SI Appendix, Materials and Methods

Dataset

Tumor center samples

DESI mass spectrometry imaging was acquired from tissue samples obtained from the tumor central region of 34 patients affected by colorectal cancer (Table 1). The samples were obtained from the patients by surgical resection.
For simplicity, we refer these as "*core samples*".
During the 6 years of observational study, all the events regarding recurrent diseases of subject disease were annotated. At the end of the observational study, 8 patients developed a metastatic recurrence within a median time of 1840 (range = [37, 2323]) days. The optical images of the hematoxylin and eosin stained tissues were acquired and reviewed by a consultant histopathologist to determine the presence of cancerous tissue. Mass spectral data was acquired in negative ion mode using a Thermo Exactive Orbitrap mass spectrometer in the range 200-1050 m/z.

Samples at 10cm from tumor center

DESI mass spectrometry imaging was acquired using tissue samples obtained from tissue sections at a distance of about 10cm from the tumor center of 29 patients affected by colorectal cancer (Table 1). The samples were obtained from the patients by surgical resection. For simplicity, we refer to these as "*10cm samples*".
During 6 years of observational study, all events regarding recurrent disease of the patients were annotated. At the end of the observational study, 6 patients developed a metastatic recurrence within a median time of 1946 (range = [310, 2254]) days. The optical images of the hematoxylin and eosin stained tissues were acquired and reviewed by a consultant histopathologist to determine the histological properties of the tissue. Mass spectral data was acquired in negative ion mode using a Thermo Exactive Orbitrap mass spectrometer in the range of 200-1050 m/z.

MS data pre-processing

An Identical procedure was applied to both the tumor center and 10cm samples.
The centroided data extracted from the RAW MS data using the ProteoWizard software (1) was saved in the imzML format (2) and imported into Matlab 2016b (The MathWorks, Inc.) using the Java package imzMLConverter (3). The spectra from each pixel were re-calibrated using the visual assessment process described in the "Results" section (code at the end of the SI) with 255.2330 m/z and 885.5499 m/z as references.
The re-calibrated data were pre-processed to obtain the intensities matrices of the peaks common to all the pixels of the images in the cohort.
Firstly, the samples belonging to one of the two groups of samples (tumor core, 10cm from the tumor) were pre-processed individually. Using the command '*binPeaks*' (MALDIquant package for R (4, 5)) with the method set to 'strict' and a tolerance equal to 1e-5 (corresponding to 10ppm) the peaks within each tissue section were matched. Peaks that were present in less than 1% of the pixels were discarded. No smoothing or baseline correction was necessary because the profiles in the RAW data (converted to centroided data) were already clean and flat.

The frequency of the matched peaks within each tissue section was used as a representative peak intensity for each tissue section and combined to generate the m/z values common to all the tissue sections of the group. For this purpose, the command '*binPeaks*' was applied to the 34 (resp. 29) representative peak profiles using the 'strict' method and a tolerance of 0.002. Only the masses present in all the sections were used for subsequent analysis.
Using this procedure, 185 m/z values were found common to the 34 tissue sections from the tumor central region, and 141 m/z values were found common to the 29 tissue sections from a region distant 10cm from the tumor center.
A median scaling normalization was applied to all the pixels. Specifically, the intensities of the pixel spectrum were divided by the median of the non-zero intensities of the spectrum. All the variable intensities were subsequently scaled in the range [0, 1] within each tissue sample, in order to individuate the presence of similar patterns across all the pixels.

WGCNA

The network adjacency matrix of WGCNA (6), representing the similarities between the ion expressions in the sample, can be defined in two ways: *signed* and *unsigned*. The main difference between these two formulations relies on the fact that the former takes into account of the sign of the similarity measure (correlation), whereas the latter considers only its absolute value.
The general definition of the signed adjacency matrix $\mathbf{A}^{signed} = \left(a_{i,j}\right)$ is (6)

$$a_{i,j} = \left(0.5 + 0.5\times \mathrm{cor}\left(\boldsymbol{y}_i, \boldsymbol{y}_j\right)\right)^{\beta}$$

where *cor* represents the correlation between the two ions relative abundances $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ across all the pixels and β is a soft power applied to reduce the effect of spurious/noisy correlations. Various similarity measures can be used, such as Spearman's correlation or bi-weight midcorrelation (7); in this study Pearson's correlation was used. For simplicity, the fact that all the adjacency matrices were signed will be omitted.
In order to determine an adjacency matrix associated with the metastatic/non-metastatic tumor core samples, the signed adjacency matrices of each tissue section $\mathbf{A}_s^{(core)} = \left(a_{s;\,i,j}^{(core)}\right), s = 1, \dots, N^{(core)}$ (with β=1) was calculated. The *consensus signed adjacency matrix* was defined as

$$\mathbf{A}_{met}^{(core)} = \left(a_{met;\,i,j}^{(core)}\right) = \min_{s=1,\dots,N_{met}^{(core)}} a_{s;i,j}^{(core)}$$
$$\mathbf{A}_{non-met}^{(core)} = \left(a_{non-met;\,i,j}^{(core)}\right) = \min_{s=1,\dots,N_{non-met}^{(core)}} a_{s;i,j}^{(core)}$$

An analogous definition was used for the 10cm samples.
Before the calculation of the consensus adjacency matrices, quantile normalisation (8) was applied to the tissue specimens adjacency matrices, in order to avoid biased results.
The final consensus signed adjacencies were therefore raised to a soft power β, defined as the smallest value that resulted in a $R^2 \geq 0.8$ for the scale-free topology for both the metastatic and non-metastatic tissue sections. For this purpose, the consensus adjacency matrices associated with the metastatic (resp. non-metastatic) samples was used as input for

the command WGCNA '*pickSoftThreshold.fromSimilarity*'. The selected soft power was the minimum integer such that both the metastatic and non-metastatic consensus adjacencies resulted in a scale free network topology.
In this way, a β_{core} = 9 for the tumor center samples and β_{10cm} = 13 for the samples at 10cm from the tumor center were used.
In the network defined by the consensus adjacency, groups of ions (network nodes) were partitioned with their mutual similarities. In this way, the ions that were expressed in similar spatial regions (highly correlated) were grouped to form a *module*. For this purpose, the signed topological overlap matrix ($\mathbf{TOM}^{signed}$) (9) was employed as a similarity measure (10). Given the signed adjacency matrix $\mathbf{A}^{signed} = (a_{ij})$, the general definition of the $\mathbf{TOM}^{signed} = (t_{ij})$ is

