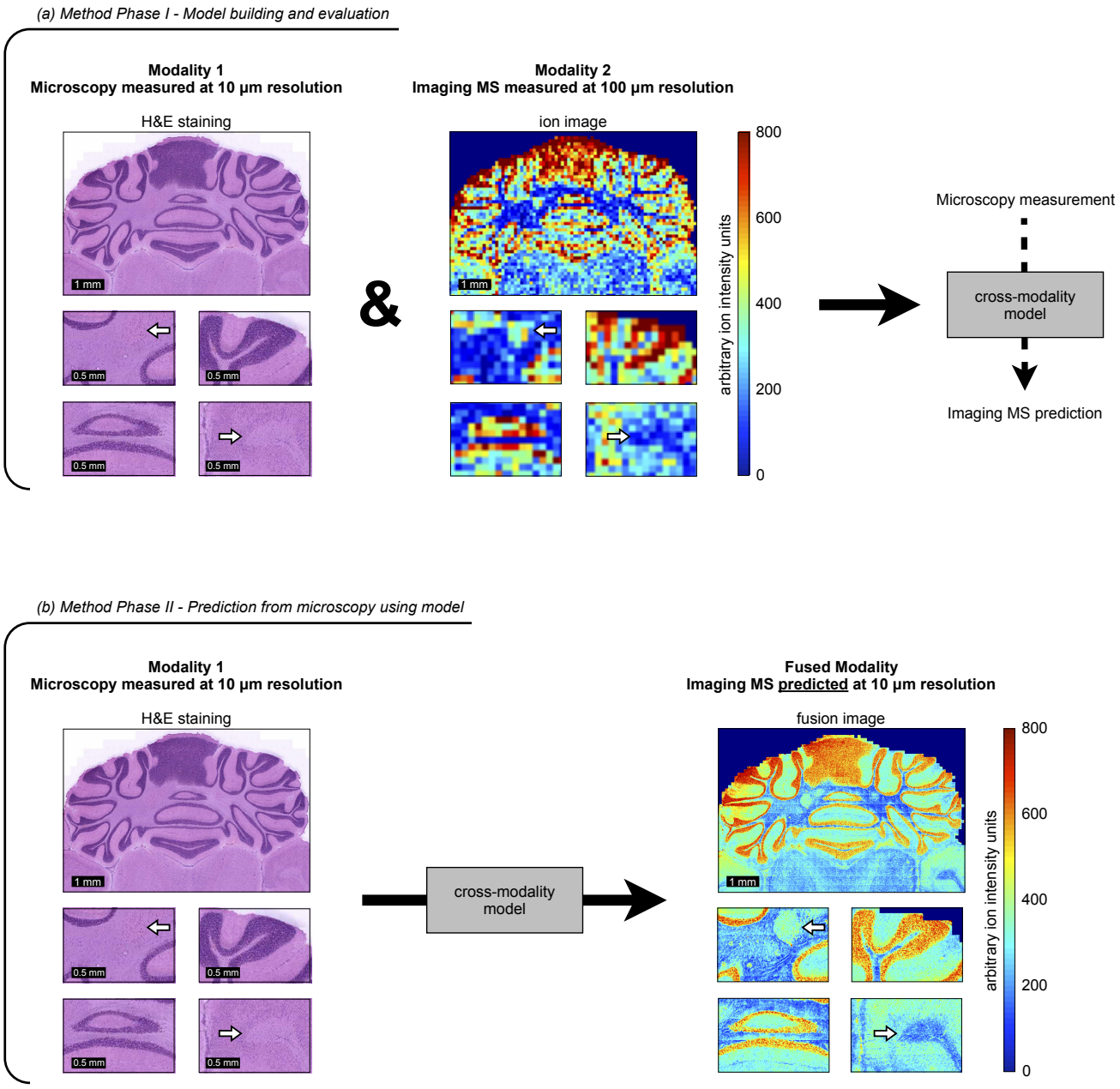
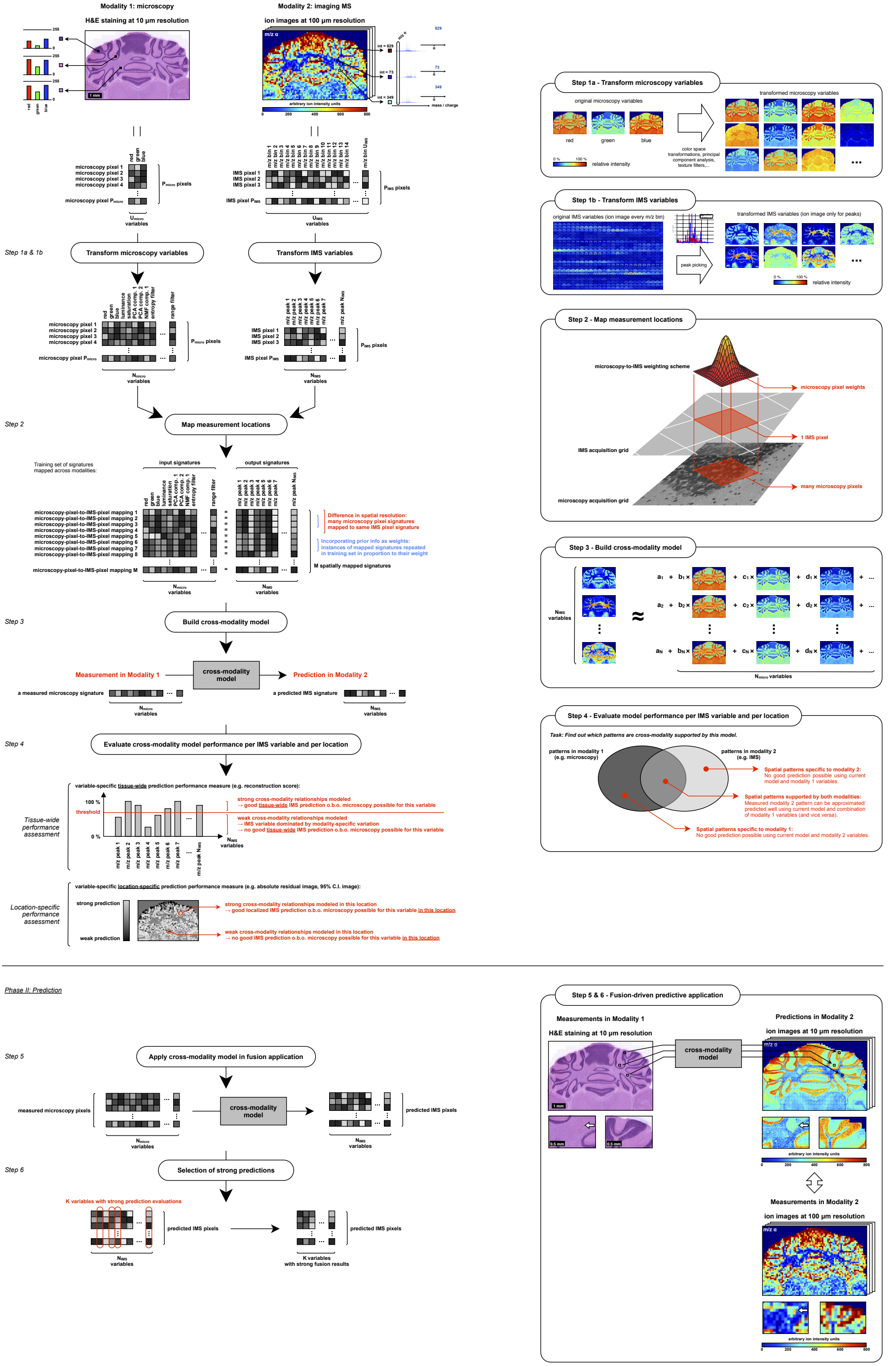
Supplementary Figures

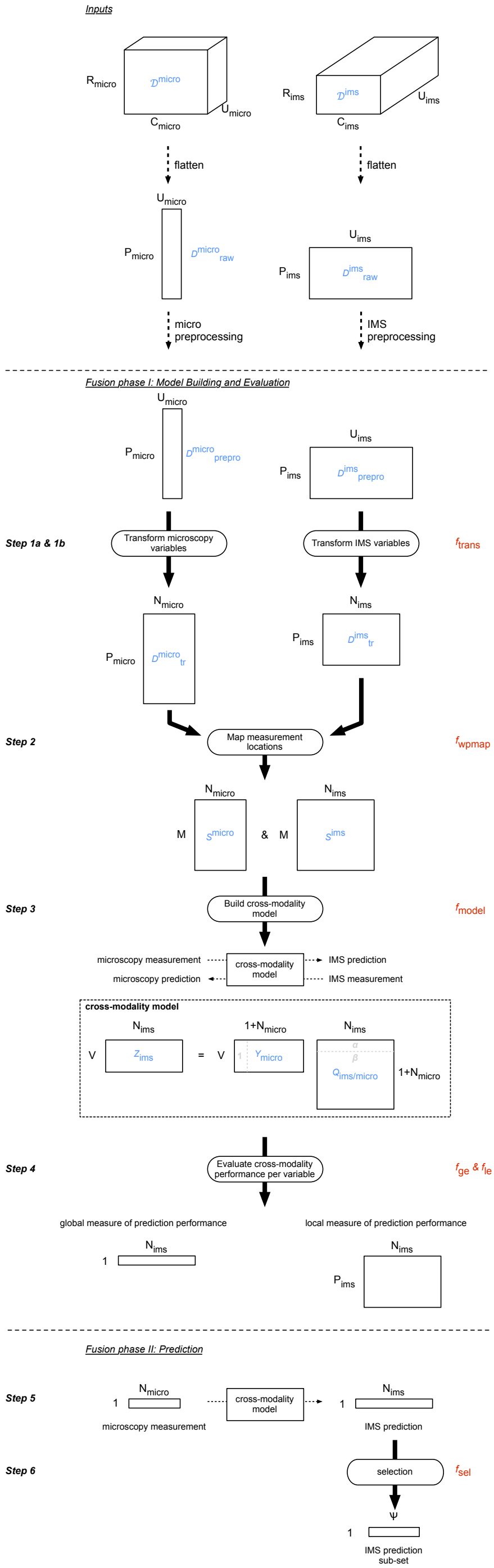


Supplementary Figure 1 Method overview. The fusion process consists of two phases. (**a**) Phase I builds a cross-modality model from the two measurement sources and evaluates for which ions good prediction is possible. (**b**) For those ions, phase II uses the model and the high-resolution microscopy measurements to predict the ion distribution at higher-than-IMS resolutions.

