.43388	1433.628
12	2183	12.45	3.3	1	186.6716	252.6042	0
13	2184	12.6	3.3	1	438.5426	420.6363	144.392
14	2185	12.75	3.3	1	118.7064	0	0

Figure 113: Format of abundance data extracted from an ROI to an Excel file.

additional rows are added to the file shown in **Figure 114** with this information. The index numbers are simply the tile locations in row-major number. Additionally, a worksheet is added to the Excel file with the tiling pattern and a list of the file names.

If you selected a file of *m/z* values as the source, you can load another file by right-clicking on the *Select a file containing an m/z list* radio button text to launch the selection dialog again.



ratbrain_info.txt - Notepad

File Edit Format View Help

```
MSiReader  v1.01  14:44:26 Sunday, 2018.10.28
C:\Users\Ken\Documents\MS Imaging\Brain delip\Brain_Delipidation_2min_09042012.imzML
Data format                          imzML File

Files in image                          1.000000
Number of scans                      5000.000000
Spots per line                        100.000000
Number of lines                        50.000000
Spot spacing (um)                     150.000000
Line spacing (um)                     150.000000
Background subtraction               none
Baseline correction                  no
Abundance display option             window max
Heatmap interpolation                none
m/z Resample option                  roi scans
m/z Center                            369.351600
m/z Tolerance (± Th)                    0.000923
m/z Tolerance (± ppm)                   2.500000
m/z Filter min (Th)                  disabled
m/z Filter max (Th)                  disabled
Abundance threshold filter              0.001000
Abundance units                      Ions/sec
Injection time (sec)                 0.0750
Normalization                        none
Normalization cutoff                    1.000000
Normalization scale factor              1.000000
Polarity                             (+)
Number of m/z values                  306.000000
Scans in region of interest          1477.000000
Area  of region of interest (mm^2)      33.232500
Parabolic Centroid threshold          100.000000
```

Ln 1, Col 1

ratbrain_peaks.txt - VEDIT

File Edit View Block Goto Misc Search Tools User Config Window Help

[1] *ratbrain_peaks.txt

[1] *C:\Users\Ken\Documents\MS Imaging\Brain delip\ratbrain_peaks.txt

Scan	1117.000000	1218.000000	1219.000000	1220.000000	1221.000000	1222.000000	1223.000000	
X	2.550000	2.700000	2.850000	3.000000	3.150000	3.300000	3.450000	
Y	1.800000	1.950000	1.950000	1.950000	1.950000	1.950000	1.950000	
Z	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	
[m/z	window	Abundance	Abundance	Abundance	Abundance	Abundance	Abund	
120.081017	0.000300	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	
121.028282	0.000303	497.089264	913.206787	878.646240	868.931458	588.462463	722.290	
121.966118	0.000305	414.656921	382.375977	404.525085	454.308014	393.652435	449.548	
130.064880	0.000325	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	
136.021347	0.000340	4567.797363	4117.329102	4400.784180	3620.265869	3321.467285	3435.338	
139.075439	0.000348	332.120880	355.624847	490.166382	396.913849	296.900269	437.305	
139.111588	0.000348	362.392670	385.088806	561.150391	499.830658	574.754211	236.896	
140.143387	0.000350	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	
144.080887	0.000360	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	
144.982346	0.000362	1157.211060	841.786438	902.773865	1000.777283	722.123291	1014.492	
146.096344	0.000365	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	
146.980576	0.000367	294.455780	618.331604	531.525818	278.961151	447.449066	451.952	
149.023544	0.000373	12154.325195	12124.501953	12617.155273	11404.149414	11140.353516	11321.445	
149.096115	0.000373	253.199829	552.170776	372.346344	453.798157	246.528381	640.693	
150.026978	0.000375	912.622620	894.675232	1088.583618	1108.381348	1244.777100	1153.055	
151.111847	0.000378	2800.941650	2794.631348	3546.447266	3186.615723	2690.124512	2579.142	
152.115173	0.000380	249.324402	180.220825	214.826904	238.553482	165.539307	329.728	
153.090942	0.000383	370.595581	304.522217	743.616394	440.859253	169.128998	263.860	

F1=Help #1 *ratbrain_pea line: 5 COL: 1 INS

Figure 114: Scan abundance data exported from an ROI to two text files. Above is the header information showing MSiReader parameters and settings and below is a portion of the abundance matrix for the selected scans.

Binned scan data export

Selecting *Build an m/z list and bin the pixels in the ROI* in the MSiExport tool allows the user to export spectral data for the selected ROI that has been binned over a uniformly spaced *m/z* range. This can be done using unnormalized data or normalized data – if the end user wishes to normalize the data, this selection must be done in the main MSiReader GUI first prior to exporting the binned data. This facilitates external processing of the ROI with a multivariate analysis tool such as MSiPCA (**§7.8.2**) or t-SNE (**§7.8.3)**. For this selection the data set may be very large and sparse and thus is always exported to a text file. A *.csv file can also be generated for use in MSiReader (not recommended) or for use in other data analysis programs. The *m/z Bins* pane in **Figure 112** is used to select the binning options. Note that in this case, the data was normalized to the TIC and that is is shown on the bottom right of the GUI so the user knows that the data is normalized and how it was normalized. Their meaning and default values are given in **Table 8**.

Table 8. Binning options.

Bin width (m/z)	full width of each bin
Bin units (ppm or Th)	units of the bin width value
m/z Low	smallest *m/z* value, lower edge of the 1st bin
m/z High	largest *m/z* value, upper edge of the last bin
Empty Bin Fill Value	the value to use for an empty bin
Abundance Threshold	abundance filter, data points with abundance lower than this value are excluded
Bin Result	selects mean, sum or max as the value saved in each bin
Plots	no plots, one plot updated in real-time as the bins are created, or a separate plot for each bin

The initial values for the bin range are the smallest and largest *m/z* values in any of the selected scans. The abundance threshold can be used to filter out low abundance noise

by including only those peaks above the entered threshold. The empty bin fill value can be any numeric value desired, including Inf (positive infinity), -Inf (negative infinity) and NaN (not-a-number). Each bin in the output file will contain either the mean abundance, sum of abundances or maximum abundance of the *m/z* values that fall in that bin. Finally,

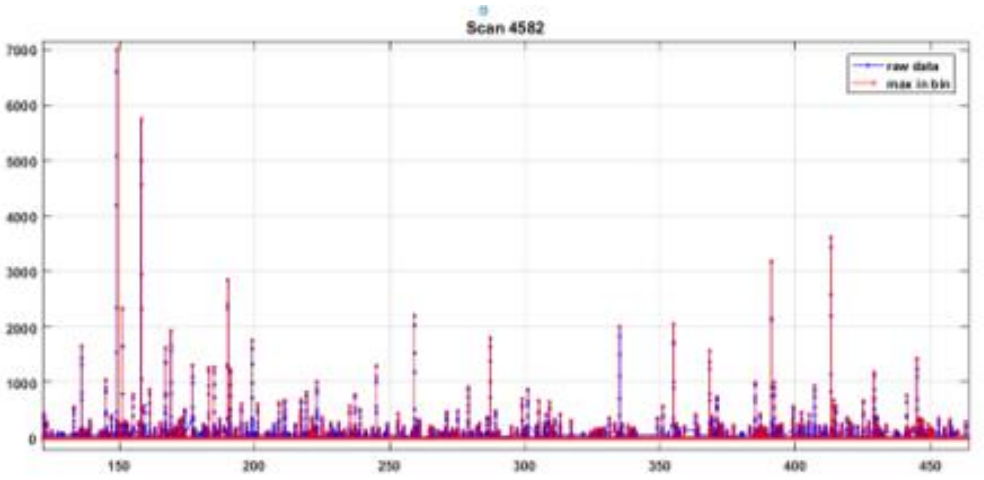


Figure 115: Plot showing original and binned spectrum for a scan.

the plot option allows the user to observe the binning progress as it is applied to each scan on a continuously updated plot similar to the one in **Figure 115**. This is useful for exploring the tradeoff between the filter threshold and the bin width. Do not select the *Separate Plots* option if the ROI has a large number of scans; this will likely cause your computer to lock up.

After the options are selected the user is prompted for a name and location to save the binned data. The file is written in a text format shown in **Figure 116** or can be output as a *.csv file. The first 36 lines of the text file contain the name of the data set, MSiReader options and parameters and the binning options, next is a vector with the *m/z* values for the center of each bin. As shown in the figure, each row of the binned abundance matrix corresponds to a scan from the ROI where the first 6 columns give the spatial location of that scan. For this example, the data was a folder of imzML files so the file number and local scan number for each scan are also given. The remaining columns contain the composite abundance values (mean, sum or max) for each bin.

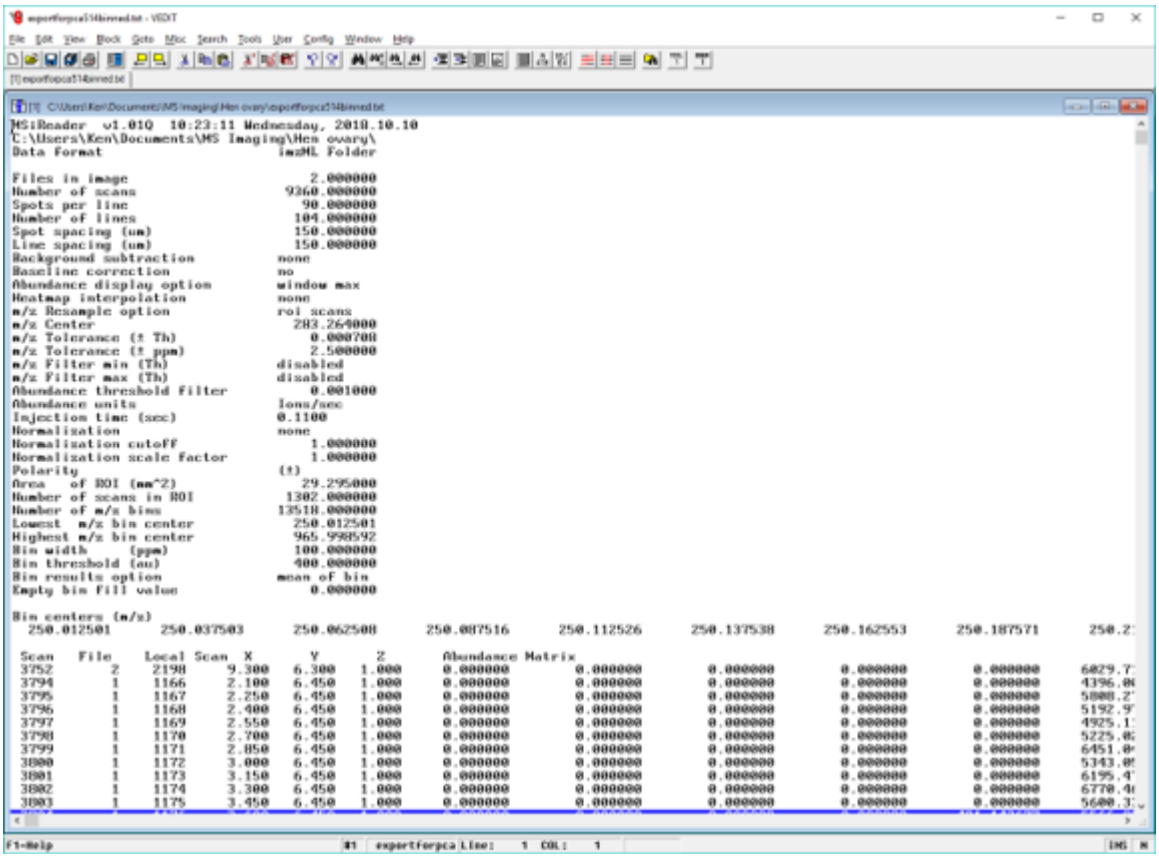


Figure 116: Example of the binned data text file format.

The exported text file may be very large and unsuitable for opening with Notepad. The text format was chosen to facilitate input for further analysis, for example MSiPCA. Two Matlab functions are included in the MSiReader installation to process a binned data export file: *getmsibindatainfo* and *loadmsibindata*. The first function can be used to query the file without opening it with a text editor and before reading it into a program. This is useful for assessing the resource requirements of a very large data set to determine if you have sufficient memory to process it or if you should export it again with a larger bin size (and thus fewer bins) or a higher filter threshold. The second function, *loadmsibindata*, loads the data set into a Matlab matrix and returns the center *m/z* for each bin, the list of scans in the ROI and the same information structure returned by

getmisbindatainfo. The binned abundance matrix can optionally be returned as a Matlab sparse matrix instead of a full rank matrix containing many zero entries. Both functions have help text that can be viewed by typing *help getmsibindatainfo* and *help loadmsibindata* in the Matlab command window.

7.7 Quantification Menu

7.7.1 Relative – MSi Peakfinder

For a **video tutorial** on how to use MSiPeakfinder, click **HERE**.

MSiReader implements a peak picking strategy to identify ions in a sample that are more abundant in a user defined interrogated ROI as compared to a reference ROI as shown in **Figure 117**. Peaks are detected by comparing the average abundance of each *m/z* value and its occurrence over both user defined ROIs. Criteria used for detection are user defined and can be easily modified within the interface. A mass spectrum plot showing the superimposed averaged signal for both the interrogated and the reference ROIs is

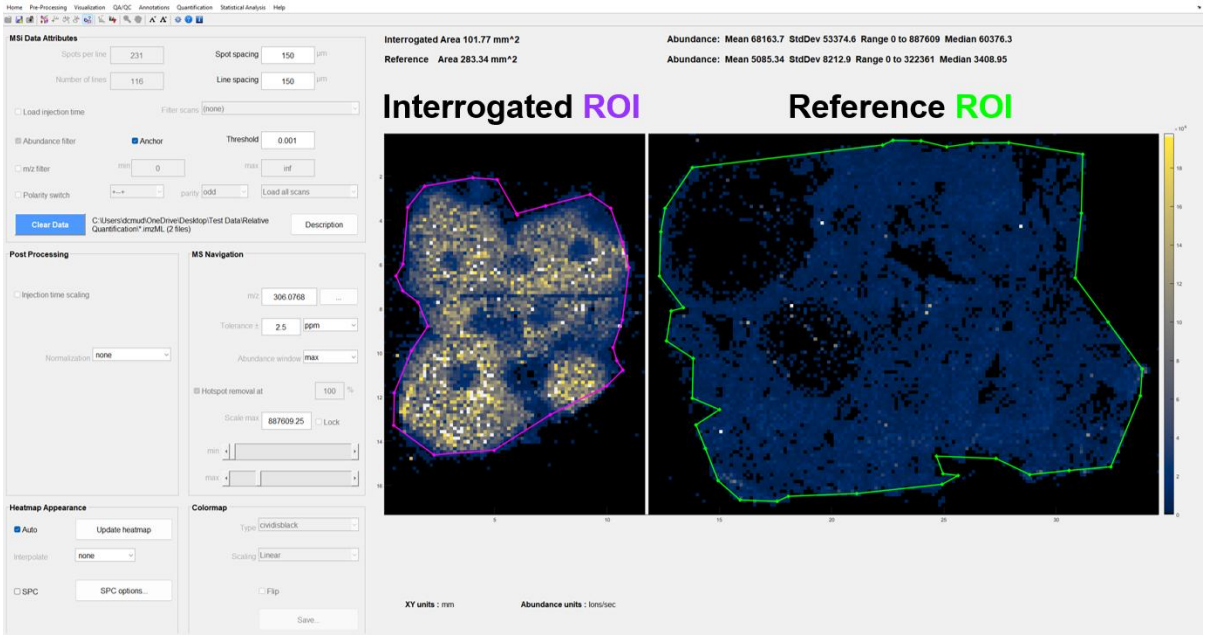


Figure 117: Automatic peak detection process using MSi Peakfinder. Start this process by selecting a reference and interrogated ROI.

optionally generated (**Figure 118 and Figure 119**) and can be used to browse through the peak list. The extracted peak list can also be sent to the clipboard or saved into an Excel workbook.

The process begins when the user presses the  toolbar icon to select the interrogated and reference ROIs using the polygon drawing tool and proceeds as follows. Alternatively, the  toolbar icon may be used to select previously saved interrogated and references ROIs.

1. Any interpolation scheme will be removed so that each pixel corresponds to a single mass spectrum.
2. The user draws the interrogated and reference ROIs or loads a saved ROI .mat file containing both the interrogated and reference ROIs.
3. When the user selects *Relative MSi Peakfinder* under the *Quantification* menu, the MSiPeakfinder GUI is called and the user selects parameters for the peak finding algorithm as shown in **Figure 118**.
4. After clicking the OK button, all spectra in the interrogated ROI are averaged and a peak list is generated. A progress bar is shown at the bottom to give the user an indication of the length of time remaining for processing.
5. Then all spectra in the reference ROI are averaged and a second peak list is generated. This is done automatically. The progress bar shown at the bottom starts over for the reference ROI to give the user an indication of the length of time remaining for processing. Steps 4 and 5 require a lot of data processing so please be patient.
6. The resulting peak list is copied to an Excel worksheet that can be used as an input file for the correlation and batch processing of images. (**§7.6.5**) If the Mass Excess template file is selected, lipid plots are automatically generated when the peaks are inserted into the worksheet.

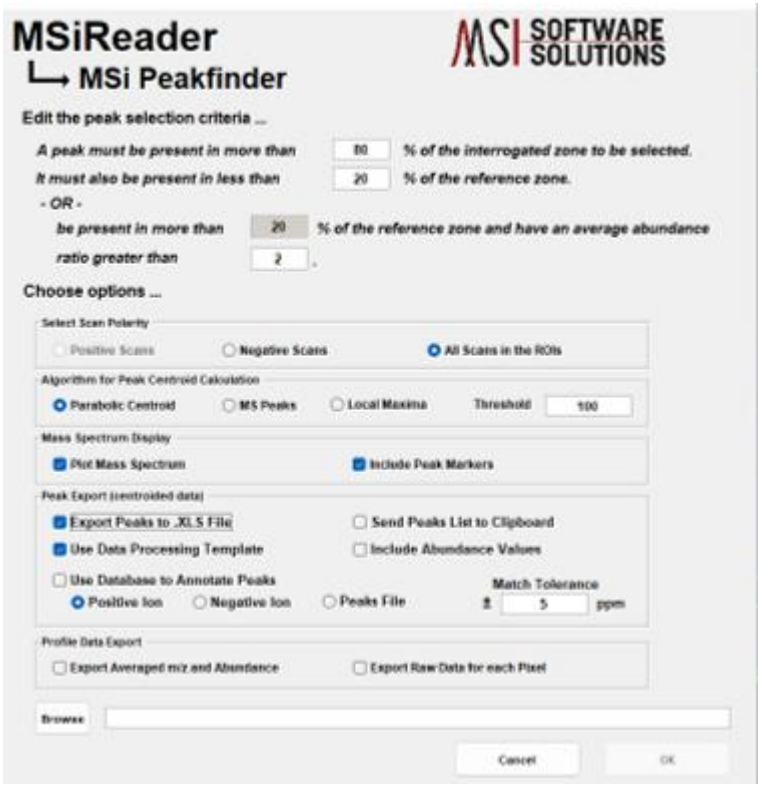


Figure 118: MSiPeakfinder GUI used to select peak picking parameters and options.

The interrogated and reference ROIs have context menus to export location information (*i.e.*, the scan numbers and X, Y location of each scan), change the line color, convert the ROI into a rectangle and to swap the reference and interrogated ROIs. This latter feature can be used to easily carry out a two-way comparison between ROIs without having to move or redraw them. If a folder of imzML or mzXML files were loaded, the ROIs can be anywhere in the image mosaic.

In the case of overlapping interrogated and reference ROIs there are several ways to treat the scans in both regions when applying the peak picking algorithm. The user is prompted to select an option as shown in **Figure 119**.

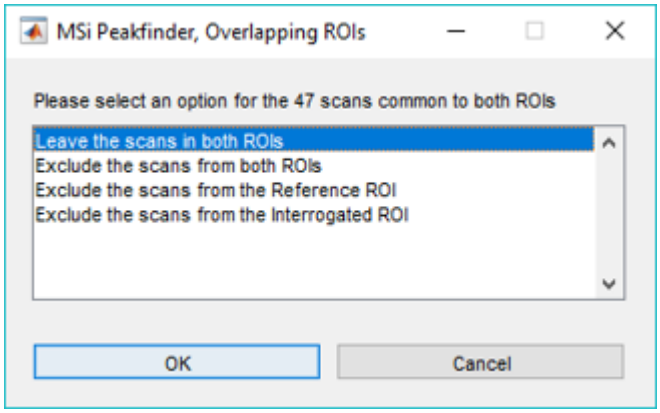


Figure 119: Overlapping ROI action dialog.

In the case of one ROI being completely enclosed inside the other (*e.g.*, the interrogated ROI is inside the reference ROI) only two of the choices make sense and the dialog in **Figure 120** is shown. If the reference ROI is inside the interrogated ROI a similar dialog is displayed. Coupled with the ability to expand an ROI to be all scans, this is a convenient way to compare all background scans (reference ROI) with tissue scans (interrogated ROI).

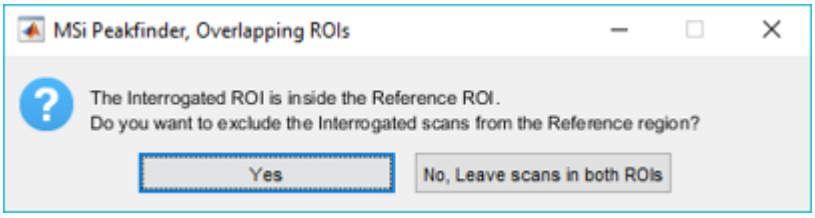


Figure 120: Dialog for the special case of an interrogated ROI that is inside the reference ROI.

Note that the tolerance for peak comparison between the interrogated and reference ROIs is specified by the user defined *m/z* window when the peak finding feature is launched. If the *m/z* window units are chosen by the user to be parts-per-million (ppm) in the MS Navigation panel, a ppm *m/z* window will also be used when peaks are compared. Information about necessary resampling steps when averaging data as well as the centroid calculation are given below.