$$t_{ij} = \frac{\left|a_{ij} + \sum_{k \neq i,j} \tilde{a}_{ik}\tilde{a}_{kj}\right|}{\min(k_i, k_j) + 1 - |a_{ij}|}$$

where $\tilde{a}_{ij} = a_{ij} \times \text{sign}\left(\text{cor}(\boldsymbol{y}_i, \boldsymbol{y}_j)\right)$ and $k_i = \sum_{l \neq i} |\tilde{a}_{il}|$.
Analogous to the calculation of the consensus adjacency matrix, the *consensus* $\mathbf{TOM}^{(met)}$ (resp. $\mathbf{TOM}^{(non-met)}$) matrix elements were calculated as the smallest value across all the tissue sections from the specific tissue region,

$$\mathbf{TOM}_{met}^{(core)} = \left(t_{met;\,i,j}^{(core)}\right) = \min_{s=1,\ldots,N_{met}^{(core)}} t_{s;i,j}^{(core)}$$
$$\mathbf{TOM}_{non-met}^{(core)} = \left(t_{non-met;\,i,j}^{(core)}\right) = \min_{s=1,\ldots,N_{non-met}^{(core)}} t_{s;i,j}^{(core)}$$

The analogous definition was used for the 10cm samples.
Before calculating the *consensus TOM*, quantile normalization was applied to all the individual *TOMs* in order to avoid biased results.
Using $1 - t_{met;\,i,j}^{(core)}$ (resp. $1 - t_{met;\,i,j}^{(10cm)}$) as a distance measure, hierarchical clustering (with average linkage) was applied to determine the ion modules. The *Dynamic Tree Cut* algorithm (11) was used to automatically identify the optimal partition from the dendrogram. The hybrid partition algorithm and a minimum number of 5 ions per module were set as parameters for the dynamic tree cut algorithm. For each module, an *eigenmetabolite* (ME), defined as the scores vector of the first principal component calculated on the module ions, was used as a representative spatial intensity of the module. Modules corresponding to similar spatial distributions (where the eigenmetabolites Pearson's correlation was larger than 0.8) were merged into a single module.
In order to check the network module differences of the non-metastatic tissue sections, additionally the *native* non-metastatic modules were identified using the same procedure using $1 - t_{non-met;\,i,j}^{(core)}$ (resp. $1 - t_{non-met;\,i,j}^{(10cm)}$) as a distance measure for the hierarchical clustering. The same parameters used for the metastatic tissue were used.

Network measures analyzed for the module preservation test

For the module preservation analysis (12), the WGCNA command '*modulePreservation*' was used with the consensus adjacency matrices as input. The metastatic related network was

used as a reference and module preservation was tested on the non-metastatic network. The number of permutations was set to 5000. The results of the preservation analysis were plotted in order to identify the degree of preservation of the metastatic modules. The modules with a $Z_{summary} < 2$ were considered not-preserved, those with $2 \leq Z_{summary} < 5$ were considered weakly preserved, those with $5 \leq Z_{summary} < 10$ were considered moderately preserved, and finally those with $Z_{summary} \geq 10$ were considered preserved in the non-metastatic network. This analysis was run on the tumor centre sections and sections at 10cm from the tumor separately.

Ion pair specificity

A differential analysis between metastatic and not-metastatic related tissue sections was performed by calculating the metastatic edge specificity (13), defined as

$$s_{ij} = \frac{t_{ij}^{(met)} / mean(TOM^{(met)})}{t_{ij}^{(met)} / mean(TOM^{(met)}) + t_{ij}^{(non-met)} / mean(TOM^{(non-met)})}$$

Ion pairs with s > 0.8 were considered present in the metastatic samples and absent in the non-metastatic samples.

LOOCV metastatic/non-metastatic classification

In order to verify the statistical relationships between the ion modules found and the "metastatic" vs "non-metastatic" condition, a leave-one-specimen-out cross validation was performed.

From the signed adjacency matrix for each tissue specimen, calculated as described in the WGCNA paragraph, the intramodular connectivity was derived as

$$k_{m,j} = \frac{\sum_i a_{ij}}{\max_j \sum_i a_{ij}}$$

where *m* is the module index, and *i, j* are the indices of the adjacency matrix.

In all the adjacencies, the same soft power value calculated from the consensus network was used. For this reason, the classification accuracy was intended only as a comparative test between the different modules and not as a measure of the real predictive power on unseen data.

In this way, a vector of features was associated with each tissue specimen representing the connectivity properties of the network module.

Using the leave-one-specimen-out cross validation, a Regularized Random Forest (RRF) classifier (14) was trained on the training samples and tested on the test sample using the intramodular connectivity of a single module. The models were trained to predict the "metastatic" and "non-metastatic" condition of the patient associated with the tissue specimen.

The following parameters were used: number of trees = 1000, mtry and coefReg (coefficient of regularization) were tuned to achieve the best Cohen's Kappa measure (15).

SI Appendix, Results.