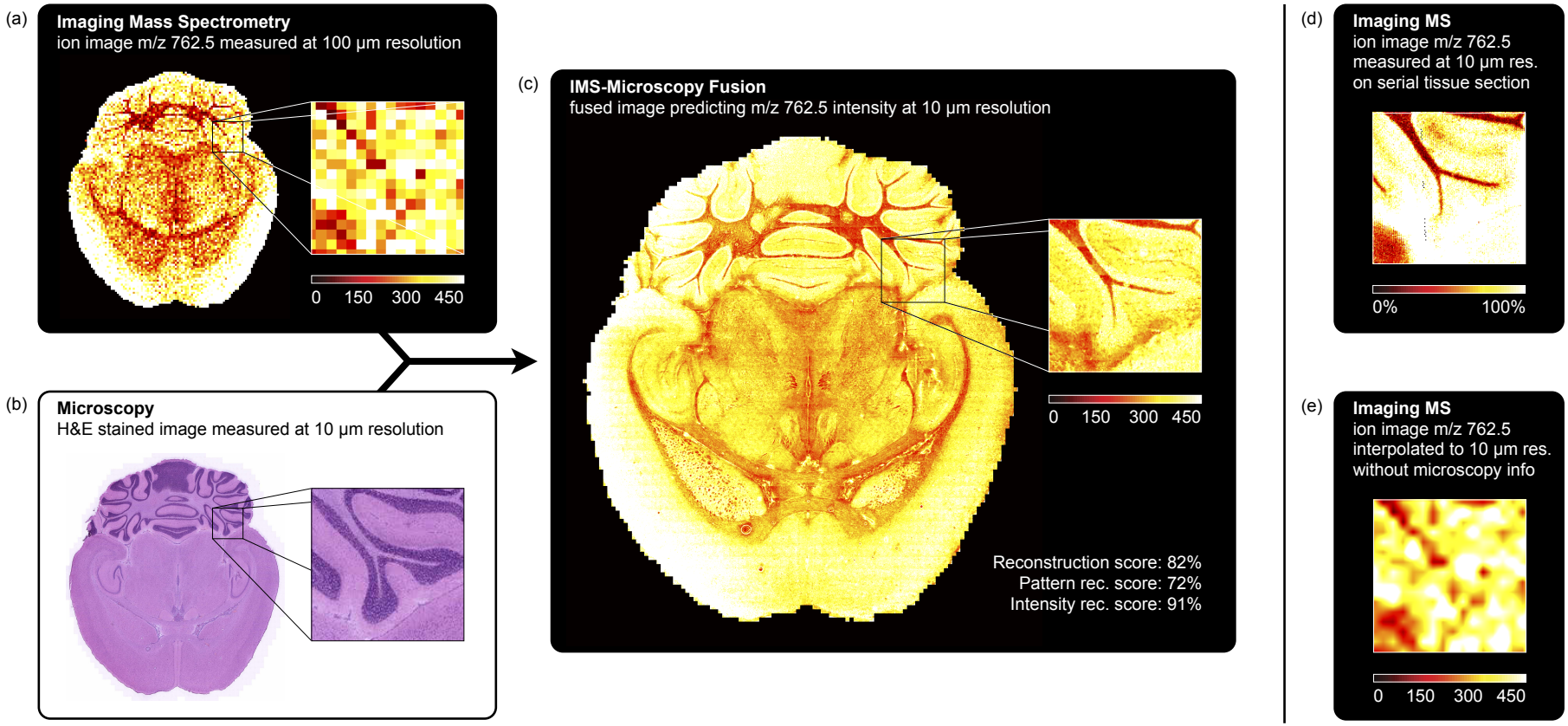


Supplementary Figure 2 Extensive step-by-step details of the fusion process. In the two-phase fusion method, phase I focuses on building and evaluating a cross-modality model between the provided modalities. It entails a transformation of the microscopy variables (step 1a) and the IMS variables (step 1b), a spatial mapping of both measurement sets (step 2), building the model (step 3), and evaluating model performance both chemically and spatially (step 4). Phase II employs the cross-modality model in a predictive application, and entails a prediction for all IMS variables (step 5) followed by a pruning of IMS variables for which predictive performance is insufficient (step 6).

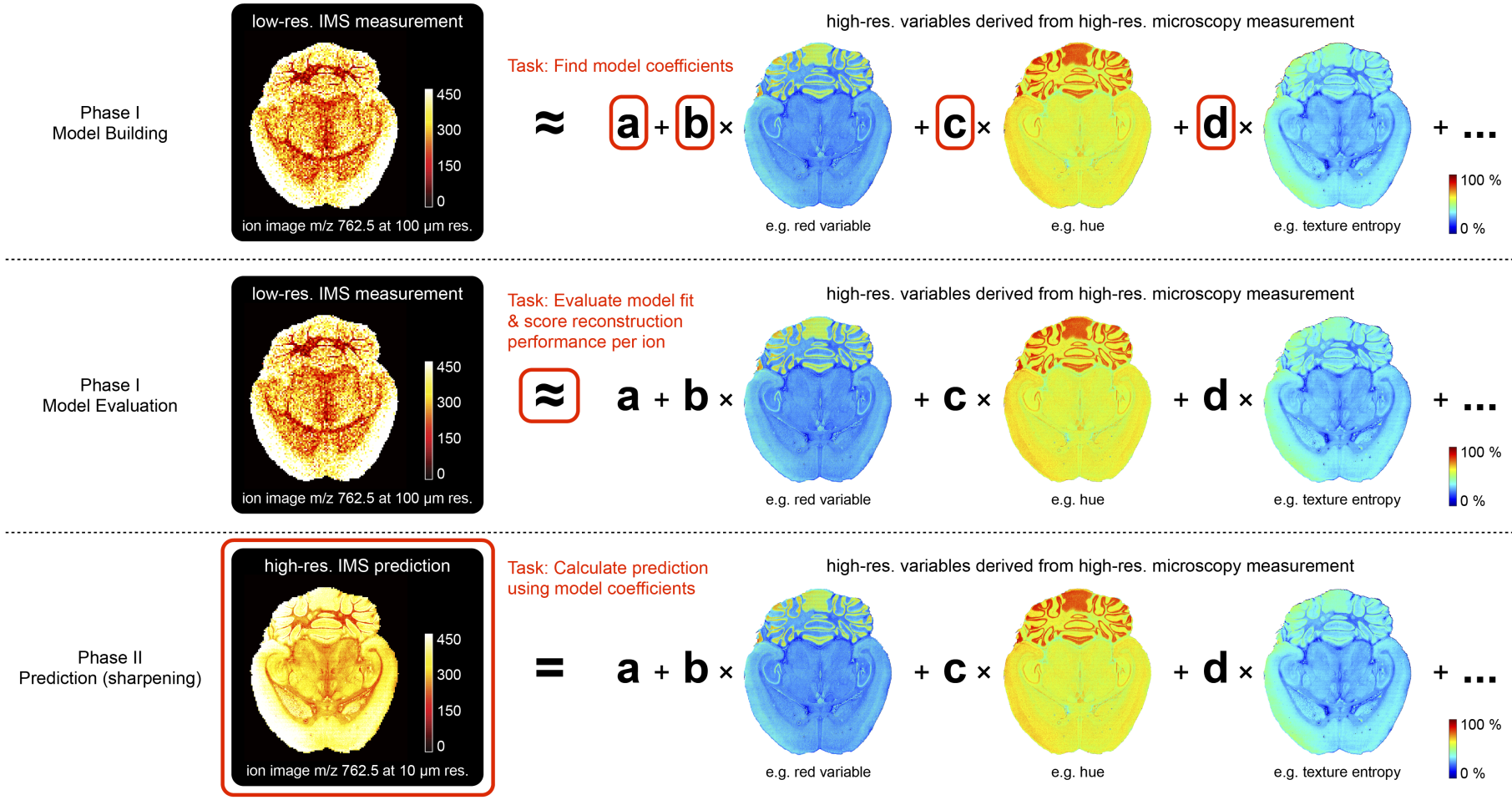
Nature Methods: doi:10.1038/nmeth.3296



Supplementary Figure 3 Method phases and steps with algebraic details. Algebraic details on the structure and size of the input and output data of each method step throughout the fusion procedure. P_{ims} = number of pixels in the IMS data source; U_{ims} = number of native variables per pixel provided by the IMS data source; N_{ims} = number of variables per pixel provided by the IMS data source after transformation; P_{micro} = number of pixels in the microscopy data source; U_{micro} = number of native variables per pixel provided by the microscopy data source; N_{micro} = number of variables per pixel provided by the microscopy data source after transformation; and M = number of mapped IMS and microscopy signatures.

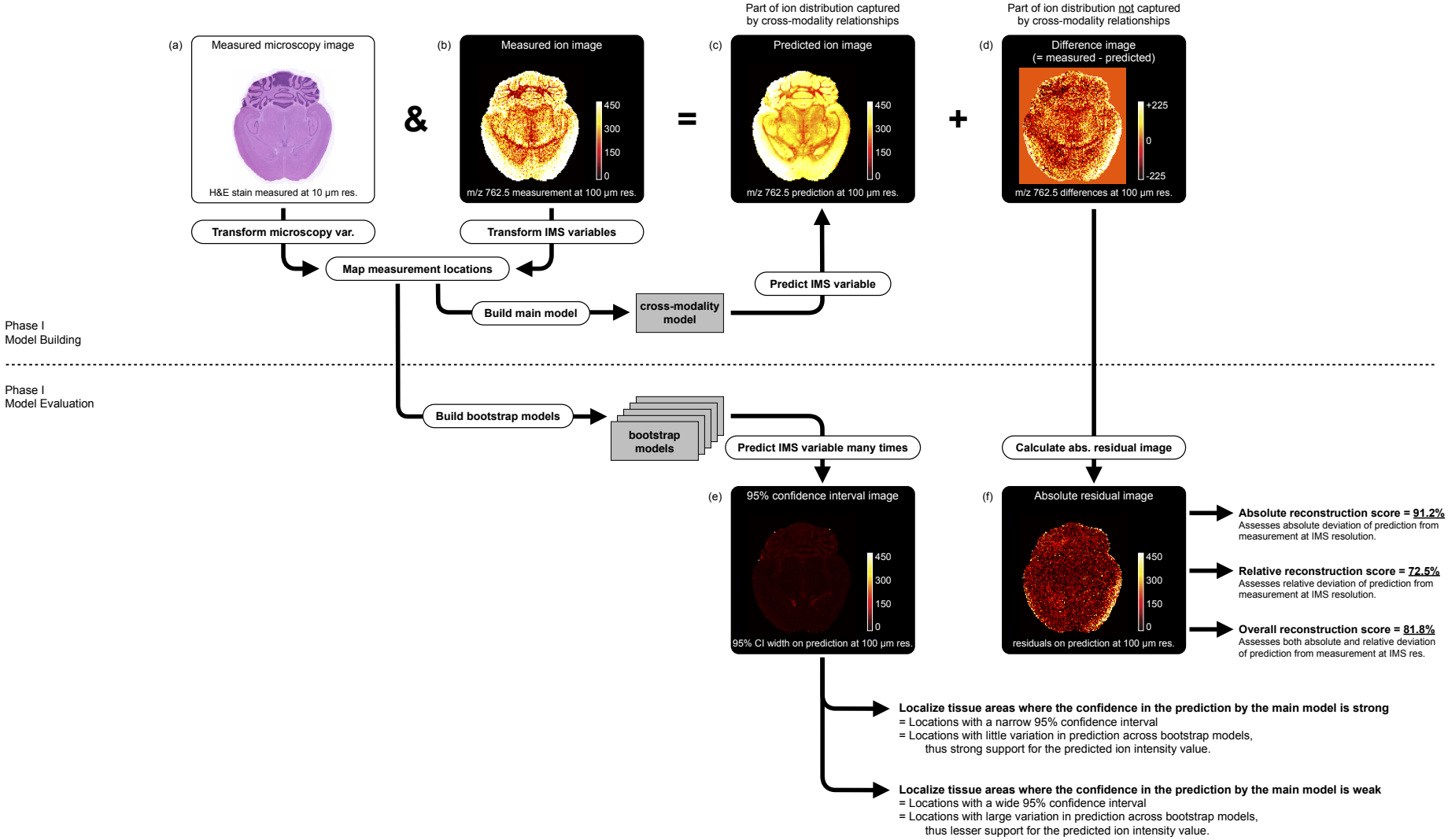


Supplementary Figure 4 Prediction of the ion distribution of *m/z* 762.5 in mouse brain at 10 µm resolution from 100 µm IMS and 10 µm microscopy measurements (sharpening). This example in mouse brain fuses a measured ion image for *m/z* 762.5 (identified as lipid PE(16:0/22:6)) at 100 µm spatial resolution (**a**) with a measured H&E-stained microscopy image at 10 µm resolution (**b**), predicting the ion distribution of *m/z* 762.5 at 10 µm resolution (reconstr. score 82%) (**c**). For comparison, (**d**) shows a measured ion image for *m/z* 762.5 at 10 µm spatial resolution, acquired from a neighboring tissue section. Additionally, (**e**) shows a 10 µm version of the *m/z* 762.5 ion image obtained through bilinear interpolation, a computational up-sampling method that does not employ information from another modality to guide its estimates.