Resampling

Some of the imaging data may be processed by the instrument vendor software or the format converters to reduce the file size (*e.g.,* zeros are omitted), and therefore the spacing between the data points on the *m/z* scale and the total number of data points are not the same for every scan. Prior to averaging spectra over a certain ROI, each individual spectrum of the ROI must be resampled to a common set of *m/z* values. This means that omitted zero values may have to be added and some data points will be interpolated. Resampling is only done when peak picking is performed or when spectra are averaged. One of three *m/z* resample options can be selected with a context-menu (right mouse click) for the Algorithm for Peak Centroid Calculation Panel in either MSiSpectrum or MSiPeakfinder as shown in **Figure 121**.

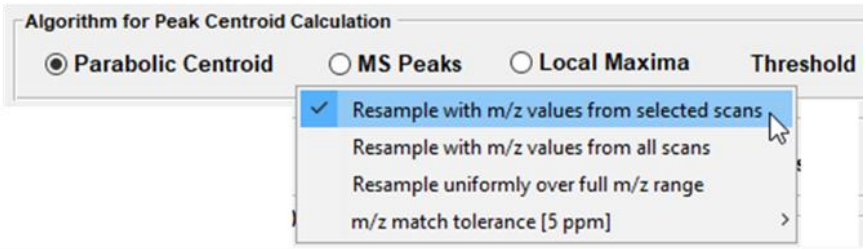


Figure 121: Context menu for selecting an *m/z* resampling algorithm.

Option 1: Resample to all existing data points on the *m/z* scale found in all the spectra in the ROI (default option).

Option 2: Resample to all existing data points on the *m/z* scale for all the spectra in the image (no matter where the ROI is).

Option 3: Resample all scans uniformly over the entire *m/z* range, regardless of the presence of a signal for a particular *m/z* in any of the scans. If the *m/z* data points are systematically different from spectrum to spectrum it is preferable to use this Option instead of Option 1 or 2, either of which may generate an extremely large number of *m/z* data points for this case.

The implementation of the resampling options in MSiReader improves performance. The new algorithm for merging *m/z* vectors from the scans in an ROI can be as much as 100 times faster. Previously, as the *m/z* values from separate scans were combined into a single vector only exact duplicates were removed. Now a tolerance for this test is used. The tolerance can be expressed as an absolute minimum difference in Th units or as a ppm value. The tolerance value and units can be changed using the final context menu item, *m/z match tolerance [5 ppm]*, in **Figure 121**. The preference INI file (**§5)** variable *mzResampleOption* can be used to select a default resampling algorithm. The tolerance value defaults to the current MSiReader Navigation panel settings when MSiSpectrum or MSiPeakfinder is launched. The upper limit on the size of the resampled *m/z* vector can be set with the *mzResampleMaxPts* variable in the preferences INI file (**§5)**. The default value is 1e7. If the number of *m/z* values needed for resampling exceeds this value, the user is prompted to confirm switching to Option 3 to resample uniformly or to abort the peak picking operation.

Algorithm for centroid calculation

Three different algorithms can be used to extract a peak list from the spectrum data (selection was made in MSiPeakfinder GUI shown in **Figure 118**). The algorithms use different approaches to calculate the centroid from the mass spectrum data. The first option is to use a Parabolic Centroid Algorithm. This algorithm was proposed by Comisarow and Marshall in their early work interpreting FTMS spectra[20,21]. For each local maximum in a mass spectrum, the centroid location will be the calculated *m/z* of the apex of a parabola fit to that local maximum and the 2 adjacent points (see **Figure 122**). The centroid locations calculated with this method are nearly identical (within a fraction of ppm) to those calculated by using instrument manufacturer software.

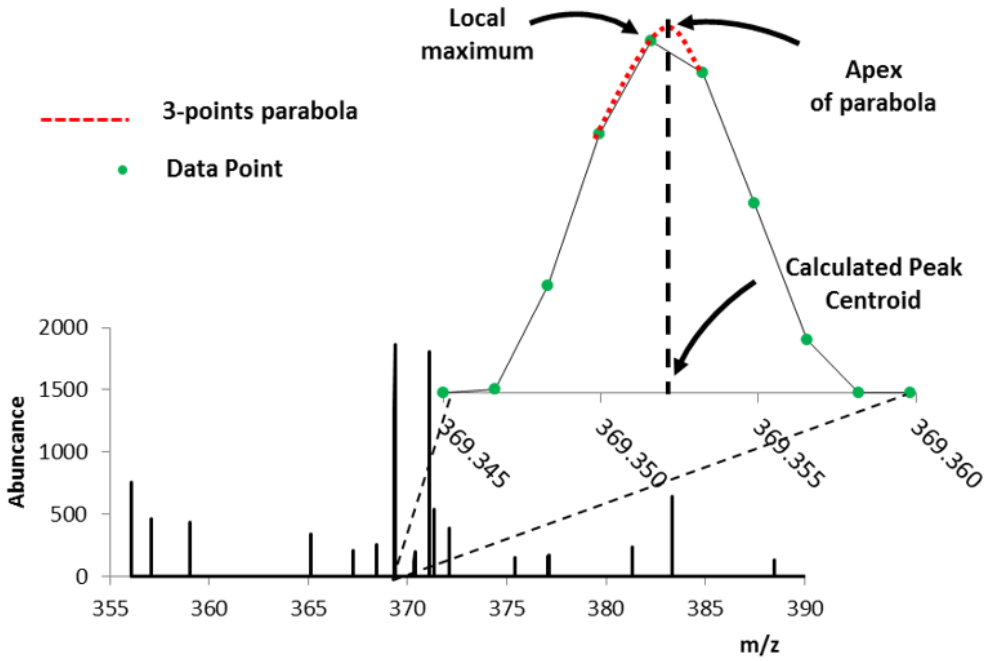


Figure 122: Calculated peak location using Parabolic Centroid Algorithm.

The second algorithm uses *mspeaks* function to calculate peak centroid locations.

The third algorithm is to select the *m/z* with the maximum abundance value in each data window as a peak. This is of course the fastest of the three algorithms and ensures that selected peaks have *m/z* values and abundances that are identical to data points in the original scan data.

The basic settings and options of this function can be easily changed using the preferences INI file (**§5)**. Also note that using *mspeaks* may increase computation time significantly.

IMPORTANT: The first two centroid algorithms may produce unexpected results if the input file is not an actual mass spectrum but a peak list (preprocessed centroid data).

Saving Results in MSi Spectrum and MSi Peakfinder

The user is given the option to export data in Excel or text format from both the MSiSpectrum GUI and the MSiPeakfinder GUI. Excel format is selected by default; however, a text format will be used if Excel is not installed, the size of the data to be exported exceeds the limits of Excel or the INI preferences variable *ExportToExcel* is set to *false*. Depending on the options selected on the GUI, here is a description of the worksheets or text files that will be generated:

Info: Contains information about the version of MSiReader used to export the data, the name of the data set, the MS Navigation and Post Processing panel parameters, the size and number of scans in the image and the region(s) of interest, the algorithm used for peak extraction and any other options selected.

Files: For the folder of imzML and folder of mzXML data formats a list of the file names is stored in this worksheet along with the number of columns and rows in the tiling pattern. The file name field in the Info worksheet contains the folder name.

Mass Excess – Lipids: By selecting the checkbox *Use Data Processing Template*, the Excel workbook template used to save the extracted peak list will already contain an

imbedded data analysis tool for tissue imaging. A mass excess plot is automatically generated from the extracted peaks (peaks in *Centroid data* spreadsheet) and compared to the mass excess distribution of lipids from the LipidMaps database. If the data is being exported to text files this option is not enabled. The user can modify the Excel template in the MSiReader folder to add custom data analysis. The name (but not the location) of the template file is stored in the preferences INI file and can thus be changed there. Users willing to share their templates with the MSI Community may contact us via email at support@msireader.com.

Centroid data: This worksheet contains information about the peak centroid found in the ROI. When the MSiSpectrum GUI is used, all peaks above the threshold that were found in the averaged spectrum of the scans in the ROI will be reported in this worksheet. When the MSiPeakfinder GUI is used, only peaks corresponding to the *Peak Selection Criteria* will be included. By selecting *Include Abundance Value*, the average abundance of those peaks will also be included in the second column of the worksheet.

Average Spectrum: When the *Export Averaged Spectrum and Abundance* checkbox is selected, the profile data of the averaged spectrum will be exported here. Note that the profile data on this worksheet is resampled data. When using MSiPeakfinder, both the Interrogated and Reference average spectra will be exported in this worksheet.

Raw Profile Data: When *Export raw data for each pixel* option is selected, profile data for each scan in the ROI (spectrum export) or Interrogated ROI (peak finding) will be exported in this sheet. Note that if background subtraction was performed, background subtracted data will be exported. Normalization does not affect the exported profile data.

7.7.2 Absolute – QMSI spatial calibration curve

A method to quantify tissue response based on a regression fit to spots of known and varying concentration is implemented by the MSiQuantification tool. The procedure for using this tool is shown in the figures and narrative below.

1. Load a data set into MSiReader, adjust the spot and line spacing, select an *m/z* center value for the heatmap display and normalize to a reference peak (optional). The dataset must contain a spatial calibration curve deposited on the tissue. Select the polygon ROI tool and draw a region of interest (in this case, the entire tissue was selected). GUI is shown in **Figure 123.**

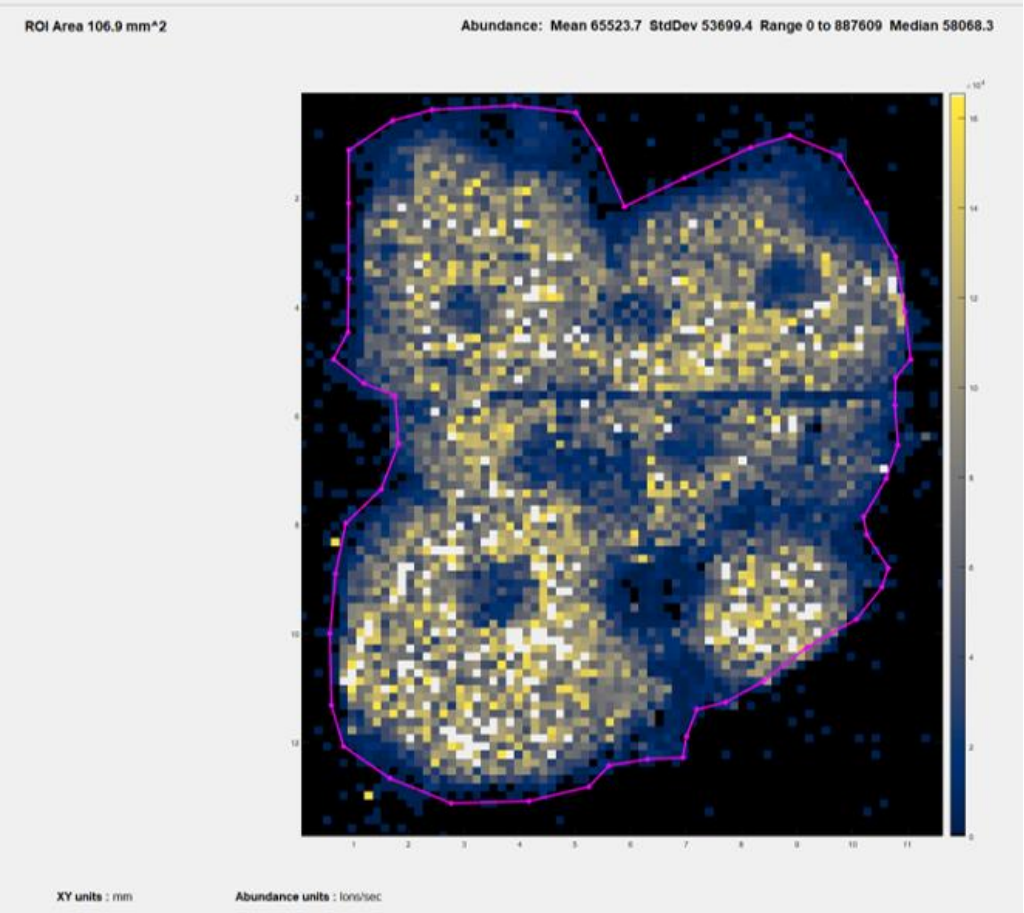


Figure 123: MSiReader GUI after loading a data set for quantification and drawing a ROI around the entire tissue.

2. Next launch the quantification tool by selecting Absolute – QMSI spatial calibration curve from the drop-down menu under Quantification. The MSiQuantification GUI is shown below in **Figure 124**. It has edit boxes for entering the calibration parameters (density, thickness, calibration *m/z* and liquid volume), buttons for creating and removing ROIs around the calibration spots and edit boxes to enter a solution concentration for each ROI. The response treatment for calibration spots and the tissue ROI can be either the mean of the scan abundances or the sum of the abundances. Unit conversions can be calculated by typing arithmetic expressions in the edit boxes.

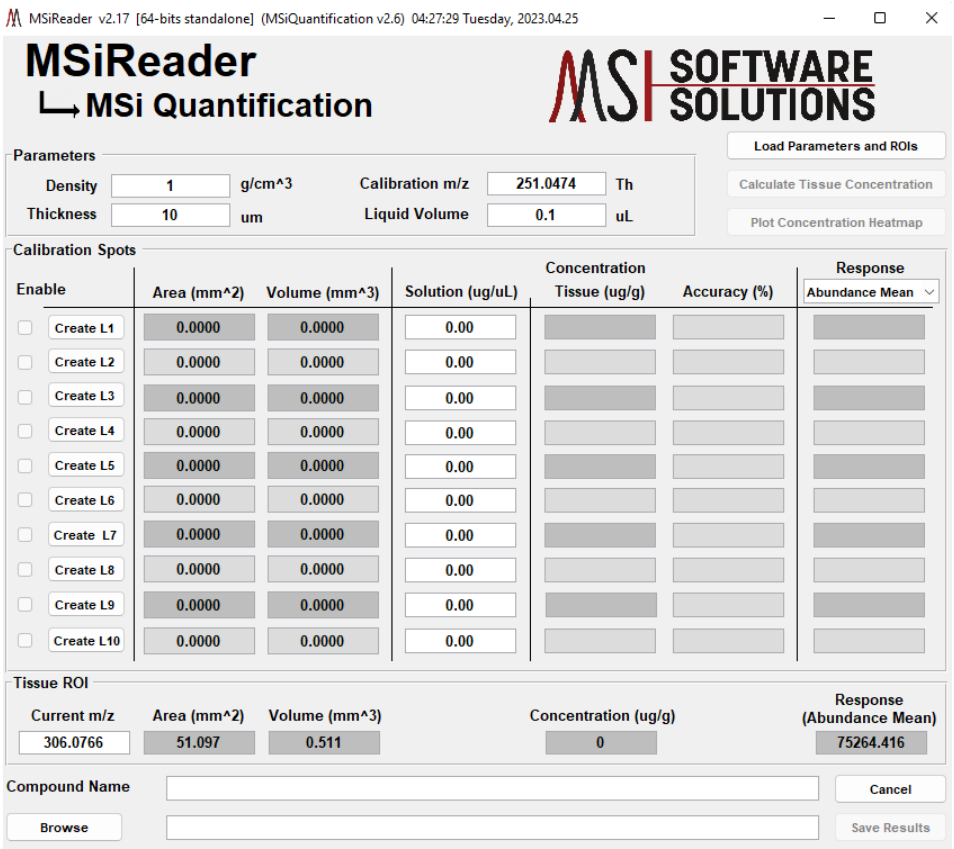


Figure 124: The MSiQuantification GUI.

The current *m/z* value in the *Tissue ROI* pane should be changed to the *m/z* of your internal standard to allow the user to draw the spots in the correct location (*i.e.*, to make the spot scans "light up" in the heatmap). In this case, our internal standard has an *m/z* = 309.0797. The colorscale edit boxes and slider controls remain active in the MSiReader main GUI as well and can be adjusted to help reveal the calibration spots. In **Figure 125** the max color scale slider has been lowered and six spots are clearly visible. There is a seventh spot with a zero-concentration solution amount.

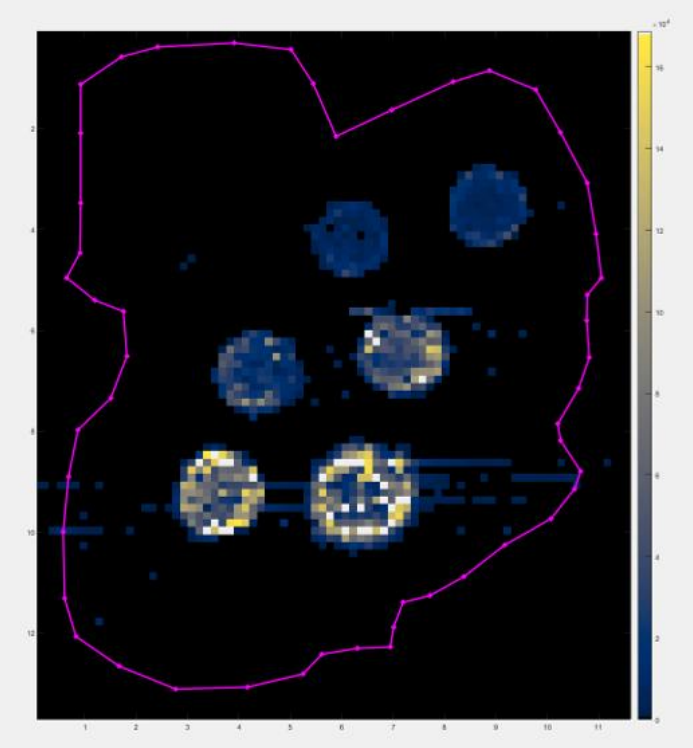


Figure 125: The *m/z* value has been changed in the Tissue ROI panel. The MSiReader colorscale and hot spot percentile were adjusted to increase the visibility of the internal standard spots.

3. Up to ten spot ROIs may be identified on the MSiReader heatmap. Clicking a *Create L1*, … button in the MSiQuantification GUI will switch window focus to the MSiReader window and enable the polygon drawing tool for the user to draw an ROI. The ROIs can be created in any order and moved and edited with the mouse. When the minimum number of ROIs for calibration have been drawn, the GUI is ready to calculate the tissue concentration. The default minimum number of spots is three, but this can be changed in the preferences INI file (**§5)** to a larger value. The corresponding spot solution amounts can be entered at any time and in any order. The enable/disable

checkboxes can be used to see the effect of removing an “outlier” spot from the calculation without deleting it. The right-mouse context menu for each spot ROI contains an identifying label, *e.g.*, L1, L2, L3, etc. In **Figure 126** seven spots have been drawn; the 6 calibration points plus an ROI to represent the zero concentration.

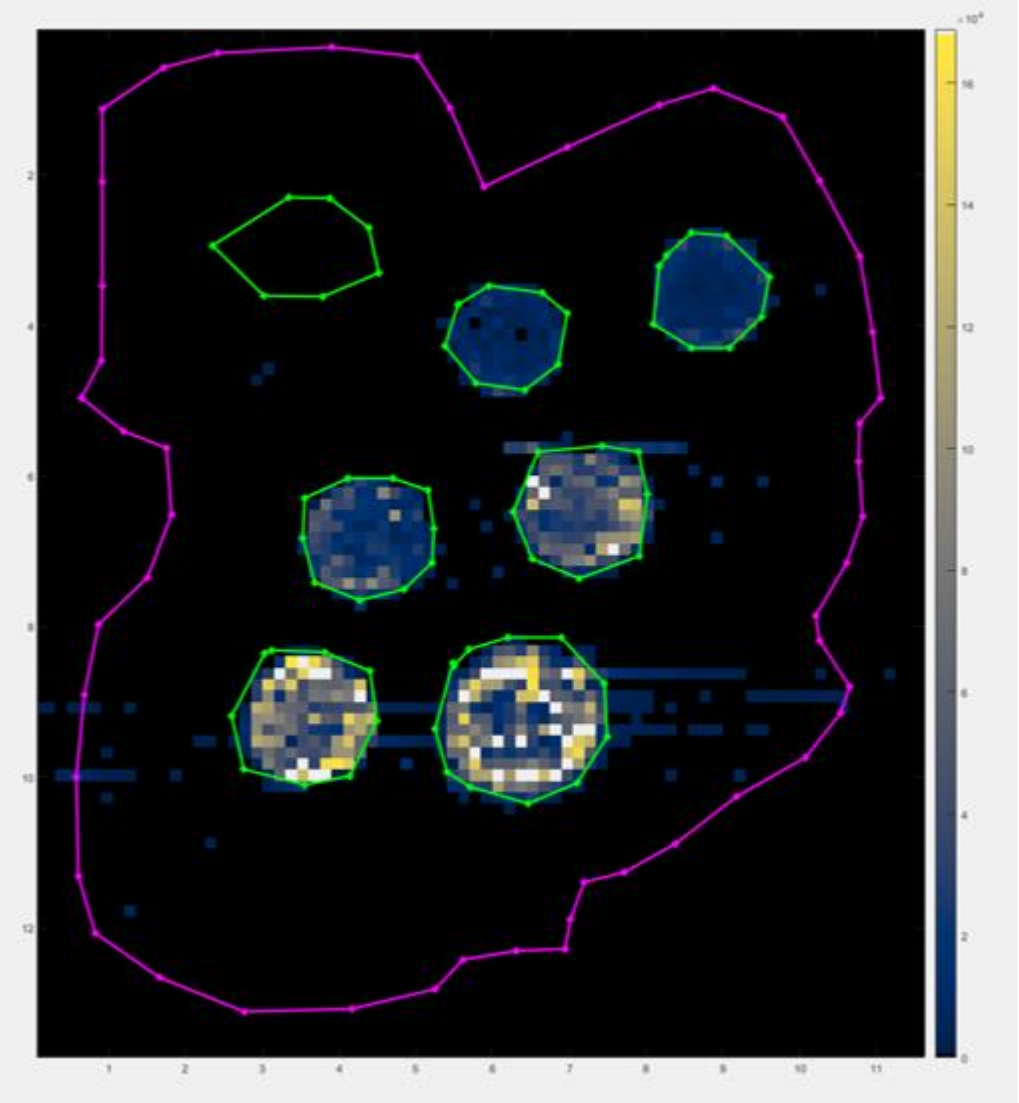


Figure 126: MSiReader after seven calibration spot ROIs have been drawn.

As an ROI is moved or edited, the area, volume and response values in the quantification table are updated automatically. If the liquid volume and solution concentration for an ROI

has been entered, the tissue concentration is also calculated for the internal standard. After enough ROIs are complete and the solution concentrations (μg/μL) are entered, the MSiQuantification GUI will appear similar to

Figure 127 and the *Calculate Tissue Concentration* button will be enabled.

