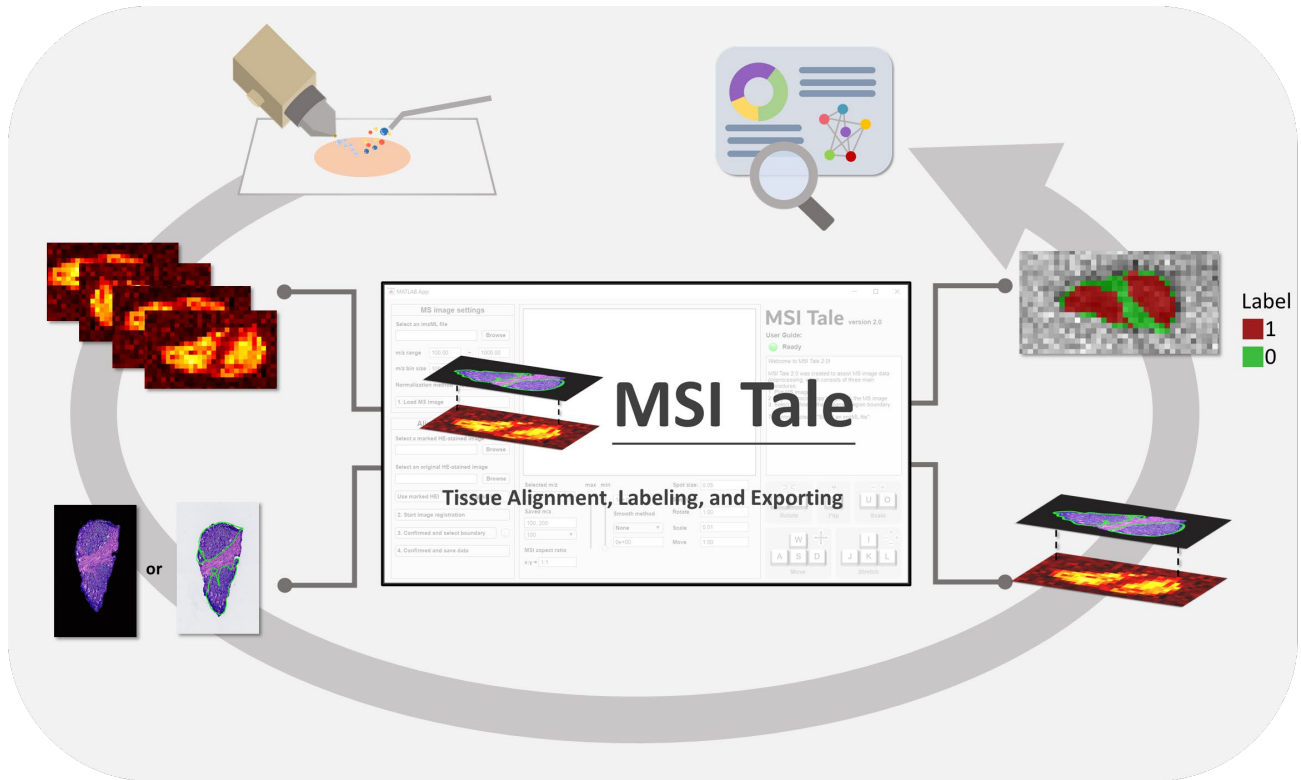


MSI Tale User Guidelines



MSI Tale is for Research Use Only

¹Yu-Hsuan Chen, ¹Wei-Kai Shao, ¹Penghsuan Huang, ¹Ying-Chen Huang, ¹Hsin-Hsiang Chung, ²Ming-Yang Wang, ¹Cheng-Chih Hsu

¹Department of Chemistry, National Taiwan University, Taipei 10617, Taiwan

²Department of Surgery, National Taiwan University Hospital, Taipei 10002, Taiwan

User guidelines

1. Installation	3
2. Tutorials	3
3. MSI Tale interface.....	3
4. Status light	3
5. Data preprocessing and visualization.....	4
6. Image registration and ROI selection	5
7. Spectra labeling and exporting.....	6

User guidelines

1. Installation

The MSI Tale software and example data are provided. Users can get access the GUI either through MATLAB applications or download MATLAB Runtime and MSI Tale package for standalone use. Before execute MSI Tale standalone app, please verify that MATLAB Runtime (R2023a) is installed. Users can follow the instruction in the readme.txt file that accompanies the application.

