Supporting Information

Materials and Methods:

Tissue Samples and Population:

The data investigated in this study is from two previous studies that used a histology-defined biomarker workflow to identify biomarkers associated with different clinical phenotypes. The first sample set consisted of 63 gastric cancer samples obtained from patients who underwent primary surgical resection and were matched to their UICC-pT status (pT=2) and Lauren's classification (intestinal type) (1). This set was used to investigate the effect of intratumor heterogeneity on the overall survival time of patients after surgery. Follow-up data was available for all gastric cancer patients (median overall survival time was 33.1 [0-53.4] months).

The second data set consisted of MSI data from breast cancer samples that were are all from invasive ductal carcinoma. This set was used to investigate metastatic potential and thus patients without (pN0, n=11) and with nodal metastases (matched to pN1, n=21) were involved (2).

The clinicopathological data of both patient series are listed in Supplementary Tables S1 and S2.

MSI Experiments and Data Preprocessing:

MSI experiments and data processing were conducted as described in Balluff *et al* (2). Briefly, MSI was used to obtain spatially-resolved mass spectrometric data from tissue sections of both patient series', in which each pixel has an associated mass spectrum that is representative of the *local* protein content of the tissue. All MSI data acquisition was performed in a black-

randomized manner using in-house established standard operating protocols, which include weekly QC checks, to minimize the impact of any known/unknown source of bias (batch effects) on the measurements and on the subsequent statistical comparisons.

Following MSI data acquisition the intact tissue sections were stained by hematoxylin and eosin, scanned with a digital slide scanner, and co-registered to the MSI data using FlexImaging (BrukerDaltonics, Bremen, Germany). In this manner each tissue section is characterized by a histological image aligned to 2D matrix of position-correlated mass spectra. A virtual-microdissection was then performed in the FlexImaging software to select tumor areas and thereby focus the subsequent analyses on tumor-specific MSI data only. This resulted in 54,833 pixels for the gastric cancer set and in ~~48,426~~ pixels for the breast cancer data set. This tumor-cell specific MSI data of each patient tissue were then read into MATLAB R2012a (MathWorks, Natick, Massachusetts) into a project-specific data cube using spatial offsets to separate the different patient samples. The project data cube contains total-ion-count-normalized (3) MSI data of all samples, with the corresponding mass spectral data (peaks) in the z-dimension, and is available as supplementary information.

Preprocessing MSI images towards m/z values

An automated routine comprising peak detection, mass spectral alignment and peak extraction was applied, which reduced the gastric cancer dataset to 82 protein peaks and the breast cancer data set to 62 peaks. Peak picking was performed on the global basepeak mass spectrum after smoothing, resampling and baseline subtraction, and was performed using an adapted version of the LIMPIC package (4), a package specifically designed for the low resolution peaks detected by linear MALDI TOF analysis of proteins (5). The basepeak spectrum displays the maximum

intensity detected in the entire imaging dataset for every peak and is more effective for detecting peaks with localized expressions (4). Peak areas were extracted from all spectra and this reduced and more computationally manageable representation of a mass spectrum was then placed, based on its original coordinate information, as a pixel into a project-specific data cube (available in supporting information, e.g. breast_cancer_dataset.mat).

Non-Linear Dimensionality Reduction:

The high dimensionality MSI data was reduced to a single (three dimensional) map representation in which every pixel of the MSI image of every patient represents one point. Dimension reduction was performed using t-distributed Stochastic Neighbor Embedding (t-SNE), which emphasizes retaining the distance between neighboring data points at the cost of faithfully representing distances between distant data points, which makes this technique a non-linear embedding. We used the recent fast implementation of t-SNE, the Barnes-Hut Stochastic Neighbor Embedding (4), which has been optimized for the analysis of large datasets (with computational complexity $N \log N$, where N is the number of data points). t-SNE uses a stochastic optimization leading to different maps when running the algorithm multiple times. To produce a fully reproducible map, we run the algorithm first using the default settings as described by van der Maaten (4) with a fixed seed point of zero, which results in the reduced map. Then, we used this result as an initialization map for the second t-SNE run to achieve global convergence. We used the freely available Barnes-Hut implementation for Matlab (http://lvdmaaten.github.io/tsne/), and mapped the data to a 3D space.

Control for systematic biases and batch effects

A quality control check was performed to check whether there are systematic biases in the data due to patient selection or introduced during MALDI MSI sample preparation and data acquisition. In the t-SNE mapped space, points where colored, for example, according to cancer subtype (Figure S7) and its measurement date (Figure S8). The results indicated that there are no prominent batch effects visible in the t-SNE mapped data distributions.

t-SNE image:

The three 3 coordinate axes of the t-SNE mapped space can be converted to colors by interpreting these dimensions as L*a*b* color system dimensions (5). Every data point (i.e. every pixel in the MSI data) can then be assigned a L*a*b* color. Hence, the position of a data point (pixel) in the t-SNE mapped can be mapped back to the imaging domain by coloring the pixels in the imaging domain according to their corresponding space L*a*b* color. This image we call the t-SNE colored image. We chose for the L*a*b* color space as it is perceptually linear, i.e. small distances in the map are perceived as small color differences (5). This way, similar colors correspond to similar mass spectra in the t-SNE image.

k-means image:

Similarly to the t-SNE colored image we can assign color to pixels in the MSI image based on the cluster to which they belong. This we denote by the *k*-means image.

Optimizing the number of clusters: spatially mapped t-SNE:

The t-SNE colored image can be considered a tissue map in which similar colors represent MSI pixels with similar mass spectra. Similarly, the k-means image represents the clusters to which pixels belong. To choose for the optimal number of clusters we require that both images should be similar, i.e. the clustered representation faithfully represents the similarity of pixels inside the cluster and dissimilarity between pixels in neighboring clusters. Instead of directly comparing the t-SNE colored image with the *k*-means image, we chose to first detect edges in these images and then compare the edge images. The reasons were to overcome the difficulty of comparing continuous (t-SNE colored image) and discrete values (k-means image), and to focus on the cluster boundaries. To detect edges in the images we used the Canny edge detector (6). Before comparing the two images we applied a low pass filter (filter of size 3*3) to reduce pixilation (7). The agreement between the two images was then determined by calculating the Pearson correlation between the vector forms of the edge representations of the two images. To find the optimal number of clusters k was varied from $k = 2$ to 20, and the value with the highest correlation selected. To test whether the method did not oversegment phenotypically homogenous regions (which would then lead to false positive identification of subpopulations) we constructed synthetic datasets of homogeneous patches of real MALDI MSI spectra, with added experimental noise (see SI Appendix fig.S15). The synthetic dataset provides a known ground truth but simulates as close as possible the spectra obtained from real biological specimens. These results demonstrated that the method finds the "ground truth" number and distribution of clusters, without over-segmenting spectrally homogeneous patches.

Association of clinical outcome with clusters:

The clinical relevance for each of the identified clusters (tumor subpopulations) was assessed by investigating their association with the clinical data of the patients. Each cluster contains pixels (data points) from different patients, and a patient's MSI dataset may contain more than one subpopulation. We associated a patient to a cluster when that cluster represented more pixels than would be expected by randomly assigning patient's pixels to the k different clusters, *i.e.* a patient is assigned to a cluster when more than $(1/k)\times 100\%$ of the pixels of that patient's MSI data are assigned to that cluster (2). In this manner every cluster has list of patients that are assigned to it, and which can be used to investigate associations with patient survival (gastric cancer) and metastasis status (breast cancer). Survival analysis was performed using the Kaplan-Meier method (8) to evaluate differences in survival times. Associations with the metastasis status were investigated using Fisher's exact test. These statistical analysis were performed using the R statistical software environment (R Foundation for Statistical Computing, Vienna, Austria).

Building the pixel classifier:

Given a set of patients we built a pixel classifier as follows. First, a spatially mapped t-SNE was run on the training set resulting in mapped data points and k clusters. Each cluster was assigned a label by associating patients to the cluster and determining clinical outcome. A significance Analysis of Microarrays (SAM) was then used to identify which m/z features significantly differentiate between the detected clusters (9). Significant features were determined by setting the FDR cutoff at 0.001. A new training set is then built in which every pixel has an outcome label based on the label of the cluster to which it was assigned and a feature vector based on the selected m/z features obtained by the SAM analysis. On this derived training set we built a k-Nearest Neighbor classifier with k=5. New (unlabeled) pixels can then be labeled by searching

for the 5 nearest pixels in the derived training set on the basis of the expression values of the SAM-selected m/z features. A majority vote on the class labels of these 5 nearest pixels in the derived training set then predicts the label of the new (unlabeled) pixel.

Building the patient-based classifier:

The pixel classifier was used to predict the label of every pixel in a previously unseen patient's MSI dataset. To arrive at an outcome prediction for the patient we integrate the pixel-based predictions as follows: if the patient's MSI image has more than *t1* pixels predicted as poor outcome (by the pixel classifier) and less than *t2* pixels predicted as good outcome (by the pixel classifier), we designate poor outcome for that patient (and vice versa for the good outcome).

For the breast cancer data set we used the clinically accepted Youden's index (10), i.e. *t1*=2% and *t2*=100%. Using these thresholds the clinical outcome was then predicted for an unseen patient. The predictions were then checked for clinical concordance by comparing predictions with the clinical ground truth for the patients using the Fisher test.

For the gastric cancer data set the Kaplan Meier analysis revealed the survival characteristics of each cluster; the cluster with greatest survival rate was designated as the favorable prognosis class, the cluster with least survival rate was designated as the poor prognosis class, and all remaining clusters were combined into a medium prognosis class. A patient with poor prognosis was found to be characterized by having *t1* = 10% of pixels that were predicted by the pixel classifier to belong to the poor survival class, and *t2* = 50% of pixels that were predicted to belong to the favorable survival class. Kaplan-Meier survival curves were then constructed of patients predicted to belong to the favorable and poor prognosis groups, and the significance of survival difference was determined using the log rank test.

Leave-one-patient-out validation of pixel and patient-based classifiers

The pixel and patient-based classifiers can be evaluated by comparing predicted labels with known labels of a patient. But we need to evaluate these classifiers in an unbiased way, i.e. information used to build the classifiers should not be shared with the evaluation (such as building the spatially-mapped t-SNE, the SAM selection or building the kNN classifier). To achieve this we made use of a Leave-One-Patient-Out (LOPO) cross validation procedure (see SI Appendix Figure S9). We removed the data of a single patient from the data set representing the complete cohort. The data of the remaining patients was then considered training data and used to build the pixel-based and patient-based classifiers, *i.e.* all aspects of building the classifiers were performed on the training data only, and did not utilize any information from the left-out patient. The performance of both classifiers can then safely be determined on the left-out patient in an unbiased way. We repeated this procedure of leaving one patient out of the cohort data set for all patients (thus every patient is left out once, training as many classifiers as there are patients). The performance of the pixel and patient-based classifiers can then be derived from the average performance over all left out patients.