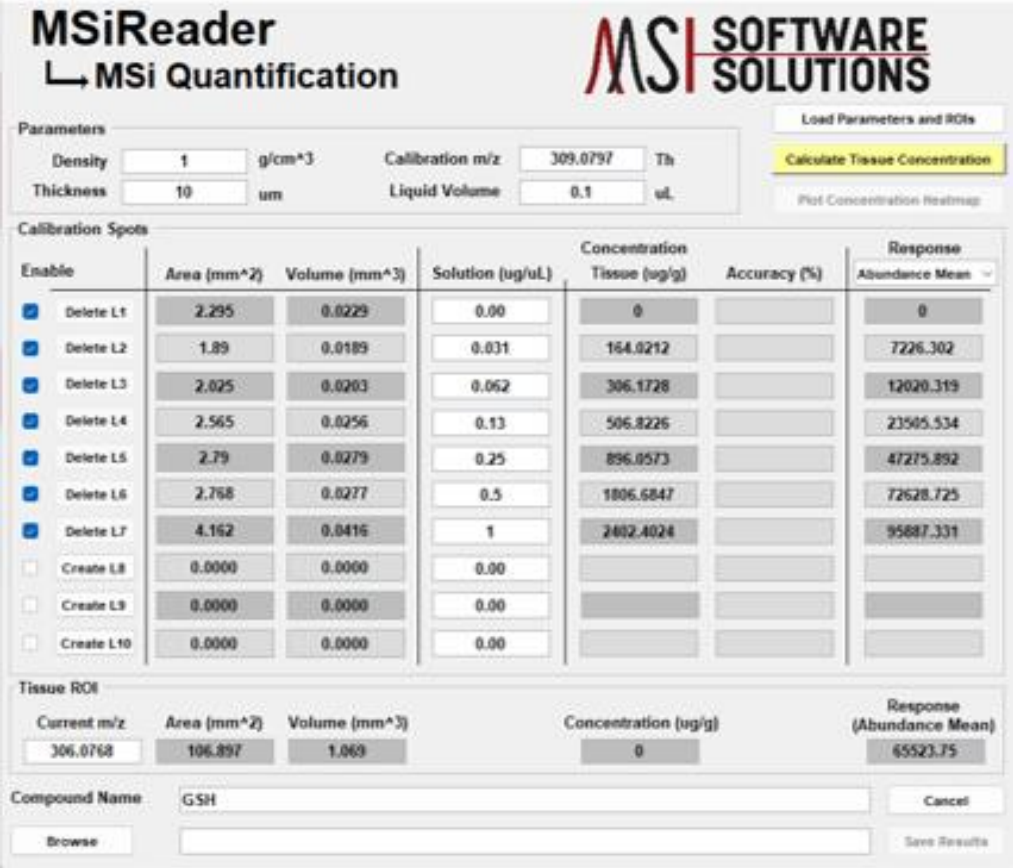


Figure 127: MSiQuantification GUI after drawing ROI spots and entering solution amounts.

4. Next change the Tissue ROI panel *m/z* value back to the correct value if it was changed in Step 2 (*m/z* = 306.0768 for this data set) and click the *Calculate Tissue Concentration* button. A linear regression of the spot response values (dependent variable) and their concentrations (independent variable) will be performed and the accuracy of the fit for each spot (100 minus percent error) calculated. The

concentration in the Tissue ROI panel for the current *m/z* value can then be calculated from the slope and intercept of the regression line and the tissue ROI response. The GUI will update the tissue concentration as shown in **Figure 128** (red circle).

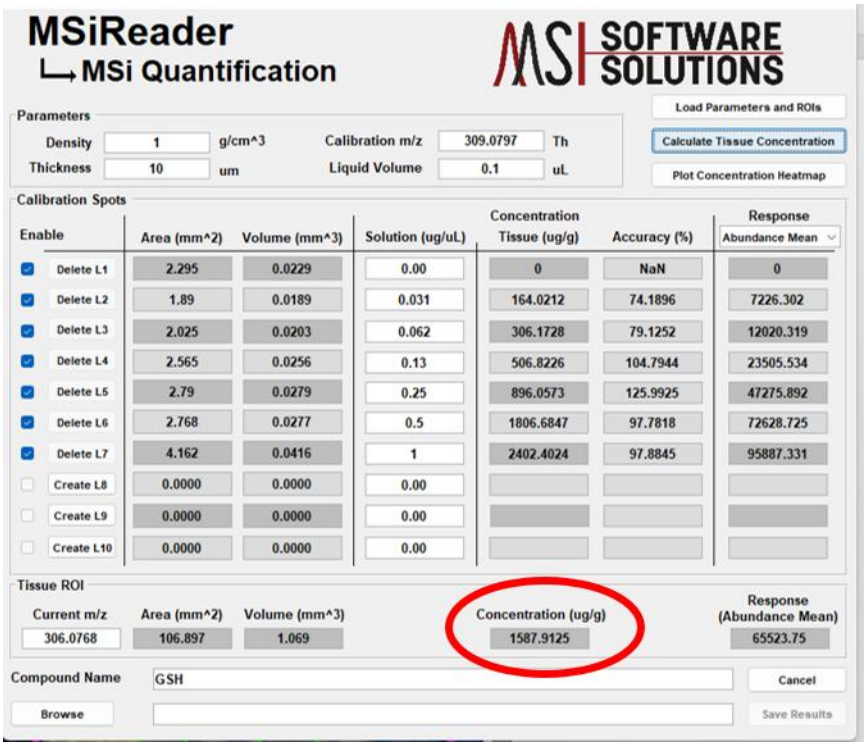


Figure 128: MSiQuantification results.

A plot of the regression results with 95% confidence limits is also displayed (see **Figure 129**) with the calibration spots shown as numbered red dots. The concentration vs. response point for the Tissue ROI is displayed as a magenta square labeled with the letter “T”. The regression equation and R^2 value are displayed on the plot and the title contains the filename and any text that was entered in the *Compound Name* box.

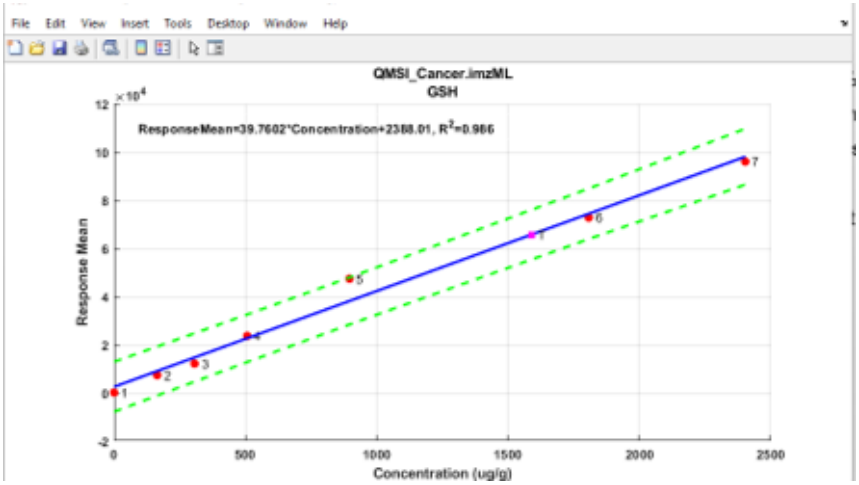


Figure 129: MSiQuantification linear regression result with 95% confidence limits.

In **Figure 130** the data cursor toolbar icon, , can be used to identify the concentration and response corresponding to a point on the heatmap. Any number of data cursor tooltips can be added to the plot. The plot figure can be edited by the user and saved for publication. The default risk factor is 0.05. It can be changed in preferences INI file (**§5)** to any value between 0.68 and 0.001.

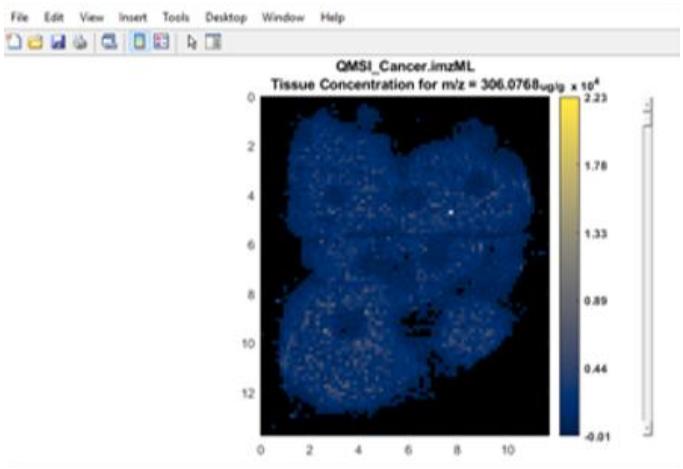
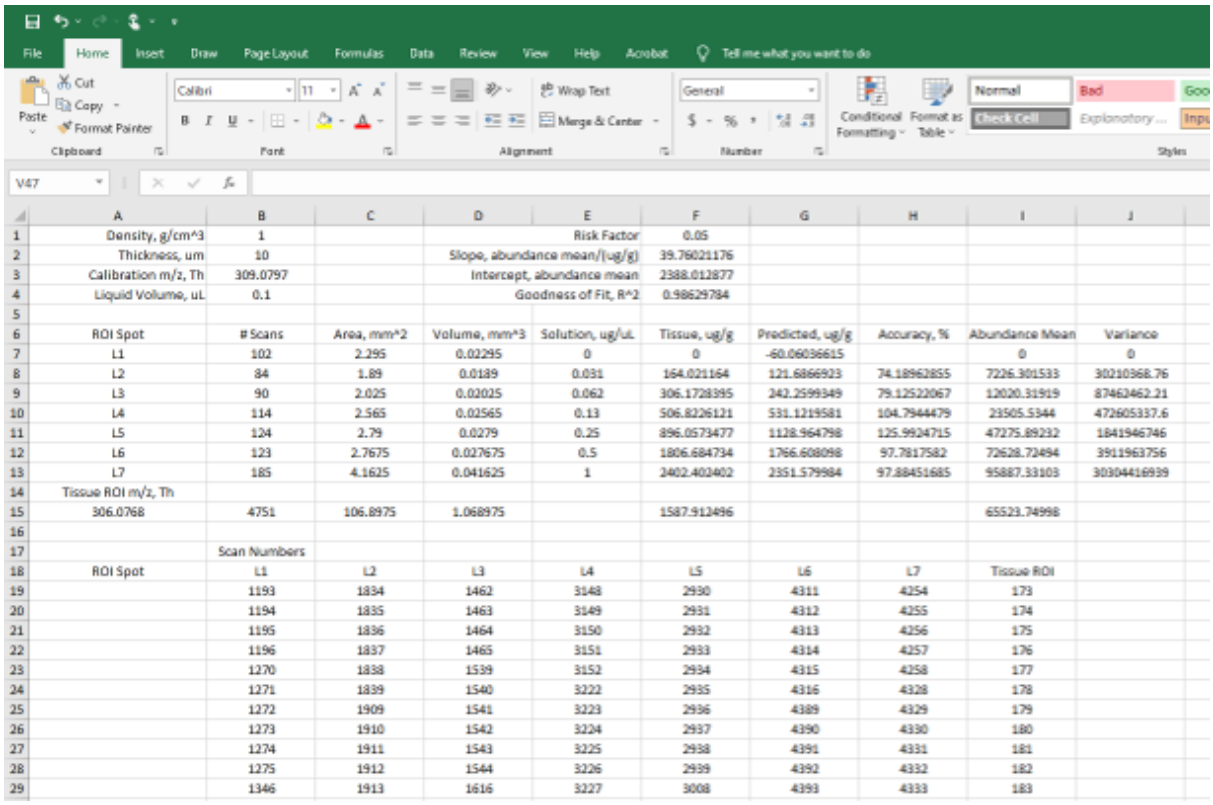


Figure 130: Heatmap plot using tissue concentration as the color scale.

5. The *Plot Concentration Heatmap* button will display the full image heatmap in a new figure using the concentration calculated from the fit results as the color scale as shown in **Figure 130**.

The user can use the slider bar to the right of the heatmap to adjust the scale (see **Figure 130**). The concentrations are listed in μg/g. Click the *Browse* button to select an output file. Then the *Save Results* button will export the parameters and results of the quantification to an Excel workbook (or to text files if Excel is not installed) and a Matlab .mat file will also be saved (see Step 7 below). The Excel workbook will contain two worksheets: one with information about the data set (Error! Reference source not found.) and another with the regression parameters and results as well as the scan numbers for each calibration spot ROI and the Tissue ROI (**Figure 131**).



	A	B	C	D	E	F	G	H	I	J
1	Density, g/cm^3	1			Risk Factor	0.05				
2	Thickness, um	10		Slope, abundance mean/[ug/g]		39.76021176				
3	Calibration m/z, Th	309.0797		Intercept, abundance mean		2388.012877				
4	Liquid Volume, uL	0.1		Goodness of Fit, R^2		0.98629784				
5										
6	ROI Spot	# Scans	Area, mm^2	Volume, mm^3	Solution, ug/uL	Tissue, ug/g	Predicted, ug/g	Accuracy, %	Abundance Mean	Variance
7	L1	102	2.295	0.02295	0	0	-60.06036615		0	0
8	L2	84	1.89	0.0189	0.031	164.021164	121.6866923	74.18962855	7226.301533	30210368.76
9	L3	90	2.025	0.02025	0.062	306.1728395	242.2599349	79.12522067	12020.31919	87462462.21
10	L4	114	2.565	0.02565	0.13	506.8226121	531.1219581	104.7944479	23505.5344	472605337.6
11	L5	124	2.79	0.0279	0.25	896.0573477	1128.964798	125.9924715	47275.89232	1841946746
12	L6	123	2.7675	0.027675	0.5	1806.684734	1766.608098	97.7817582	72628.72494	3911963756
13	L7	185	4.1625	0.041625	1	2402.402402	2351.579984	97.88451685	95887.33103	303044169 39
14	Tissue ROI m/z, Th									
15	306.0768	4751	106.8975	1.068975		1587.913496			65523.74998	
16										
17		Scan Numbers								
18	ROI Spot	L1	L2	L3	L4	L5	L6	L7	Tissue ROI	
19		1193	1834	1462	3148	2930	4311	4254	173	
20		1194	1835	1463	3149	2931	4312	4255	174	
21		1195	1836	1464	3150	2932	4313	4256	175	
22		1196	1837	1465	3151	2933	4314	4257	176	
23		1270	1838	1539	3152	2934	4315	4258	177	
24		1271	1839	1540	3222	2935	4316	4328	178	
25		1272	1909	1541	3223	2936	4389	4329	179	
26		1273	1910	1542	3224	2937	4390	4330	180	
27		1274	1911	1543	3225	2938	4391	4331	181	
28		1275	1912	1544	3226	2939	4392	4332	182	
29		1346	1913	1616	3227	3008	4393	4333	183	

Figure 131: MSi Quantification results exported to Excel, regression worksheet.

6. The saved results MAT file can be used to reload the MSi Quantification parameters, spot and tissue ROIs, the tissue *m/z*, and normalization parameters in the same or a new MSiReader session. This allows the user to replicate an analysis on the same data set and get exactly the same result. To do this, click the button “Load Parameters

and ROI's" on the MSi Quantification GUI on the top right. (**Figure 128**). The user will be prompted if the data set name, its dimensions or other parameters do not match and given the opportunity to continue or cancel (**Figure 132**).

Figure 132: MSi Quantification load parameters dialog if loading incorrect .mat file.

If the data sets do not match or the image dimensions are different continuing is likely to cause errors or produce meaningless results; however, it is permitted and could be a handy shortcut for analyzing multiple tissue samples with similar parameters. The ROIs can easily be "tweaked" and moved after they are automatically drawn. The user is also prompted to restore the Tissue ROI or keep the current ROI (**Figure 133**). Note that an ROI must be drawn somewhere on the tissue as described in Step 2 to enable the button that launches MSi Quantification.

Figure 133: MSiQuantification tissue ROI dialog.

7.7.3 Absolute – QMSI Voxel-by-Voxel Quantification

This approach to absolute quantification is where a tissue is mounted on a slide and then an internal standard of known concentration is sprayed on top of the slide / tissue at a known flow rate and total area covered. While this is a single point calibration curve, this approach significantly decreases variability from pixel-to-pixel (or voxel-to-voxel) due to tissue heterogeneity. An important point about this approach is some trial and error to determine the concentration required to have the ion abundance of the internal standard be as similar as possible to the ion abundance from the endogenous compound.

In this example, the spot spacing and line spacing were both 150 μm and the goal was to quantify dopamine (*m/z* = 154.0863) in the brain using stable isotope labeled dopamine *m/z* = 160.1064) as the internal standard. First, the user must use the polygon ROI tool and draw around the area that they wish to quantify the molecule of interest – in this case, dopamine. The GUI for the V×V quantification tool is shown in **Figure 134**. The values

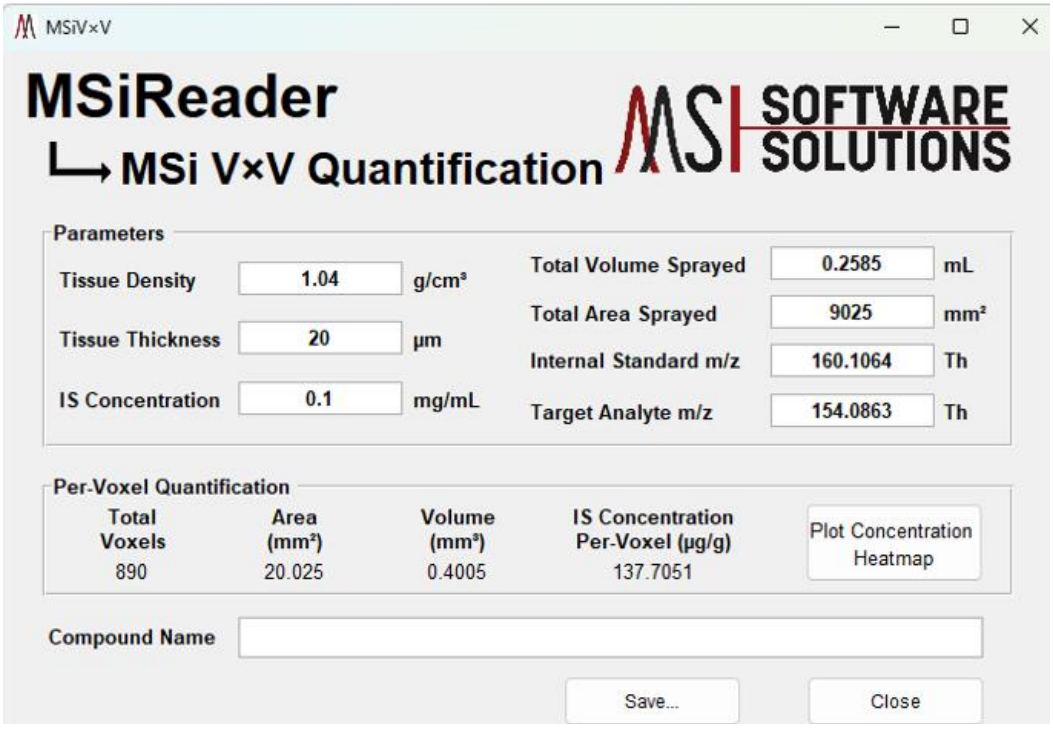


Figure 134: The V×V GUI for the absolute quantification of a targeted analyte.

in the parameter box are all entered by the user based on the experimental conditions they used. Using these data and that from the ROI drawn by the user, the Per-Voxel Quantification information is calculated. The user can input the compound name (optional) if they want it displayed on top of the output images.

Once the experimental data is entered in, the user can click on Plot Concentration Heatmap. The GUI shown in **Figure 135** is now displayed showing only the pixels (voxels) that were selected by the ROI tool. The user can click on any location and the coordinates, scan# and the absolute concentration will be displayed. The user can output the image from this GUI as well.

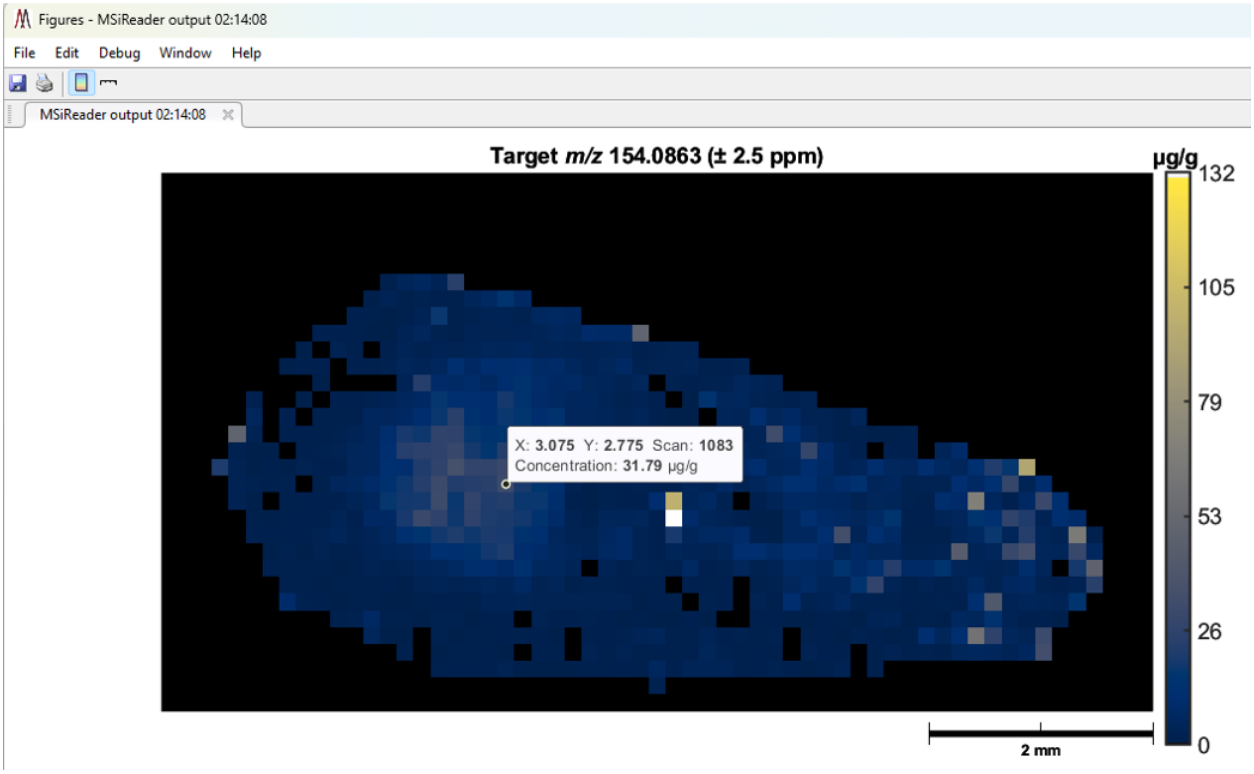


Figure 135: The interactive GUI which displays the concentration heatmap.