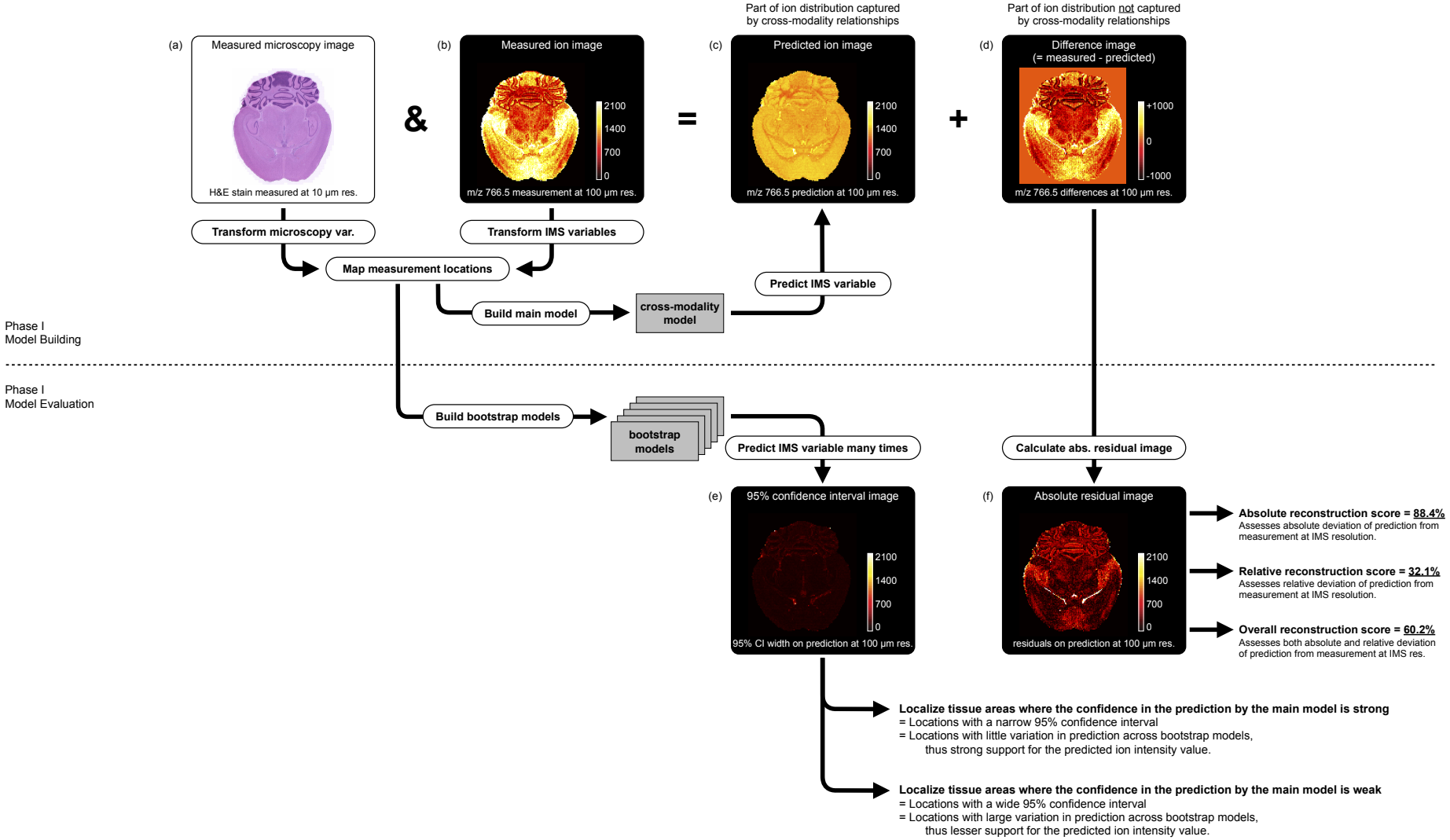


Supplementary Figure 5 Modeling details - Each IMS variable is approximated by a linear combination of microscopy-derived variables. (**top**) Model building step: the best linear equation coefficients are calculated. (**middle**) Model evaluation step: the fit of the linear sub-model to the measurements is determined and summarized as a reconstruction score. (**bottom**) Sharpening-specific prediction step: For those IMS variables with a high reconstruction score, apply the linear equation on the high-resolution microscopy-derived variables and calculate a high-resolution ion intensity prediction.