	Gastric carcinoma	Breast carcinoma
Number of patients	63	32
Primary tumor extension		
pT1	0	13
pT2	63	13
pT3	0	2
pT4	0	4
Regional lymph nodes metastasis		
pN0	18	11
pN1	24	21
pN2	16	0
pN3	5	0
Resection status		
R0	53	28
R1	9	1
Rx	1	3
Distant metastasis		
M0	54	32
M1	9	0

Table S1 Clinicopathological information for the patient series

Her2_positive	**Luminal_A**	**Luminal_B**	**Triple_Negative**
1	23	5	2
3%	74%	16%	6%

Table S2. Clinical data of the breast cancer dataset

k=3	Supopulation2	Supopulation3
Supopulation1	0.0215*	0.4570
Supopulation2	-	0.0868

Table S3.1. p-values of comparing survivals of different subpopulations in the gastric cancer dataset for k =3; numbers refer to data shown in the top row of Figure S3. * indicates $p<0.05$.

k = 6	Cluster#2	Cluster#3	Cluster#4	Cluster#5	Cluster#6
Cluster#1	0. 3800	0. 2630	0.1920	0.3610	0.0771
Cluster#2	-	0.0719	0.0351	0.0740	0.0084*
Cluster#3	0.0719	-	0.8150	0.6620	0.7190
Cluster#4	0.0351*	0.8150	-	0.5400	0.8210
Cluster#5	0.3610	0.6620	0.5400	-	0.3160

Table S3.2. p-values of comparing survivals of different subpopulations in the gastric cancer dataset for k = 6; numbers refer to data shown in the middle row of Figure S3. * indicates $p<0.05$.

k = 8	Cluster#2	Cluster#3	Cluster#4	Cluster#5	Cluster#6	Cluster#7	Cluster#8
Cluster#1	0.3400	0.9710	0.3990	0.3900	0.6200	0.5560	0.1840
Cluster#2	-	0.2570	0.0714	0.0246*	0.1110	0.0892	0.0050*
Cluster#3	0.2570	-	0.3940	0.2720	0.5570	0.5940	0.1430
Cluster#4	0.0714	0.3940	-	0.9650	0.9040	0.6630	0.6860
Cluster#5	0.0246*	0.2720	0.9650	-	0.7510	0.6320	0.5980
Cluster#6	0.1110	0.5570	0.9040	0.7510	-	0.8650	0.4980
Cluster#7	0.0892	0.5940	0.6630	0.6320	0.8650	-	0.4980

Table S3.3. p-values of comparing survivals of different subpopulations in the gastric cancer dataset for k = 8; numbers refer to data shown in the bottom row of Figure S3. * indicates $p<0.05$.

Metastasis subpopulation	Patient without metastasis	Patient with metastasis
<= 2%	5	2
> 2%	6	19
Fisher test: p-value = 0.0318		

Table S4. Clinical evaluation of LOPO cross validation results for breast cancer patients. A threshold-based classifier on the pixel contribution of the metastasis-associated subpopulation was created using Youden's index (Youden, Cancer, 1950) to find the optimum threshold (2%). This threshold was then applied to each unknown patient of each LOPO run and the predictions were tested for concordance with the ground truth using the Fisher test.

References:

1. Balluff B*, et al.* (2011) MALDI Imaging Identifies Prognostic Seven-Protein Signature of Novel Tissue Markers in Intestinal-Type Gastric Cancer. *Am. J. Pathol.* 179:2720-2729.
2. Balluff B*, et al.* (2015) De novo discovery of phenotypic intra-tumor heterogeneity using imaging mass spectrometry. *J. Pathol.* 235:3-13
3. Deininger SO*, et al.* (2011) Normalization in MALDI-TOF imaging datasets of proteins: practical considerations. *Anal. Bioanal. Chem.* 401:167-181.
4. van der Maaten L (2014) Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* 15:3221-3245.
5. Hunter RS (1948) Photoelectric Color-Difference Meter. *JOSA* 38:661.
6. Canny J (1986) A Computational Approach To Edge Detection. *IEEE Trans. Pattern Analysis* 8:679-714.
7. McDonnell LA, van Remoortere A, van Zeijl RJM, & Deelder AM (2008) Mass Spectrometry Image Correlation: Quantifying Co-Localization. *J. Proteome Res.* 7:3619-3627.
8. Kaplan EL & Meier P (1958) Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assn.* 53:457-481.
9. Tusher VG, Tibshirani R, & Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98(9):5116-5121.
10. Youden WJ (1950) Index for Rating Diagnostic Tests. *Cancer* 3(1):32-35.

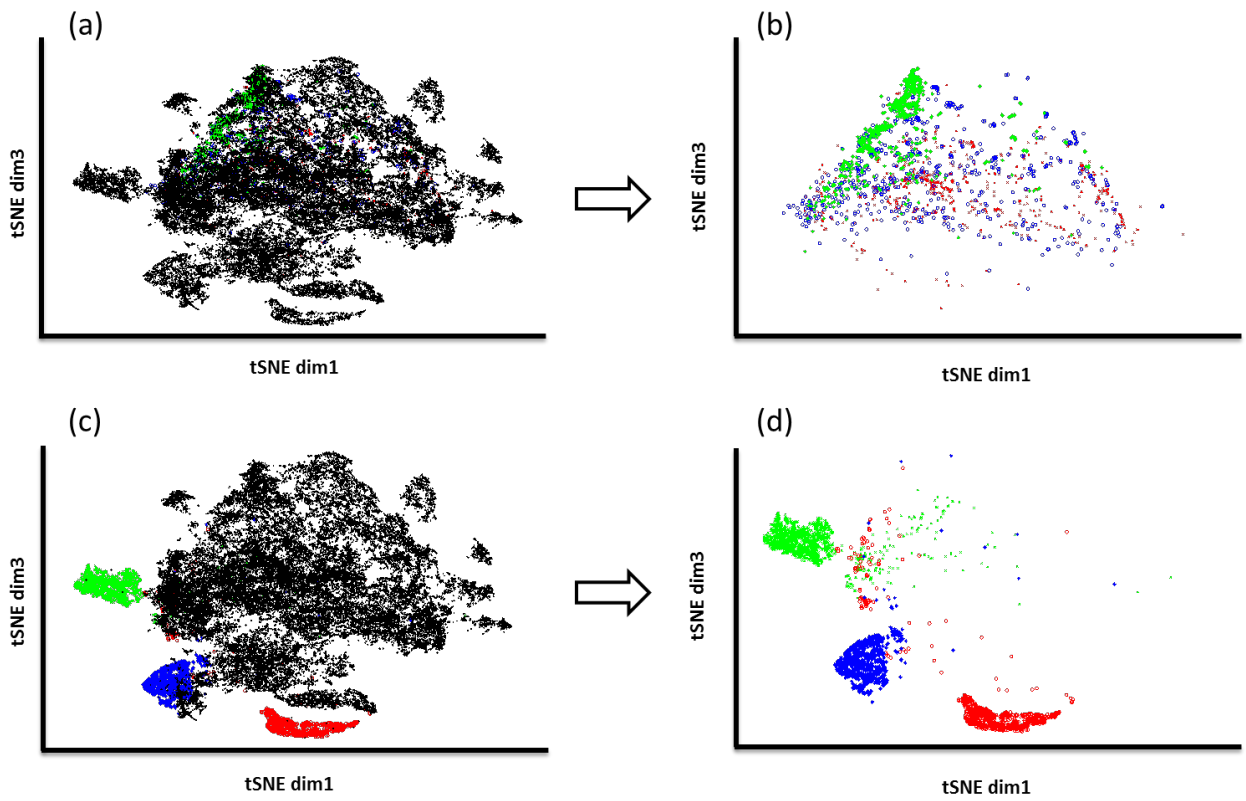


Figure S1. Visualization of the contributions to the molecular heterogeneity in the gastric cancer MSI data due to intratumor heterogeneity (top row) and patient variability (bottom row). In (a) the data points from one patient have been colored green, another blue, and another red; the data from the remaining 60 patients are colored black. When the data from the remaining 60 are made invisible (b) it can be seen that the molecular signals from some tumors are highly dispersed indicating a high intratumor heterogeneity. Figures (c) and (d) show the same plots but for patients that are less heterogeneous and whose MSI signatures were found to be quite distinct (inter-patient heterogeneity).

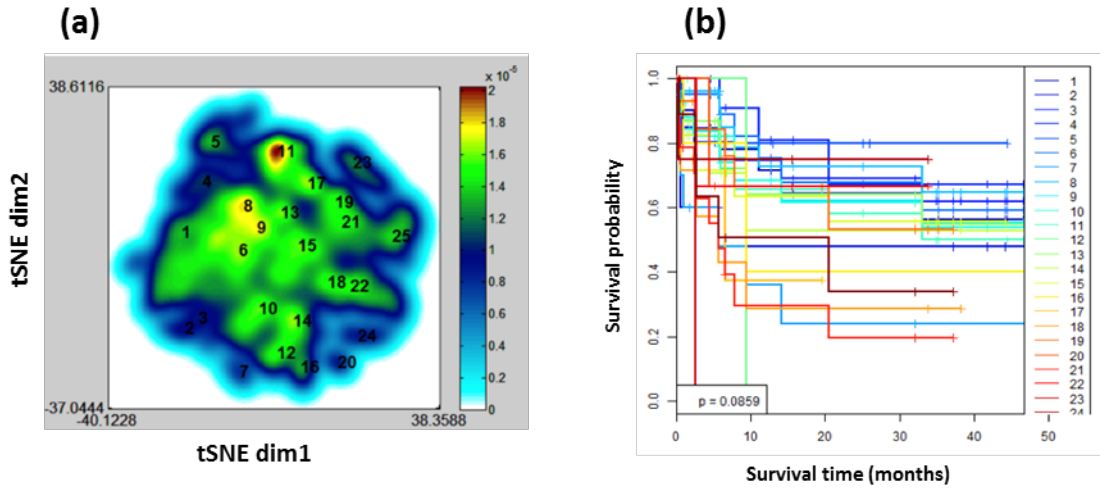


Figure S2. ACCENSE, a density-based analysis of the data points in the t-SNE space found 25 clusters (a); while their corresponding Kaplan-Maier curves, (b), displayed apparent trends the low number of patients contributing to each cluster led to the results lacking statistical significance.