If the user would like to work up the data more, click on the **Save** button shown in the bottom of **Figure 134**. This will save in one worksheet the parameters of the experiment and in a second worksheet, more parameters and the concentration of the analyte by X,Y coordinates as well as scan number. This is a new tool in MSiReader and we are working

on a statistical method to determine the error associated with each concentration and this will be added and explained in the near future.

7.7.4 Functional Mass Spectrometry Imaging PIE tool

In this section, the manual will provide a step-by-step example of how to use the Percent Isotope Enrichment (PIE) tool in MSiReader.

1. After launching MSiReader, load an appropriate fMSI dataset as shown in **Figure 136**.

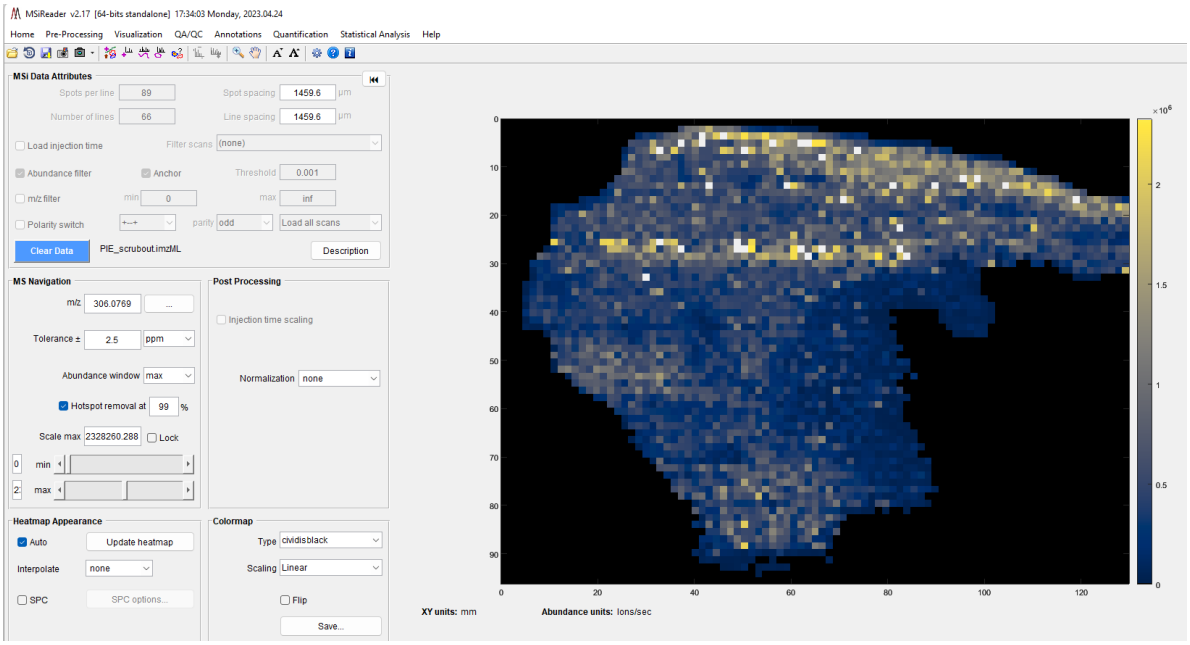


Figure 136: MSiReader after a fMSI dataset has been loaded into the main GUI.

Note: If you don't see this image using the test dataset from the www.msireader.com website, change the *m/z* value in the *MS Navigation* panel to 306.0765.

2. Normalize to the reference peak 320.0922. This is done in the main MSiReader GUI in the *Post-Processing* pane with the result shown in **Figure 137**. This is optional and depends on the experimental design.

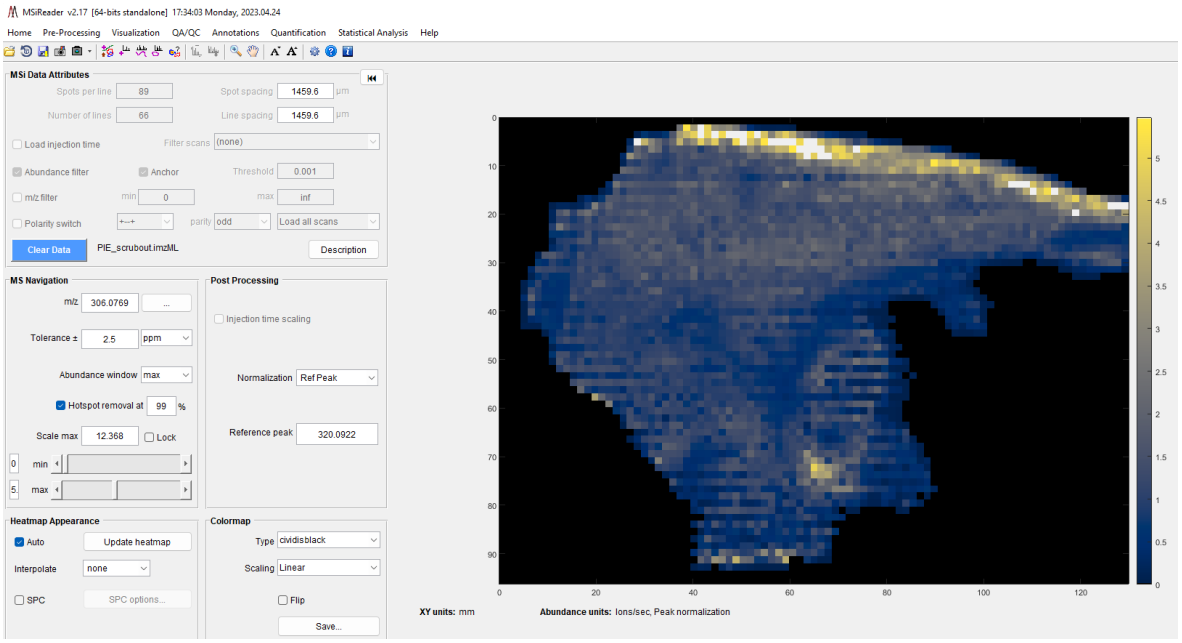


Figure 137: Heatmap after normalization to a compound that was homogenously coated across the entire sample.

3. Select the toolbar icon to enable the polygon tool and draw around the tissue. Save the ROI by right clicking on the heatmap and selecting *Export ROI location info*. In this way, you can reload the ROI for a future session and get the exact same ROI / result. Next, you can *create a binary mask* by right clicking on the magenta ROI line. Then choose scan scrubber under the pre-processing tools and remove pixels outside the ROI and then save the new imzML file (extension will add _scrubout) to the original filename. The area outside of the ROI will become black in the heatmap image.

4. From the main menu, go to Quantification and then select fMSI PIE tool. Select ROI Scans in the dialog as shown in **Figure 138**.

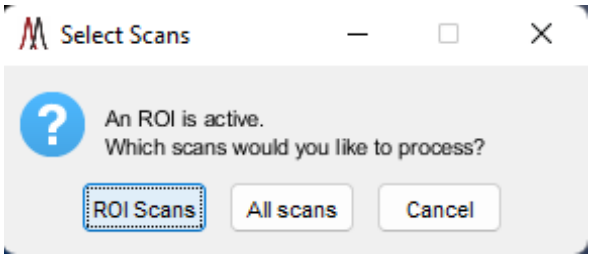


Figure 138: ROI selection dialog box when you launch the fMSI PIE Tool.

5. The next step is to load an example matrix file. In this example, a file is provided with the filename Glutathione Probability Matrix (can be found in the test data sets at www.msireader.com). Then, select the worksheet from the dialog box that pops up.
6. Once you chose the probability matrix, another dialog box will appear as shown in **Figure 139**. Leave the default options as they are and click the OK button.

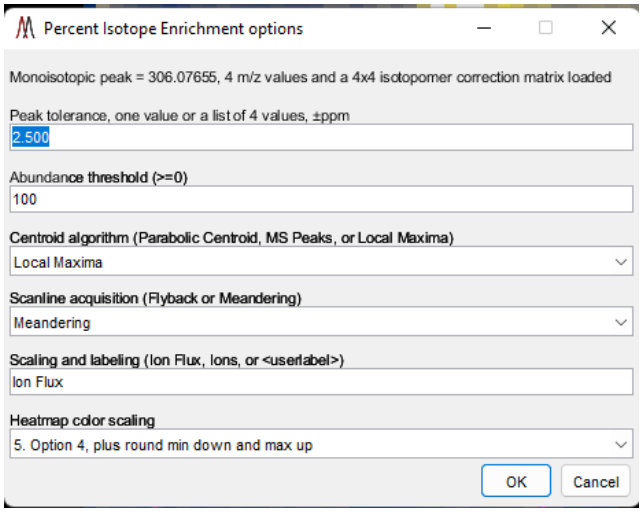


Figure 139: Options dialog box for the PIE tool. Leave the default settings as they are for this example. Heatmap color scaling options are shown in the GUI under “scaling and labeling”.

Eight plots will automatically be produced. There are two plots for each *m/z* value in the isotopologues matrix: percent enrichment and normalized corrected abundance. The corrected abundance heatmap is normalized only if reference peak normalization is selected in MSiReader when the PIE tool is launched. Other normalization strategies are not supported. You can save each plot (heatmap) by using the menu bar in each plot.

A workbook will be also created and the user will automatically be prompted to choose a location and filename. It will contain five worksheets: *Info*, *PIE Summary*, *PIE Abundance, PIE Abundance Normalized*, and *PIE Results*. If normalization to a reference peak is not selected in MSiReader when the PIE tool is launched the *PIE Abundance Normalized* worksheet is not included in the results file.

7.8.1 Preparing MSI Data for Statistical Analysis

A **video tutorial** on how to prepare MSI and HTS/phenotypic screening data for downstream statistical analysis can be found **HERE**.

An interactive multivariate analysis tool is provided in MSiReader. MSiPCA is based on singular value decomposition and loads an abundance matrix saved with the MSiExport sub-GUI, calculates PCA loadings and scores and plots the results for user selected components. The plots are 1) a heatmap showing the PCA score distribution for any number of user selected components (optional), 2) a biplot showing loading and scores for two or three components (and the end-user can select which PC's to plot); and 3) an interactive PCA loadings plot as a function of *m/z*. The rows of the input matrix are observations (*i.e.*, mass spectrometry scans, and the columns are variables, *i.e.*, *m/z* values. The PCA results: component loadings, scores, latent variances, T-squared values, and explained variance percentages can be saved into a single Excel file, each in its own worksheet. Moreover, a scree plot is also automatically generated. Examples and more details will be shown later in this section.

Important Note: PCA algorithms are limited by the number of degrees of freedom (DOF) in the data you are working with; thus, if the # of scans (mass spectra) exceed the number of *m/z* values you can use all the *m/z* bins in the analysis. However, this is not typical with HRAM data and thus, the DOF will be limited to the number of scans in your data. In the PCA algorithm, it will use all the *m/z* binned data and then limit the output to those PC's which describe most of the variance in the data and put in null values for *m/z* bins that exceed the number of DOF. Alternatively, prior to using the MSiPCA tool, the end-user can pre-process the data to bring the number of *m/z* values to be less than the number of scans. There are several ways to meet this criterion and more than one, or all of them, can be used. It depends on the types of data that the end-user is trying to analyze. The ways that the user can ensure that the # scans > # *m/z* values are listed here.

- Select larger ROI's for MSI data as this inherently increases the number of scans but, to a first approximation, the number of *m/z* values does not change.
- Centroiding data. Since a profile peak contains many *m/z* values to describe the entire peak, by centroiding, one can reduce of *m/z* values in the dataset.
- Abundance thresholding data. Given that each instrument is different in terms of the values it reports for ion abundance, the end-user should threshold the data so noise and low abundant peaks (high variance) are filtered out. This will substantially reduce the number of *m/z* values. It is recommended to set this value at an abundance that is 10× the limit-of-detection.
- Filter out narrower *m/z* ranges prior to data export. Not recommended as a first approach because the goal of PCA and other multivariate methods is to use all the data that describes the data.

Below, an example is shown using a healthy tissue image and a cancerous tissue image where one ROI is drawn for each tissue and then the data is extracted using MSiExport.

Checking Data Attributes and Quality

It is important to check your data attributes and quality prior to selecting abundance threshold and the tolerance for peak exclusion. To check for the abundance threshold, the user should load a typical file in the study and then using the single pixel ROI, select an on-tissue pixel. Right click and "plot *m/z* spectrum". Expand the spectrum using the mouse wheel and then look for low abundant peaks by moving the cursor on top of the peak. The pop-up window will give you the *m/z* value and the abundance. Look for the low abundance peaks and this will give the user an idea of what the threshold should be. In both the MSI and HCS data sets, the low abundance peaks are about 12,000 in ion abundance; thus, to be conservative the user could input 12,000 or to include additional lower abundance peaks, the user could use 8000.

The next step for QC would be to make a plot of the MMA for some representative data. The user can do this by loading a file and then going to the drop-down QA/QC menu and choosing “mass measurement accuracy”. In the main GUI, enter 10 ppm for starters into the tolerance and then make the plot. If the data is showing that it is all better than 5 ppm, then your tolerance for peak exclusion should be 5 ppm. If the data does not have the expected MMA and you have some known *m/z* values in the different images, the data should be mass corrected **before** doing anything else. This process can be found in **§7.3.1**.

Preparing data prior to using MSiPCA

1. Unload your data files from MSiReader.
2. Launch the centroid data function under preprocessing drop-down menu. It will automatically be in batch mode since no data files are loaded. The dialogue box as shown in **Figure 140** will appear. Once you set your parameters, click OK and it will open a file explorer to load the imzML files that you will use in the PCA.

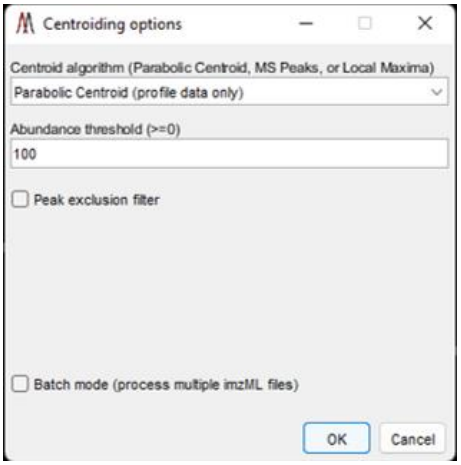


Figure 140: Centroid Data Options Panel in MSiReader

If your dataset is profile data, it is recommended to centroid your data using the parabolic centroid. If the data is already centroided, please select local maxima in the drop-down

menu. Next, select an abundance threshold that is specific to the instrument; 100 is the default but in most cases, this will be much higher.

Inevitability, mass spectrometry data has peaks in the mass spectra that are not sample related. For example, in MALDI, the ions from the organic matrix are present all over the tissue but are not tissue-specific. In ESI-post ionization methods and DESI, ambient molecules interact with the charged droplets and produce ambient background ions. The file PCAPeakstoRemove.txt has a single *m/z* that is not tissue specific; this file can be used in this example. It is critical that as many non-tissue specific ions for MSI and HCS data are removed from the data prior to PCA analysis because these could drive the PCA to an incorrect conclusion. See **§7.3.4** for how to classify ions as being background or sample-specific. When peak exclusion filter is checked, the user can browse for .txt file containing the theoretical *m/z* values of the ions that are not sample related (make sure to include *m/z* values for ions with abundant A+1 and A+2 peaks). These will be excluded if the *m/z* is detected in the mass spectrum with a tolerance set by the end user. It is important to know the MMA of your dataset prior to setting this tolerance; otherwise, it will remove non-tissue peaks in some spectra that have high MMA but not in those that fall outside this tolerance. Click OK and new imzML files will be generated with "_centroided" automatically added to those files. If you loaded a *.raw file, the user can save the processed data as a *.mim file.

It is important to note that a user can also carry out the above steps for an entire folder of imzML files without ever having to load the files into MSiReader. To do this, clear the data and then choose Centroid Data under the Pre-Processing menu and enter in the dialogue box what function(s) needs to be carried out and click OK. It will then open a file explorer to choose the imzML files that need to be processed. Batch processing of *.raw files is not available but will be in a future release.

Open the processed data files that will be used to carry out PCA. Launch the MSiExport tool under the *Annotations* drop-down menu, then *Data Export* then *Export Abundance Data*. Draw a single ROI around all the data OR draw an interrogated and reference ROI. The user will be asked to select which scans to consider – select scans for the ROI(s). Next, choose *Build and m/z list and bin the pixels* in the ROI. Next, enter in values that correspond to your data; for high resolution accurate mass (HRAM) data, 5 ppm is recommended. Click the Browse button and enter a filename and folder for the .txt file to be prepared. Click OK. The .txt file will be the data that is entered into the PCA function.

For export to text format two files named <*name*>*_info.txt* and <*name*>*_peaks.txt* were saved. If the data was from a tiled image mosaic an additional file, <*name*>*_files.txt* was also saved. All of these are required to enable the full capabilities of MSiPCA and must be in the same folder. The <*name*>*_peaks.txt* file should be selected from the input dialog. While the user can use the *.csv file, we don't recommend that as input to the multivariate tools in MSiReader but it is provided so users can use other software programs of their choice.

7.8.2 Principle Component Analysis (PCA)

A **video tutorial** on how to use the PCA tool can be found **HERE**.

7.8.2.1 Description of PCA

Principle Component Analysis (PCA) is a statistical technique used to simplify complex datasets by reducing the number of variables and retaining the most important information in a set of new, uncorrelated latent variables. PCA identifies the most important of these new variables or "principle components (PCs)" based on the amount variation they account for in a dataset. While there are as many PCs as there are original variables, the

first PC accounts for the most variation in the data, the second accounts for the second most, etc., such that a small number of PCs may explain a large percentage of the overall variation found in the data. While the original variables may be heavily correlated, the resultant PCs are uncorrelated with one another.

PCA works by transforming the original variables into a new set of variables that are linear combinations of the originals. If variables are scaled (such that the variation of the variables is approximately equal), the value of a coefficient explains the impact that the original variable has on a PC. If a coefficient is extreme, the original variable is more highly correlated with that specific PC. This helps users identify which original variables are most important in explaining variation and potentially most important at differentiating between samples or groups.

PCA can be useful for analyzing mass spectrometry imaging data because it can help to identify the most important mass-to-charge ratios (*m/z*) in large datasets of spectral features. It can be useful for HCS data because in these cellular assays, the goal is to determine mode of action of different drugs. Both of these data types can be complex and high-dimensional with many variables representing different chemical components and their abundances. The number of *m/z* values, as well as correlation between scans and the identified peaks can make spectral datasets difficult to analyze directly. PCA can reduce the dimensionality of the datasets by converting the *m/z* values into smaller numbers of principle components that can explain large portions of the variation between scans. This is especially useful when comparing spectra from different samples and identifying similarities and differences between them. If PCs can be identified that separate samples or groups of samples, the chemical components that are responsible for the differences in PCs can be investigated.

7.8.2.2 PCA Details

Once samples are imported and ROI specified, PCA ready datasets are created. Each scan creates an observation with identified *m/z* values representing original variables and the abundances at each *m/z* value being the signal recorded for that variable. The *m/z*

values are binned using a user defined range such that similar values across scans are labeled as the same *m/z* value (*e.g.*, 212.0601 detected in one scan and 212.0603 detected in another may be binned together as the same *m/z*). The result is a set of scans with values recorded at each *m/z* found across all scans. If specific peaks were not identified in all scans, the observations missing these peaks are filled with 0's for the missing values. Information about the sample, spatial location or well plate number of the scan, and possible group (*e.g.*, treated vs. untreated, drug 1 vs. drug 2) are also retained.

Because variance in chemical abundances can be far greater at some *m/z* values than others, each *m/z* value is auto-scaled (mean centered and scaled by the standard deviation) prior to PCA, such that the mean and variance for each is approximately equal. Scaling *m/z* values is performed independently for each *m/z* and uses values from all scans and all samples.

Load the .txt file that was created in **§7.8.1**. A progress bar showing the loading of scans along with the total number of scans is shown. Principal component analysis is immediately performed when the OK button is clicked. The method used is singular value decomposition. As mentioned, prior to doing PCA, all the *m/z* measurements need to be on a similar abundance scale so that we can compare the variation attributed to each variable. Each *m/z* should be mean centered and scaled by the standard deviation. To accomplish this, the ion abundance is first converted to a scaled abundance (*a**) for measurement *i* at a specific *m/z*:

$$a^{*}_{mz,i} = \frac{a_{mz,i} - mean(a_{mz})}{sd(a_{mz})} \text{ (Auto-scaled)}$$

where a_{mz} represents the vector of all abundances for that *m/z*.

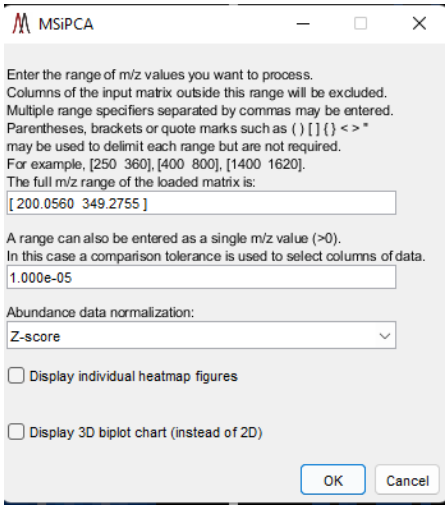


Figure 141: Selecting input columns by *m/z* range and abundance normalization (Z-score or Pareto). Checkboxes allow for individual heatmaps to be generated for each PC and an option to make a 3D plot.

Once all the scans are loaded into MSiReader, the dialogue box will appear as shown in **Figure 141**. This allows the user to select the full *m/z* range of the data or a narrower range(s). The user must then select abundance data normalization (Z-score or Pareto). The first checkbox is whether or not the end user wants to display individual heatmaps for each PC. We don't recommend doing this as a default since datasets are quite large in the field. The second checkbox allow the user to change to a 3D biplot after the computations are done (PC1, PC2 and PC3 will be plotted).