2. Tutorials

The step-by-step guidelines will be presented in the user guide panel. Users can follow this guidance to process mass spectrometry imaging (MSI) data. Additionally, supplementary instructions on feature usage and more detailed explanations will be provided. Furthermore, video tutorials for operating the MSI Tale can be accessed here: <https://www.youtube.com/watch?v=DscESZURzvI>

3. MSI Tale interface

Figure S1 shows the MSI Tale's interface, comprising of the MS image settings panel, align image settings panel, image panel, and a user guide panel that provides comprehensive step-by-step operation guideline. The MSI Tale is compatible with various MSI data from different brands for allowing to import imzML format files. The processed data were exported in CSV format file, enabling users easier get access to the data.

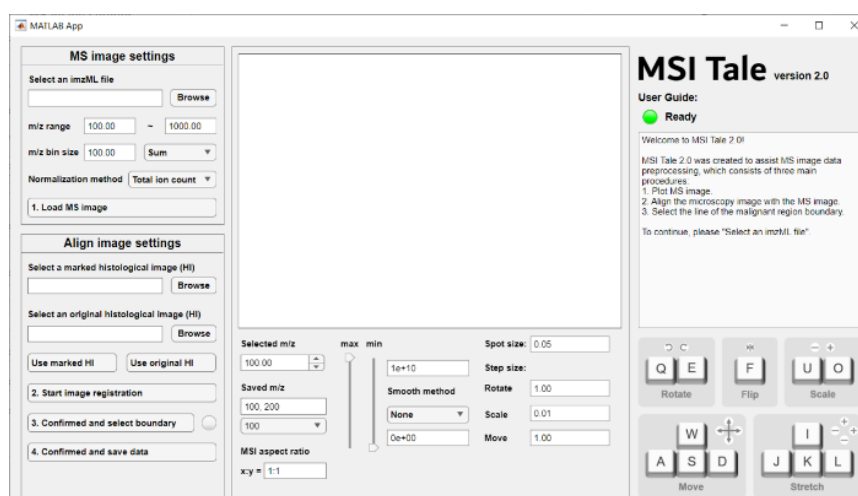


Figure 1 The interface of MSI Tale. MSI Tale's interface includes the MS image setting panel (Upper left), align image setting panel (Down left), image panel (Middle), and user guide panel (Right).

4. Status light

The MSI Tale GUI includes a status light for reporting the progress of the MSI Tale. The MSI Tale employs a color-coded lighting system to convey different operational states (Figure S2). When the graphical user interface is prepared for subsequent actions, a green light is displayed. During

processing operations, a yellow light illuminates, accompanied by text that details the ongoing task. If an error or malfunction occur, a red light appears, along with an adjacent message explaining the nature of the issue. In this example, the user may not correctly placed or prepared the file intended for reading.

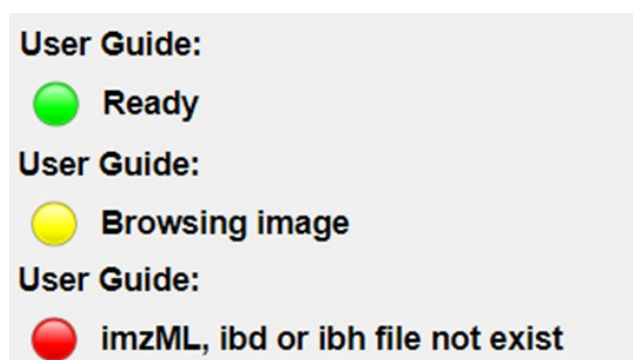


Figure 2 The status lights on the MSI Tale use different colors to indicate various states: a green light indicates that the graphical user interface is ready for the next step; a yellow light shows that the GUI is processing, with accompanying text describing the current processing status; a red light signifies that an improper operation has occurred, with an error message displayed next to the light.

5. Data preprocessing and visualization

In MSI Tale, the preprocessing step includes browsing imzML file, setting parameters, and loading MSI data. MSI data preprocessing will be done in the MS image setting panel. After browsing MSI data, it is possible to specify the settings before loading MS images, including the desired mass range, m/z bin size, and the method used to calculate the abundance for each bin (Figure S3). There are two options for calculating the abundance: The “Sum” method accumulates all the abundances in the selected bin, while the “Max” method extracts only the maximum abundance in the chosen bin. Notably, the format of the exported data will follow the settings established in this process. Next, the normalization options, including “none”, “total ion count (TIC)” and “base peak”, can be conducted to assist subsequent MSI data visualization.