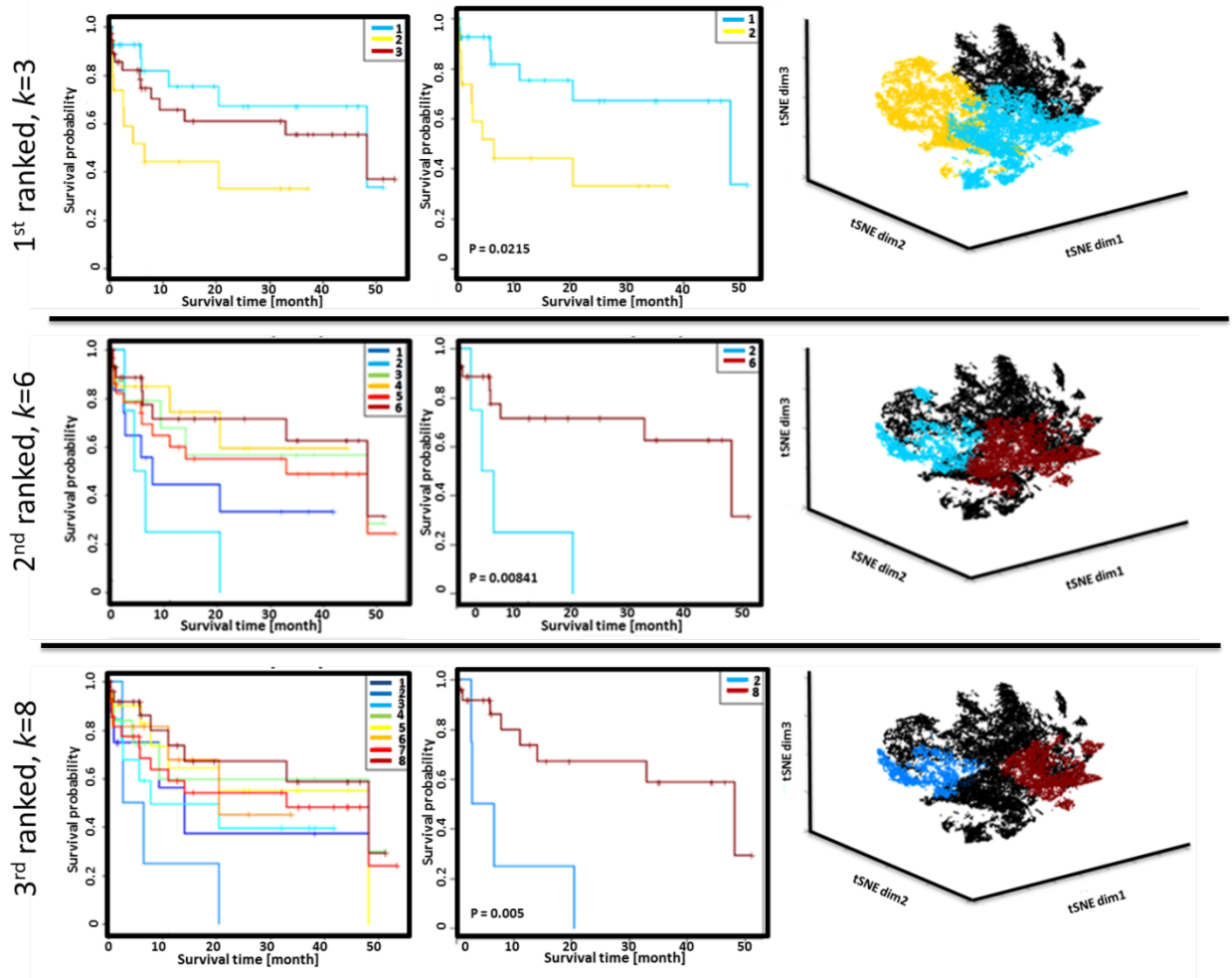


Figure S3. Kaplan-Meier survival analysis of the clusters found for the three highest ranking values of k for the gastric cancer MSI dataset: k = 3, top row; k = 6, middle row; k = 8, bottom row. There are significant differences in survival associated with clusters 1 and 2 (k = 3, top row); clusters 2, 4, and 6 (k = 6, middle row); and cluster 2, 5, and 8 (k = 8, bottom row).

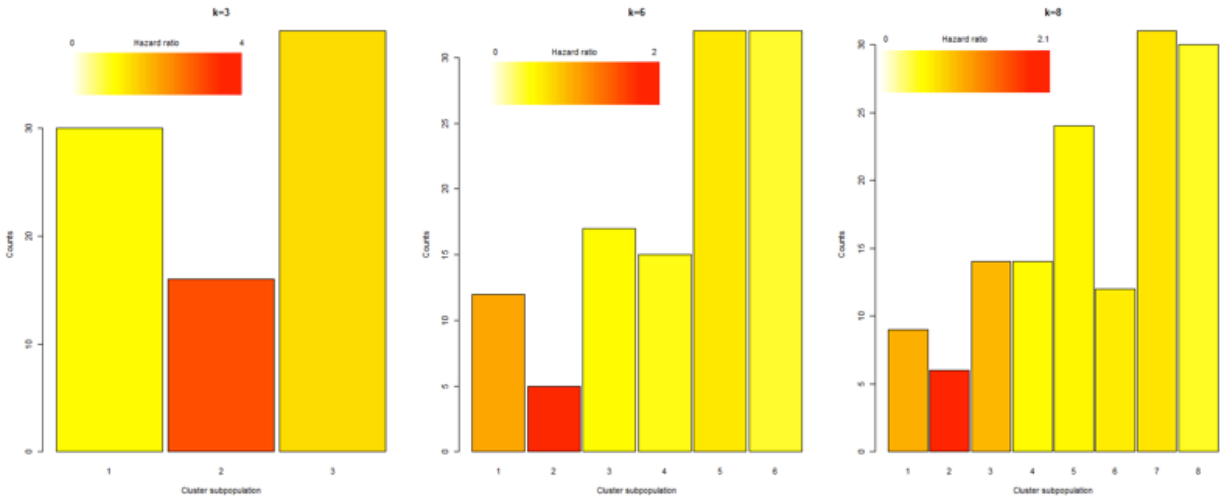


Figure S4. Survival analysis of all subpopulations in the gastric cancer dataset for the *k*-means cluster representations that exhibited the highest gradient correlations to the t-SNE map. The results are presented as bar plots showing the number of patients contributing to the different subpopulations, in which the bar is colored according to the hazard ratio (survival) of the patient group.

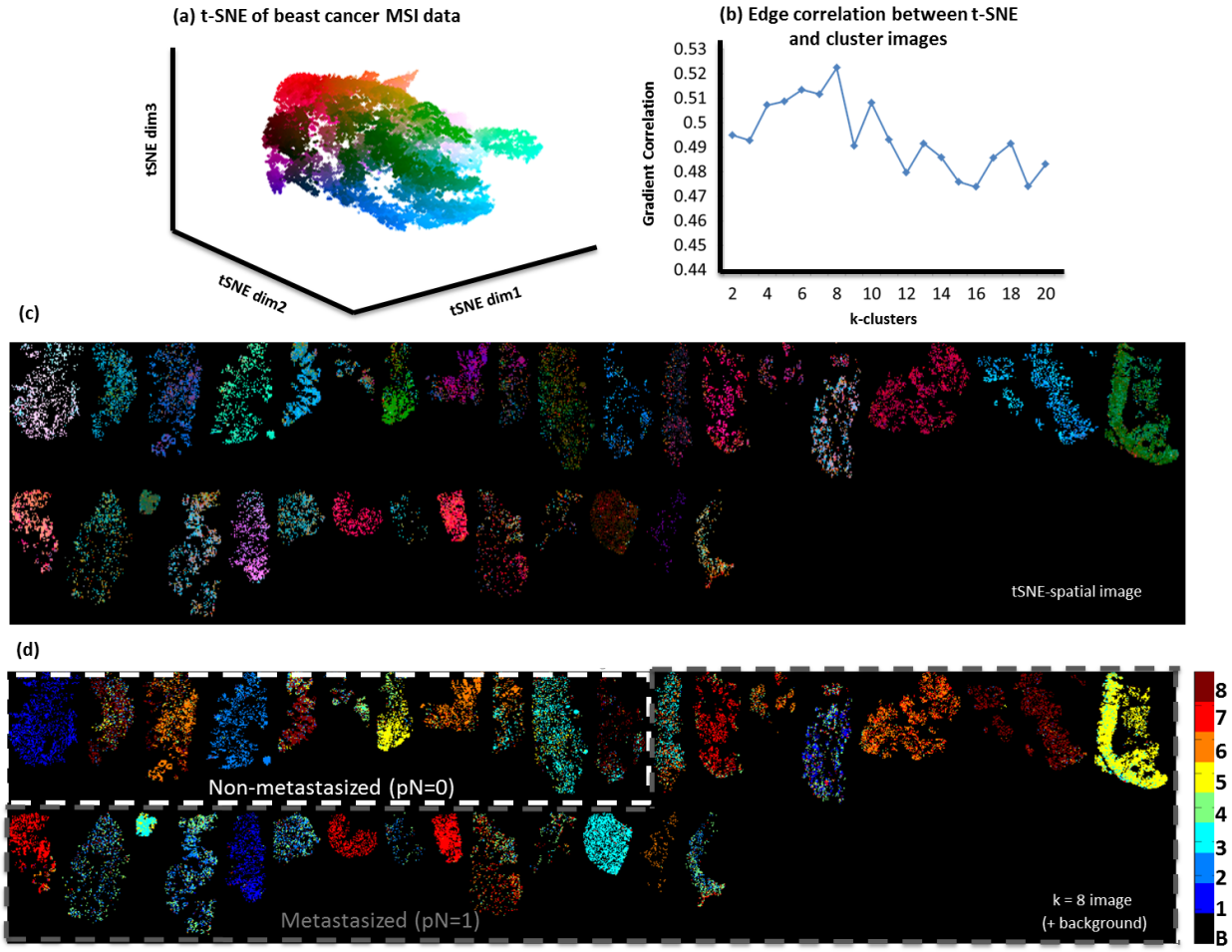


Figure S5. Non-linear visualization of tumor cell specific MSI data from 32 breast cancer patients using t-SNE (a). An edge-based image correlation is then used to determine the discrete representation with the highest correlation, (b), here k=8. The t-SNE image, (c), is formed by coloring each pixel according to its location in the t-SNE space. The k=8 discrete approximation of the 32-tumor sample t-SNE image is shown in (d).