Note: for two or more images, empty scans are inserted between the rows and columns to delineate the file boundaries and make the tilling pattern plaid. These scans are not removed before running the SVD so that the scan ordering and numbering of the data set as seen in the MSiReader heatmap plot is preserved. Instead, their abundance is set to NaN (not-a-number) and thus they are treated as missing data by the SVD algorithm and have no impact on the results.

As stated above, the PCA algorithm that is being used is MSiReader is SVD. It is important to note that when you have a n by p matrix of data, n being the rows (or scans) and p being the *m/z* values, the degrees of freedom are limited to n-1. Thus, if you load a data set with 500 mass spectra with each mass spectrum containing 100,000 *m/z* values, the maximum number of factors that can be returned is 499. When you have a wide (more columns than rows) matrix, PCA will calculate the covariance matrix of your input matrix first, resulting in a p by p covariance matrix. Next, the eigenvalue of the covariance matrix will be calculated. Since the input matrix is wide, the number of nonzero eigenvalues will not be larger than the degrees of freedom, which is n-1 in this case. Since the rest of the eigenvalues are zero, only the first n eigenvectors take part in the singular value decomposition, and the remainder being multiplied by zero. Only those eigenvectors not multiplied by zero are computed for decomposition and returned.

7.8.2.3 PCA Output Plots and Data

Loading the data and then running the MSiPCA tool will first prompt the user with how they would like the save the PCA results – this can be *.csv, *.txt or *.xlsx format. This saves the PCA scores data in an easy-to-use format for further analysis. Only the top 20 PC's are exported.

Once the PCA data is saved, the default is to generate a 2D biplot; it is good to start with just the 2D biplot until you are more familiar with these. A 3D biplot can be selected if the checkbox in **Figure 141** was selected.

The score plot is a representation of the dataset in the reduced dimensional space defined by one or more PCs. Scores from PCA represent the value of each PC for each observation. The score of a specific PC from a single scan is the linear combination of all *m/z* values recorded for that scan. A score plot displays these scores in 2 or 3-dimensional space with each point representing a single scan.

Each axis of the scores plot is a user chosen PC, such that two or three principle components can be viewed as a scatter plot. By default, the first principle component is displayed on x-axis and the second on the y-axis, though other PCs can be substituted

by the user. Three-dimensional score plots plot the third PC on the z-axis. Scans with similar values for the displayed PCs are plotted near each other, whereas scans with largely different values for a PC are separated.

The score plot can be useful to identify patterns and relationships in the dataset. Clear separation between scans or samples can suggest that these principle components are capturing important differences between groups. If specific PCs are identified that differentiate between samples or scans, a loading plot can be used to detect the impact that each *m/z* value has on the PC.

An example of a two-dimensional biplot is shown in **Figure 142** and will appear automatically. Note that the observations from each file in the data set are assigned the same color (in this case, red is ovarian cancer tissue and green is healthy ovary tissue). The colors are automatically selected by MSiPCA from the default plot color order if there are fewer than eight files or from a random permutation of the current colormap if there

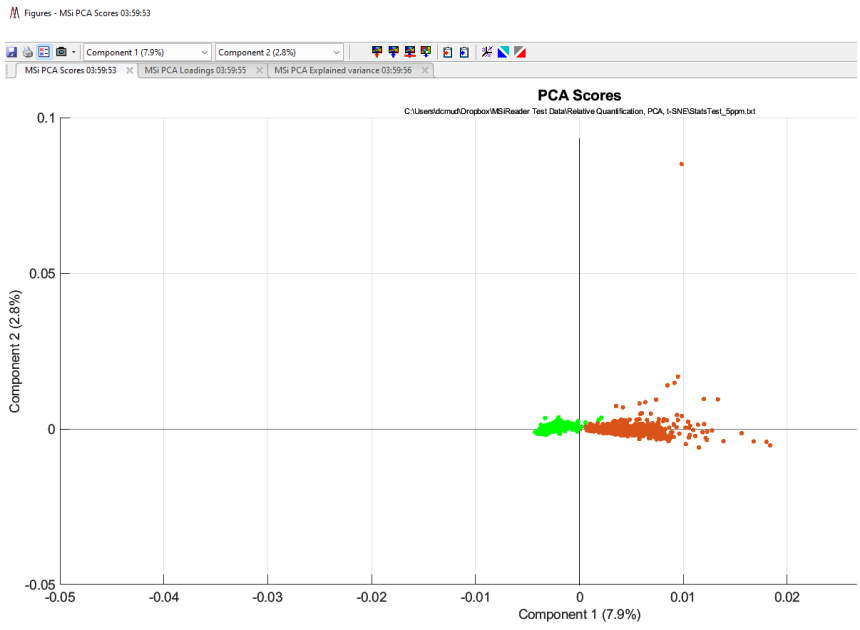


Figure 142: PCA of two small ROI's derived from MSiExport of a healthy and a cancer tissue.

are more than seven files. The data cursor, rotate, pan and zoom tools are to the upper right of the plot and will appear after the mouse is hovered over the plot for a few seconds.

The observation data cursor display text has been customized to include the scan location for an observation in the image mosaic and in the original file. The variable data cursor display text shows the *m/z* value. The component values are also shown in both types of data cursor text.

Note that at the top of **Figure 142**, Component 1 and Component 2 are listed with their amount of explained variance. These are drop down windows and the end-user can then make plots of PC1 versus PC3 for example, and so on. Also note there are two other tabs, one is PCA loadings plot (**Figure 143**) and the other is a scree plot ().

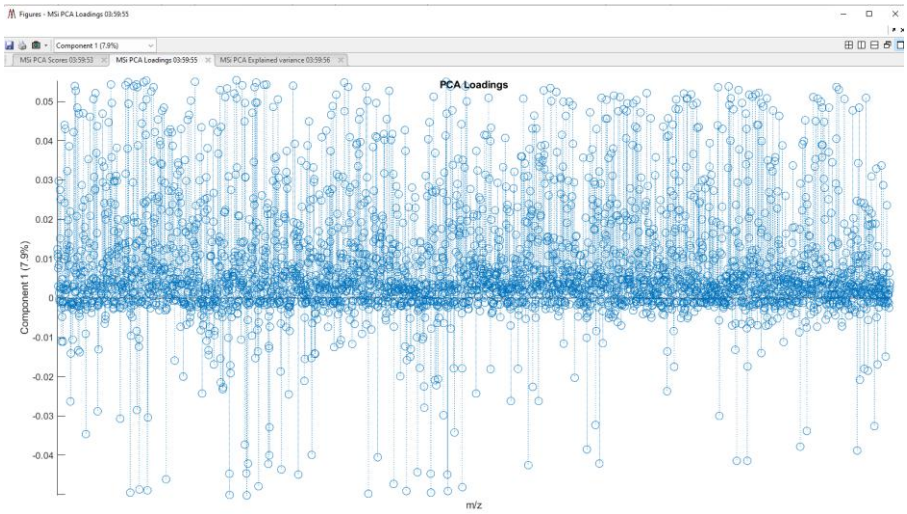


Figure 143: PCA Loadings Plot for the data shown in **Figure 142**.

The loadings plot is a graphical representation of the relationship between the mass-to-charge ratios and the resultant principle components. Each of the original *m/z* values has a loading for each PC, representing that *m/z*'s coefficient from the linear combination of all *m/z*. A loading that is high relative to others indicate that a strong positive correlation between abundances at that *m/z* and the PC, whereas loadings that are low compared to others represent strong negative correlation.

For loadings plot, a PC is chosen by the user. The x-axis contains each *m/z* (or possibly a subset of *m/z*'s if filtered) detected in the dataset. The y-axis contains the coefficient associated with that *m/z* value for the PC.

If one or more PCs are identified that separate or describe the data well, the loadings plot can help identify the mass-to-charge ratios that have the highest impact on these PCs.

The PCA loadings plot (**Figure 143**), one can use the mouse wheel to expand and contract the x-axis and the cursor will indicate which *m/z* value describes that particular loading for PC1. Note that the upper left-hand corner, the user can select which PCA loadings plot they want. It is a pull-down menu and also indicates the percentage of explained variance.

The scree plot for the data is shown in **Figure 144**. Note that the number of principal components is limited to 2077 since only 2078 scans were used to generate this test dataset.

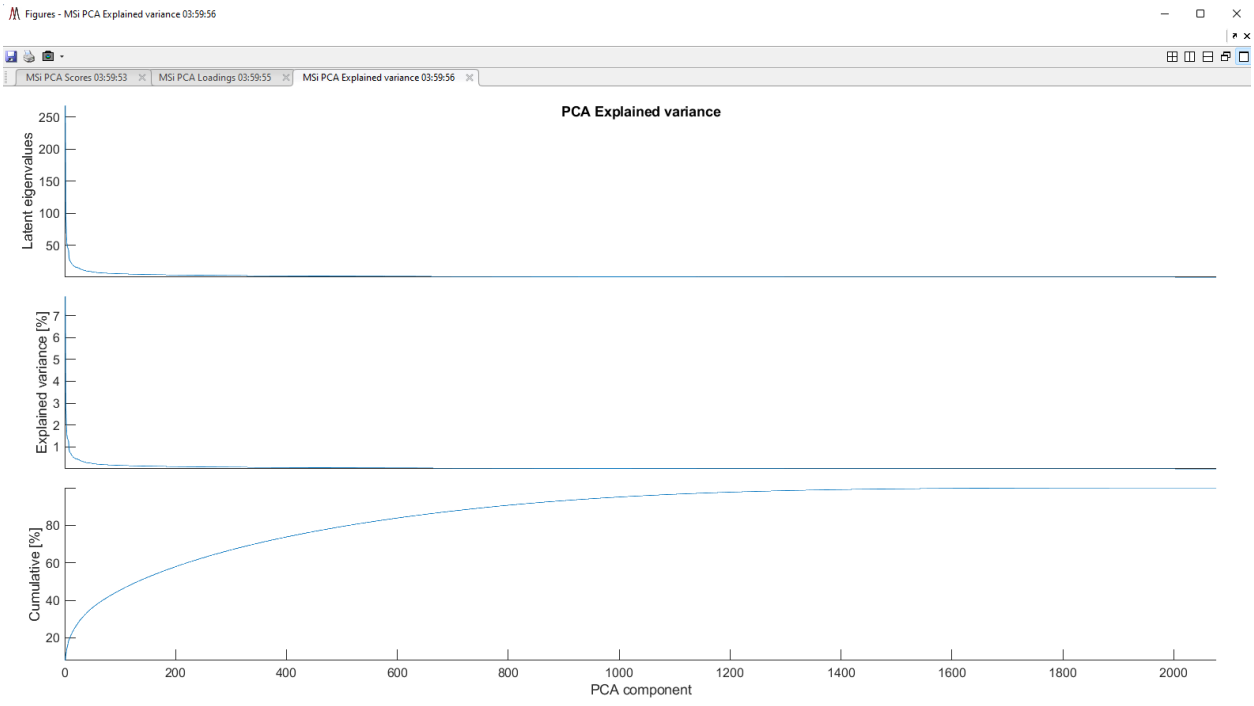


Figure 144: Scree plot for the data shown in **Figure 142**

Four icons are on the biplot toolbar; the function of each icon is described below.

 The *m/z* values for all current variable data cursors are appended to the clipboard. The icon is always enabled on the biplot toolbar.

 The scan numbers for all data cursors attached to an observation are appended to the clipboard. The icon is enabled when the observation scores are plotted.

 Visibility of observation scan markers on the biplot is toggled. The icon is enabled when the observation scores are plotted. If the data set is for a folder of files, a selection dialog is launched with an entry for each file. Markers for the de-selected files are hidden on the biplot. Any data cursor text boxes attached to these scans are also hidden but not deleted. Click the icon again to restore the hidden markers and data cursors.

 Takes the average of all scores for each sample type. In this case there are only two samples and thus, plotting all of the data (each voxel is a point for each sample) made more visual sense.

The text files or the single Excel file with multiple worksheets also contain important information about the PCA analysis – please export these values using the *.txt or *.xlsx format while the *.csv format only includes the scores for further downstream analysis using other programs.

pca_mz:

A list of the *m/z* values identified across all scans and used for PCA.

pca_scores:

The scores for all observations and all PCs. Each row represents a scan and columns represent the score for each PC.

pca_coeff:

The loadings (coefficients) for each *m/z* value on the principle components. Each row represents an original variable (*m/z*) and each column represents a PC. The score of a specific PC is determined by multiplying each coefficient by the abundance for the associated *m/z* and summing these across all *m/z*'s.

pca_latent:

The eigenvalue associated with each PC. This represents the total amount of variation in that dataset that is explained by each PC on its original scale.

pca_varpct:

The percentage of variation of the data described by each PC. This is determined by dividing each of the values in pca_latent by the sum of all values.

pca_tsquare:

A list of Hotelling t-squared values from each observation. These measure the distance that each observation is from the center of the dataset based on the principle components. High t-squared values can indicate that an observation is far away from the center of the dataset and may be an outlier.

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

A **video tutorial** on how to use the t-SNE tool can be found **HERE**.

7.8.3.1

Similar to PCA, t-SNE is used to understand high-dimensional data and visualize it in lower-dimensional spaces. While PCA is a useful tool to analyze datasets with linear relationships between variables, t-SNE is a dimensionality reduction tool that relaxes these assumptions and can project linearly non-separable data into reduced dimensions.

Please make sure you read through **§7.8.1** to make sure your data is properly prepared prior to using this statistical tool.

t-SNE works by constructing a probability distribution for pairs of observations of high dimensions (many *m/z* values) such that observations that are similar have a relative high probability assigned and observations that are dissimilar have low probabilities assigned. This probability distribution incorporates all *m/z* values found across a dataset. A second probability distribution is created in lower dimensional space (typically 2 or 3 dimensions). The observations are adjusted in the low-dimensional space so that the second probability distribution matches the first probability distribution as closely as possible. This allows visualization of complex, high-dimensional datasets in lower dimensional space while preserving the underlying structure of the dataset (similarities and differences between scans or samples).

7.8.3.3 t-SNE Details

Once samples are imported and ROI specified, t-SNE ready datasets are created. Each scan creates an observation with identified *m/z*'s representing original variables and the abundances at each *m/z* value being the values recorded for that variable. The *m/z* values are binned using a user defined range such that similar values across scans are labeled as the same *m/z* value (*e.g.*, 212.0601 detected in one scan and 212.0603 detected in another may be binned together as the same *m/z*). The result is a set of scans with values recorded at each *m/z* found across all scans. If specific peaks were not identified in all scans, the observations missing these peaks are filled with 0's for the missing *m/z*'s. Information about the sample, spatial location of the scan, and possible group (*e.g.*, treated vs. untreated) are also retained.

Because variance in chemical abundances can be far greater at some *m/z* values than others, each *m/z* value is auto-scaled (mean centered and scaled by the standard deviation) prior to t-SNE, such that the mean and variance for each is approximately

equal. Scaling *m/z* values is performed independently for each *m/z* and uses values from all scans and all samples.

Next the t-SNE algorithm is applied to generate observations in lower dimensionality. Pairwise similarities are determined between each pair of scans based on the distance measure chosen (*e.g.*, Euclidian distance). This similarity is represented as the conditional probability that a pair are neighbors in the high-dimensional space (i.e., high probability indicates scans are similar whereas low probability indicates significant differences between scans). A probability distribution is constructed based on all pairwise similarities in the high-dimensional space, using a Gaussian kernel to define each probability.

Next, a similar probability distribution is constructed in a lower dimensional space and using a Students-t kernel. Observations are initially scattered at random in the lower dimensional space. The lower dimension distribution is then optimized by minimizing the Kullback-Leiber divergence between the higher and lower dimensional distributions. Points are adjusted such that observations that are similar in higher dimensions are also similar in the lower dimension distribution.

7.8.3.3 Input and Output for t-SNE

The input for t-SNE is the same as for MSiPCA. Launch t-SNE by going to the menu Statistical Analysis and then select t-SNE. When the tool is launched, a file explorer will open asking the end-user to select a .txt file to load. In this example, it is the same .txt file that was used for MSiPCA. Load the .txt file – a progress bar will appear showing the user the scan number being loaded (*e.g.*, 1000 out of 5069) and a percentage of the total

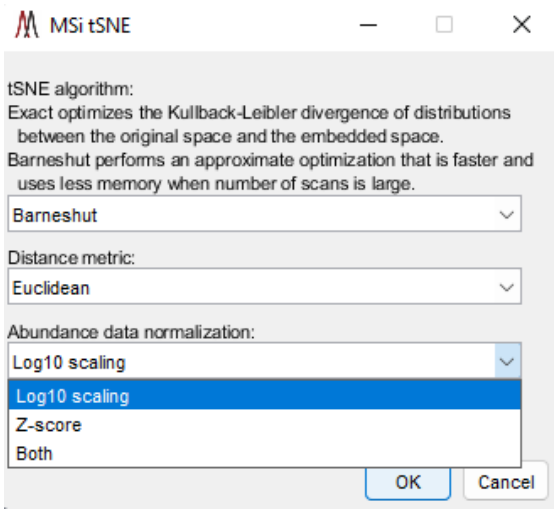


Figure 145: t-SNE dialog box. It is recommended that Barneshut and Euclidean (defaults) are used but other options are available in the pull-down menus. Choose the scaling function (Log10, Z-score or Both which is Log10 followed by Z-score). t-SNE can also make use of a seeding function which will change the plots. To use the seeding function, check the box re-randomize (not visible) so that each time you process the file, a different set of random seed functions are applied.

number of scans. Next, the dialogue box shown in **Figure 145** will be displayed. Please see figure caption for the choices that the user can make. While the computations are being carried out, a window is displayed indicating “Computing tSNE may take some time, please be patient”.

The user will be prompted to save the tSNE output data (scan number, source, DIM1 and DIM2) for the analysis in either *.csv format or *.xlsx format. Choose which option and where the data file should be stored.

MSiReader's t-SNE algorithm generates a 2-dimensional plot of the dataset with a single point for each scan. The units and scale of these dimensions are unimportant as they do not directly correspond the units of the original dataset. Instead, the relationship and relative distances between points can be observed. Points that are close to one another are more similar in high-dimensional space. Clusters of points may identify similar data points in high dimensions, where the more tightly points are clustered, the more similar the abundances detected at each *m/z* are among scans.

Data points can be colored based on sample or group (*e.g.*, Treated vs. Untreated). Clear clustering based on these criteria indicate that the dataset can be separated based on some set of *m/z* values. Plots may also identify other important clusters or relationships such as similar sample types (*e.g.*, type of tissue being scanned) or outliers that can be removed by redefining the ROI.

t-SNE can identify non-linear relationships between scans in high-dimensional space and help determine if data points can be clustered or separated. However, the reduced dimensions do not have a defined relationship with the original variables making interpretation of variables that are important difficult. If relationships or clusters are identified using t-SNE, alternative analytical methods should be used to further investigate variables responsible.

When the t-SNE computations have been completed, a plot of t-SNE Dimension 1 (x-axis) and t-SNE Dimension 2 (y-axis) are displayed as shown in **Figure 146**.

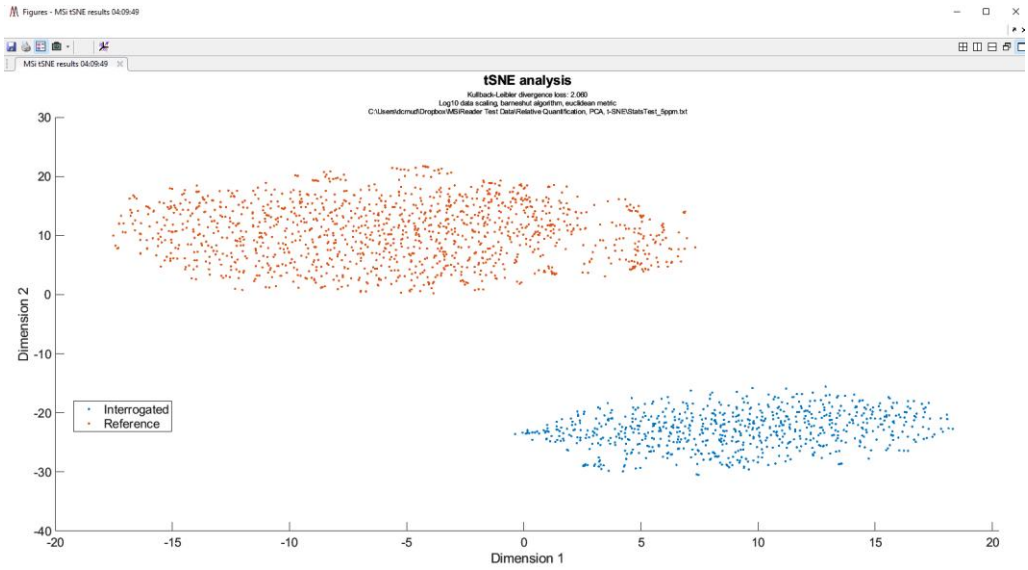


Figure 146: t-SNE 2D output plot. Clearly, the cancer and healthy samples separate well using this multivariate approach. Notice that on the top, the choices the user made are listed under t-SNE analysis.

Part IV: User Guide - BioPharma

8 BioPharma Mode to Support HTS and HCS Well-Plate Analysis

8.1 The MSiReader Main Graphical User Interface (GUI) in BioPharma Mode

A **video tutorial** on navigating the main GUI of MSiReader in BioPharma Mode can be found **HERE**.

The main GUI in MSiReader BioPharma Mode contains 4 panes and includes: **1)** Plate Attributes; **2)** Post-Processing; **3)** MS Navigation; and **4)** Colormap. Collectively, these serve as a simple and effective interface to efficiently begin to look at your HTS/HCS data with a large plate map display on the right. Below, each of these panes will be discussed. Please note that when you load MSiReader for a given session, only the Plate Attributes pane is shown until a data file is loaded. The overarching GUI with menus, sub-menus, and context menus were discussed in §4 and the description and function of the MSiReaderPrefs.INI were presented in §5. In this section, details will be provided to guide you through the process.

Recall that you should adjust your font sizes to match your display resolution as described in §4.2 using the “A” ICONS in the taskbar for an improved user experience. This can also be done by clicking on the “Visualization” tab and going down and clicking on “Increase font size” or “Decrease font size” repeatedly until the GUI is appropriated sized for your display. You can also set this in the preferences .INI file (§5) default value = 9.