Users can select the desired m/z value, adjust the minimum and maximum abundance for the displayed m/z , and control the MSI pixel aspect ratio. Smoothing methods for the MS image can be applied if necessary. Frequently used m/z values can be saved for efficient future use. Any changes made to these settings immediately update the MS image, allowing users to refine and optimize the image for better alignment and registration in subsequent analyses (Figure S4). Upon altering any of the functions, the MS image will be immediately updated MS image, which enables the user to optimize the most favorable MS image for successive image registration.

Figure 3 GUI for MS image setting, allowing the selection of an imzML file and specification of parameters for MSI data processing.

Figure 4 The interface for fine-tuning MS image settings.

6. Image registration and ROI selection

Users can align MSI data with histological images and either directly select or delineate ROIs for comprehensive analysis. During this step, the align image setting panel will be activated, allowing the user to import histological images and successive operations. For precise alignment and targeted selection of ROI, the MSI Tale enhances insights into spatial molecular distribution and tissue morphology, building upon the preprocessing steps.

Upon completion of adjustments to MS imaging, users can import histological images for aligning with MSI data within the MSI Tale. The registration process is facilitated in the align image panel, which enables users to browse for marked or original HE-stained image files, respectively. Both the marked and original histological images are available for browsing before the registration procedure, allowing users to choose either image to align subsequently. Besides, the size of the MS image will be rescaled to approximately match that of the histological image to prevent blurring of the histological image during registration. For utilizing the original histological image, delineating areas of interest

will be needed while marked histological images can skip this step. Users can firstly click the image panel to activate the function and press the space key to start delineating the ROIs. After the ROIs are decided, the image registration step will be available. The image panel should be clicked so that the registration procedure can be activated. Through intuitive keyboard controls, users can manipulate histological images by translating, rotating, flipping, and scaling them to achieve precise spatial registration with the MSI data, ensuring accurate alignment.

After successfully aligning images, users subsequently move on to the ROI selection procedure. The MSI Tale provides interactive tools within the interface, enabling users to draw boundaries and select ROIs with precision. For marked histological images, the user can select the color that was used to delineate the ROIs. Hence, it is suggested that the ROIs should be marked with conspicuous markers so that the ROIs can be recognized more easily. On the other hand, the ROIs decided in the original histological image will be recognized automatically. Whether through manual delineation or threshold-based selection methods, users navigate a user-friendly interface to identify and extract relevant regions for further analysis.

7. Spectra labeling and exporting

As noise within MSI backgrounds lacks statistical significance or value for analysis, it should exclude all background pixels when further data processing. To effectively extract essential information, the MSI Tale enables the removal of background region (non-tissue region) spectra based on the selected histological image. Subsequently, it facilitates the labeling of the mass spectra that correspond to the tissue regions. The process involves labeling regions within ROIs as 1 (colored in red), while the remaining spectra that are on the tissue are labeled as 0 (colored in green) (Figure S5).

Firstly, it is imperative to standardize the size of the histological images, MS images, and ROI mask. Given that the ROI mask is created at the same scale as the selected histological image and covers the regions of ROI, which will be utilized as a criterion of labeling. The scale of histological image should be reverted to match that of the MS image for that each histological information can one-to-one correspond to MS pixel. Subsequently, if there exist the discrepancies between histological image and MS image, the non-overlapping areas will be recognized as background regions. Hence, it is notable that the users should ensure tissue boundaries do not exceed the boundaries of histological image.

Although previously removal of background spectra by recognizing non-overlapping areas in two images, there still exists background spectra since the histological image still contains background. For completing the removal of background spectra, we further utilize RGB (Red, Green, Blue) and HSV (Hue, Saturation, Value) parameters aid in distinguishing between tissue and background based on common tissue staining colors in histological images. Pixels with saturation below 10% are predominantly grayscale, and when all three RGB parameters fall outside the range of 8% (20/255) to 78% (200/255), they are deemed close to pure black or white. Histological image backgrounds range from pure white to light gray, or potentially pure black. Nevertheless, the complete exclusion of background spectra can be achieved by abiding by these criteria.

Following background spectra exclusion, the remaining spectra will be considered as tissue regions. Given the coverage within the ROI mask as 1, pixels marked as 0 outside the ROI mask. After data labeling, data storage excludes all background pixels, retaining only classification labels (Malignant or Healthy), pixel coordinates (x, y), and intensity information of spectral peaks (Figure S6).