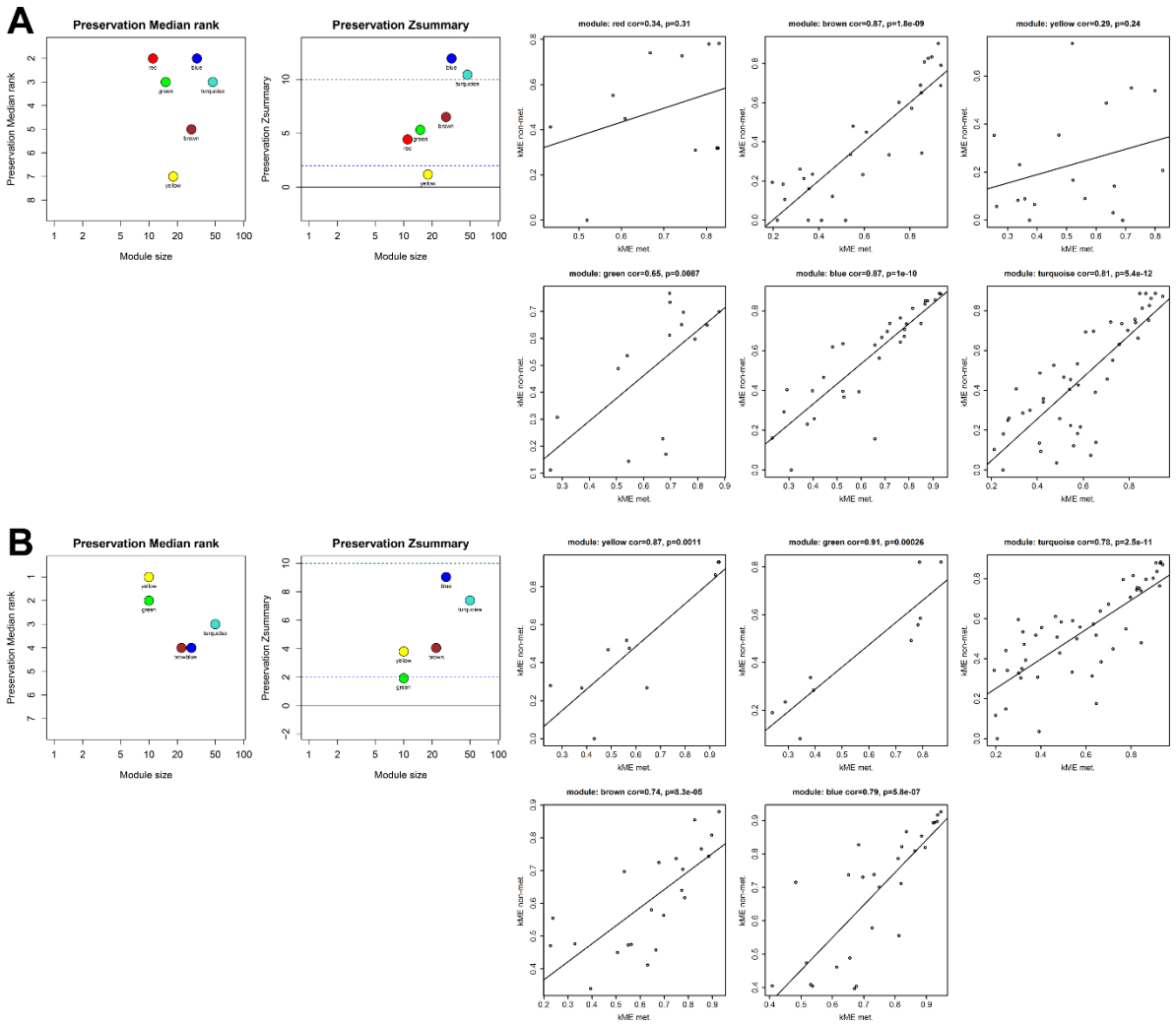


Fig. S1 – Results of the module preservation for the tumor core samples (A) and 10cm samples (B). A $2 \leq Z_{summary} \leq 10$ *is interpreted as a weakly/moderate preservation. In the tumour core samples, the tissue unrelated "yellow" module is not preserved, whereas modules "red" and "green" are considered weakly preserved (*$Z_{summary} \lesssim 5$*). Among those, only the "green" module is detected in the tissue. The correlation plots (right) confirm that the module membership (kME) of the metastatic and non-metastatic weakly preserved modules is different. In the 10cm tissue sections, all the metastatic modules are weakly/moderately preserved in the non-metastatic network, with only the "brown" module expressed in the tissue and weakly preserved. Again, the relative correlation value between the metastatic and non-metastatic module membership confirms that the "brown" module is weakly preserved.*

Analysis of tumor center samples.

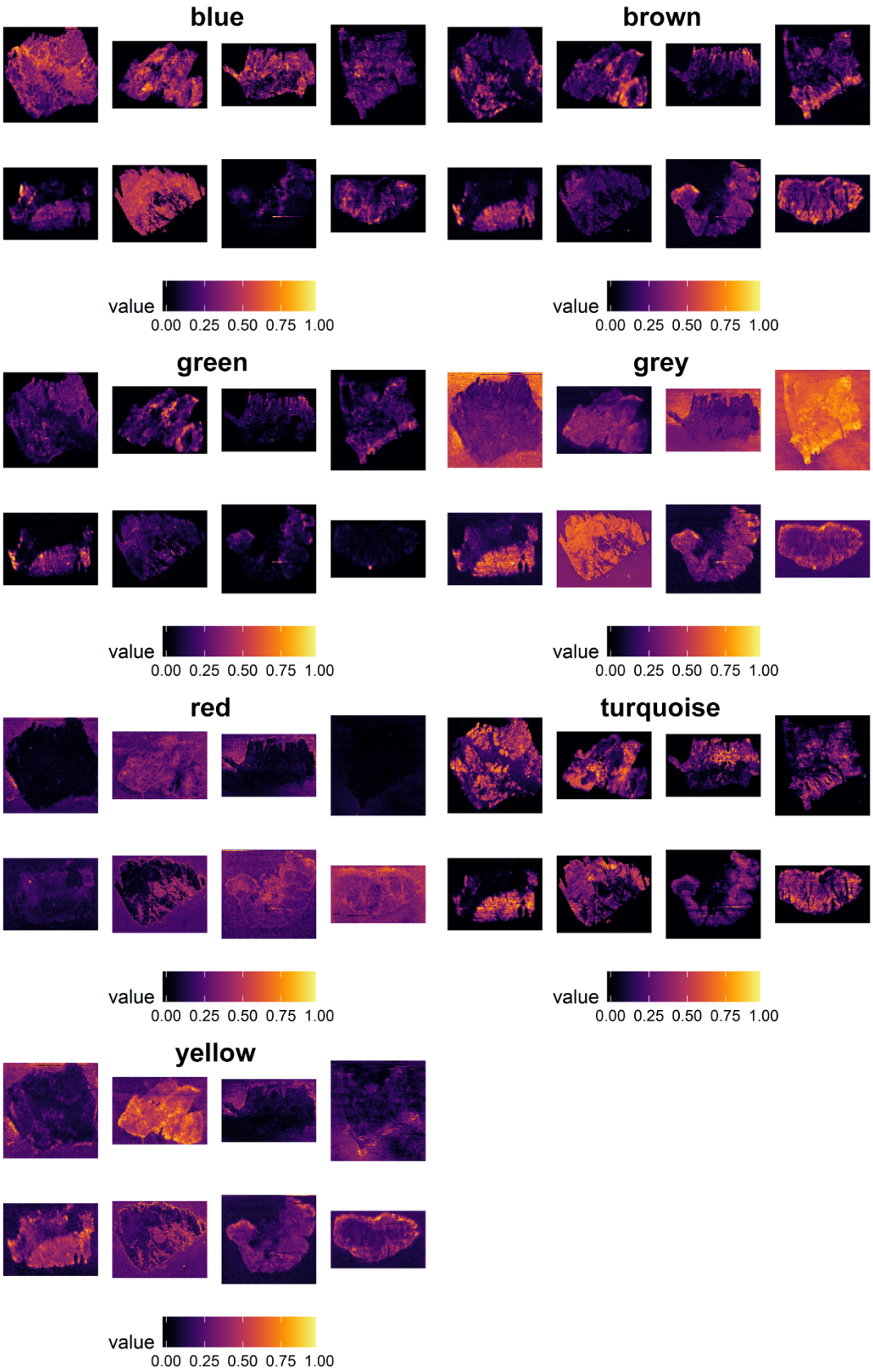


Fig. S2 – Spatial distribution of the metastatic related data ME for from the tumor core tissue section. Modules blue, brown, green and turquoise are expressed within the tissue of all the patients. All the intensities were scaled to [0, 1] within the tissue pixels.