8.1.1 Plate Attributes

Figure 147 shows the Plate Attributes Pane; when you first load MSiReader for a session, this is the only pane that is displayed until a file is loaded. However, **prior** to loading a specific data set, you can still make changes to these default checkboxes and values (*e.g.*, abundance filter and the threshold the user wishes to specify). Below the details of each of these attributes are described

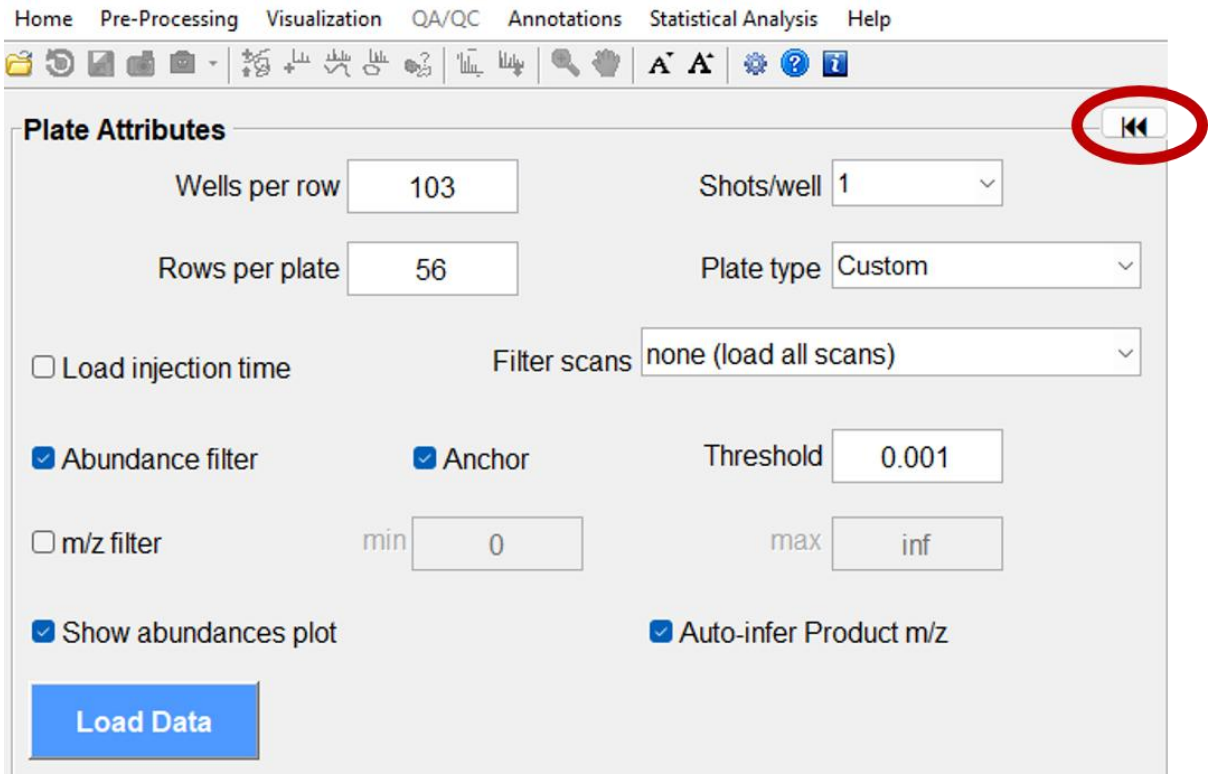


Figure 147: The Plate Attributes pane displays the values that were imported from the imzML file, a *.json file, or a *.raw file. If you are re-loading a previous session using a *.mss or a *.mim file format, you will **not** need to recall the attributes of the data as all of that is saved in these custom MSiReader data formats. Once the user has selected or verified the plate attributes, carried out normalization, etc the user can collapse this pane and the post processing pane by clicking the double arrow in the red circle in **Figure 147**. The pane can be recovered by clicking the Settings button that will appear on the top left of this pane.

8.1.1.1 Characteristics of the HTS or HCS Data

The *wells per row* and *rows per plate* change automatically when a user selects the well-plate format that they used for the experiment in the pull-down menu *plate type* (*e.g.*, a 384 well plate will automatically populate 24 wells per row and 16 rows per plate). The shots per well is either entered in manually (from 1-9) OR this is read directly from a *.json file that must be in the same folder as the data. If a user selects “custom” under *plate type*, the user can enter in the values specific to their plate. When a common plate type is selected (*e.g.*, 384 well-plate), these values cannot be changed.

8.1.1.2 *Loading imzML and Native File Formats in BioPharma Mode*

In BioPharma mode there are different ways one can read in these types of data and then save them as MSiReader custom file formats for improved performance and UX. Please see **§8.2** for the different scenarios for loading (and saving) these data streams.

8.1.1.3 Injection time scaling

Heatmap abundance can be loaded and subsequently scaled by injection time with a checkbox in the Plate Attributes pane “Load injection time”. Injection times will either be read from the data file directly or, if not found in the file, the user will be prompted to enter a value during the load process. When the load injection time box is checked, the injection time is read into (or an injection time is manually entered in the dialog box). All of the scans in a plate do not have to have the same injection time. For example, an imzML file that is a “stitched” composite of multiple data sets or a folder of imzML or mzXML files. How to use the injection time values was discussed in **§7.2.2**.

8.1.1.4 Filtering data

Data sets can be filtered during loading in a number of ways including **1)** using an ROI location file (**§2.4.3**) or a bespoke scan pattern; **2)** abundance threshold (**§2.4.4**); and **3)**

m/z range. These filtering approaches will be described here individually; some can be carried out simultaneously (*e.g.,* abundance threshold and *m/z* range) while others are mutually exclusive (*e.g.,* ROI location and bespoke scan pattern). Using an ROI location file when loading data was described in **§2.4.3** and setting an abundance threshold (including the meaning of the anchor checkbox) was described in **§2.4.4** and thus, these will not be discussed here.

Unwanted scans that follow a regular pattern can be filtered from a data set as it is read with a bespoke scan filter by selecting "using bespoke scan pattern" in the pull-down menu to the right of filter scans. When the load button is clicked the user will be prompted with the dialog shown in **Figure 148** to describe the scan pattern. The pattern specifies the scans to keep from each pattern replication across rows of the image. If the pattern length is not an integer multiple of the number of columns in the image, the last pattern replicate can be trimmed and the pattern will start again on the next row (*Trim*), rows of the image can be padded with empty scans to fulfill the pattern (*Pad*), or the pattern can be wrapped around to the next row (*Wrap*).

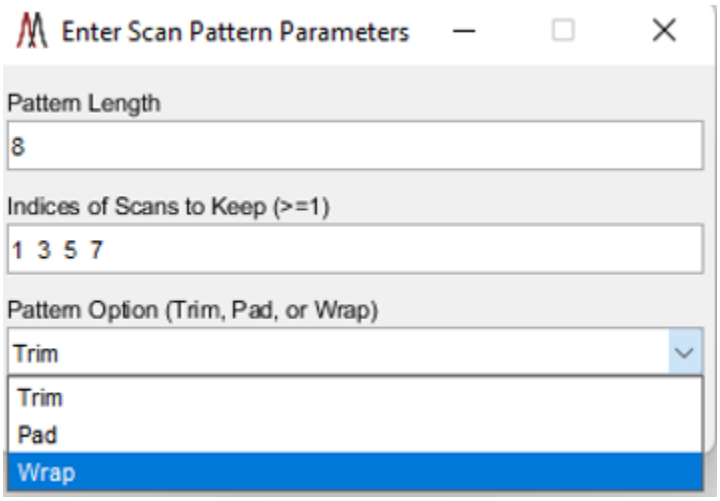


Figure 148: Bespoke scan filter dialog box.

While the file is being loaded scans that are filtered from the image are set to empty. For the parameters shown in **Figure 148** the odd numbered scans in each row would be read

and saved while the even numbered scans would be skipped. After loading is finished, rows and columns that are completely empty will be removed from the image if the preferences INI file variable *SqueezeROIEmptyScans* is *true* (**§5**). Only those rows and columns outside of a bounding box around the non-empty scans are removed if the variable *SqueezeROIBorderScansOnly* is also *true* (**§5**).

m/z range filtering can be carried out for all file formats except *.mss; the scans are filtered by *m/z* value as they are read. In the Plate Attributes pane shown in **Figure 147**, one can check the *m/z* filter which will allow the user to set the minimum and maximum values allowed which are zero and infinity, respectively. Data pairs (*m/z*, abundance) outside of this range will not be saved in the loaded image. The default values for the filter (0 and infinity) can be changed in the INI preferences file (**§5**). This filtering can be done to break a large file into several smaller ones or perhaps, a user collected 100 images from *m/z* = 200 to *m/z* = 2500 and upon inspection of all the data, it is observed that there are no analyte peaks between *m/z* 1000 and 2500. In either scenario, *m/z* range filtering will reduce the demand for physical memory.

The “auto-infer Product *m/z*” checkbox selects the most abundant peak in the mass spectra that were loaded and populates that value in the “Product *m/z*” field in the MS Navigation pane. It does **not** mean that it is the product of the reaction, it just ensures that the plate heatmap plots an *m/z* value that was found to be abundant in the mass spectra.

“Show abundances plot” checkbox will prompt MSiReader to open another window that plots out the abundances of the different *m/z* channels (TIC, Product *m/z*, Material *m/z* – short for starting material, and internal standard). An example of this output is shown in **Figure 149**.

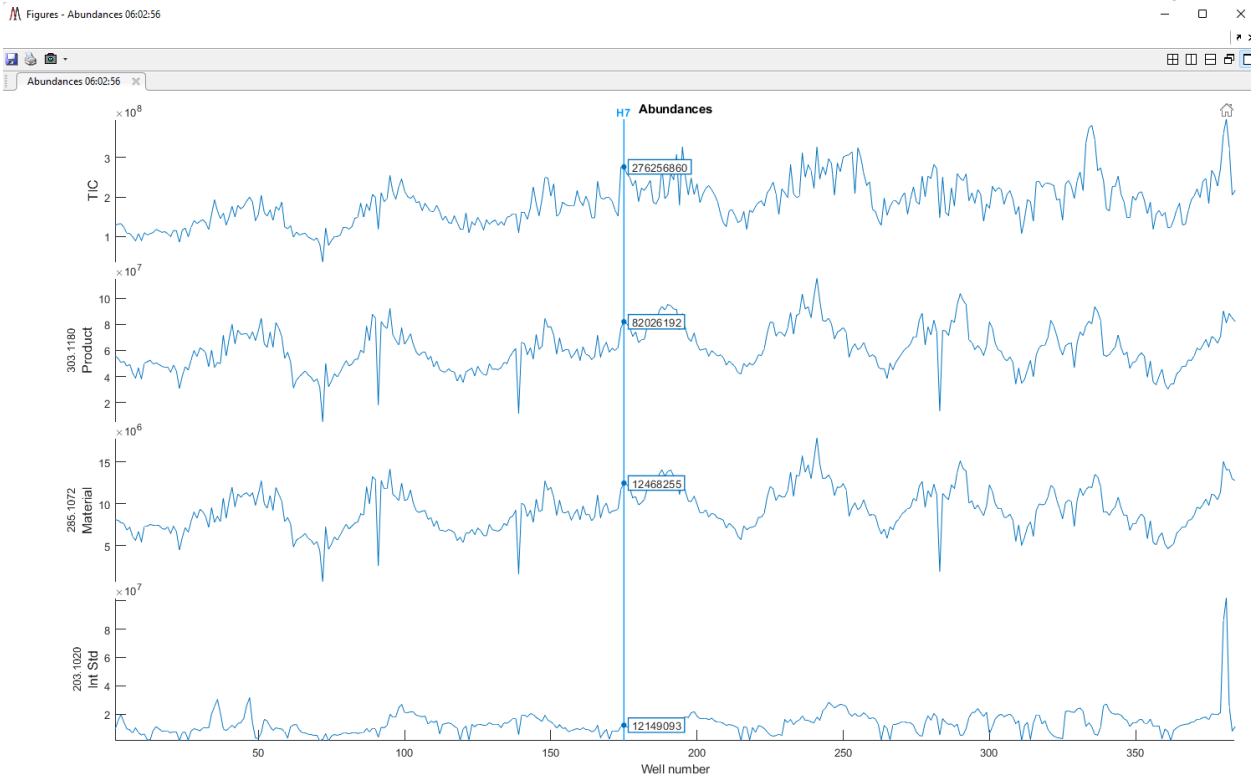


Figure 149: Abundances plot as a function of scan number for 4 different channels including TIC, the product, starting material and internal standard. Note that the x-axis is scan number. If the user puts their cursor on abundances plot, notice that on the top it gives the well plate ID (H7 in this case on a 384 well-plate). If a user puts the cursor on the well-plate of interest, another window will open displaying the mass spectrum from that well. Use the icons in the top right corner of the window to enable displaying of multiple windows in different orientations.

When a user updates the product, internal standard and/or starting material *m/z* values in the main GUI, the abundances chart automatically updates. If the user clicks on the abundances plot for a specific well, a mass spectrum will appear; an example is shown in **Figure 150**. This is an expanded region of the mass spectrum showing the internal standard along with the *m/z* tolerance (width of purple highlight). If the user wants to see the internal standard, starting material or the product, simply click on the analyte in the legend and it will automatically move to that *m/z* value which is taken from the values entered into the main GUI in BioPharma mode. The dashed line is the centroided data.

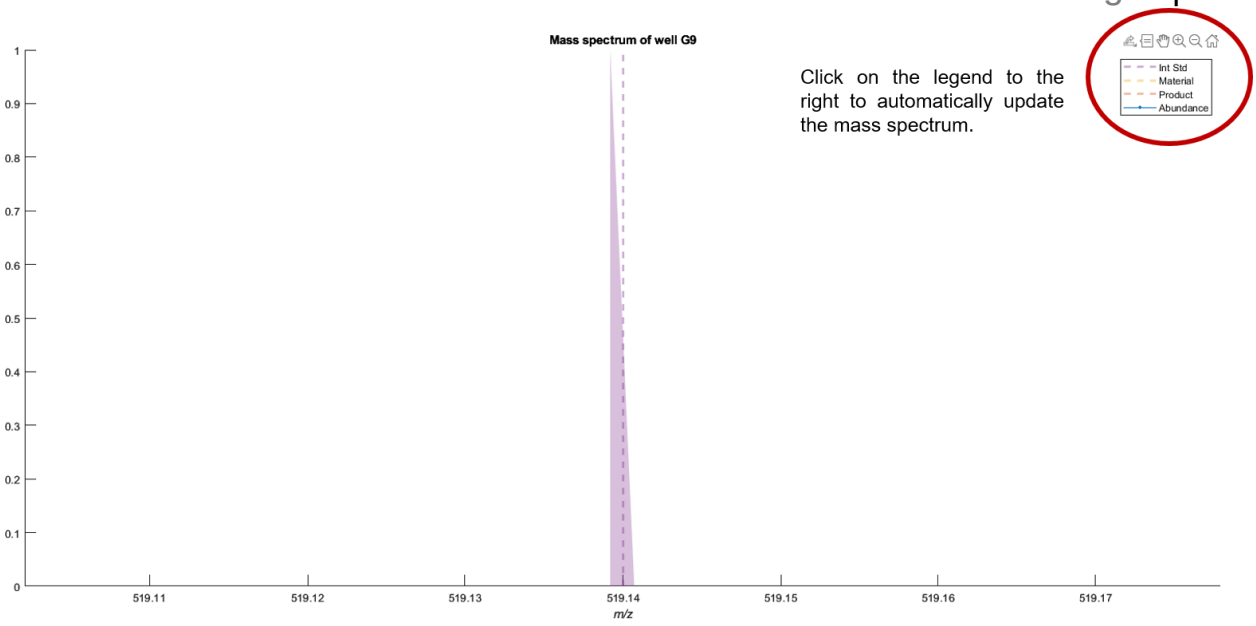


Figure 150: Mass spectrum displayed after clicking on the abundance plot to show well G9 and then clicking on the legend to view the internal standard *m/z*. The dashed line is the mass spectrometry data and the purple tapered object is the tolerance where the peak should occur.

8.1.2 Post-Processing

The post-processing pane is data context dependent. For example, if a user loads 1-shot-per-well data, only normalization choices will be selectable while “display” and “grouping” pull-down menus are hidden. The post-processing pane shown in **Figure 151** is for HTS data that was collected with 3-shots per well. The display choices will be to show all plates, unified (depends on grouping) or the ability to display shot #1, #2 or #3. The plate map automatically updates.

Once the data is loaded and perhaps the user selects “unified”, what is displayed in the plate map is dictated by “group” which can be minimum, average, median or maximum. Again, the plate map is automatically updated.

The last set of choices relate to normalization of the data. Given that the internal standard *m/z* is already entered into the MS Navigation pane, selecting “Ref Peaks” under normalize with a single *m/z* value will normalize all the data to this specific *m/z* value. If the user wishes to normalize segments of the *m/z* range to a different *m/z* reference peak,

click “multiple refs” and then *m/z* Bounds – a table will appear to define the segments of the data (*m/z* ranges) and the specific *m/z* reference value to be used for each segment. A user can also normalize to the TIC, local TIC, Max, Mean, Sum, Median, Midpoint or a Custom Normalization map.

The last entry under normalization, which is unique to the BioPharma mode, is %Conversion which is actually a calculation. This value is determined by taking the abundance of the product *m/z* and dividing it by the sum of the product *m/z* and the starting material *m/z* and expressing it as a percentage. In simple terms, if the product *m/z* abundance is very small (*e.g.*, 100) and the starting material very large (*e.g.*, 10^6), the percentage conversion would be calculated to be 0.01%. Keep in mind this calculation assumes equal ionization efficiencies of the starting material and the product. However, this calculation quickly allows a researcher to determine which drugs are promoting or perhaps inhibiting conversion.

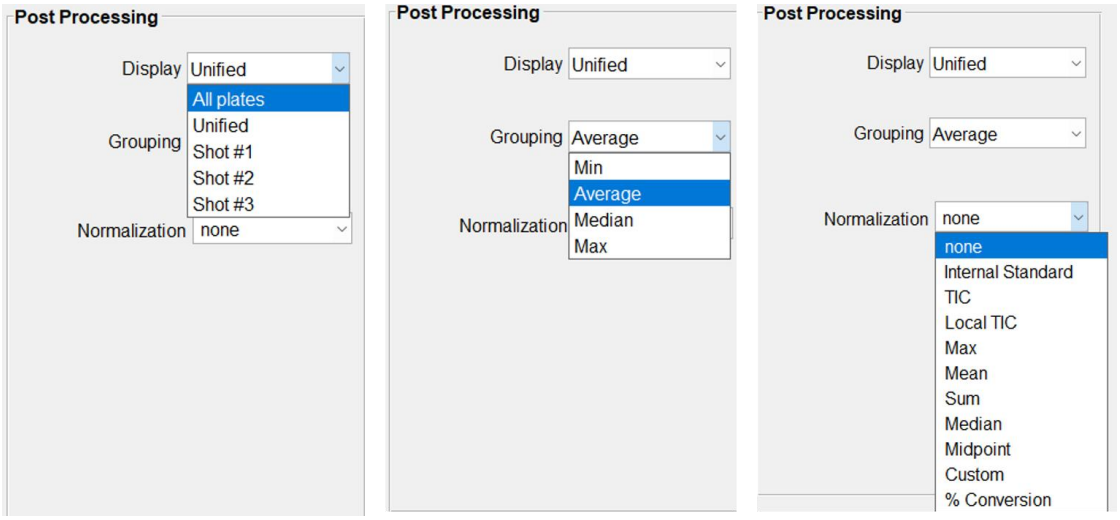


Figure 151: The post-processing pane in MSiReader BioPharma mode when 3-shots-per-well were loaded. With these data, “display”, “grouping” and “normalization” will be visible.

If the user selects “all plates” – which is a display of each individual shot per well, and an example is shown in **Figure 152**. Since these data were collected with 3-shots-per-well, MSiReader automatically inserts a “blank plate”. This allows a user to quickly visualize if the replicate analysis for a given well was consistent or not. For example, inspection of

wells D19, F19, L19, C24 and O24 all look to have low ion abundances relative to the rest of the plate. Conversely, well P24 is consistently higher in ion abundance. If it was determined that the low ion abundance data should be removed prior to statistical analysis (as one example), the user could use scan scrubber and remove them from the data set. It could also suggest to the user that they should normalize their data to account for the variation in the ion abundances.

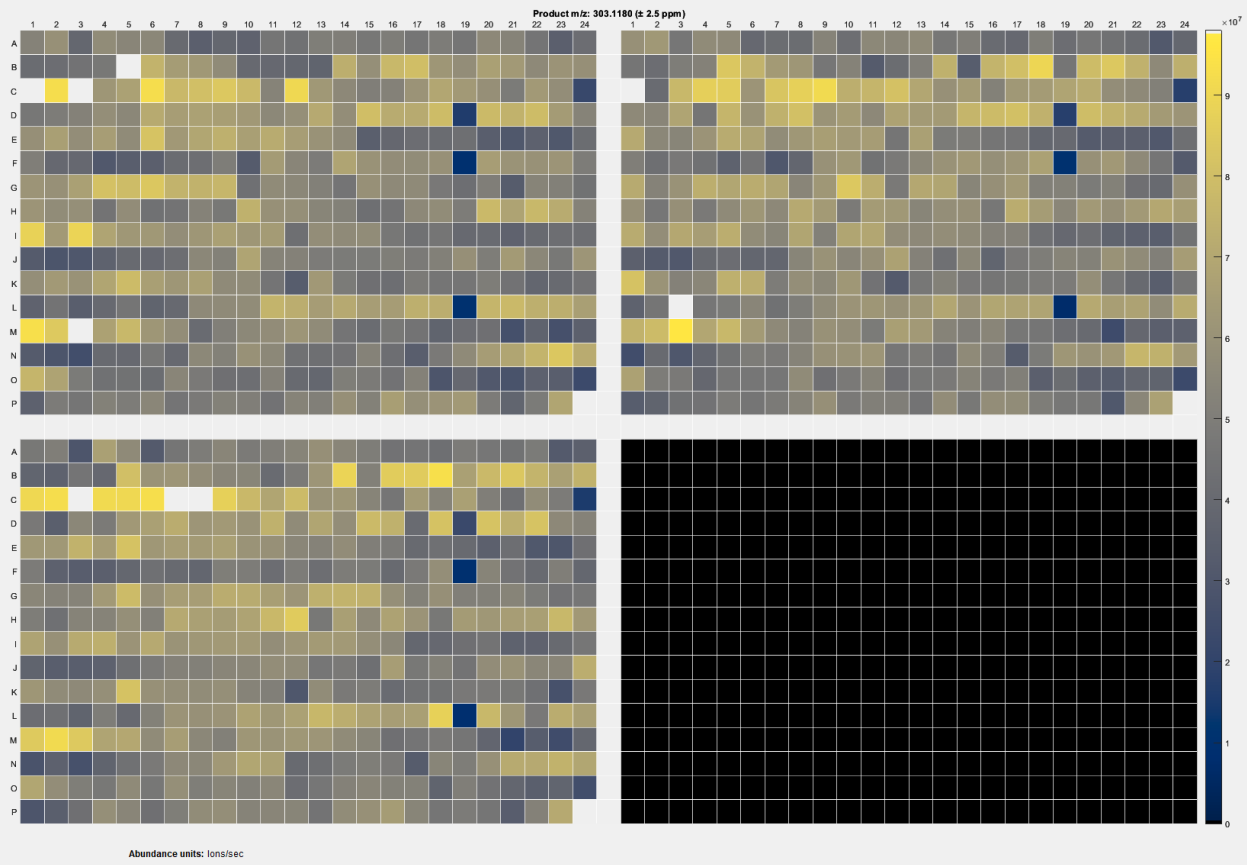


Figure 152: Loading of 3-shots-per-well high throughput screening data.