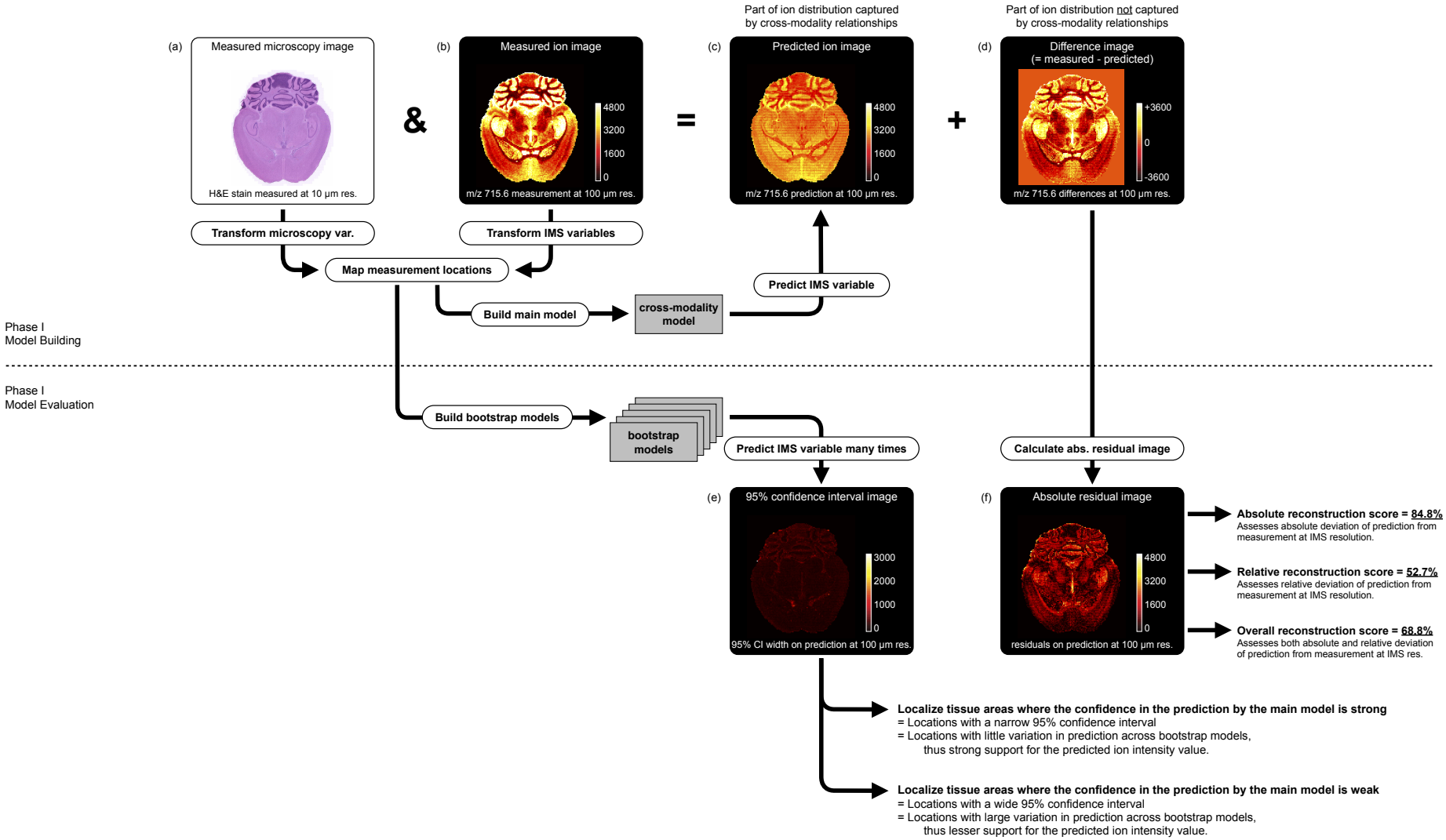
Nature Methods: doi:10.1038/nmeth.3296



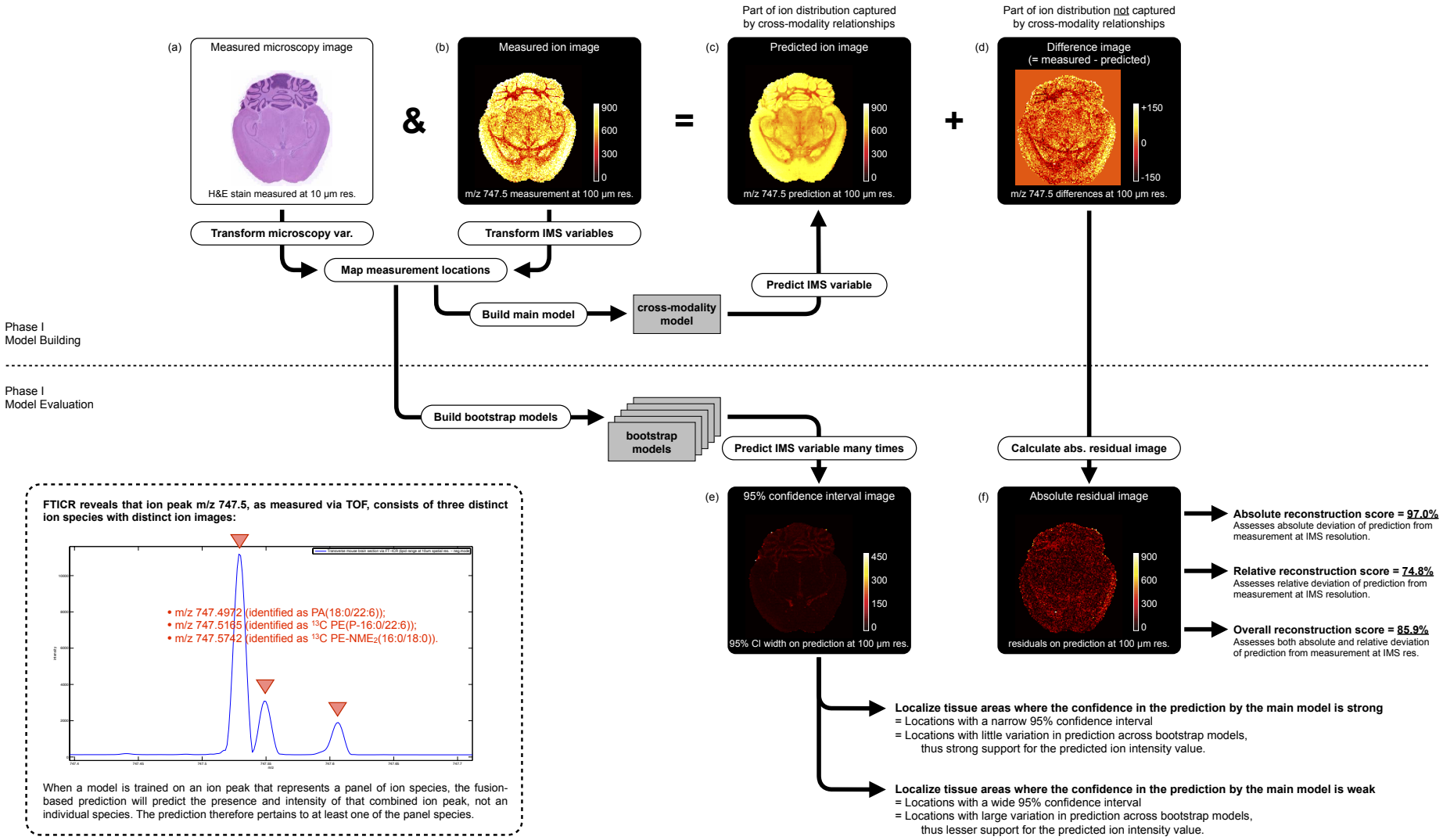
Supplementary Figure 6 Model building and evaluation for *m/z* 762.5. (**top**) Model building delivers for each IMS variable linear model coefficients (**Supplementary Figure 5**). The linear equation pertaining to *m/z* 762.5, its 'sub-model', separates the measured ion image into two images: a microscopy-predicted approximation (the cross-modally supported part of the *m/z* 762.5 measurement) and a difference image, which encodes the size and location of ion variation not captured by the sub-model (the IMS-specific part of the *m/z* 762.5 measurement). (**bottom**) Model evaluation assesses the sub-model strength for *m/z* 762.5 across the entire tissue section by summarizing the content of the difference image in a quality measure called the 'reconstruction score' (higher values mean better prediction). The location-specific prediction performance is reported by two evaluation images: the absolute residuals image (low valued areas are well predicted using microscopy) and the 95% confidence interval image (low valued areas are predicted robustly). All model evaluation happens at the native IMS resolution (100 µm).



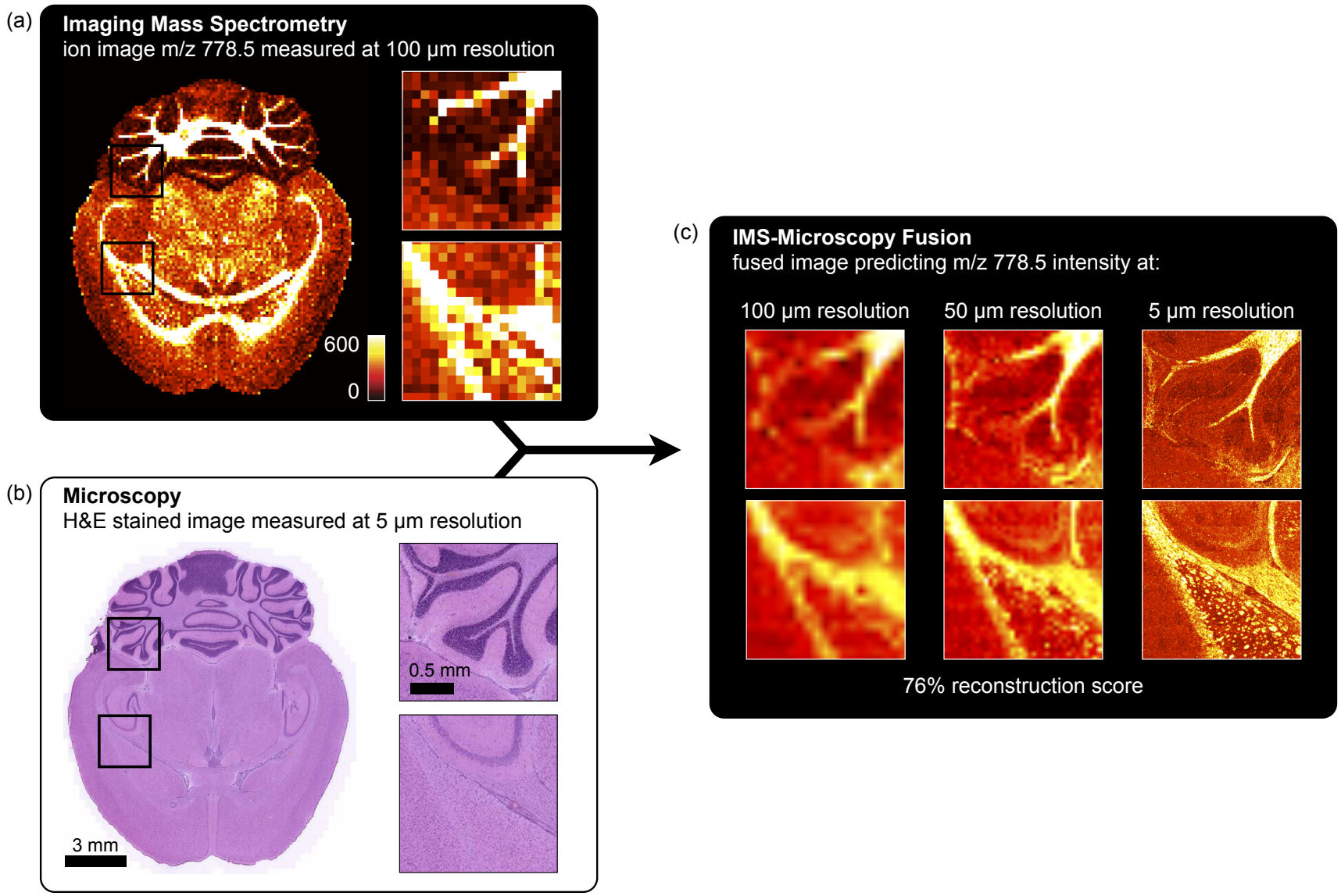
Supplementary Figure 7 Example of ion peak with low confidence cross-modality prediction, *m/z* 766.5 (identified as PE(18:0/20:4)). Although the average absolute peak intensity across the tissue is well approximated (absolute reconstruction score of 88%), the specific distribution pattern is not supported by the microscopy-derived patterns (relative reconstruction score of 32%). The overall reconstruction score therefore indicates a relatively low value of 60%, arguing against fusion-driven applications for this ion given the available microscopy measurements. (Further diagram details provided in **Supplementary Figure 6**).



Supplementary Figure 8 Example of ion peak with low confidence cross-modality prediction, *m/z* 715.6 (identified as PE-Cer(d16:1/22:0)). Although the average absolute peak intensity across the tissue is well approximated (absolute reconstruction score of 85%), the specific distribution pattern is not supported by the microscopy-derived patterns (relative reconstruction score of 53%). The overall reconstruction score therefore indicates a relatively low value of 69%, arguing against fusion-driven applications for this ion given the available microscopy measurements. (Further diagram details provided in **Supplementary Figure 6**).

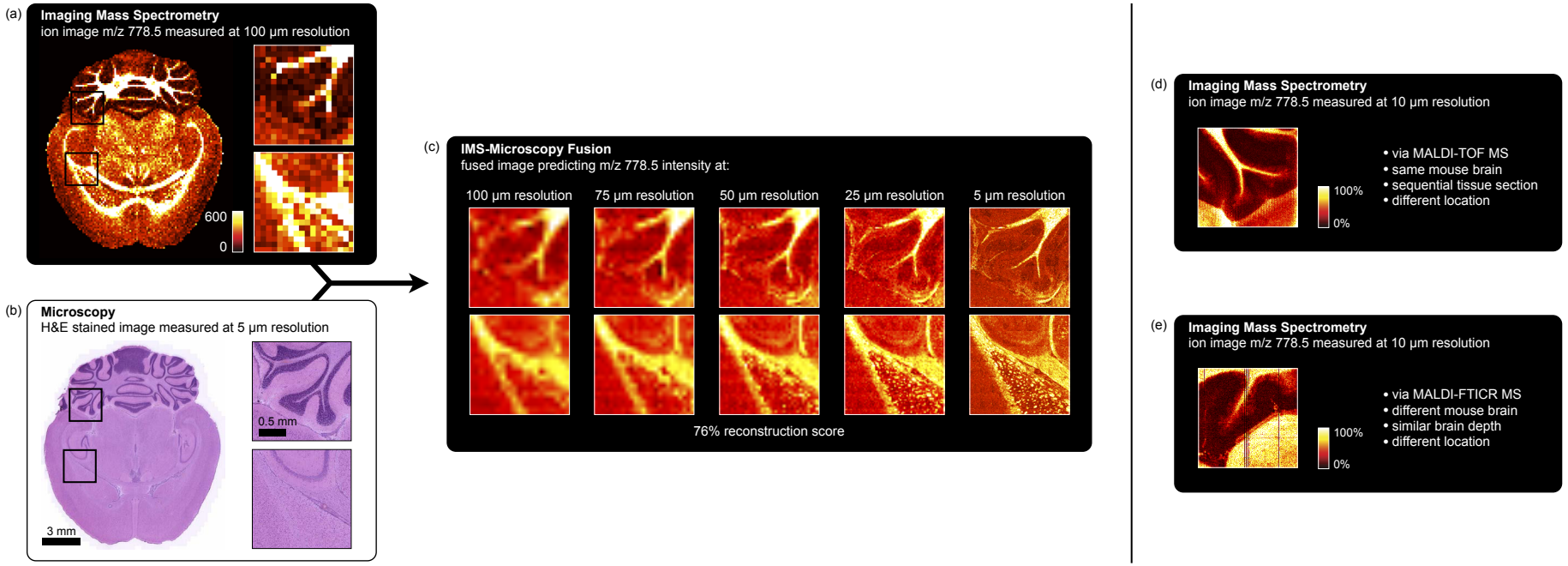


Supplementary Figure 9 Example of ion peak with high confidence cross-modality prediction, *m/z* 747.5. The high overall reconstruction score for this ion peak indicates that its tissue presence and intensity can be predicted with high confidence using H&E microscopy. However, through the advanced mass resolution provided by MALDI Fourier transform ion cyclotron resonance (FTICR) mass spectrometry, we know that ion peak *m/z* 747.5 is actually a superposition of multiple ion species (**inset**): ^{13}C PE-NME$_2$(16:0/18:0) at *m/z* 747.5742, ^{13}C PE(P-16:0/22:6) at *m/z* 747.5165, and PA(18:0/22:6) at *m/z* 747.4972. This example demonstrates that our method not only predicts for individual molecular species, but also works for a panel of species without requiring them to be uniquely resolved. (Further diagram details provided in **Supplementary Figure 6**).