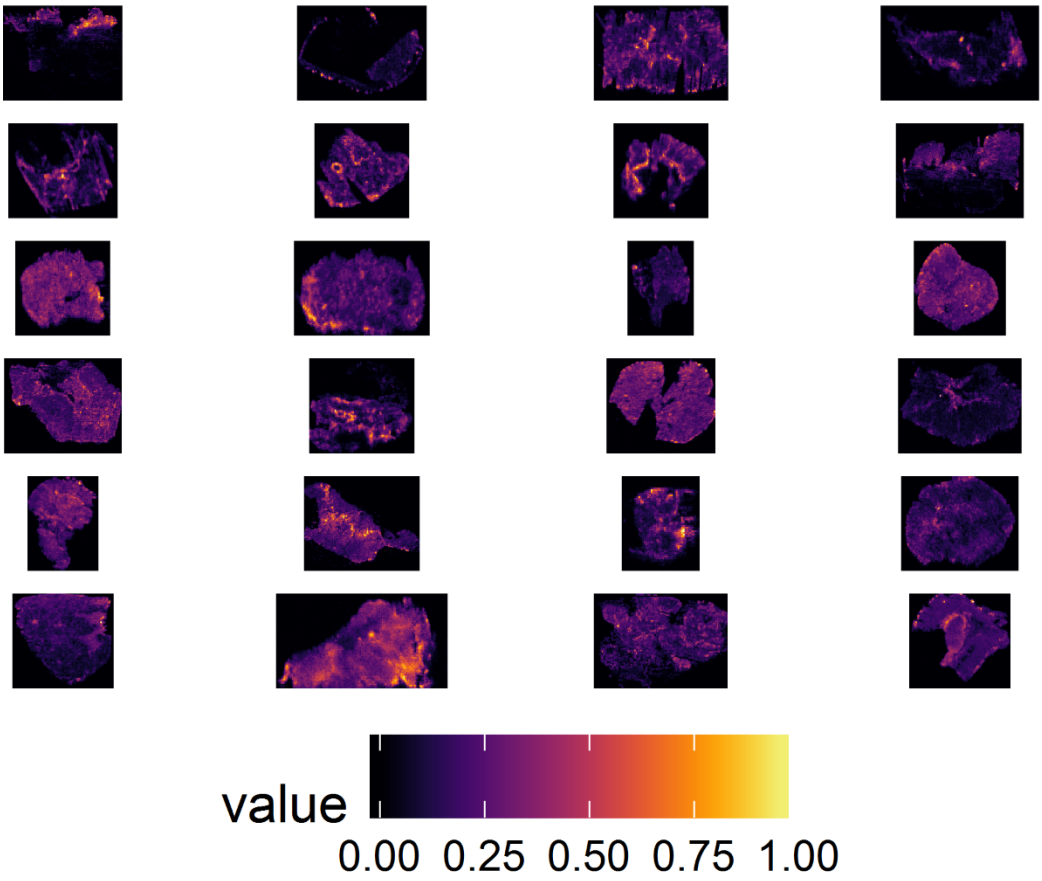


Fig S3 – Spatial distributions of the weakly preserved green module ME of the non-metastatic samples from the tumor core tissue show that the selected ions were also expressed in these samples. All the intensities were scaled to [0, 1] within the tissue pixels.