8.1.3 MS Navigation

The MS navigation pane in BioPharma mode allows end users to specific the m/z values for the product, material, internal standard (if used) and chose which m/z channel is being plotted by selecting in the pull-down menu under “Active *m/z*”. The pane is shown in **Figure 153**. The user can also change the tolerance in ppm or Th unite, change the abundance to the max, mean or sum. More options include hotspot removal with percentile, scale max and if the user wishes to lock it, as well as slide bars for the heatmap in terms of min and max abundances. Finally, the heatmap default is to auto update when changes are made but unchecking the box will not allow this. The user then can make changes and click “update heatmap”.

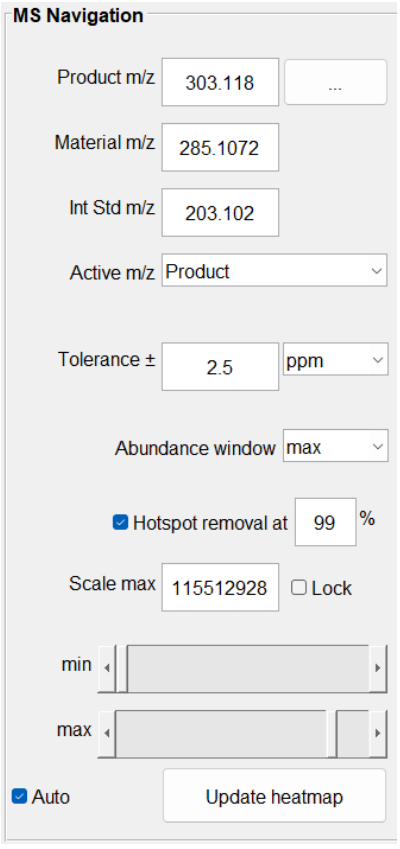


Figure 153: MS Navigation Pane in MSiReader’s BioPharma Mode.

8.1.4 Colormap

The default colormap is *cividisblack* which is color vision deficiency compliant[3,4] and presents a heatmap that is representative of the data. It is a perceptually linear colormap instead of a "rainbow" style colormap like the previous default, *jet,* which has long been considered misleading for the presentation of scientific data[3], especially when converted to grayscale and printed.

The scaling is a simple a way to better display large dynamic range data in the heatmap when you have an analyte that varies over orders of magnitude in abundance within your image. The user can choose from linear, log base 10, log base 2, and log base e. If you wish to "flip" which color is most abundant and which is least abundant, check the "flip" checkbox.

8.2 Loading imzML and Native File Formats into MSiReader in BioPharma Mode

There are many different scenarios in which one can collect HTS/HCS data. In this section, several different scenarios will be described to load data but then also, after loading, the options to save the data in a new file format that will not require the user to enter in other metadata.

*8.2.1 Loading a Thermo *.raw File Format **without** a *.json file*

MSiReader can read in the *.raw file directly. However, before loading the *.raw file, the user must select under plate type, the size of the plate (*e.g.*, 384) and also the number of shots per well. In the MSiReader Test Data folder under Biopharma then RAW Files, select the 1-shot-per-well.raw file. Select any other filtering options such as abundance thresholding or *m/z* filter. Click "Load Data" once all your selections are made. It will prompt you to select "meander" or "flyback". These data were collected in meander mode so select that. There will be an error that says there was no text file in the folder with the

data. It is telling the user that no other metadata was provided and thus, it can only use the information that was input by the user. MSiReader does check to make sure the number of scans matches the information the user put in (single shot 384 well plate should have 384 scans). If it does not match or if the user has collected multi-shot data, it will give feedback. Once you load the data, the plate map shown in **Figure 154** should appear.

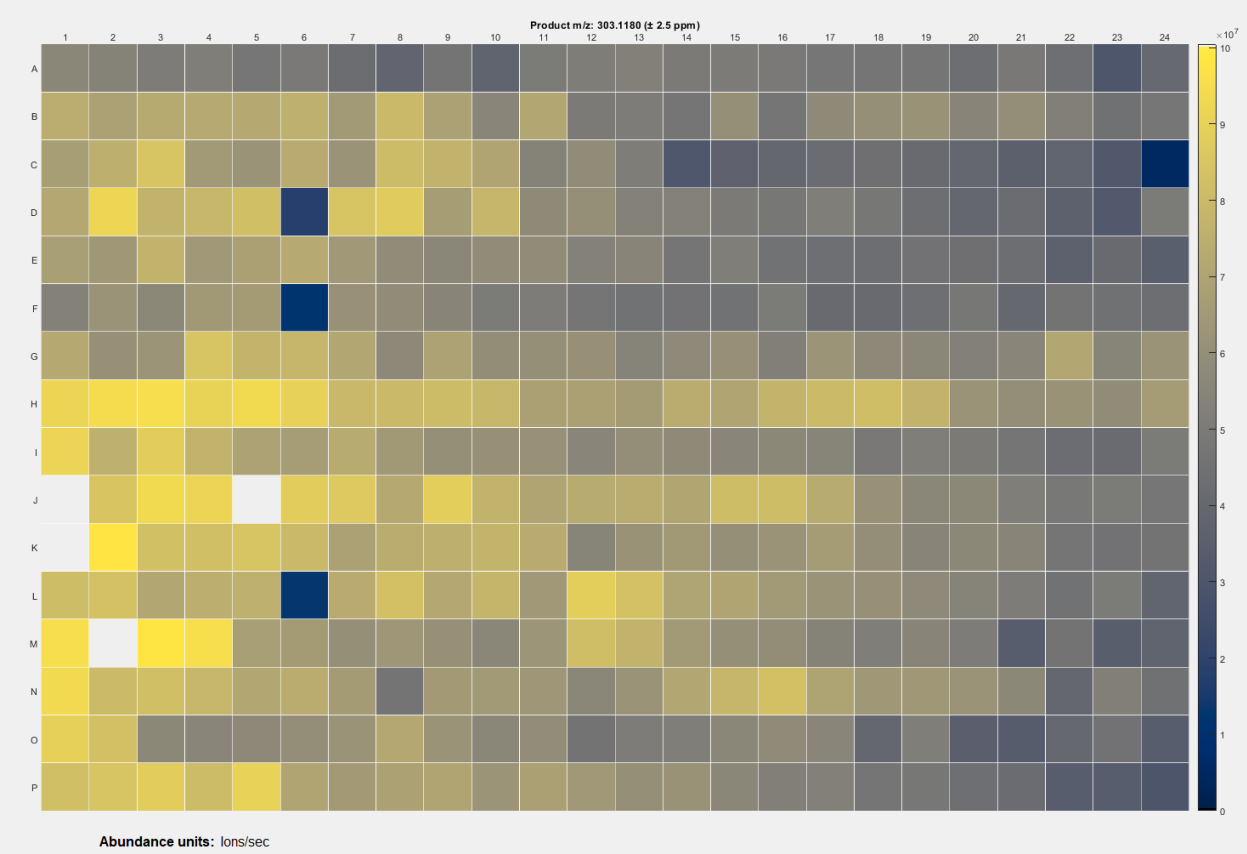


Figure 154: Plate map with abundance of *m/z* = 303.1180 being displayed. The plate map is interactive and the user can use the single pixel ROI tool to select a well, display the mass spectrum. If the user selects a single well and then right clicks, they can plot the mass spectrum. In real-time, the user can drag the cursor around the plate map and the mass spectrum automatically updates.

Once the *.raw data is loaded, it can be saved as a *.mss or *.mim format. MSiReader cannot save the data into a new *.raw file as that is a proletary Thermo format. Our custom file formats obviate the need to have all files in the same file (*e.g.*, a series of *.raw files and associated *.json files) and our formats load significant faster than reading imzML and *.raw files. To save your loaded data, go under the Home menu and select “save

current session". A dialogue box will appear and you can change the name of the file and use the pull-down under "file type" to change the format to either *.mss or *.mim. You can also access these file saving options by selecting the icon -  - third from the left in the toolbar; mouse over tooltip will state "save current session".

*8.2.2 Loading a Thermo *.raw File Format **with** a *.json file*

In the same folder as the *.raw file, if a *.json file is present, a user can load more complex data structures (*e.g.*, 3-shots-per-well). As an example, there is a 3-shots-per-well.raw file and a 3-shots-per-well *.json file in the RAW files folder. **The *.json file must have the exact same filename** as the *.raw file but have *.json as the extension. Load the 3-shots-per-well.raw file and MSiReader automatically reads in the metadata from the associated *.json file. Note that in this case, there were 3 extra scans at the end of this run and the software recognized that and gave the user feedback if they wanted to continue.

8.2.3 Loading HTS/HCS using the imzML File Format

MSiReader can read in a single or multiple imzML files at the same time. In the case of a single imzML file, simply click "Load Data" and select the imzML file you wish to analyze. Even when the user selects the wrong plate type (*e.g.*, the data is 384 but the user selected 1536, MSiReader knows this and automatically changes the plate type to 384.

To convert your single-shot per well data to an imzML file, start with the *.raw file and convert it from *.raw to *.mzML. Next, open the imzML convertor and set pixels in x = 24 and pixels in y = 16 (for a 384 well-plate – adjust accordingly for other plate types). The file organization should be "image per file". Click convert. This file is now readable into MSiReader.

If you are collecting multiple shots per well, a location file is necessary so that the software knows that these are replicates (*e.g.*, scans 1-3 go to well position A1). The user will need to generate location files for each of the mzML files using the tool under Home called

"prepare plate location files". When this tool is selected, the dialog box as shown in **Figure 155**

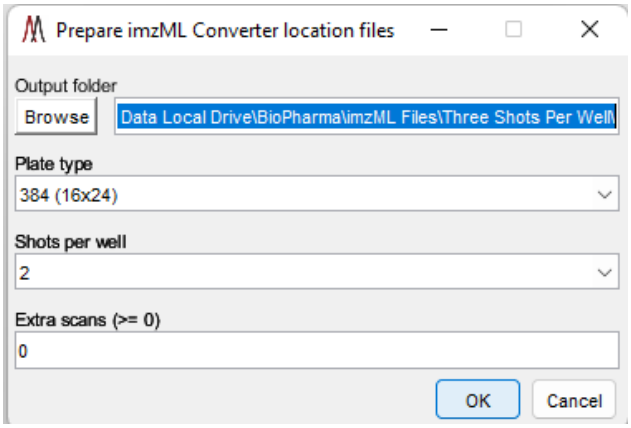


Figure 155: Preparing plate location files. The user should enter in the folder where the data is located, choose the plate type, enter in how many shots per well and enter in if there are any extra scans. This will automatically generate location files equal to the number of shots per well the user entered. These location files will be needed in a subsequent step.

Next, convert your multiple-shot per well data to an imzML file, start with the *.raw file and convert it from *.raw to *.mzML. This is identical to the process for single-shot per well data. Next, open imzML converter and click ADD and then select each mzML files to be converted (one at a time). Next, set pixels in x = 24 and pixels in y = 16 and file organization should be "image per file". For the first shot-per-well, select loc384_1 for the spectra location file and when prompted for filename, add _1 to denote that this is shot #1. Repeat this step leaving all settings the same except this time, select loc384_2 for the spectra location file. Be sure to add _2 to denote that this is shot #2. Repeat for the number of shots that you have in your data. The imzML files that were just created can now be read into MSiReader. Important note, if you are working with 384 well plates and regularly conduct 3 shots per well, you do NOT need to regenerate the location files. They are specific to the plate type and the # of shots-per-well – once created they can be reused.

8.3 Preparing HTS and Phenotypic Data for Multivariate Analysis

A **video tutorial** on how to prepare MSI and HTS/phenotypic screening data for downstream statistical analysis can be found **HERE**.

An interactive multivariate analysis tool is provided in MSiReader. MSiPCA is based on singular value decomposition and loads an abundance matrix saved with the MSiExport sub-GUI, calculates PCA loadings and scores and plots the results for user selected components. The plots are 1) a heatmap showing the PCA score distribution for any number of user selected components (optional), 2) a biplot showing loading and scores for two or three components (and the end-user can select which PC's to plot); and 3) an interactive PCA loadings plot as a function of *m/z*. The rows of the input matrix are observations (*i.e.*, mass spectrometry scans, and the columns are variables, *i.e.*, *m/z* values. The PCA results: component loadings, scores, latent variances, T-squared values, and explained variance percentages can be saved into a single Excel file, each in its own worksheet. Moreover, a scree plot is also automatically generated. Examples and more details will be shown later in this section.

Important Note: PCA algorithms are limited by the number of degrees of freedom (DOF) in the data you are working with; thus, if the # of scans (mass spectra) exceed the number of *m/z* values you can use all the *m/z* bins in the analysis. However, this is not typical with HRAM data and thus, the DOF will be limited to the number of scans in your data. In the PCA algorithm, it will use all the *m/z* binned data and then limit the output to those PC's which describe most of the variance in the data and put in null values for *m/z* bins that exceed the number of DOF. Alternatively, prior to using the MSiPCA tool, the end-user can pre-process the data to bring the number of *m/z* values to be less than the number of scans. There are several ways to meet this criterion and more than one, or all of them, can be used. It depends on the types of data that the end-user is trying to analyze. The ways that the user can ensure that the # scans > # *m/z* values are listed here.

- Select a larger number of wells to include for HTS/HCS data as this inherently increases the number of scans but, to a first approximation, the number of *m/z* values does not change.
- Centroiding data. Since a profile peak contains many *m/z* values to describe the entire peak, by centroiding, one can reduce of *m/z* values in the dataset.
- Abundance thresholding data. Given that each instrument is different in terms of the values it reports for ion abundance, the end-user should threshold the data so noise and low abundant peaks (high variance) are filtered out. This will substantially reduce the number of *m/z* values. It is recommended to set this value at an abundance that is 10× the limit-of-detection.
- Filter out narrower *m/z* ranges prior to data export. Not recommended as a first approach because the goal of PCA and other multivariate methods is to use all the data that describes the data.
-

Checking Data Attributes and Quality

It is important to check your data attributes and quality prior to selecting abundance threshold and the tolerance for peak exclusion. To check for the abundance threshold, the user should load a typical file in the study and then using the single pixel ROI, select a representative well. Right click and “plot *m/z* spectrum”. Expand the spectrum using the mouse wheel and then look for low abundant peaks by moving the cursor on top of the peak. The pop-up window will give you the *m/z* value and the abundance. Look for the low abundance peaks and this will give the user an idea of what the threshold should be. In the test HTS and HCS data sets, the low abundance peaks are about 12,000 in ion abundance; thus, to be conservative the user could input 12,000 or to include additional lower abundance peaks, the user could use 8000.

The next step for QC would be to make a plot of the MMA for some representative data. The user can do this by loading a file and then going to the drop-down QA/QC menu and choosing “mass measurement accuracy”. In the main GUI, enter 10 ppm for starters into the tolerance and then make the plot. If the data is showing that it is all better than 5 ppm, then your tolerance for peak exclusion should be 5 ppm. If the data does not have the

expected MMA and you have some known *m/z* values in the different images, the data should be mass corrected **before** doing anything else. This process can be found in §7.3.1.

Preparing data prior to using MSiPCA

1. Unload your data files from MSiReader.
2. Launch the centroid data function under preprocessing drop-down menu. It will automatically be in batch mode since no data files are loaded. The dialogue box as shown in **Figure 140** will appear. Once you set your parameters, click OK and it will open a file explorer to load the imzML files that you will use in the PCA.

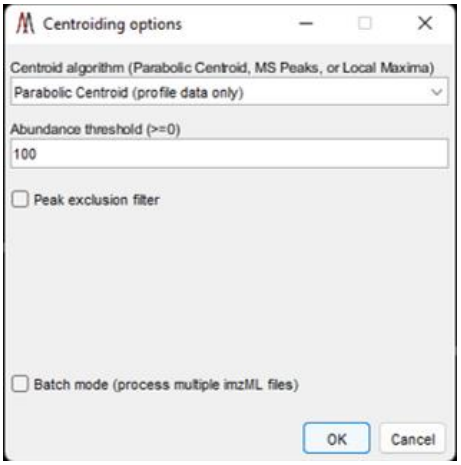


Figure 156: Centroid Data Options Panel in MSiReader

If your dataset is profile data, it is recommended to centroid your data using the “parabolic centroid” algorithm. If the data is already centroided, please select “local maxima” in the drop-down menu. Next, select an abundance threshold that is specific to the instrument; 100 is the default but in most cases, this will be much higher.

Inevitability, mass spectrometry data has peaks in the mass spectra that are not sample related. For example, in MALDI, the ions from the organic matrix are present but are not

related to the HTS/HCS screen. In ESI-post ionization methods and DESI, ambient molecules interact with the charged droplets and produce ambient background ions. The file PCAPeakstoRemove.txt has a single *m/z* that is not tissue specific; this file can be used in this example. It is critical that as many non-tissue specific ions for HTS/ HCS data are removed from the data prior to PCA analysis because these could drive the PCA to an incorrect conclusion. See **§7.3.4** for how to classify ions as being background or sample-specific. When the peak exclusion filter is checked, the user can browse for a .txt file containing the theoretical *m/z* values of the ions that are not sample related (make sure to include *m/z* values for background ions that have abundant A+1 and A+2 peaks). These will be excluded if the *m/z* is detected in the mass spectrum with a tolerance set by the end user. It is important to know the MMA of your dataset prior to setting this tolerance; otherwise, it will remove non-tissue peaks in some spectra that have high MMA but not in those that fall outside this tolerance. Click OK and new imzML files will be generated with "_centroided" automatically added to those files. If you loaded a Thermo *.raw file, the user can save the processed data as a *.mim file.

It is important to note that a user can also carry out the above steps for an entire folder of imzML files without ever having to load the files into MSiReader. To do this, clear the data and then choose Centroid Data under the Pre-Processing menu and enter in the dialogue box what function(s) needs to be carried out and click OK. It will then open a file explorer to choose the imzML files that need to be processed. Batch processing of *.raw files is not available but will be in a future release.

Exporting Data for Statistical Analysis using MSiExport

Open the processed data files that will be used to carry out PCA. Launch the MSiExport tool under the *Annotations* drop-down menu, then *Data Export* then *Export Abundance Data*. Draw a single ROI around all the data OR draw an interrogated and reference ROI. The user will be asked to select which scans to consider – select scans for the ROI(s). Next, choose *Build and m/z list and bin the pixels* in the ROI. Next, enter in values that correspond to your data; for high resolution accurate mass (HRAM) data, 5 ppm is

recommended. Click the Browse button and enter a filename and folder for the .txt file to be prepared. Click OK. The .txt file will be the data that is entered into the PCA function.

For export to text format two files named <*name*>*_info.txt* and <*name*>*_peaks.txt* were saved. If the data was from a tiled image mosaic an additional file, <*name*>*_files.txt* was also saved. All of these are required to enable the full capabilities of MSiPCA and must be in the same folder. The <*name*>*_peaks.txt* file should be selected from the input dialog.

8.4 PCA and t-SNE Analysis of High Throughput and Phenotypic Screening Data

A **video tutorial** on how to use the PCA tool can be found **HERE**.

A **video tutorial** on how to use the t-SNE tool can be found **HERE**.

This section is a step-by-step example using phenotypic screening data and applying PCA and t-SNE to it. The background and details of how to prepare data for PCA and t-SNE was provided above in **§7.8.1**.

8.4.1 Load the Data into MSiReader BioPharma Mode and Check Data Attributes.

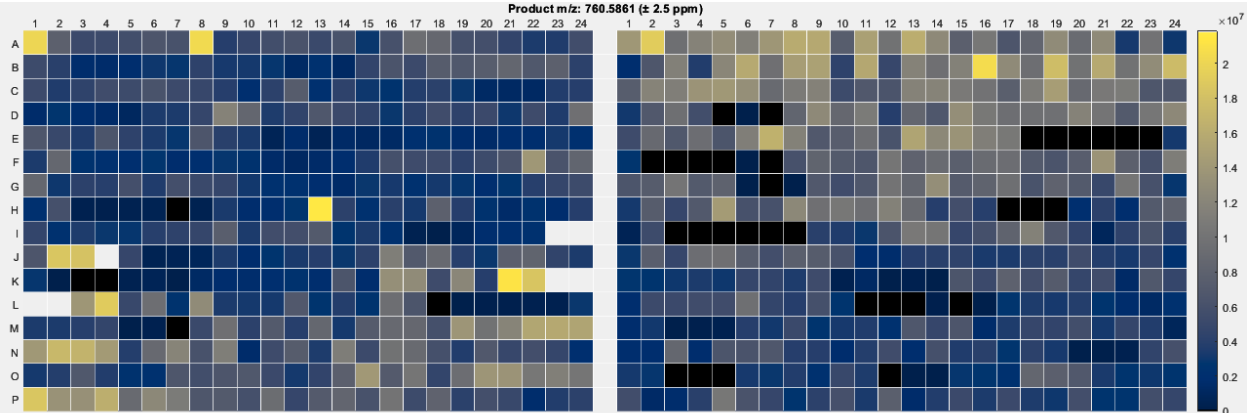


Figure 157: HCS replicate data from a 384-well-plate run twice; plate was flipped before the second run. Notice immediately that the upper part of run 1 (left plate) has lower abundance which is mirrored by run 2 in the bottom part (right plate).