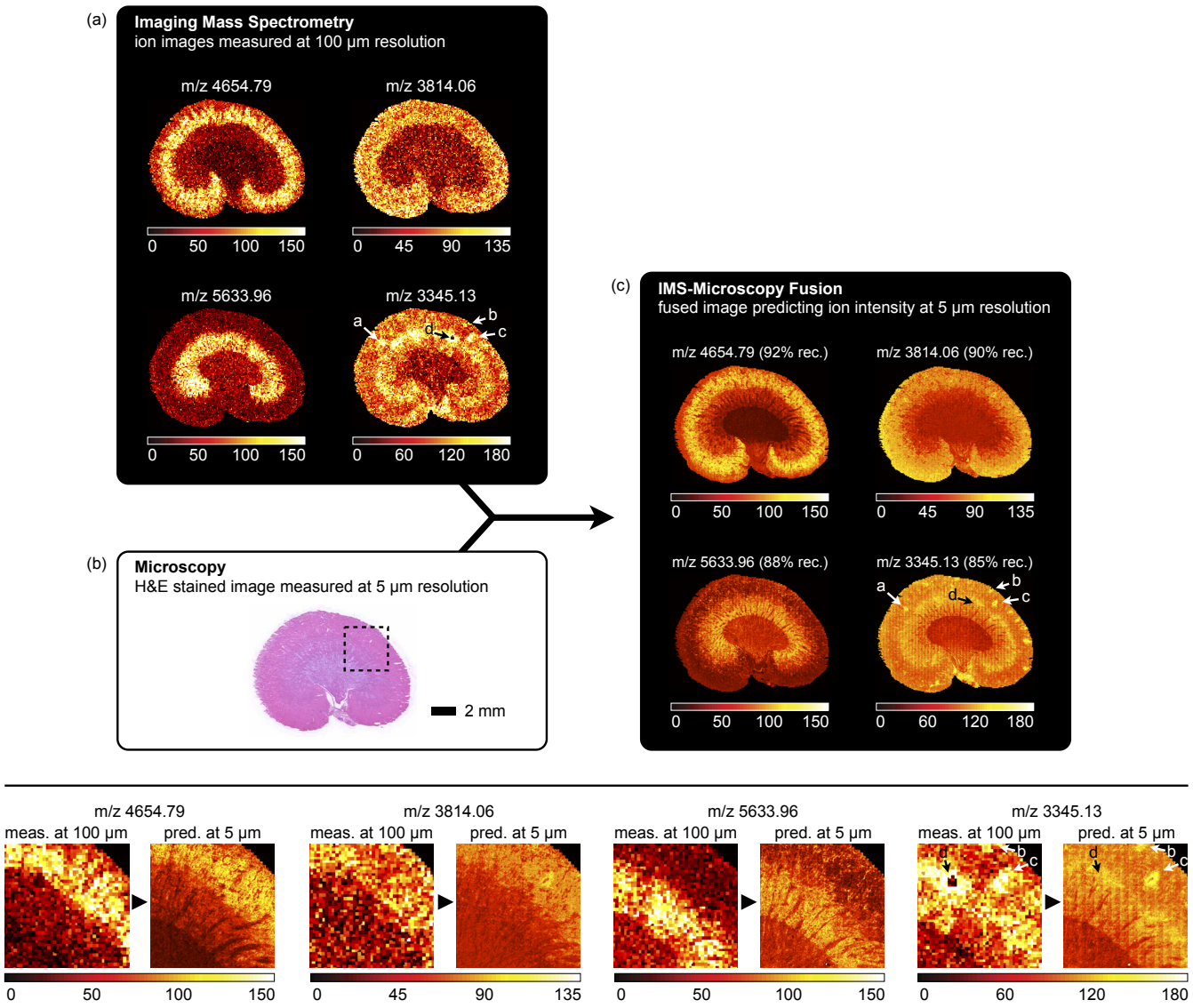


Supplementary Figure 10 Prediction of the ion distribution of *m/z* 778.5 in mouse brain at different target resolutions from 100 µm IMS and 5 µm microscopy measurements (sharpening). An IMS-microscopy model fuses information from an ion image for *m/z* 778.5 (identified as lipid PE(P-40:4)) measured at 100 µm spatial resolution (**a**), with that of an H&E-stained microscopy image measured at 5 µm resolution (**b**) (reconstr. score 76%). Combined with the microscopy measurements, the fusion model is then used to predict the ion distribution of *m/z* 778.5 at 100, 50, and 5 µm resolution (**c**).