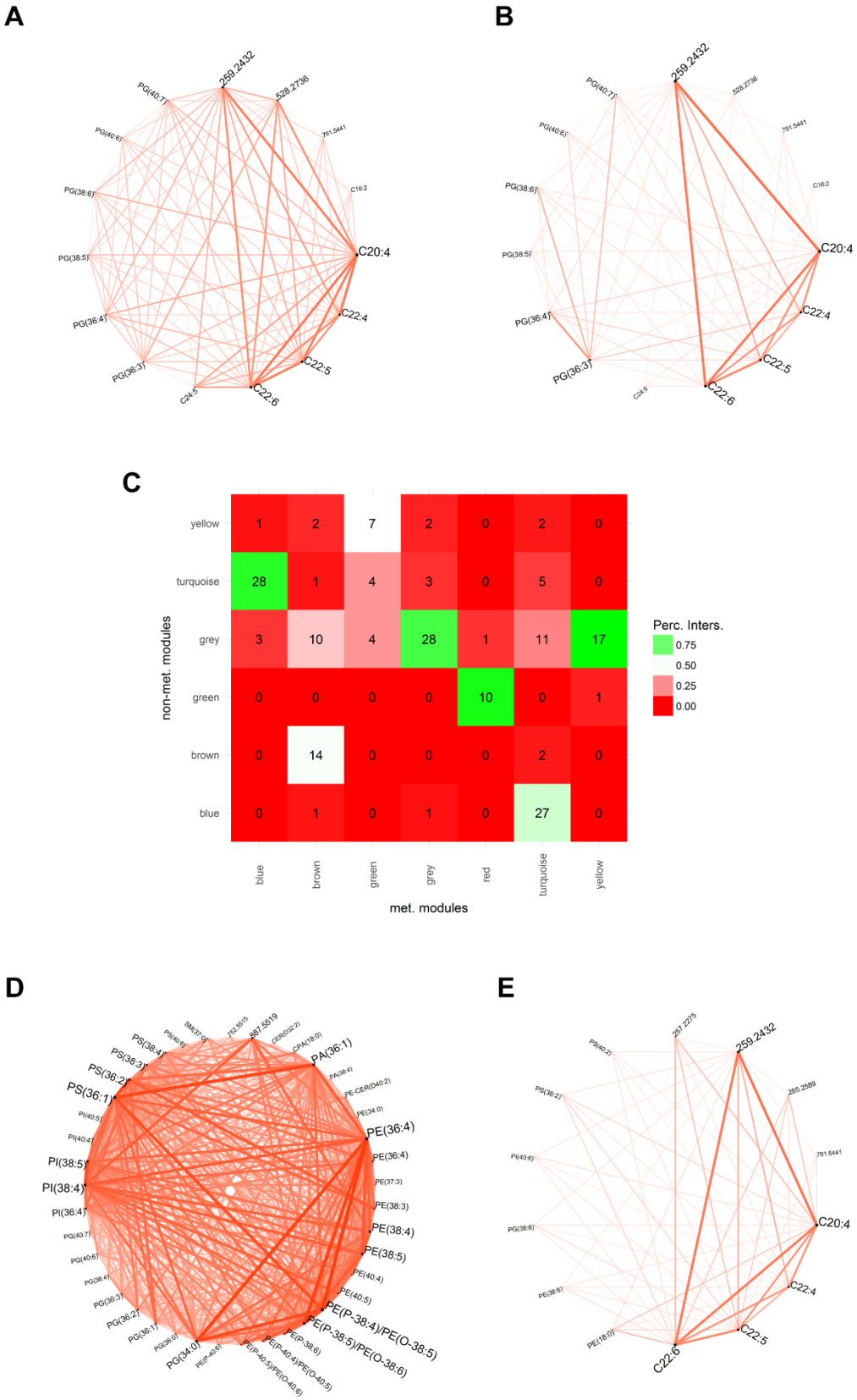
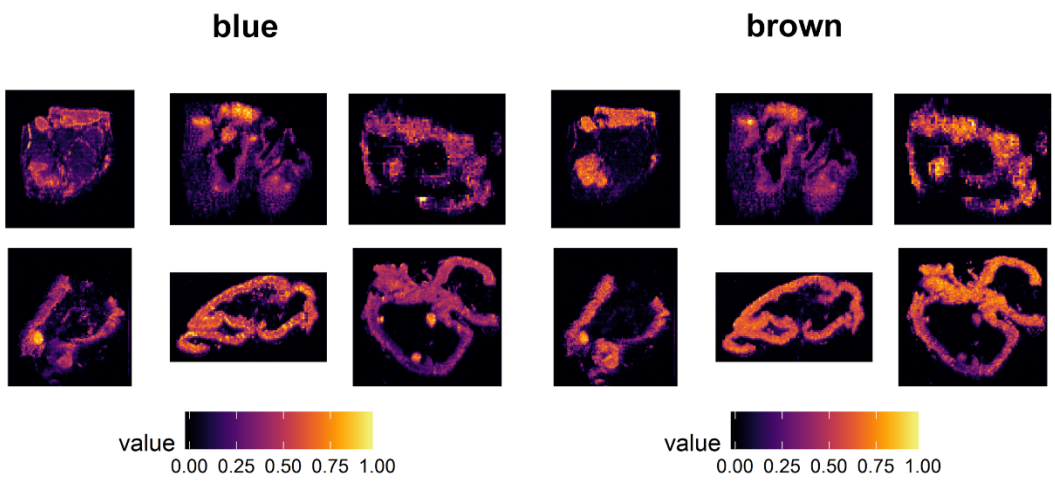
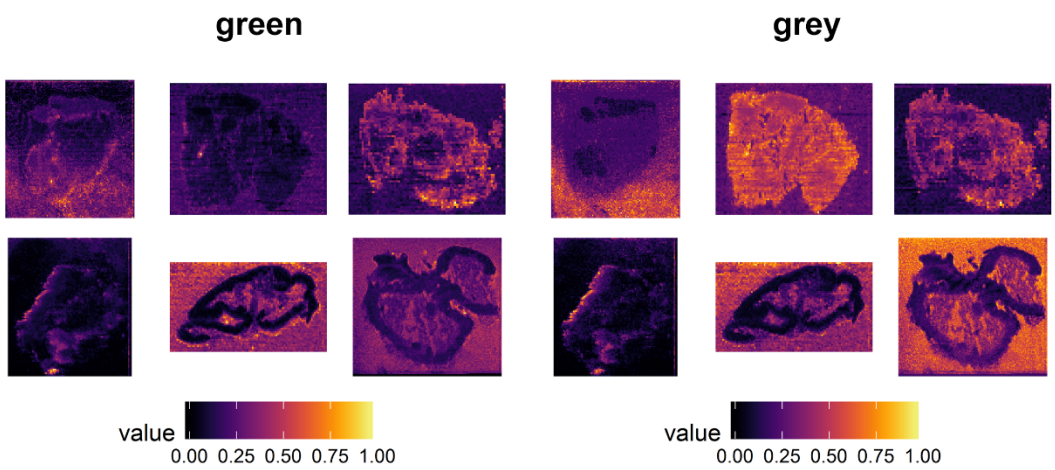


Fig S4 – Circle plots showing the TOM values for the weakly preserved "green" module using the metastatic (A) and the non-metastatic related (B) tumor center tissue sections. The module ions are represented as dots (whose size is proportional to the number of edges), and the lines represent the TOM values (the thickness is proportional to the TOM value).

In the metastatic data, the FA and PG appear to be more interconnected than in the non-metastatic data. Running WGCNA on the non-metastatic data, the "green" module ions were split into 3 modules (C): "turquoise" (4 ions, 25% intersection), "yellow" (7 ions, 50% intersection) and "grey" (4 ions, 25% intersection). The circle plots of the non-metastatic "turquoise" and "yellow" modules show that PGs are assigned together with other phospholipids (D), whereas FA are co-expressed mostly with PE and PS (E).

Analysis of the tissue sections at 10cm from tumor core.