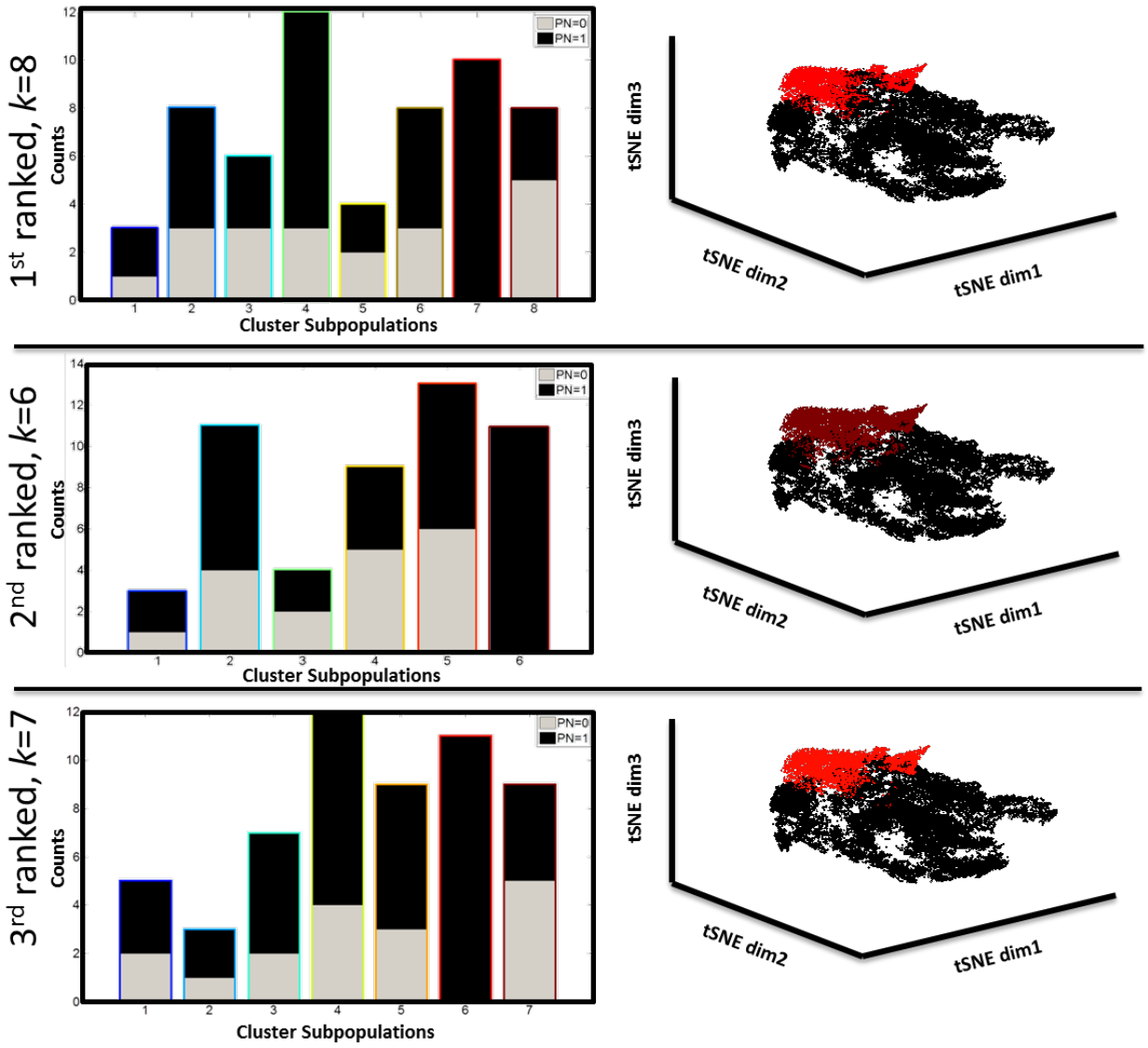


Figure S6. Visualization of the metastasis-associated subpopulations in the breast cancer MSI data as revealed by t-SNE. The top-row (k = 8) shows the contributions of metastatic (black) and non-metastatic patients (grey) to the eight clusters in a grouped histogram. A statistical analysis detected cluster subpopulation 7 to be exclusively associated with metastasis. This subpopulation is highlighted in red in the t-SNE scatterplot. The middle and bottom rows show the results for second and third ranked k, namely k = 7 and 6, respectively (re. Figure S5b).

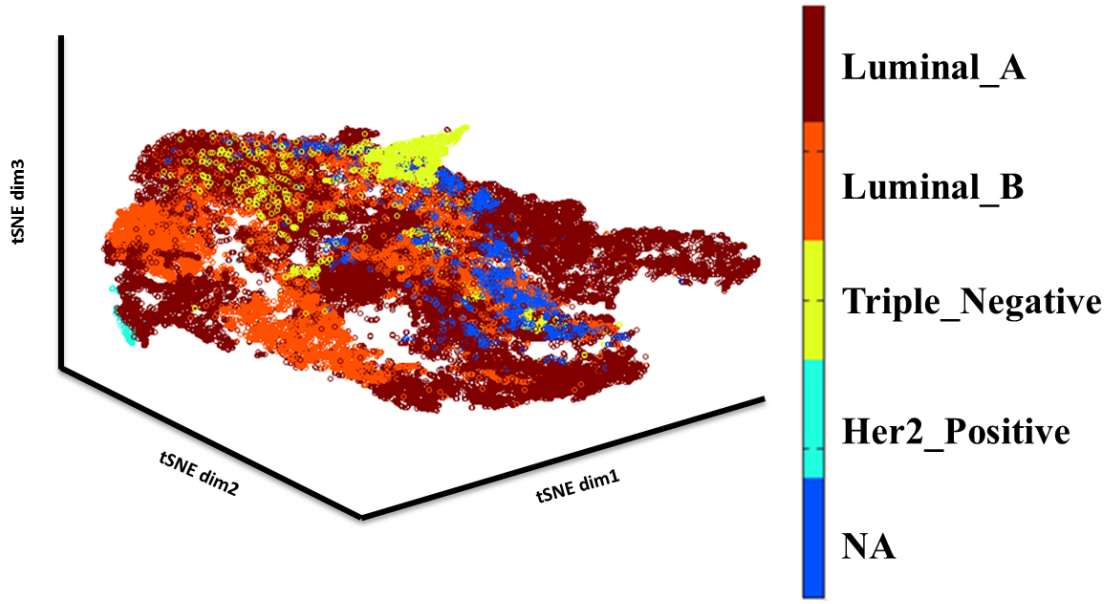


Figure S7. tSNE plot of the breast cancer MALDI MSI data in which the data points are colored based on the clinical data of the breast cancer patients, demonstrating the data structure is not defined by the subtypes.

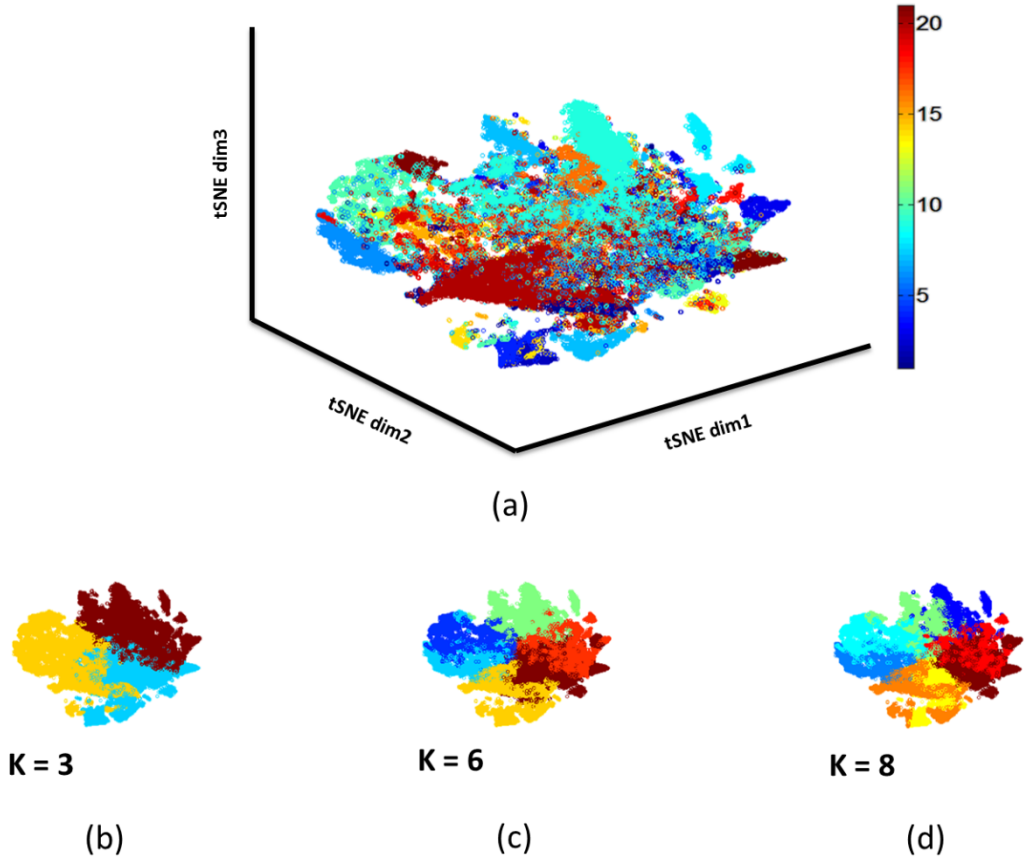


Figure S8. (a) tSNE map is colored based on the MSI data measurement date (labeled simply as 1-20). Clustering analysis with different clustering numbers (k = 3, 6 and 8) showing different clusters hold data from different batches.

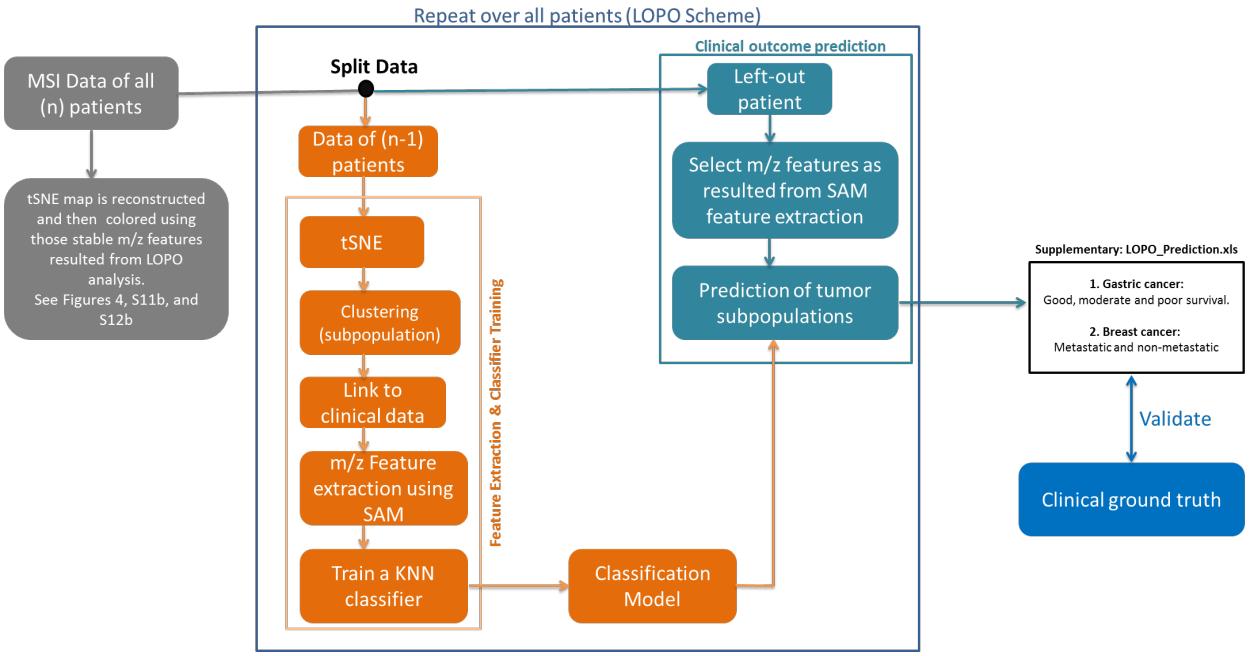


Figure S9. Schematic diagram for the LOPO cross validation in which the left-out patient is not seen during the feature extraction and the classifier training and thus clinical outcome prediction of that unseen patient is unbiased. The LOPO predication results were validated by comparing it with the clinical ground truth(see Figure S10 and Table S4).