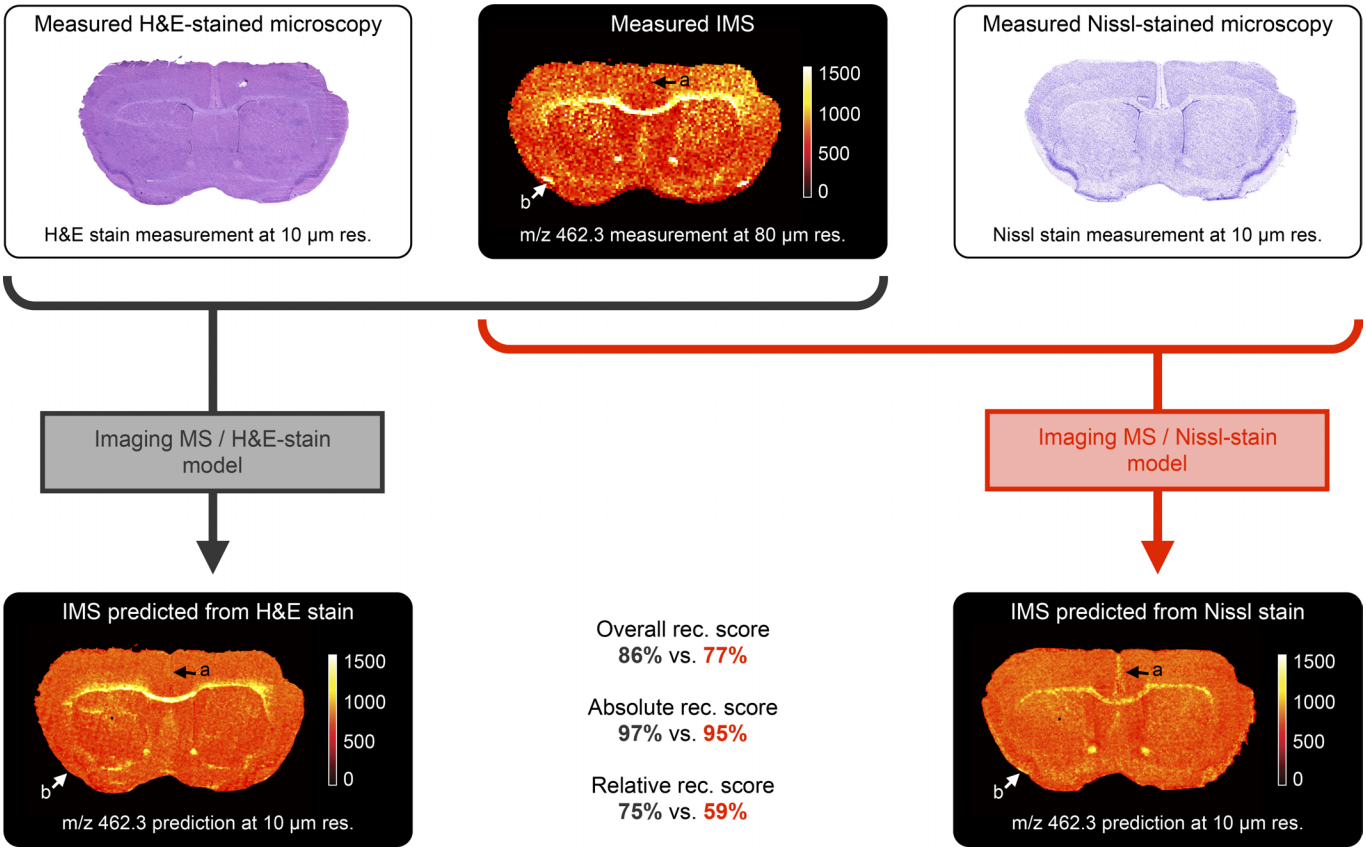
Nature Methods: doi:10.1038/nmeth.3296



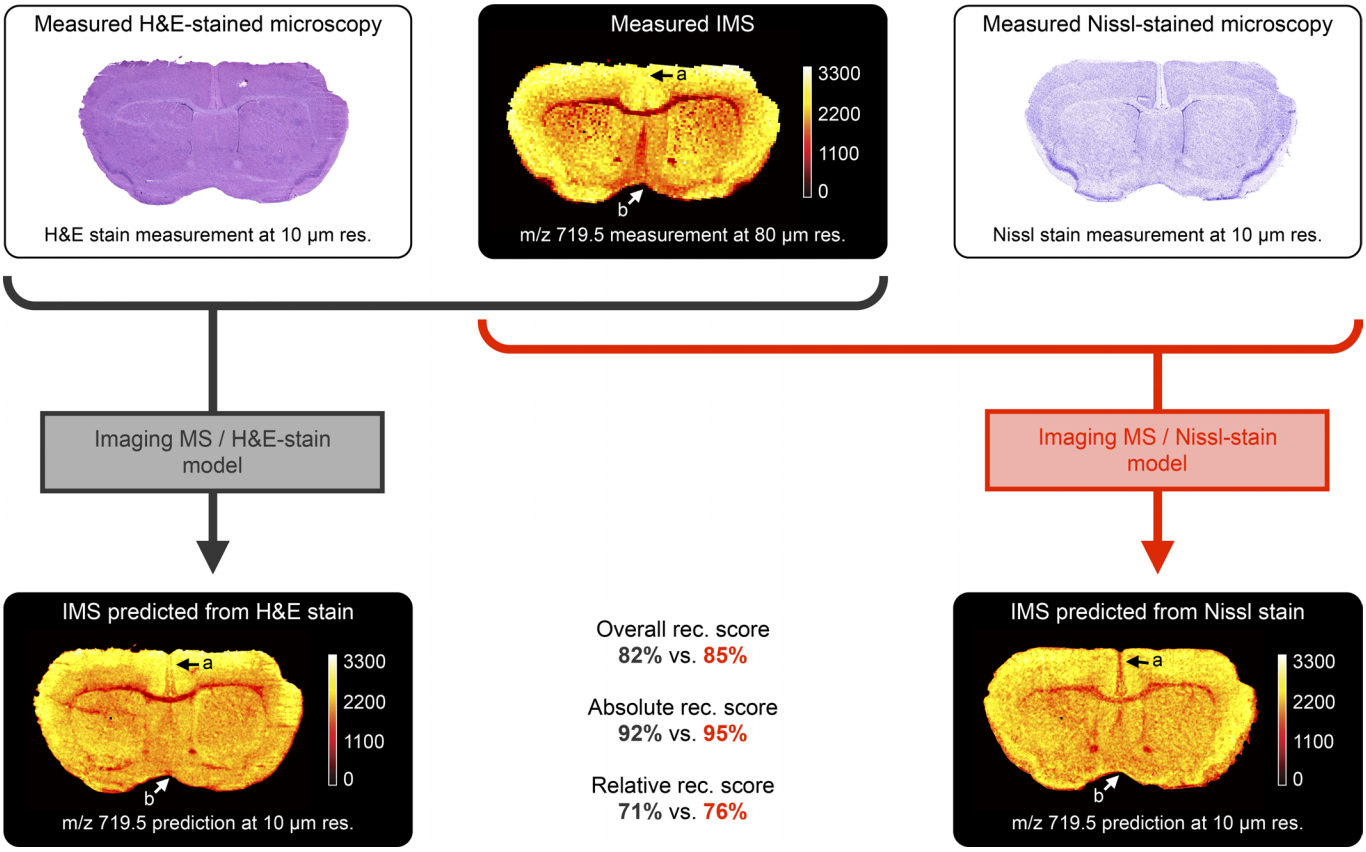
Supplementary Figure 11 Prediction of the ion distribution of *m/z* 778.5 in mouse brain at different target resolutions, with comparison to measured TOF and FTICR ion images (sharpening). An IMS-microscopy model fuses information from an ion image for *m/z* 778.5 measured at 100 µm spatial resolution (**a**), with that of an H&E-stained microscopy image measured at 5 µm resolution (**b**) (reconstr. score 76%). Combined with the microscopy measurements, the fusion model is then used to predict the ion distribution of *m/z* 778.5 at 100, 75, 50, 25, and 5 µm resolution (**c**). For comparison: an ion image for *m/z* 778.5 measured at 10 µm resolution by TOF-based IMS from a neighboring tissue section (**d**), and an ion image for *m/z* 778.5 measured at 10 µm resolution by FTICR-based IMS from a different mouse brain at a similar brain depth (**e**).



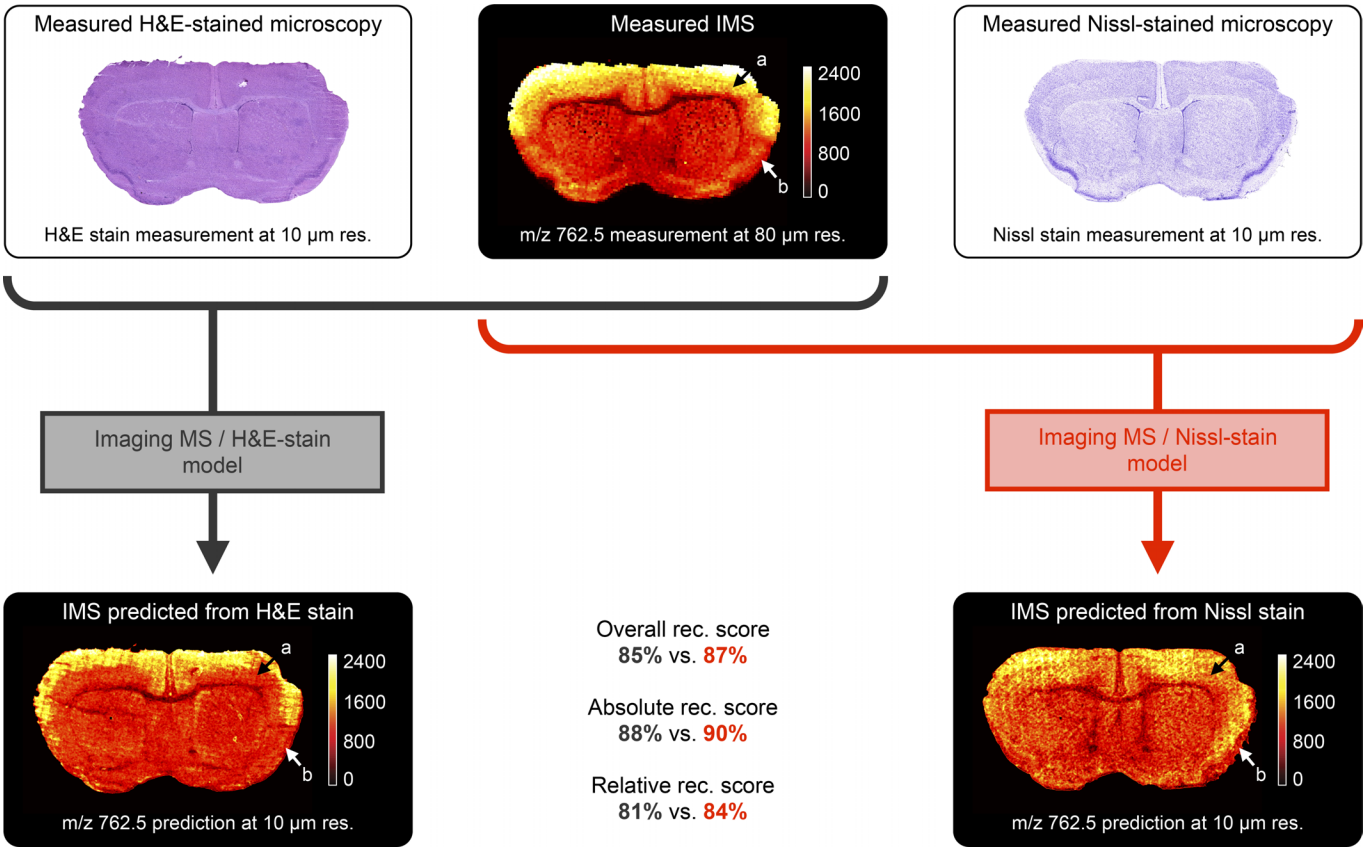
Supplementary Figure 12 Prediction of the ion distributions in rat kidney for *m/z* 4,655, *m/z* 3,814, *m/z* 5,634, and *m/z* 3,345 at 5 µm resolution (sharpening & enrichment). This example in rat kidney fuses a measured ion images acquired via IMS at 100 µm spatial resolution (**a**) with a measured H&E-stained microscopy image at 5 µm resolution (**b**), predicting the ion distributions at 5 µm resolution (**c**). The IMS-microscopy model achieves for *m/z* 4,655, *m/z* 3,814, *m/z* 5,634, and *m/z* 3,345 an overall reconstruction score of respectively 92%, 90%, 88%, and 85% at the native IMS resolution (100 µm). (**bottom**) Enlarged views of the measured 100 µm and predicted 5 µm ion distributions. Annotations *a*, *b*, and *c* are examples of ion image features that could be considered matrix homogeneity artifacts if only IMS is considered, but the fusion procedure shows that they have a support base in the microscopy as well. These features are thus corroborated across different technologies, and are more likely to be of genuine tissue origin than apparent from IMS measurements alone. Annotation *d* is an example of an ion image feature that is not supported by microscopy, and thus appears to be a modality-specific feature (and potential IMS artifact).



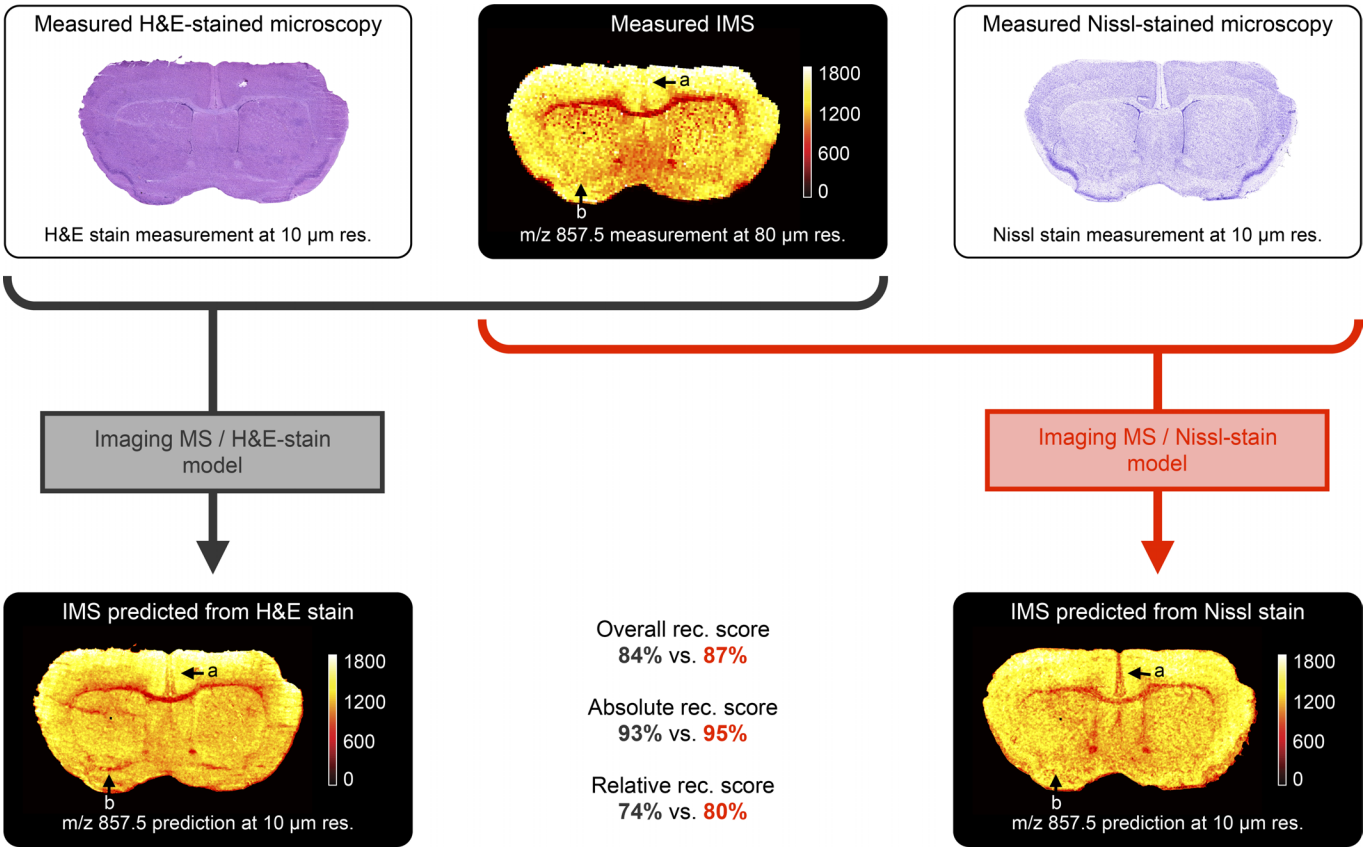
Supplementary Figure 13 Prediction of the ion distribution of *m/z* 462.3 in mouse brain at 10 µm resolution from 80 µm IMS and 10 µm H&E versus Nissl stained microscopy measurements (sharpening). (**left**) A measured ion image for *m/z* 462.3 at 80 µm spatial resolution is fused with a measured H&E stained microscopy image at 10 µm resolution, predicting the ion distribution of *m/z* 462.3 at 10 µm resolution (reconstr. score 86%). (**right**) An identical fusion procedure using a measured Nissl stained microscopy image at 10 µm resolution, instead of the H&E stain, delivers an ion distribution prediction for *m/z* 462.3 at 10 µm resolution with a reconstruction score of 77% at the native IMS resolution (80 µm).