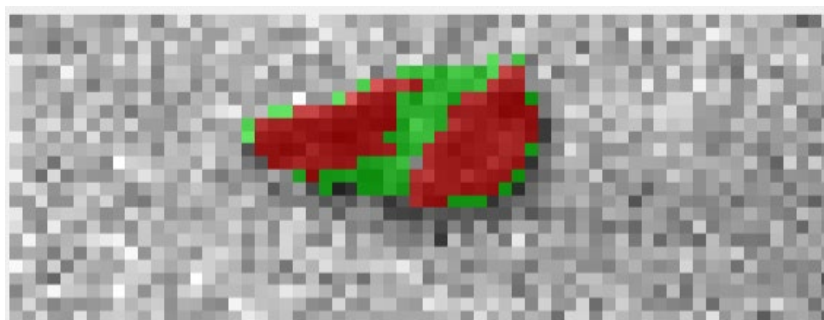


Figure 5 Example of MSI after the removal of background noise. Data in red regions will be labeled as 1 while data in green region will be labeled as 0.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Malignant	x	y	100	200	300	400	500	600	700	800	900
2	1	22	7	248.801991	384.727943	2082.5605	290.080872	318.487293	2748.8476	302.077637	484.521816	2185.75975
3	1	22	8	250.758774	144.313854	1083.41533	369.447132	152.506165	1095.4263	292.791477	188.735767	2323.92299
4	1	22	9	96.1975632	180.942318	876.130554	137.886337	141.387123	598.6339	101.222717	144.679276	561.723995
5	1	22	10	1826.66365	45.0388107	373.156765	1166.28963	294.757938	100.527054	2005.48906	335.225544	474.555038
6	1	23	5	686.250877	134.516052	384.103298	737.938999	503.402199	327.191029	1094.96453	359.098484	142.605904
7	1	23	6	299.816601	194.045521	193.465374	320.997349	55.5688553	292.710495	547.198441	173.488144	244.079693
8	1	23	7	193.30529	1411.96276	374.492451	42.3771477	1510.34031	48.1548271	104.825974	2128.59462	51.1893463
9	1	23	8	248.32724	1267.81958	159.275017	99.783062	880.133976	247.118015	463.163616	1070.80742	162.790798
10	1	23	9	217.481365	920.663723	237.447788	190.379925	648.796127	295.708057	157.508503	450.3783	0
11	1	23	10	372.883701	221.056187	1950.29418	409.808609	337.158749	2595.33256	381.061562	279.431862	1607.18184
12	1	23	11	200.589966	289.111019	1118.71894	272.945339	477.878506	945.010838	301.643841	280.056454	1263.53833
13	1	23	12	106.012032	110.189449	503.209652	267.890148	72.9099426	681.187527	194.308624	196.279308	545.051723
14	1	24	6	240.80476	205.570484	426.949867	49.5332909	141.497334	231.358532	144.596283	255.586025	698.055016
15	1	24	7	1923.61821	207.121265	189.464394	1187.85707	146.633335	249.315239	2327.92321	93.4511414	475.890579
16	1	24	8	880.891548	488.216606	336.97147	928.761135	153.63266	322.076534	1159.53355	55.2626038	465.902401
17	1	24	9	306.591637	154.692921	207.422413	564.040173	0	145.227131	550.291424	172.524509	382.749615
18	1	24	10	191.943535	1556.68618	103.785488	280.424908	1840.10735	209.116188	194.646935	1707.96099	247.303673
19	1	24	11	94.0761223	734.240669	49.6790924	260.117085	670.337223	288.539528	91.8868752	1001.76148	306.049423
20	1	24	12	51.7755661	570.280251	44.9409943	238.101738	490.56498	54.1112175	97.0770226	671.631985	48.6058006
21	1	24	13	149.588596	405.020386	1405.84032	144.089504	422.075569	2060.87204	205.155285	384.976608	1838.06083
22	1	25	6	253.008274	308.301792	219.820129	0	285.632774	398.361912	211.82579	586.589905	148.455555
23	1	25	7	147.614754	383.707859	1355.09316	406.302544	237.479633	2103.6175	198.32159	185.079556	1713.54271
24	1	25	8	358.541386	78.306572	1049.09648	288.544727	150.987518	4569.01093	200.342644	305.882874	2416.0997
25	1	25	9	142.211735	170.265251	679.202999	52.0607758	77.1791077	3358.34048	98.4812965	144.12114	490.450119
26	1	25	10	1684.99623	157.456123	328.199837	2343.87616	90.2421494	1037.02073	1340.54076	397.833889	457.079788

Figure 6 Format of abundance data extracted from the tissue region to an Excel file.