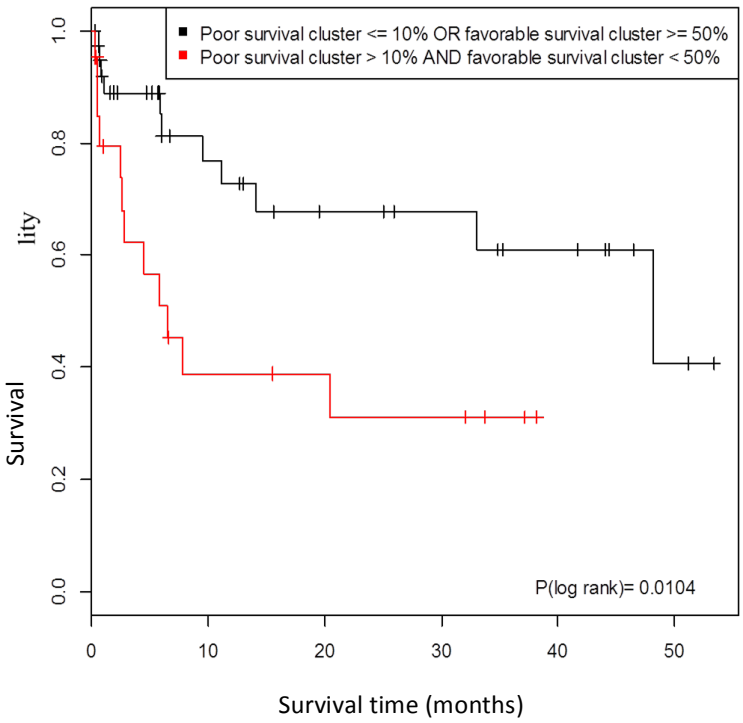


Figure S10. Clinical evaluation of LOPO cross validation results for gastric cancer patients. A rule-based classifier on the pixel contributions was determined, >10% poor survival class and <50% favorable survival class, which were then applied to each totally unknown patient of each LOPO run. The predictions resulted in significant survival time differences as shown by Kaplan-Meier survival analysis and the log rank test (p=0.0104).

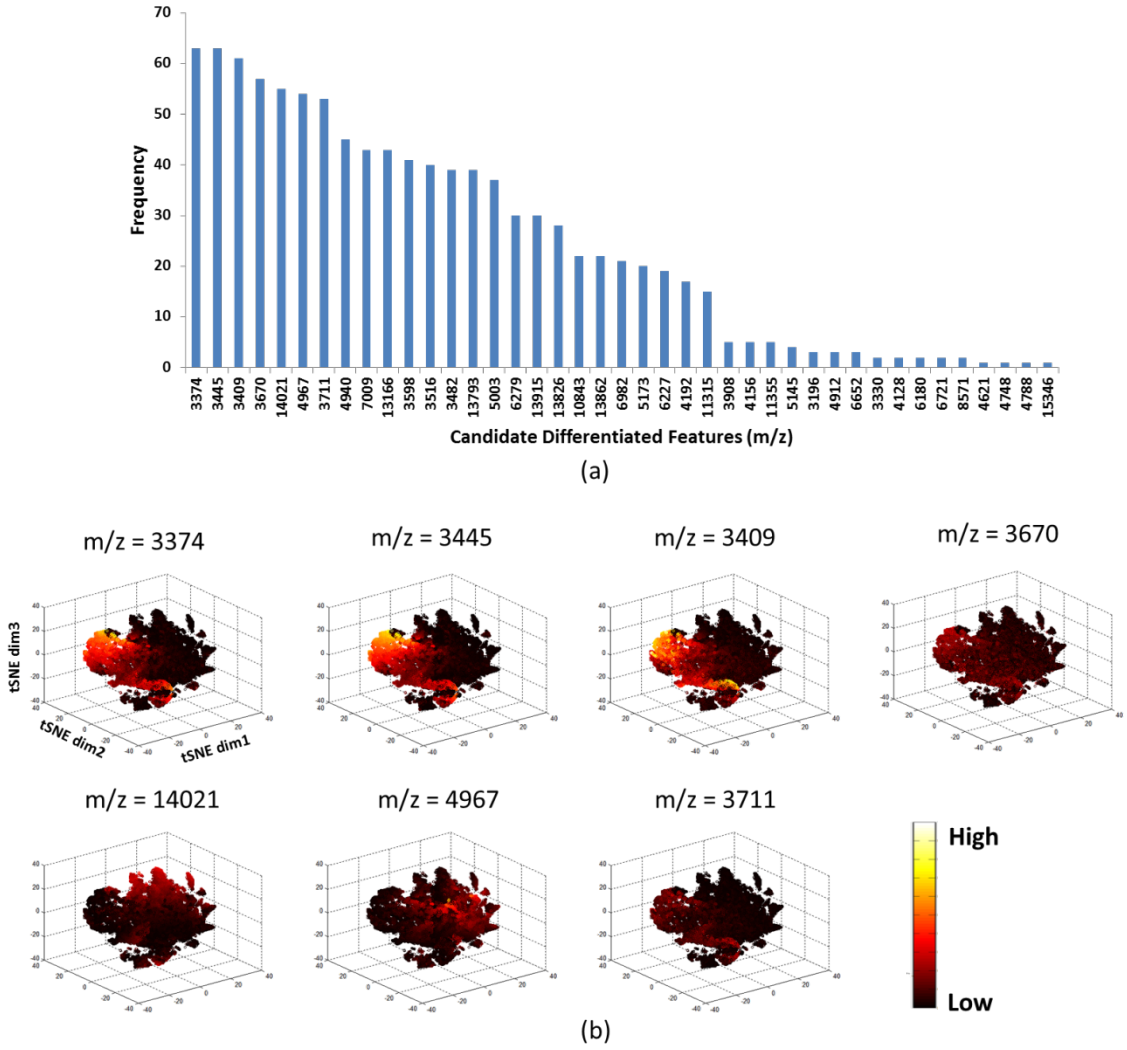


Figure S11. (a) Bar plot showing the frequency of SAM-identified protein ions for the LOPO cross validation performed on MALDI MSI dataset of gastric cancer; (b) Coloring 3D-tSNE map of the full dataset (see Figure S3) with intensity profiles of highly stable prognostic signatures – m/z features that are commonly detected by the SAM analysis in at least 80% of all LOPO experiments. It can be seen that these 7 protein ions preserve the separation between the clinically significant subpopulations, for example m/z features of 3374 and 3445, were detected in 100% of all LOPO runs, are highly discriminating for the poor survival cluster with exclusive over-expression (yellow cluster in top-panel of Figure S3).

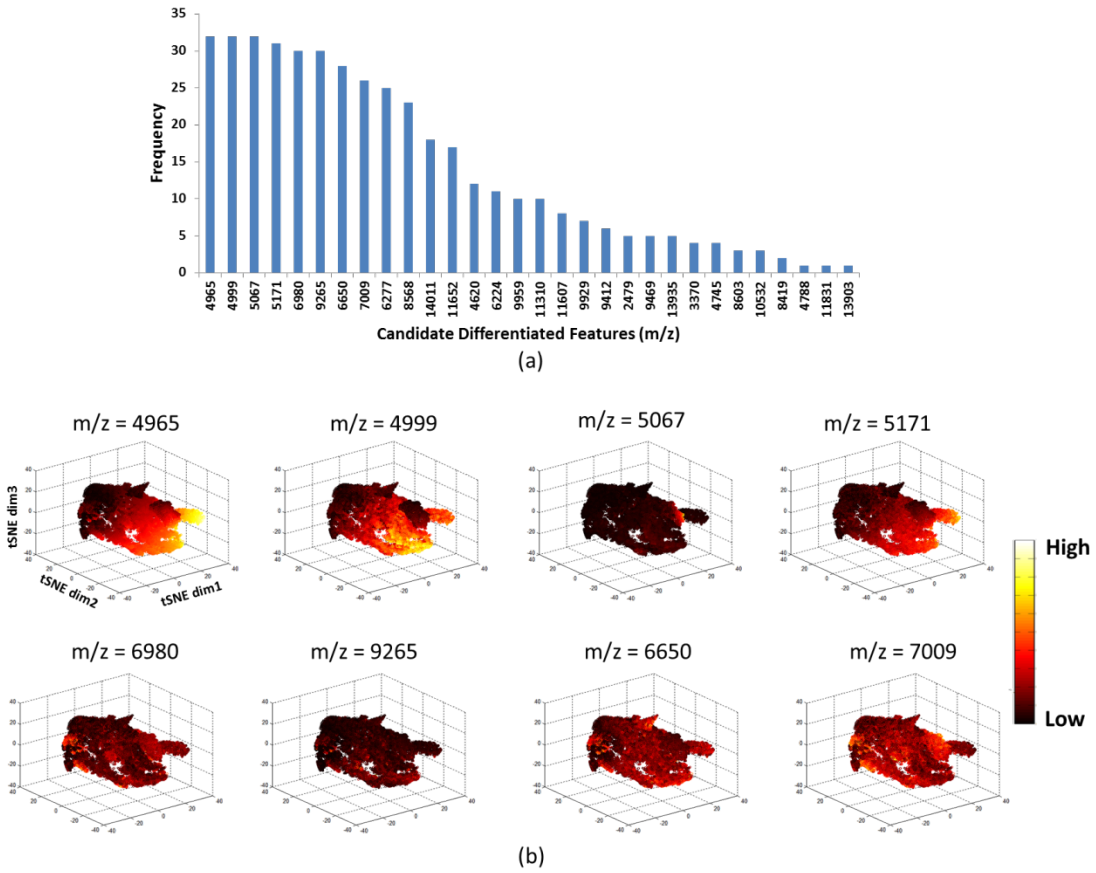
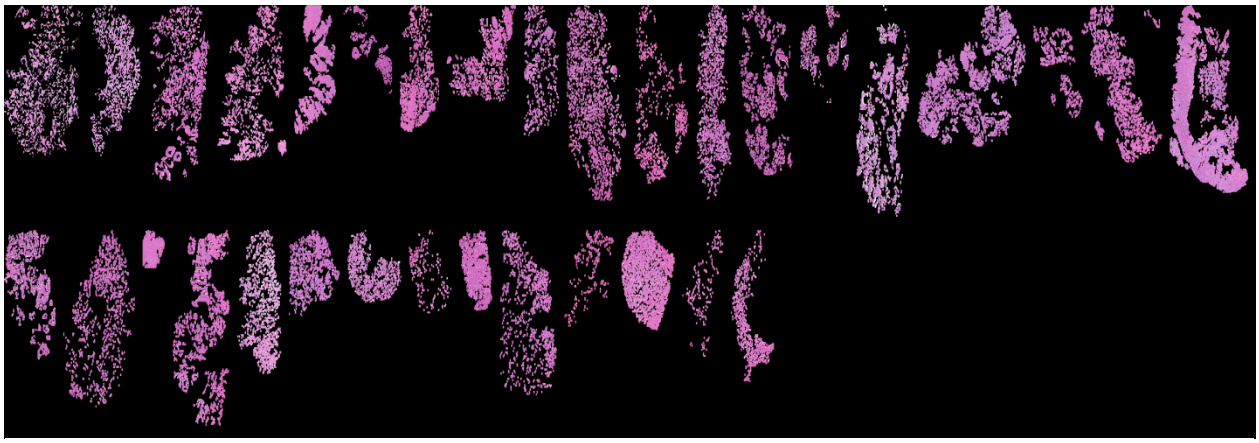
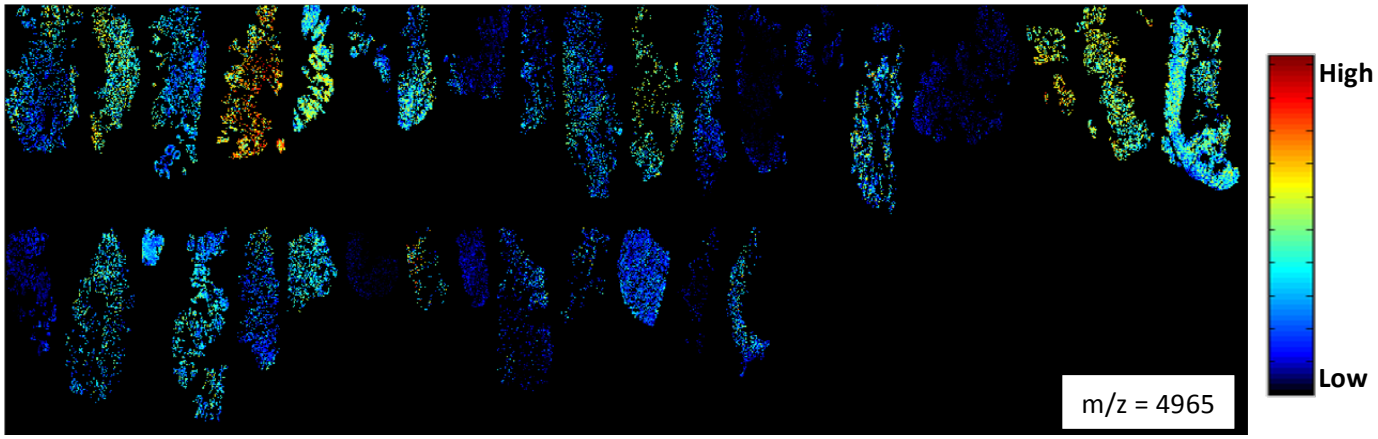