Supplementary Figure 14 Prediction of the ion distribution of *m/z* 719.5 in mouse brain at 10 µm resolution from 80 µm IMS and 10 µm H&E versus Nissl stained microscopy measurements (sharpening). (**left**) A measured ion image for *m/z* 719.5 at 80 µm spatial resolution is fused with a measured H&E stained microscopy image at 10 µm resolution, predicting the ion distribution of *m/z* 719.5 at 10 µm resolution (reconstr. score 82%). (**right**) An identical fusion procedure using a measured Nissl stained microscopy image at 10 µm resolution, instead of the H&E stain, delivers an ion distribution prediction for *m/z* 719.5 at 10 µm resolution with a reconstruction score of 85% at the native IMS resolution (80 µm). Ion *m/z* 719.5 is identified as PA(38:6).

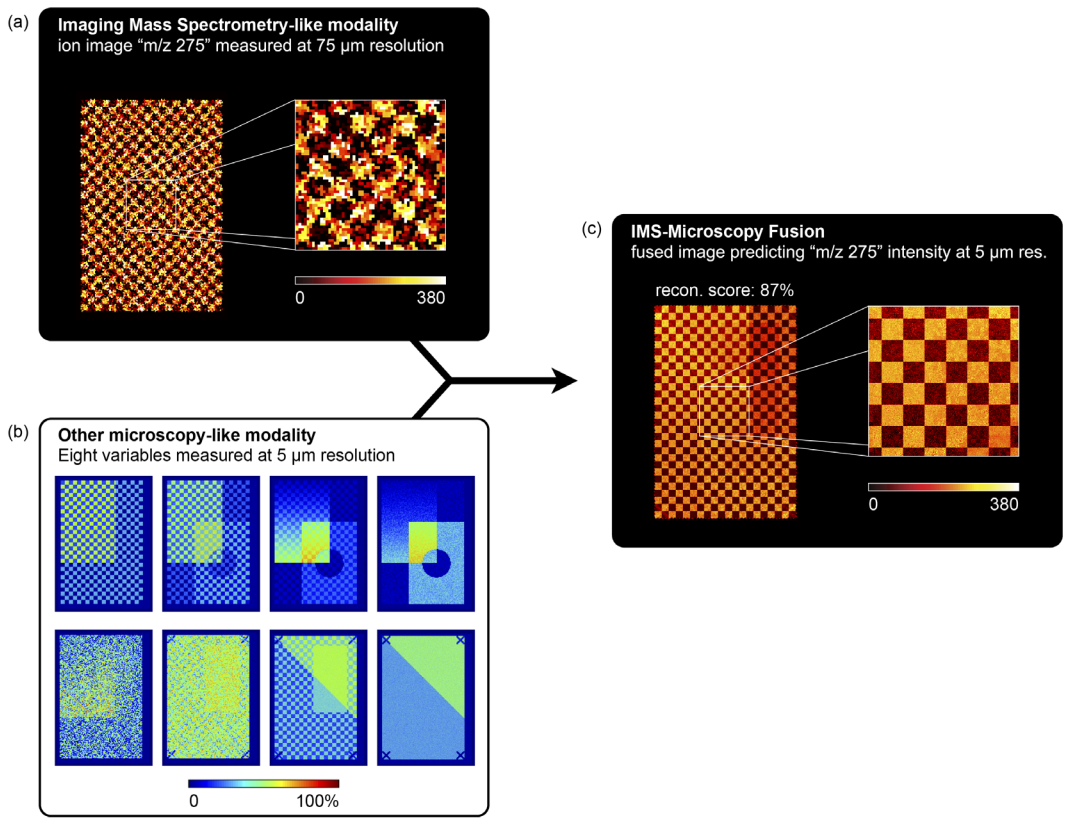


Supplementary Figure 15 Prediction of the ion distribution of *m/z* 762.5 in mouse brain at 10 µm resolution from 80 µm IMS and 10 µm H&E versus Nissl stained microscopy measurements (sharpening). (**left**) A measured ion image for *m/z* 762.5 at 80 µm spatial resolution is fused with a measured H&E stained microscopy image at 10 µm resolution, predicting the ion distribution of *m/z* 762.5 at 10 µm resolution (reconstr. score 85%). (**right**) An identical fusion procedure using a measured Nissl stained microscopy image at 10 µm resolution, instead of the H&E stain, delivers an ion distribution prediction for *m/z* 762.5 at 10 µm resolution with a reconstruction score of 87% at the native IMS resolution (80 µm). Ion *m/z* 762.5 is identified as PE(16:0/22:6).

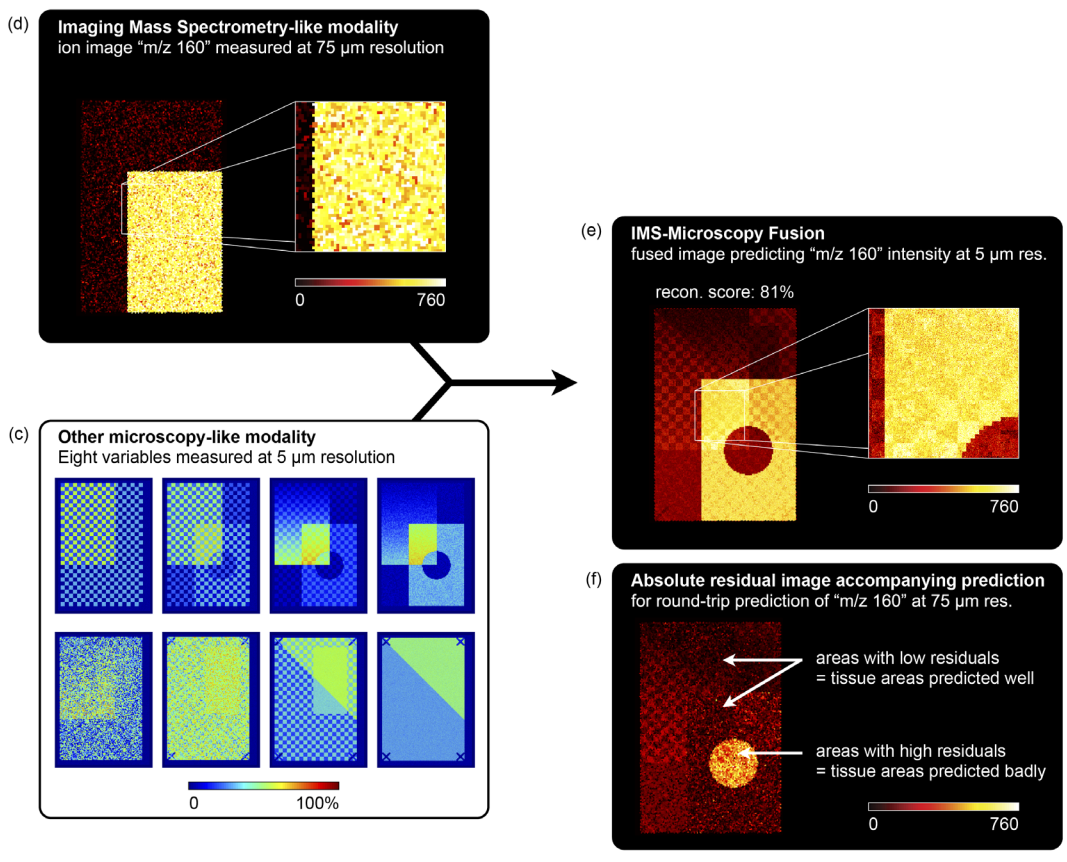


Supplementary Figure 16 Prediction of the ion distribution of *m/z* 857.5 in mouse brain at 10 µm resolution from 80 µm IMS and 10 µm H&E versus Nissl stained microscopy measurements (sharpening). (**left**) A measured ion image for *m/z* 857.5 at 80 µm spatial resolution is fused with a measured H&E stained microscopy image at 10 µm resolution, predicting the ion distribution of *m/z* 857.5 at 10 µm resolution (reconstr. score 84%). (**right**) An identical fusion procedure using a measured Nissl stained microscopy image at 10 µm resolution, instead of the H&E stain, delivers an ion distribution prediction for *m/z* 857.5 at 10 µm resolution with a reconstruction score of 87% at the native IMS resolution (80 µm). Ion *m/z* 857.5 is identified as PI(16:0/20:4).

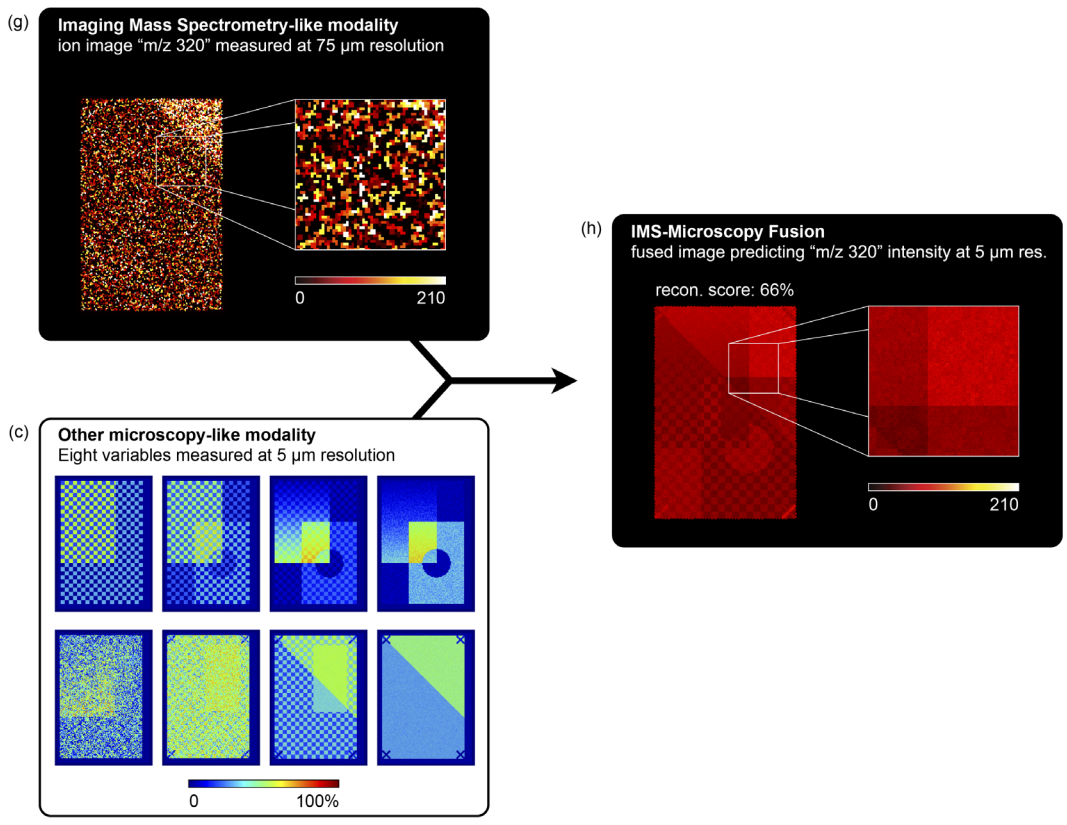
Fusion for cross-modally supported IMS pattern with strong multi-modal relationships (reconstruction score of 87%):