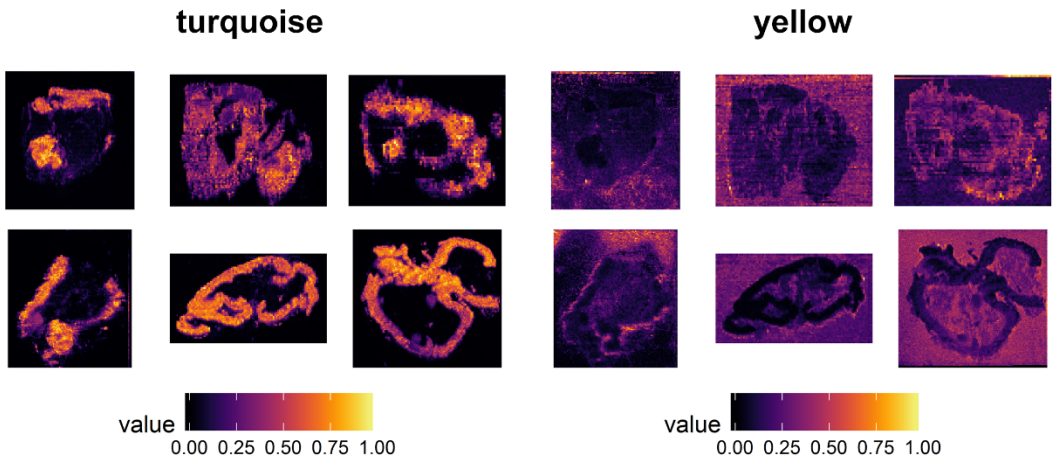
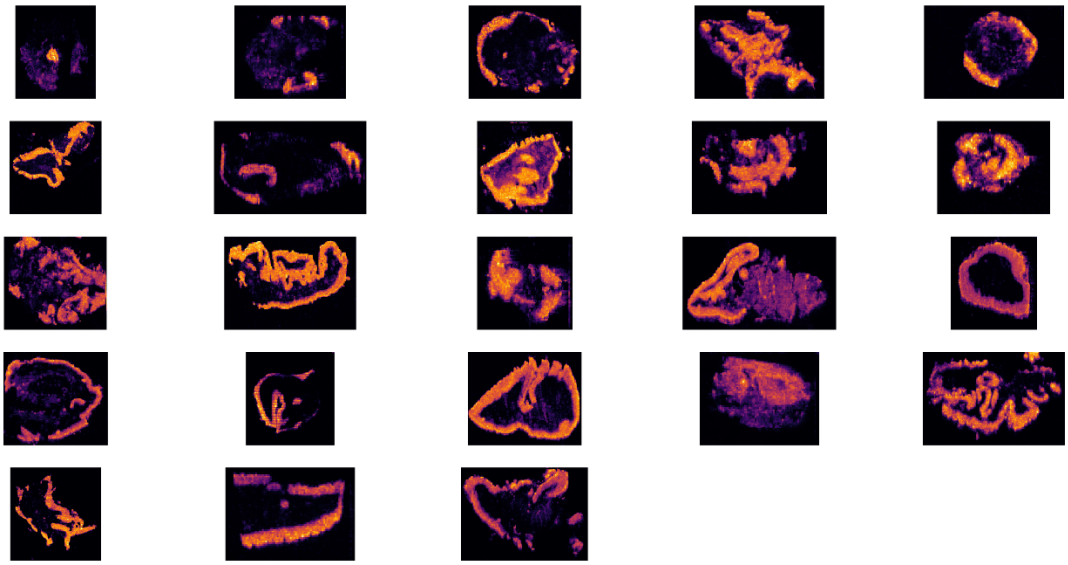


Fig S5 – Spatial distribution of the eigenmetabolites associated to the WGCNA modules in the metastatic related data. Only the modules blue, brown and turquoise were expressed within the tissue. Intensities were scaled at [0, 1] within each tissue section.



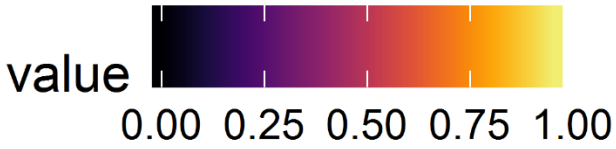


Fig S6 –Spatial distribution of the 23 non-metastatic samples ME associated with the weakly preserved brown module ions show that these ions were also expressed in the test samples, confirming the importance of the different co-expression levels.

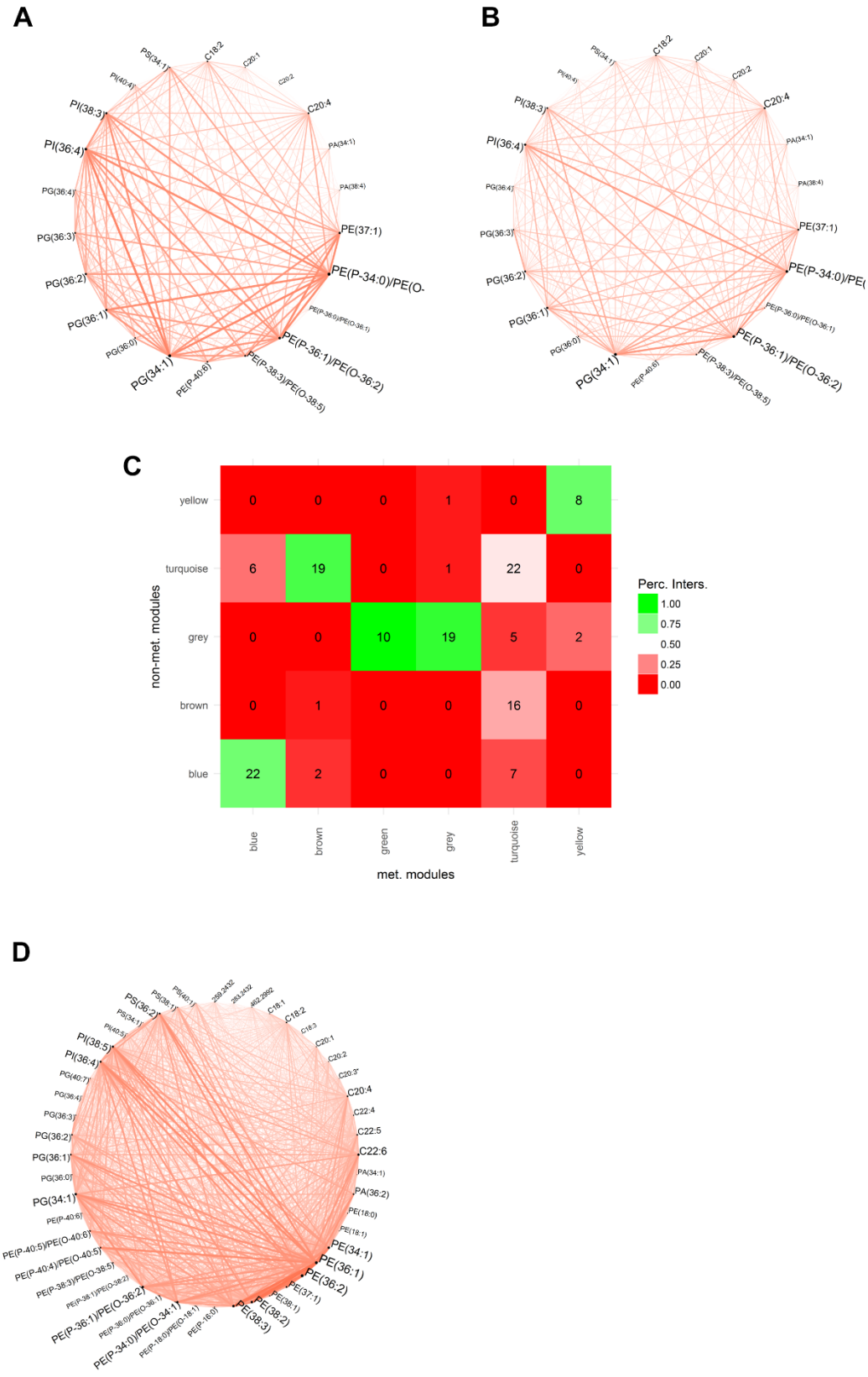


Fig S7 – Circle plots showing the TOM values for the weakly preserved "brown" module using the metastatic (A) and the non-metastatic (B) related tissue sections at 10cm from the tumor center. In the metastatic data, the phospholipids appear to be more strongly interconnected than in the non-metastatic data. Running WGCNA on the non-metastatic data, the "brown"

module ions were split into 2 modules (C): "turquoise" (19 ions, 90% intersection) and "blue" (2 ions, 10% intersection). The circle plot of the non-metastatic module "turquoise" shows that the "brown" module phospholipids are co-expressed with several other phospholipids and FA (D).