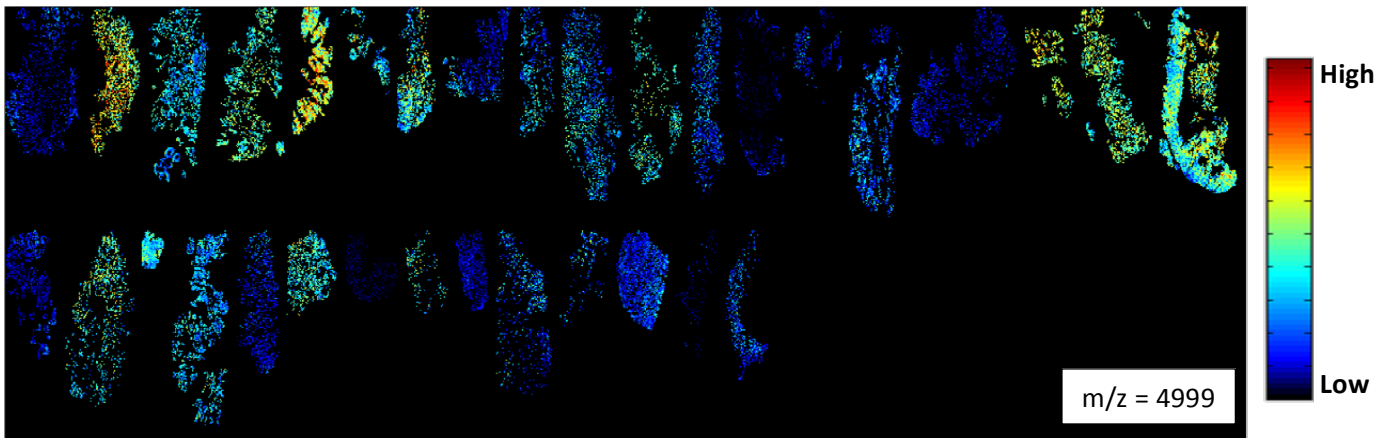
Figure S12. (a) Bar plot showing the frequency of SAM-identified protein ions for the LOPO cross validation performed on MALDI MSI dataset of breast cancer; (b) Coloring 3D-tSNE map of the full dataset (see Figure S6) with intensity profiles of highly stable prognostic signatures – m/z features that are commonly detected by the SAM analysis in at least 80% of all LOPO experiments. It can be seen that these 8 protein ions preserve the separation between the clinically significant subpopulations, for example m/z features of 4965 and 4999, were detected in 100% of all LOPO runs, are highly discriminating for the metastatic exclusive cluster with under-expression profile (see location of metastasis-exclusive cluster in Figure S6).



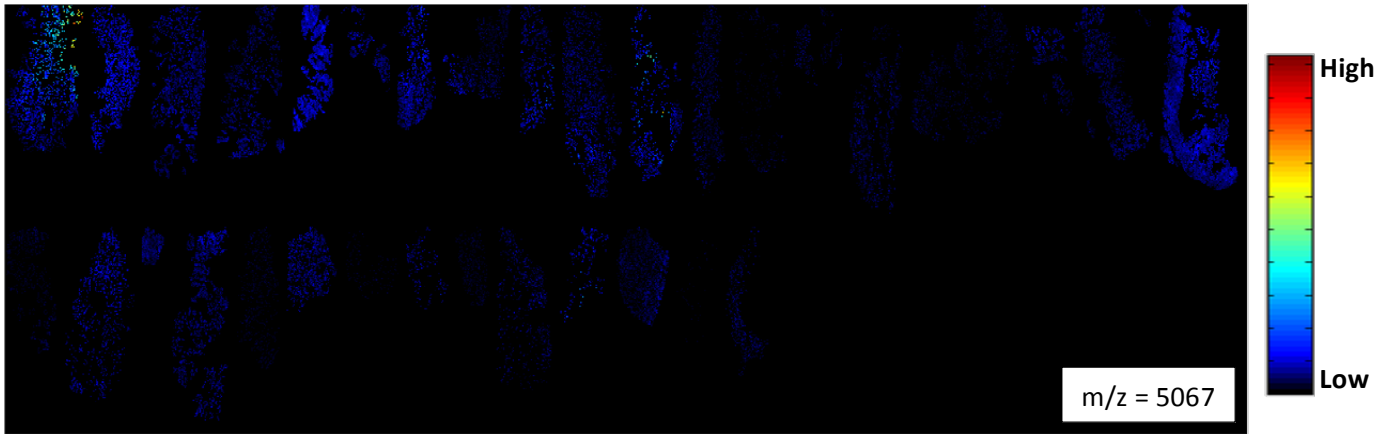
(a)



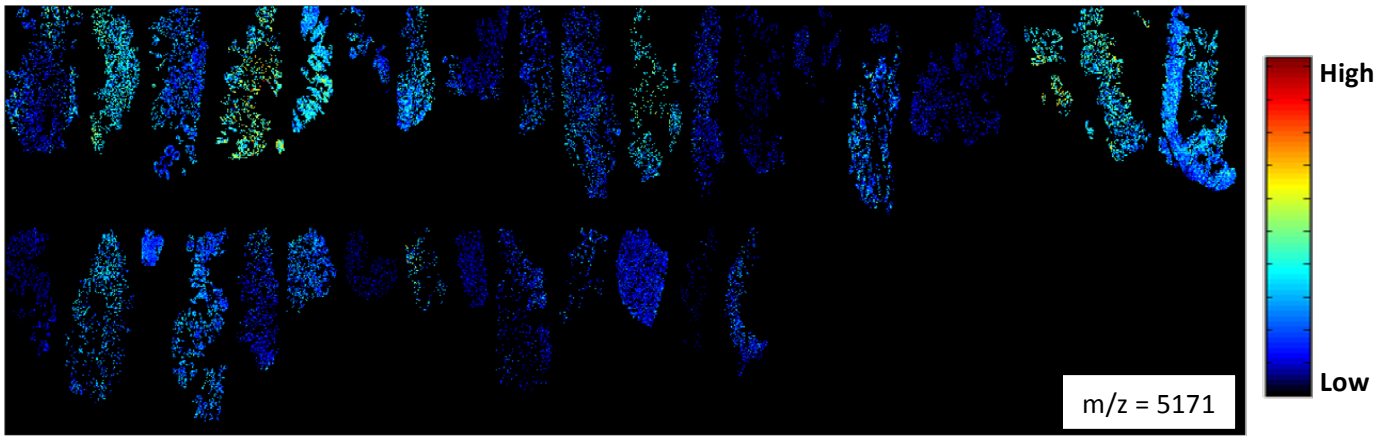
(b)



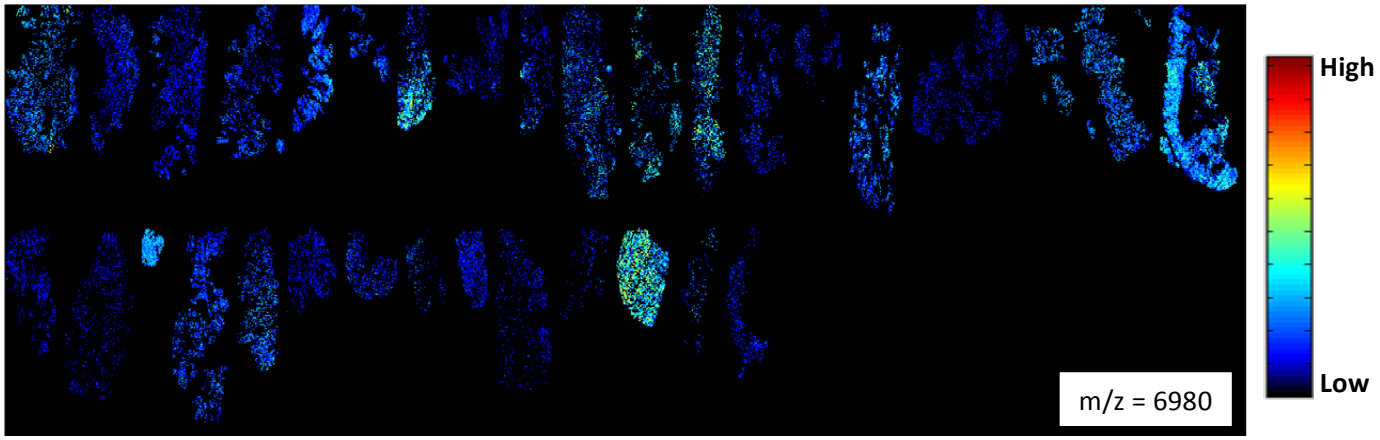
(c)



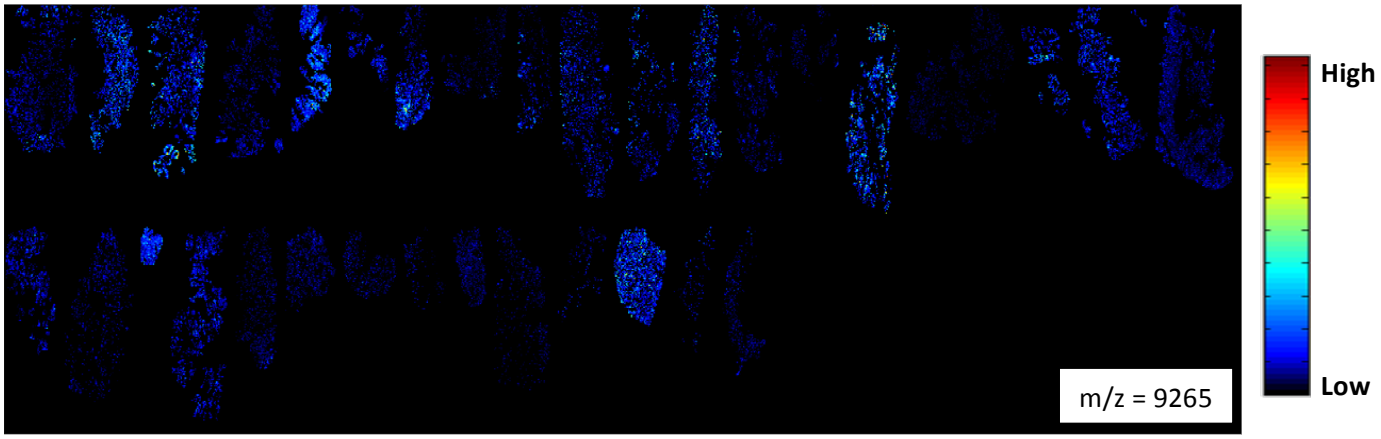
(d)