In the example data shown in **Figure 157** are loaded from 2 384-well plates; they are the same samples but before the plate was run a second time, it was flipped. These data

(imzML format) can be found in the test data folder BioPharma, subfolder statistics. The first step is to do a check on the abundance of the peaks in the spectra. To do this, simply click the single pixel ROI icon and drag it to a well. Right click on the location and select "plot *m/z* spectrum). Expand the mass spectrum in a region with lower abundance and then place your cursor on the centroided peak – the abundance will appear. In this example, low abundant peaks were about 12,000 so entering in a threshold abundance of about 7500 would be conservative to avoid losing information.

The next check is to determine if the MMA is sufficiently good or does it need to have a mass correction applied. To determine this, simply go to the QA/QC menu and then select from the drop down under MMA, select, plot the MMA for the current m/z. In doing so, the data will appear as shown in **Figure 158**.

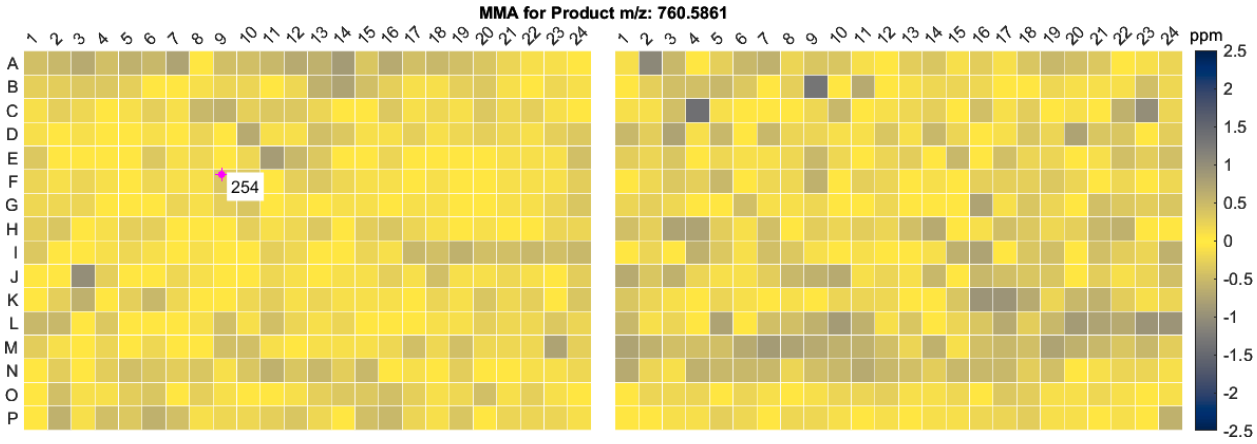


Figure 158: MMA heatmaps for *m/z* = 760.5861. The MMA of this analyte across wells is mostly sub-ppm and is not a function of time; however, searching with a tolerance of +/- 2.5 ppm is recommended.

The next step could be to apply the threshold abundance filter of 7500 to these data and then save them as new imzML files. However, this is not essential since PCA and t-SNE will reduce the dimensionality of the data. We will skip this step in this example. If you do want to threshold the data, go to Pre-Processing, Centroid Data and then choose local maxima (since these data were collected natively as centroided data), enter in 7500 as

the abundance threshold and if you have a list of peaks that you wish to remove (*e.g.*, backgrounds ions), check peak exclusion filter and then upload the .txt file with the *m/z* values you wish to remove. Important to note that if the end-user collected data natively as centroid and they do not wish to apply a peak exclusion filter, the abundance threshold can be applied during MSiExport.

The next step is to export the data – this data export file will be used for both PCA and t-SNE. Go to Annotations, Data Export and then Select MSiExport. MSiReader assumes that the user will process all wells in each plate and all plates. This will prompt the MSiExport tool explained previously. Select "build a binned *m/z* list" and enter in other values that are appropriate for your data. In this example, the data was binned at 20 ppm. Click browse and choose your location and filename for the output. There are a total of 784 scans in these data; extra due to the white border between the two well plates (these will not affect downstream analysis). At 20 ppm bin width, the number of bins = 69,312 which is far greater than the number of scans. A progress bar at the bottom is displayed.

The next step is to load the data export *.txt file and run PCA. Select PCA under the Statistical Analysis menu item. It will prompt the user to select the data export file that was just created and give another progress bar as it loads that data into memory. The program will then request user input with respect to the *m/z* range and abundance data normalization. Other choices are also available. The PCA of these data resulted in scores plot shown in **Figure 159**. Since there are replicate analysis of the same plate, they should not separate well. However, within a sample, there is separation and that is due to the fact that each well has the same drug but at a different concentration or has a different drug in it.

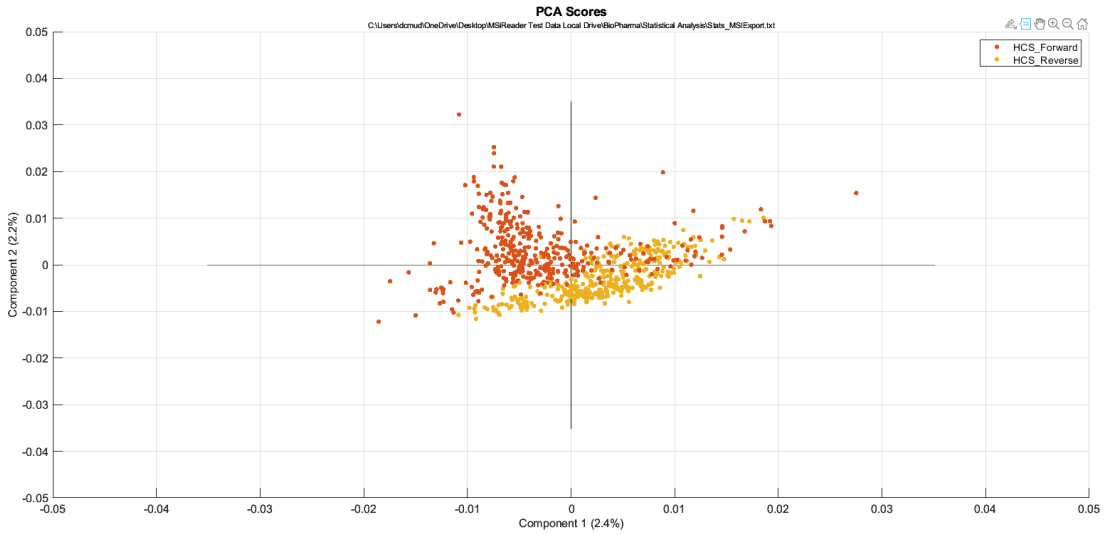


Figure 159: PCA Scores plot upon analysis of the HCS replicate data. These should not separate very much from one another as they are "replicate" data aside from one plate being flipped. Note the explained variance from PC1 (2.4%) and PC2 (2.2%). What is more important here is within plate analysis as each well is the same cells but the same drug at different concentrations or different drugs.

t-SNE analysis of the same HCS data MSiExport file is shown in **Figure 160**. While the separation is different than PCA, closer inspection will reveal their similarities. This is a simple example about how to use t-SNE to reduce the dimensionality of the data.

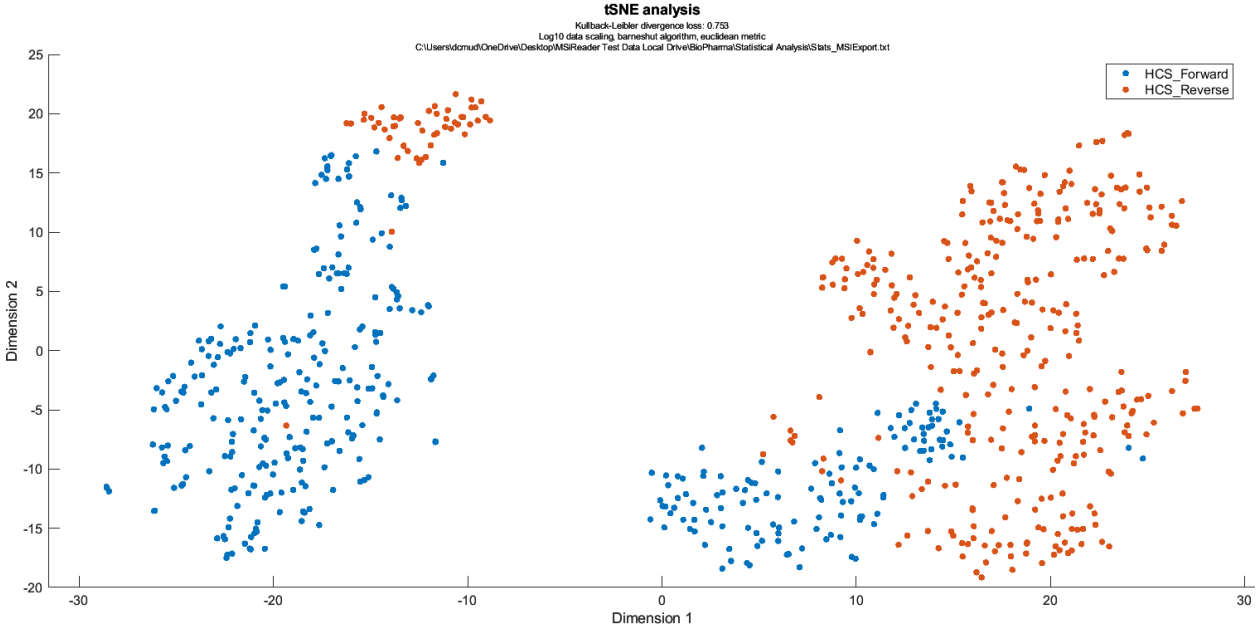


Figure 160: t-SNE analysis of the HCS data. This analysis reveals separation within a plate and between plates. A deeper look at the data should reveal that there is separation within a plate based on the drug present or the drug concentration.

One more example would be to load up multiple shots per well HTS data where each well is the exact same sample. In this way, one can look at intra- and inter-well variability of the method. In this example, 3-shots-per-well of a 384-well plate were loaded into MSiReader and the data was exported as above using MSiExport. In this case, since each well in each plate is the same reaction and there are 3 replicates per well, the average of each plate should be identical. The result of each plate, once averaged for the 3 replicates, are essentially identical and are at the origin of the PCA plot as shown in **Figure 161**.

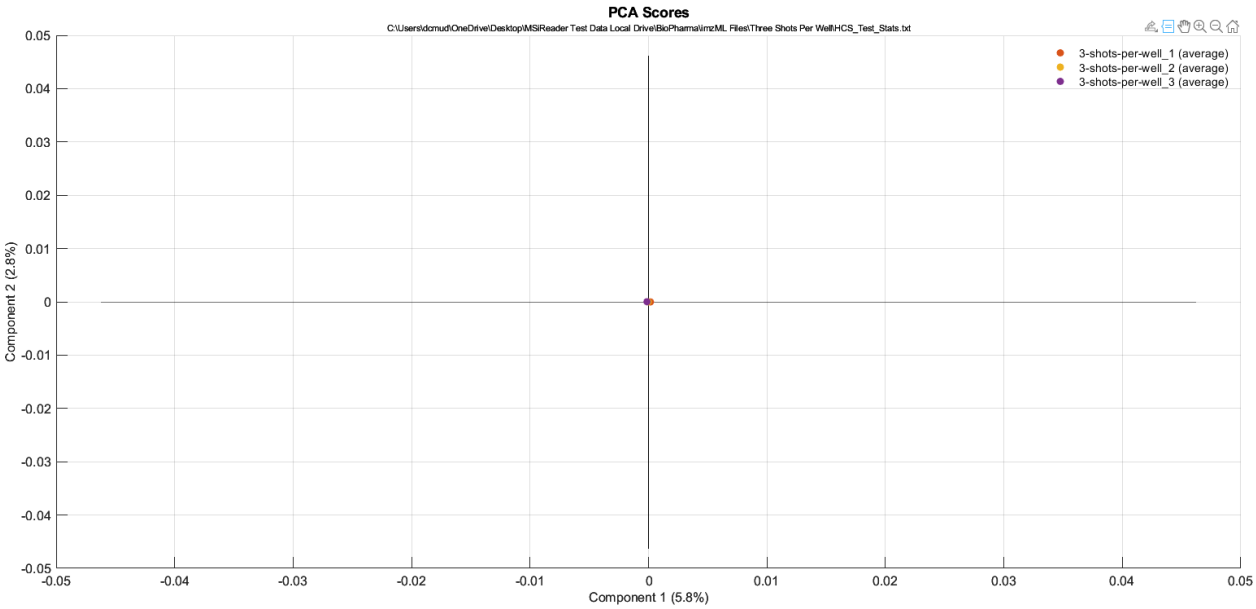


Figure 161: PCA plot of HTS data, same reaction in each of the 384 wells and each well was measured in triplicate. The scores plot shows the average each of the three replicates at the origin. This is the expected result for a repeatable method.

t-SNE analysis of the same HTS data that was shown above in **Figure 161**. Remember that every well and every replicate scan should be 100% identical if there was not any analytical variability. Thus, the results in **Figure 161** and **Figure 162** should not separate out using PCA or t-SNE.

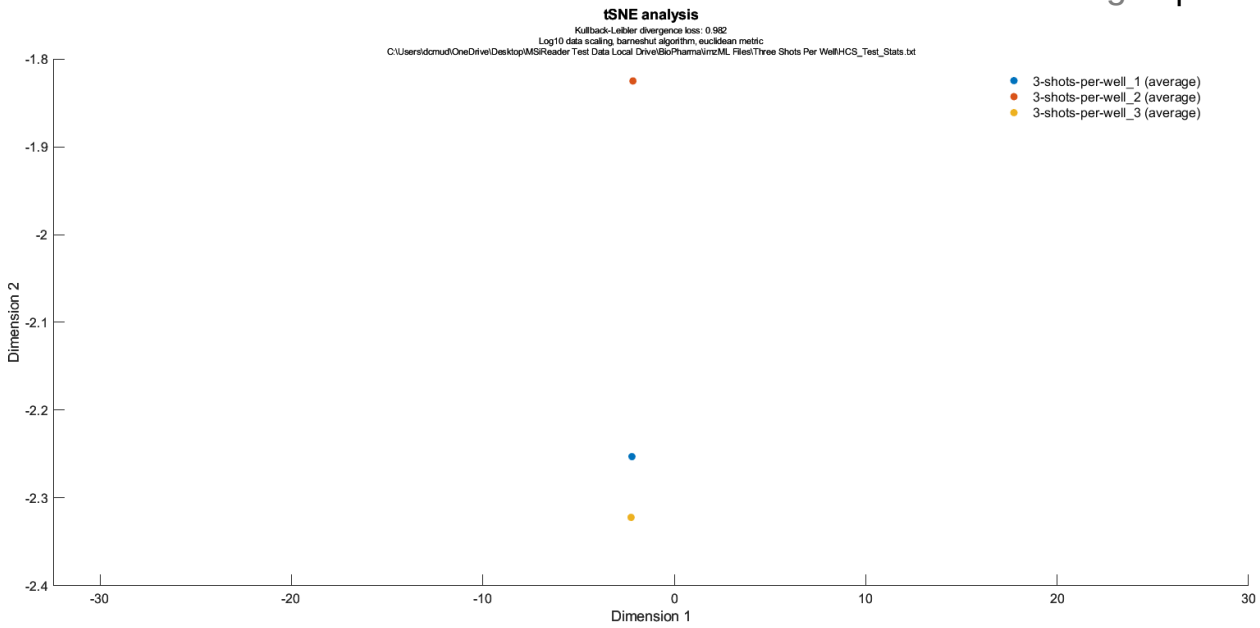


Figure 162: t-SNE plot of HTS data, same reaction in each of the 384 wells and each well was measured in triplicate. The plot shows the average each of the three replicates in dimension 1 is identical while there is modest separation in dimension 2. This is the expected result for a repeatable method.

More examples of using statistical methods for the analysis of MSI data will be provided at a later date. Moreover, new statistical tools are currently in development and will be released in the latter half of 2023. Please let us know if you have any questions: support@msireader.com.

9 Literature Cited

(1) Robichaud, G.; Garrard, K. P.; Barry, J. A.; Muddiman, D. C. MSiReader: An Open-Source Interface to View and Analyze High Resolving Power MS Imaging Files on Matlab Platform. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (5), 718–721.

(2) Bokhart, M. T.; Nazari, M.; Garrard, K. P.; Muddiman, D. C. MSiReader v1. 0: Evolving Open-Source Mass Spectrometry Imaging Software for Targeted and Untargeted Analyses. *J. Am. Soc. Mass Spectrom.* **2017**, *29* (1), 8–16.

(3) Nuñez, J. R.; Anderton, C. R.; Renslow, R. S. Optimizing Colormaps with Consideration for Color Vision Deficiency to Enable Accurate Interpretation of Scientific Data. *PLOS ONE* **2018**, *13* (7), e0199239. https://doi.org/10.1371/journal.pone.0199239.

(4) Knizner, K. T.; Kibbe, R. R.; Garrard, K. P.; Nuñez, J. R.; Anderton, C. R.; Muddiman, D. C. On the Importance of Color in Mass Spectrometry Imaging. *J. Mass Spectrom.* **2022**, *57* (12). https://doi.org/10.1002/jms.4898.

(5) Pedrioli, P. MSConvert Tool from Proteowizard. http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML.

(6) Pedrioli, P. G. A.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R. A Common Open Representation of Mass Spectrometry Data and Its Application to Proteomics Research. *Nat. Biotechnol.* **2004**, *22* (11), 1459–1466. https://doi.org/10.1038/nbt1031.

(7) Kibbe, R. R.; Muddiman, D. C. Achieving Sub-Parts-per-Million Mass Measurement Accuracy on an Orbitrap Mass Spectrometry Imaging Platform without Automatic Gain Control. *J. Am. Soc. Mass Spectrom.* **2023**, jasms.3c00004. https://doi.org/10.1021/jasms.3c00004.

(8) Schramm, T.; Hester, Z.; Klinkert, I.; Both, J.-P.; Heeren, R. M. A.; Brunelle, A.; Laprévote, O.; Desbenoit, N.; Robbe, M.-F.; Stoeckli, M.; Spengler, B.; Römpp, A. ImzML--a Common Data Format for the Flexible Exchange and Processing of Mass Spectrometry Imaging Data. *J. Proteomics* **2012**, *75* (16), 5106–5110. https://doi.org/10.1016/j.jprot.2012.07.026.

(9) *imzML / imzML Validator · GitLab*. GitLab. https://gitlab.com/imzML/imzMLValidator (accessed 2022-09-09).

(10) Race, A. M.; Römpp, A. Error-Free Data Visualization and Processing through ImzML and MzML Validation. *Anal. Chem.* **2018**, *90* (22), 13378–13384. https://doi.org/10.1021/acs.analchem.8b03059.

(11) *METASPACE annotation platform*. https://metaspace2020.eu/ (accessed 2022-09-09).

(12) Pace, C. L.; Garrard, K. P.; Muddiman, D. C. Sequential Paired Covariance for Improved Visualization of Mass Spectrometry Imaging Datasets. *J. Mass Spectrom.* **2022**, *57* (7). https://doi.org/10.1002/jms.4872.

(13) Muddiman, D. C.; Rockwood, A. L.; Gao, Quanyin.; Severs, J. C.; Udseth, H. R.; Smith, R. D.; Proctor, Andrew. Application of Sequential Paired Covariance to Capillary Electrophoresis Electrospray Ionization Time-of-Flight Mass Spectrometry:

Unraveling the Signal from the Noise in the Electropherogram. *Anal. Chem.* **1995**, *67* (23), 4371–4375. https://doi.org/10.1021/ac00119a027.
(14) Muddiman, D. C.; Huang, B. M.; Anderson, G. A.; Rockwood, A.; Hofstadler, S. A.; Weir-Lipton, M. S.; Proctor, A.; Wu, Q.; Smith, R. D. Application of Sequential Paired Covariance to Liquid Chromatography-Mass Spectrometry Data Enhancements in Both the Signal-to-Noise Ratio and the Resolution of Analyte Peaks in the Chromatogram. *J. Chromatogr. A* **1997**, *771* (1–2), 1–7. https://doi.org/10.1016/S0021-9673(97)00069-1.
(15) Eisenberg, S. M.; Knizner, K. T.; Muddiman, D. C. Development of an Object-Based Image Analysis Tool for Mass Spectrometry Imaging Ion Classification. *Anal. Bioanal. Chem.* **2023**. https://doi.org/10.1007/s00216-023-04764-x.
(16) Khodjaniyazova, S.; Nazari, M.; Garrard, K. P.; Matos, M. P. V.; Jackson, G. P.; Muddiman, D. C. Characterization of the Spectral Accuracy of an Orbitrap Mass Analyzer Using Isotope Ratio Mass Spectrometry. *Anal. Chem.* **2018**, *90* (3), 1897–1906. https://doi.org/10.1021/acs.analchem.7b03983.
(17) Rockwood, A. L. Relationship of Fourier Transforms to Isotope Distribution Calculations. *Rapid Commun. Mass Spectrom.* **1995**, *9* (1), 103–105. https://doi.org/10.1002/rcm.1290090122.
(18) Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image Quality Assessment: From Error Visibility to Structural Similarity. *Image Process. IEEE Trans. On* **2004**, *13*, 600–612. https://doi.org/10.1109/TIP.2003.819861.
(19) Ekelöf, M.; Garrard, K. P.; Judd, R.; Rosen, E. P.; Xie, D.-Y.; Kashuba, A. D. M.; Muddiman, D. C. Evaluation of Digital Image Recognition Methods for Mass Spectrometry Imaging Data Analysis. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (12), 2467–2470. https://doi.org/10.1007/s13361-018-2073-0.
(20) Verdun, F. R.; Ricca, T. L.; Marshall, A. G. Beating the Nyquist Limit by Means of Interleaved Alternated Delay Sampling: Extension of Lower Mass Limit in Direct-Mode Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Appl. Spectrosc.* **1988**, *42* (2), 199–203.
(21) Giancaspro, C.; Comisarow, M. B. Exact Interpolation of Fourier Transform Spectra. *Appl. Spectrosc.* **1983**, *37* (2), 153–166. https://doi.org/10.1366/0003702834633984.