m/z	name	adduct	delta ppm	alt name
251.2017	C16:2	[M-H]-	0	Hexadecadienoic acid
259.2432				
303.2332	C20:4	[M-H]-	0	Arachidonic acid
327.2332	C22:6	[M-H]-	0	Cervonic acid
329.2489	C22:5	[M-H]-	0	Clupanodonic acid
331.2645	C22:4	[M-H]-	0	Adrenic acid
357.2803	C24:5	[M-H]-	1	Tetracosapentaenoic acid
528.2736				
769.5035	PG(36:4)	[M-H]-	1	
771.5193	PG(36:3)	[M-H]-	1	
791.5441				
793.5039	PG(38:6)	[M-H]-	1	
795.5202	PG(38:5)	[M-H]-	2	
819.5199	PG(40:7)	[M-H]-	2	
821.5365	PG(40:6)	[M-H]-	3	

Table S1 - Annotated m/z values for the "green" module of the tumour core samples.

m/z	name	adduct	delta ppm	alt name
257.2275				
259.2432				
285.2589				
303.2332	C20:4	[M-H]-	0	Arachidonic acid
327.2332	C22:6	[M-H]-	0	Cervonic acid
329.2489	C22:5	[M-H]-	0	Clupanodonic acid
331.2645	C22:4	[M-H]-	0	Adrenic acid
480.3099	PE(18:0)	[M-H]-	0	
762.5094	PE(38:6)	[M-H]-	1	
791.5441	PI(O-33:0)	[M-H2O-H]-	0	
793.5039	PG(38:6)	[M-H]-	1	
814.5624	PS(38:2)	[M-H]-	2	
842.593	PS(40:2)	[M-H]-	1	
909.5508	PI(40:6)	[M-H]-	1	

Table S2 - Annotated m/z values for the "yellow" native module of the non-metastatic tumour core samples.

m/z	name	adduct	delta ppm
419.2573	CPA(18:0)	[M-H]-	1
646.6152	Cer(d32:2)	[M-H]-	1
701.5137	PA(36:1)	[M-H]-	1
718.5402	PE(34:0)	[M-H]-	1
722.5141	PE(36:4)	[M-H]-	1
723.4979	PA(38:4)	[M-H]-	1
736.5302	PE(37:3)	[M-H2O-H]-	2
746.5149	PE(P-38:6)	[M-H]-	2
748.5309	PE(P-38:5)/PE(O-38:6)	[M-H]-	2
749.5336	PG(34:0)	[M-H]-	0
750.5453	PE(P-38:4)/PE(O-38:5)	[M-H]-	1
752.5515			
764.5248	PE(38:5)	[M-H]-	1
768.5565	PE(38:3)	[M-H]-	2
768.5565	PE(38:3)	[M-H]-	2
769.5035	PG(36:4)	[M-H]-	1
771.5193	PG(36:3)	[M-H]-	1
773.5348	PG(36:2)	[M-H]-	1
774.548	PE(P-40:6)	[M-H]-	4
775.5502	PG(36:1)	[M-H]-	0
776.5628	PE(P-40:5)/PE(O-40:6)	[M-H]-	3
777.5653	PG(36:0)	[M-H]-	0
778.5771	PE(P-40:4)/PE(O-40:5)	[M-H]-	1
779.58	SM(37:0)	[M+Cl]-	5
786.5302	PS(36:2)	[M-H]-	1
788.5459	PS(36:1)	[M-H]-	1
792.5565	PE(40:5)	[M-H]-	2
793.5594	PE-Cer(d40:2)	[M+Cl]-	4
794.5721	PE(40:4)	[M-H]-	1
810.5301	PS(38:4)	[M-H]-	1
812.5465	PS(38:3)	[M-H]-	2
819.5199	PG(40:7)	[M-H]-	2
821.5365	PG(40:6)	[M-H]-	3
834.5325	PS(40:6)	[M-H]-	2
857.5196	PI(36:4)	[M-H]-	1
883.5354	PI(38:5)	[M-H]-	1
885.5511	PI(38:4)	[M-H]-	1
887.5519			
911.5673	PI(40:5)	[M-H]-	1
913.5834	PI(40:4)	[M-H]-	2

Table S3 - Annotated m/z values for the "turquoise" native module of the non-metastatic tumour core samples.

Module	red	brown	yellow	green	blue	turquoise
Bal. acc.	0.7083	0.7917	0.7500	0.8958	0.7500	0.8125

Pred. / Ref.	met	non-met.
met.	7	2
non-met.	1	22

(Fisher test p-value = 8e-5)

Table S4 – Classification results for the tumor center tissue sections. Top: "green" intramodular connectivity can predict the metastatic/non-metastatic class of tissue g with the best balanced accuracy (16). Bottom: The confusion matrix using the "green" intramodular connectivity and the Fisher's test p-value.

m/z	name	adduct	delta ppm	alt name
279.233	C18:2	[M-H]-	0	Linoleic acid
303.233	C20:4	[M-H]-	0	Arachidonic acid
307.2644	C20:2	[M-H]-	0	Eicosadienoic acid
309.28	C20:1	[M-H]-	0	Eicosenoic acid
673.4818	PA(34:1)	[M-H]-	0	
700.5292	PE(P-34:0)/PE(O-34:1)	[M-H]-	0	
723.4973	PA(38:4)	[M-H]-	0	
728.5607	PE(P-36:1)/PE(O-36:2)	[M-H]-	1	
730.5763	PE(P-36:0)/PE(O-36:1)	[M-H]-	0	
747.5184	PG(34:1)	[M-H]-	0	
752.5609	PE(P-38:3)/PE(O-38:5)	[M-H]-	1	
760.514	PS(34:1)	[M-H]-	0	
771.5188	PG(36:3)	[M-H]-	0	
773.5342	PG(36:2)	[M-H]-	0	
774.5472	PE(P-40:6)	[M-H]-	3	
775.5495	PG(36:1)	[M-H]-	0	
794.5478	PE(37:1)	[M+Cl]-	0	
857.5191	PI(36:4)	[M-H]-	0	
887.5708	PI(38:3)	[M-H]-	5	
913.5829	PI(40:4)	[M-H]-	1	