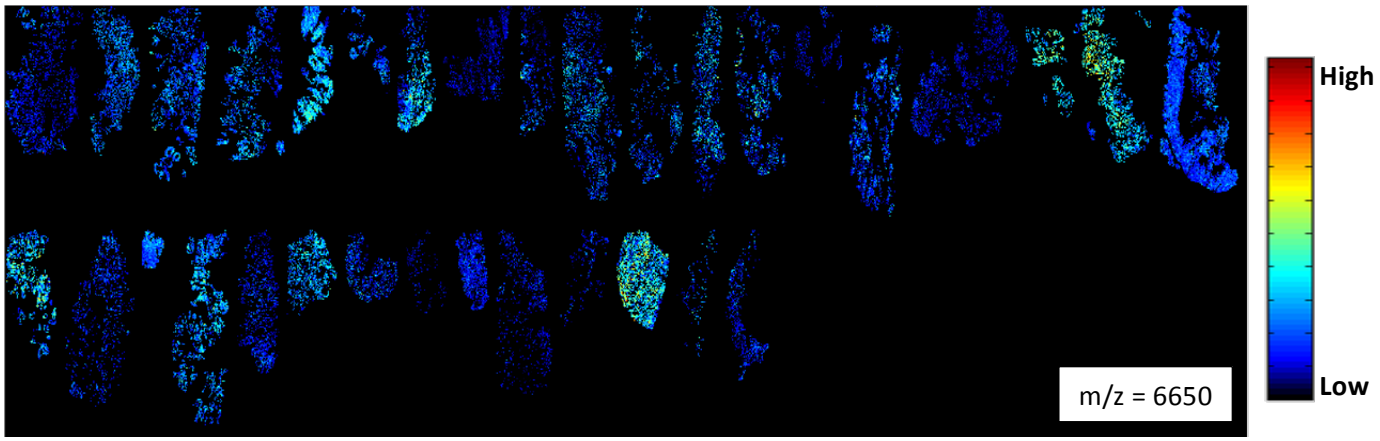
(e)



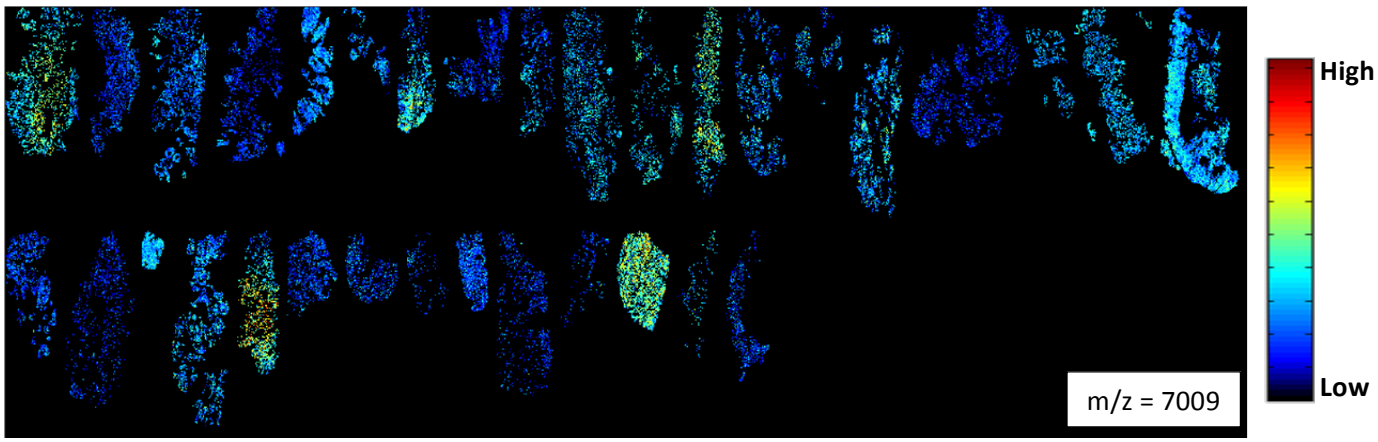
(f)



(g)



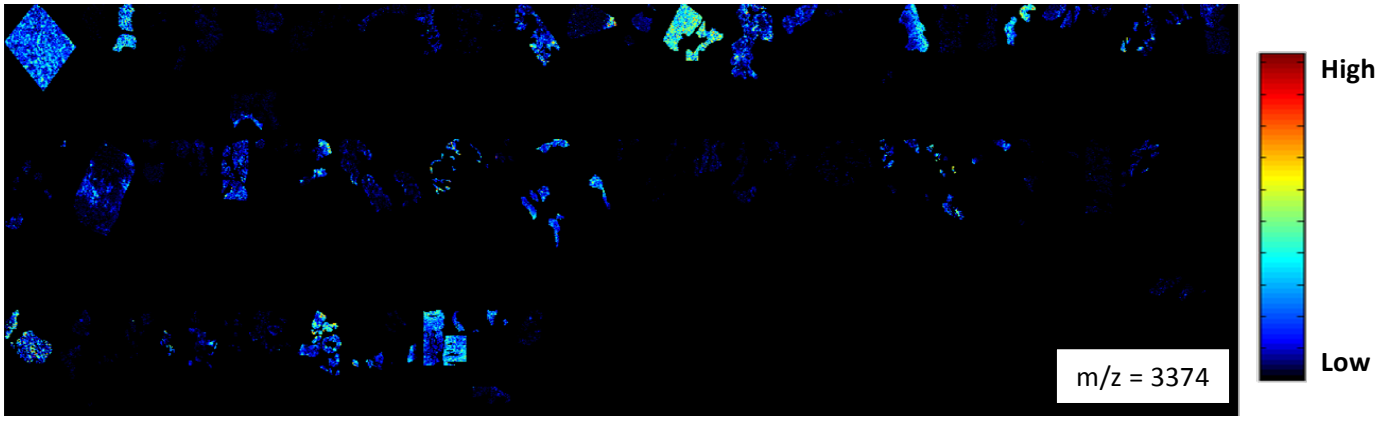
(h)



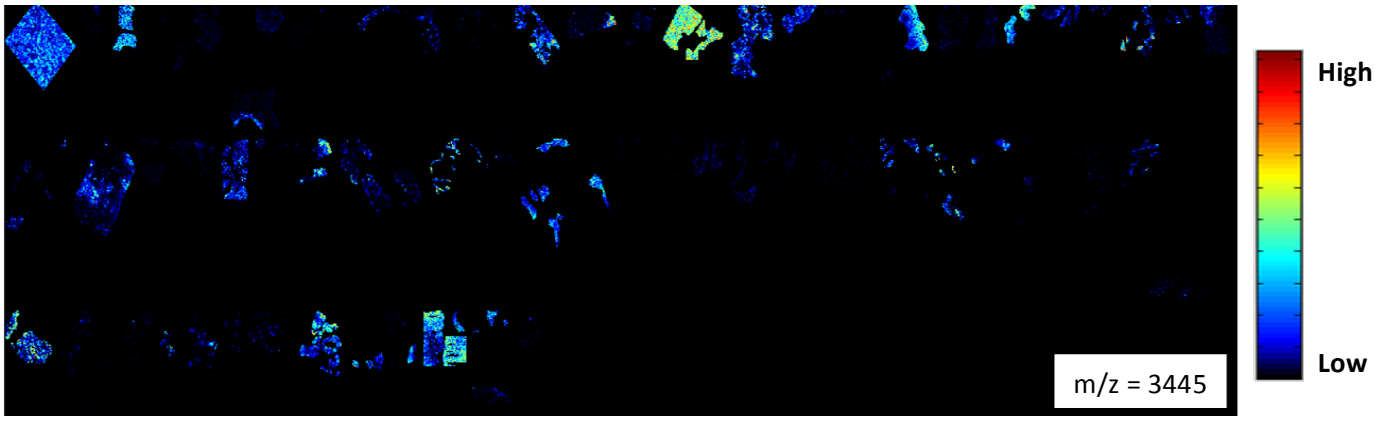
(i)

Figure S13. Breast cancer tissues: H&E stained histology (a), and ion intensity expressions of the potential prognostic m/z features (b-i).

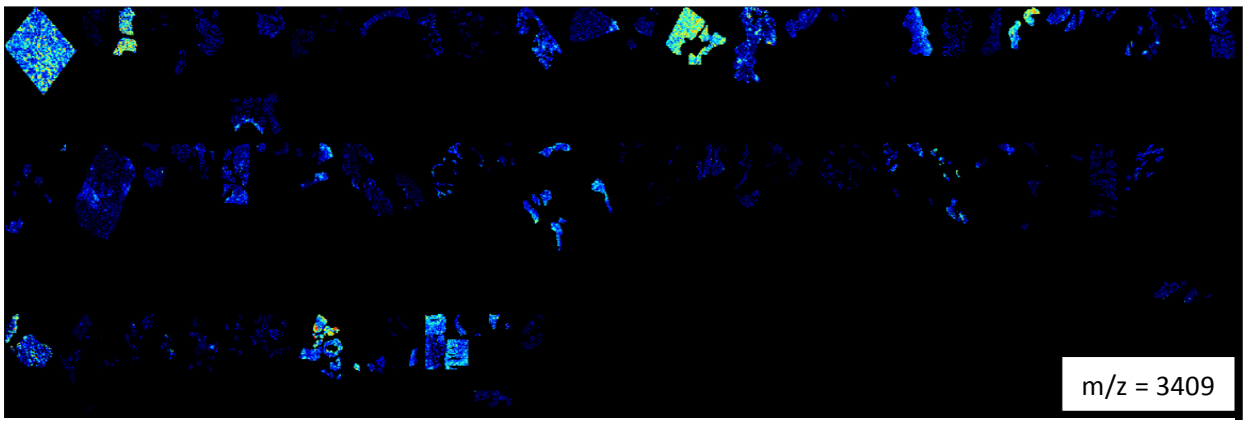
(a)