Table S5 - Annotated m/z values for the "brown" module of the samples at 10cm from the tumour centre.

m/z	name	adduct	delta ppm	alt name
259.2432				
277.2174	C18:3	[M-H]-	0	alpha/gamma-Linoleic acid
279.233	C18:2	[M-H]-	0	Linoleic acid
281.2486	C18:1	[M-H]-		Oleic acid
283.2432				
303.233	C20:4	[M-H]-	0	Arachidonic acid
306.2521	C20:3*	[M-H]-	0	Dihomo-gamma-linoleic acid
307.2644	C20:2	[M-H]-	0	Eicosadienoic acid
309.28	C20:1	[M-H]-	0	Eicosenoic acid
327.233	C22:6	[M-H]-	0	Cervonic acid
329.2487	C22:5	[M-H]-	0	Clupanodonic acid
331.2644	C22:4	[M-H]-	0	Adrenic acid
436.2834	PE(P-16:0)	[M-H]-	0	
462.2992				
464.3146	PE(P-18:0)/PE(O-18:1)	[M-H]-	0	
478.294	PE(18:1)	[M-H]-	0	
480.3097	PE(18:0)	[M-H]-	0	
673.4818	PA(34:1)	[M-H]-	0	
699.4976	PA(36:2)	[M-H]-	0	
699.4976	PA(36:2)	[M-H]-	0	
700.5292	PE(P-34:0)/PE(O-34:1)	[M-H]-	0	
701.5133	PA(36:1)	[M-H]-	0	
716.5242	PE(34:1)	[M-H]-	0	
728.5607	PE(P-36:1)/PE(O-36:2)	[M-H]-	1	
730.5763	PE(P-36:0)/PE(O-36:1)	[M-H]-	0	
747.5184	PG(34:1)	[M-H]-	0	
752.5609	PE(P-38:3)/PE(O-38:5)	[M-H]-	1	
756.5921	PE(P-38:1)/PE(O-38:2)	[M-H]-	1	
760.514	PS(34:1)	[M-H]-	0	
768.5559	PE(38:3)	[M-H]-	1	
769.503	PG(36:4)	[M-H]-	0	
770.5713	PE(38:2)	[M-H]-	1	
771.5188	PG(36:3)	[M-H]-	0	
772.5874	PE(38:1)	[M-H]-	1	
773.5342	PG(36:2)	[M-H]-	0	
774.5472	PE(P-40:6)	[M-H]-	3	
775.5495	PG(36:1)	[M-H]-	0	
776.5616	PE(P-40:5)/PE(O-40:6)	[M-H]-	2	
777.5644	PG(36:0)	[M-H]-	0	
778.5766	PE(P-40:4)/PE(O-40:5)	[M-H]-	1	
786.5297	PS(36:2)	[M-H]-	0	
794.5478	PE(37:1)	[M+Cl]-	0	
816.5771	PS(38:1)	[M-H]-	1	

819.5195	PG(40:7)	[M-H]-	1	
844.6082	PS(40:1)	[M-H]-	1	
857.5191	PI(36:4)	[M-H]-	0	
883.5348	PI(38:5)	[M-H]-	0	
911.5664	PI(40:5)	[M-H]-	0	

*Table S6 - Annotated m/z values for the "turquoise" native module of the non-metastatic samples at 10cm from the tumour centre. *306.2521 m/z was annotated as the isotopic form of C(20:3); it was not removed during the preprocessing because the mono-isotopic corresponding m/z value was found in the common m/z vector.*

Module	**yellow**	**green**	**turquoise**	**brown**	**blue**
Bal. acc.	0.5906	0.6449	0.5833	0.8080	0.7065

Pred. / Ref.	**met**	**non-met.**
met.	5	5
non-met.	1	18

(Fisher test p-value = 0.0105)

Table S7 – Classification results for the 10cm tissue sections. Top: "brown" intramodular connectivity can predict the metastatic/non-metastatic class of tissue sections with the best balanced accuracy. Bottom: The confusion matrix using the "brown" intramodular connectivity and the Fisher's test p-value.

References

1. Kessner D, Chambers M, Burke R, Agus D, & Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics (Oxford, England)* 24(21):2534-2536.
2. Schramm T*, et al.* (2012) imzML—a common data format for the flexible exchange and processing of mass spectrometry imaging data. *J Proteomics* 75.
3. Ding Y, Bi S, Fang G, & He L (2004) [Relationship between occurrence of Dryocosmus kuriphilus and development of cecidum]. *Ying Yong Sheng Tai Xue Bao* 15(1):108-110.
4. Gibb S & Strimmer K (2012) MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* 28(17):2270-2271.
5. Ihaka R & Gentleman R (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* 5(3):299-314.
6. Zhang B & Horvath S (2005) A General Framework for Weighted Gene Co-Expression Network Analysis. in *Statistical Applications in Genetics and Molecular Biology*.
7. Wilcox RR (2011) *Introduction to robust estimation and hypothesis testing* (Academic Press).
8. Bolstad BM, Irizarry RA, Åstrand M, & Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2):185-193.
9. Fuller T, Langfelder P, Presson A, & Horvath S (2011) Review of weighted gene coexpression network analysis. *Handbook of Statistical Bioinformatics*, (Springer), pp 369-388.
10. Zhang B & Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4(1):1128.
11. Langfelder P, Zhang B, & Horvath S (2008) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24(5):719-720.
12. Langfelder P, Luo R, Oldham MC, & Horvath S (2011) Is My Network Module Preserved and Reproducible? *PLOS Computational Biology* 7(1):e1001057.
13. Oldham M, Horvath S, & Geschwind D (2006) Conservation and Evolution of Gene Co-expression Networks in Human and Chimpanzee Brains. *Proc Natl Acad Sci USA* 103.
14. Deng H & Runger G (2013) Gene selection with guided regularized random forest. *Pattern Recognition* 46(12):3483-3489.
15. Cohen J (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20(1):37-46.
16. Brodersen KH, Ong CS, Stephan KE, & Buhmann JM (2010) The balanced accuracy and its posterior distribution. *Pattern recognition (ICPR), 2010 20th international conference on*, (IEEE), pp 3121-3124.