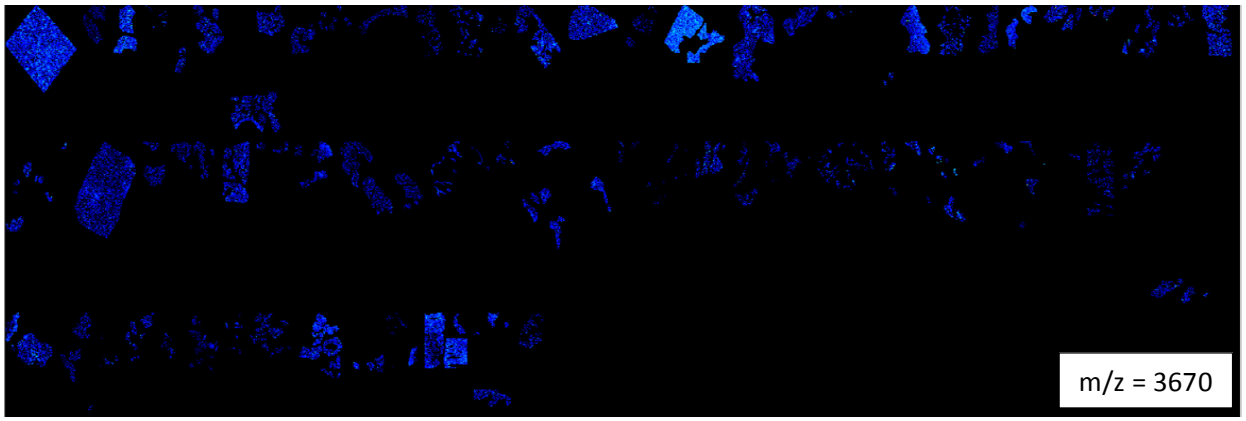
(b)



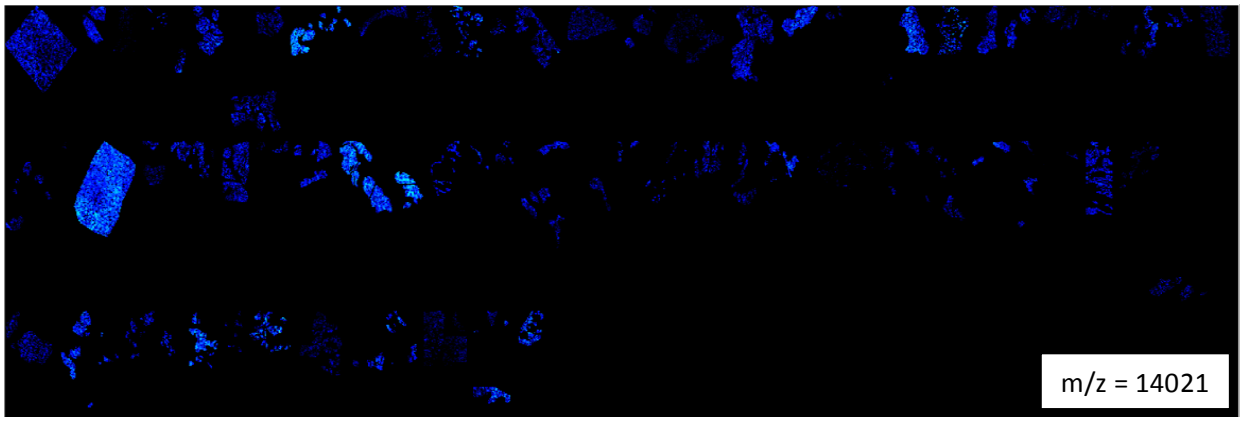
(c)



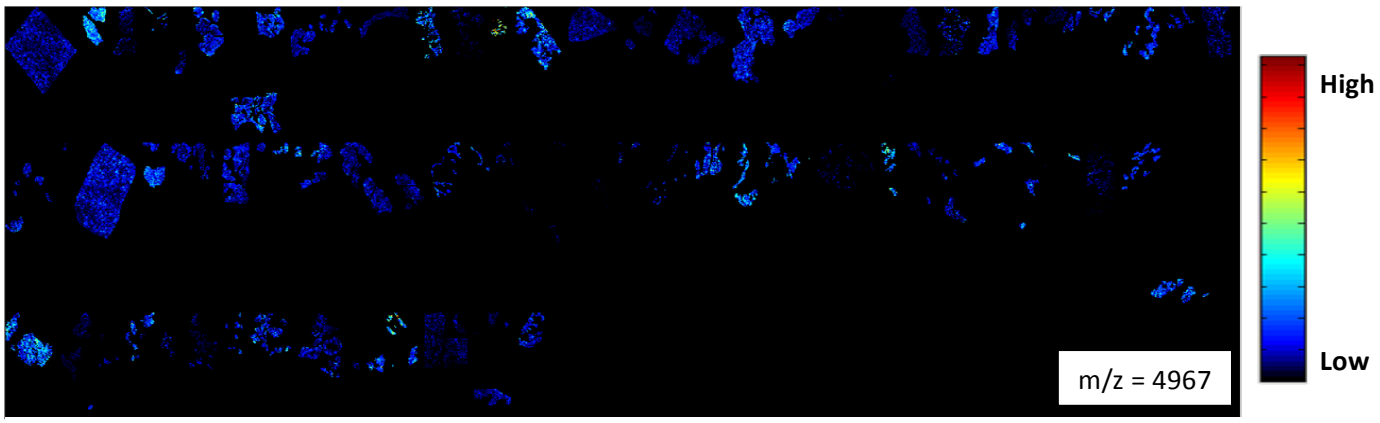
(d)



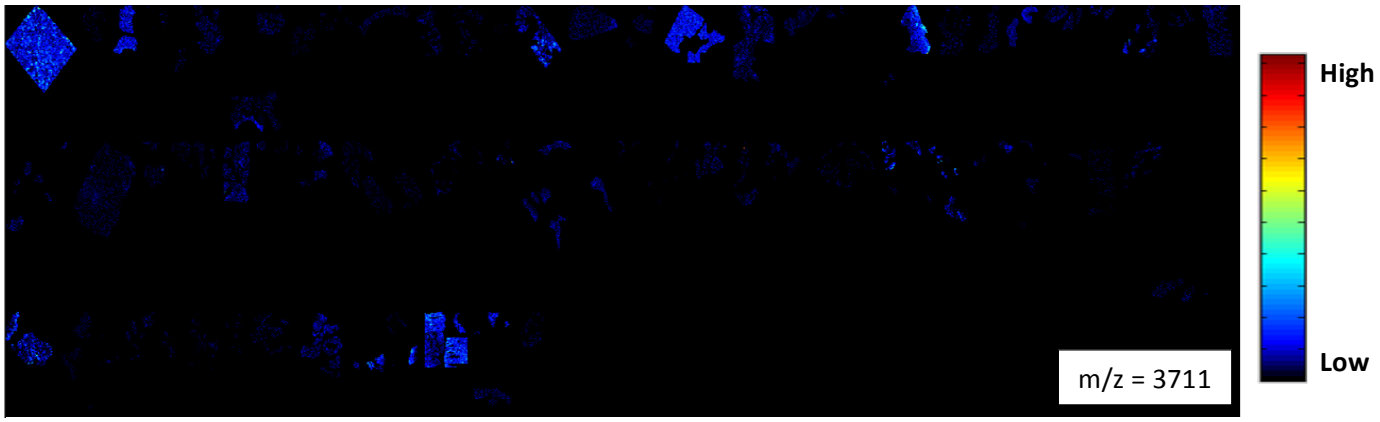
(e)



(f)



(g)



(h)

Figure S14. Gastric cancer tissues: H&E stained histology (a), and ion intensity expressions of the potential prognostic m/z features (b-h).

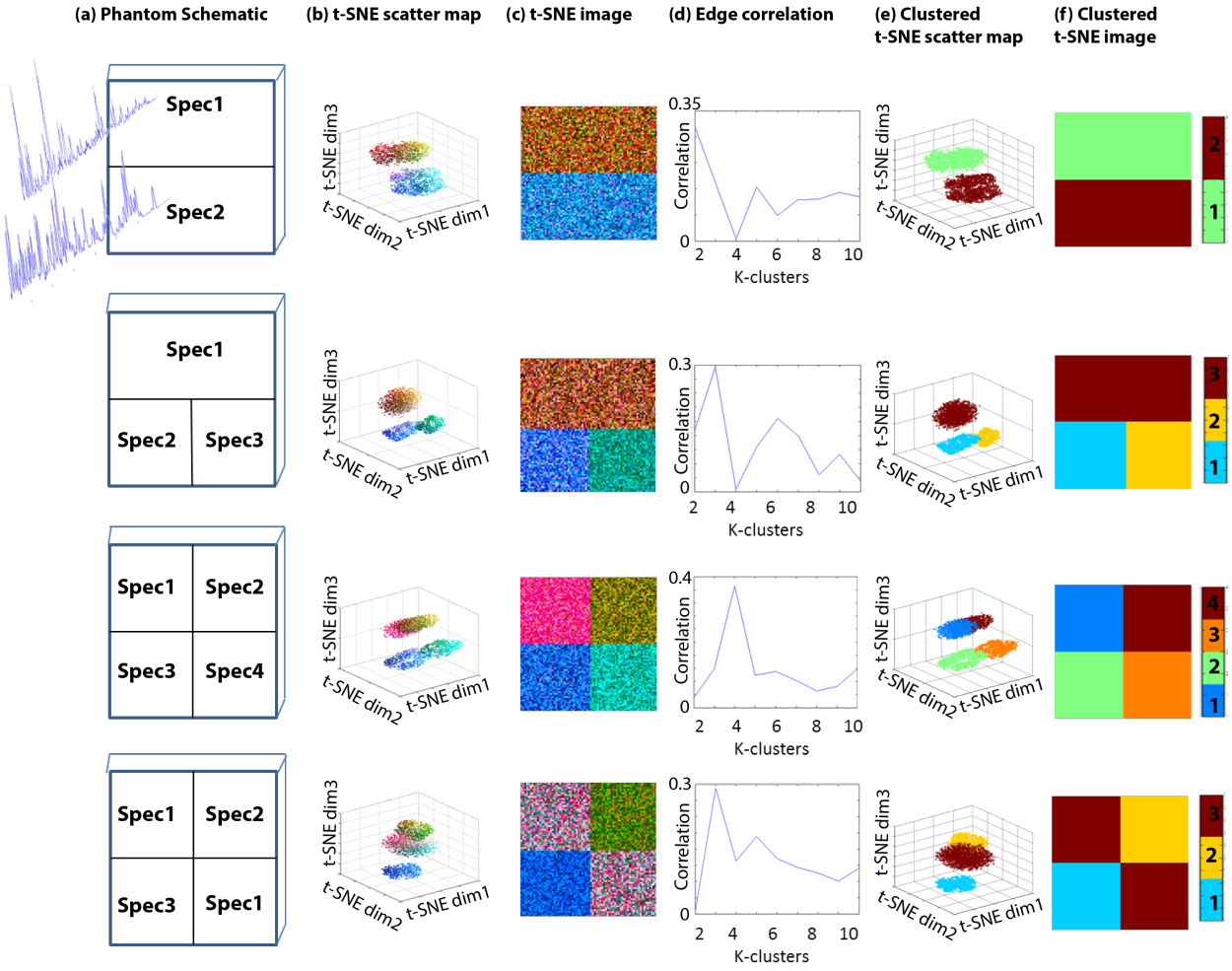


Figure S15. Four synthetic datasets (different rows), each with a size of (80*80*389) and containing patches with different real MALDI-MSI spectra with added experimental noise from a uniform distribution. The proposed pipeline could detect the "ground truth" number of piece-wise homogeneous regions (b-f). The last row shows results of a dataset in which the number of patches are greater than the number of phenotypes (i.e. here, similar spectrum), and the algorithm could separate the different regions properly.