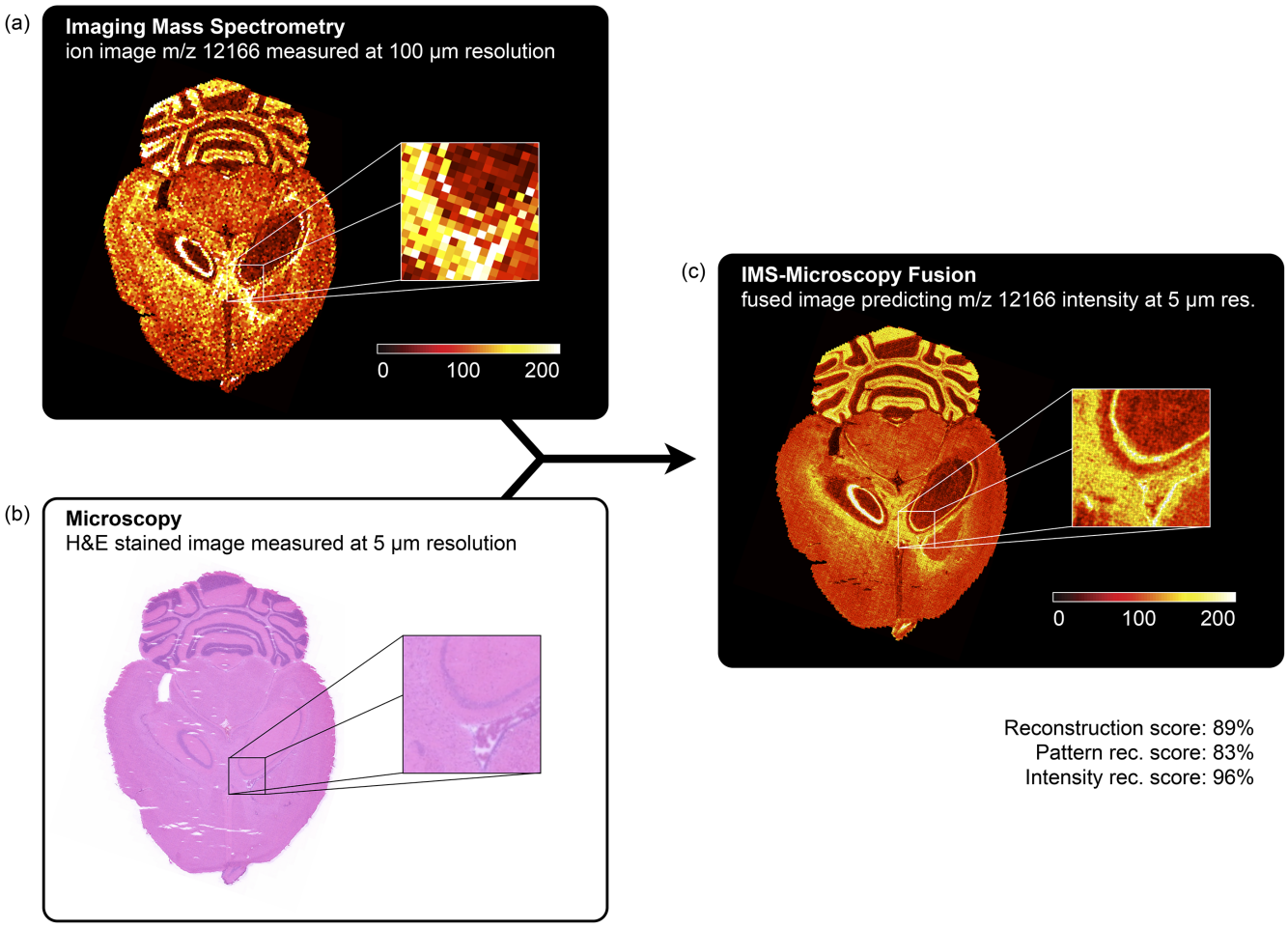
Fusion for cross-modally supported IMS pattern with weak global but strong local multi-modal relationships (recon. score of 81%):



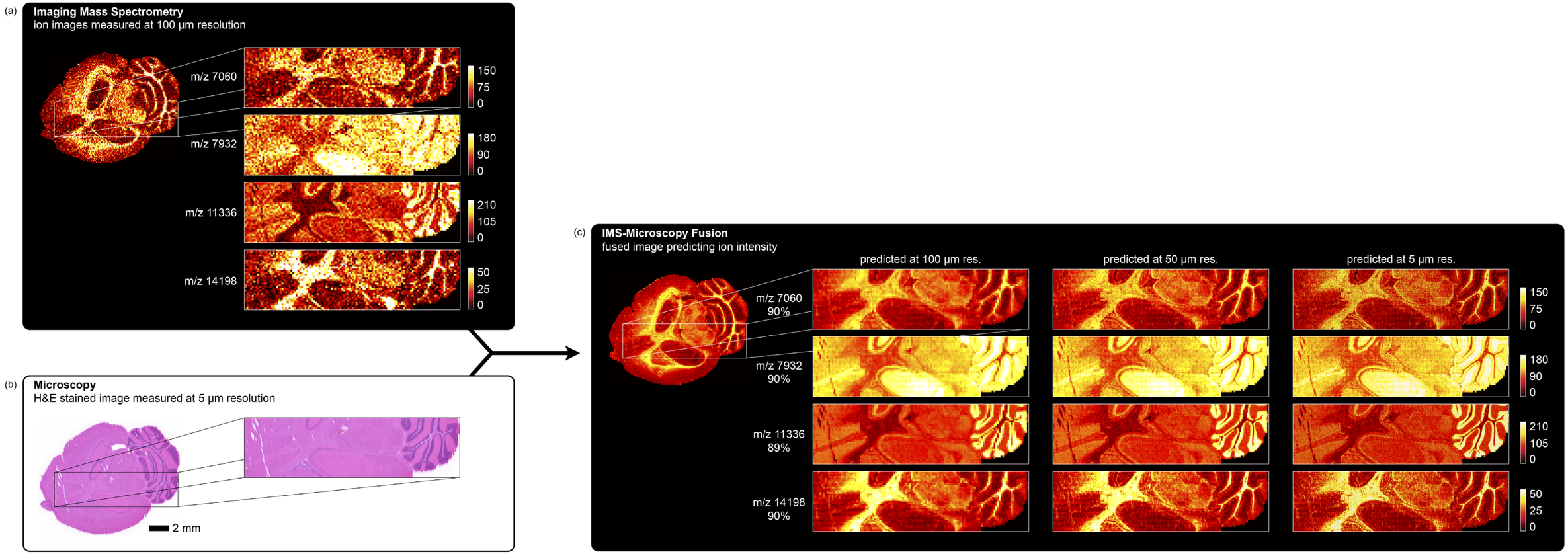
Fusion for modality-specific IMS pattern with no multi-modal relationships (reconstruction score of 66%):



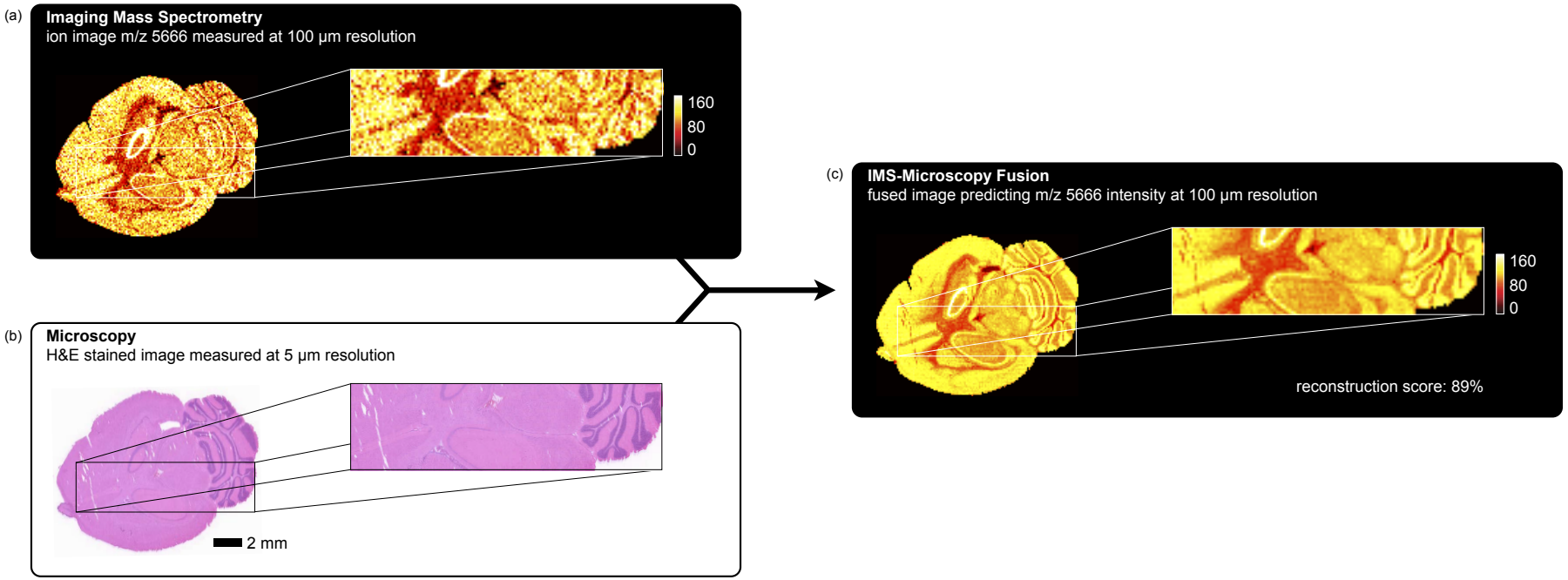
Supplementary Figure 17 Prediction of ion distributions from a synthetic multi-modal data set with known cross-modal and modality-specific patterns (sharpening). The fusion task consists of integrating an IMS-like modality at 75 µm spatial resolution with a microscopy-like modality acquired at 5 µm (both with noise added), and to sharpen the IMS-like patterns to 5 µm. The fusion method successfully finds all embedded cross-modal patterns and provides prediction for (**top**) a pattern with strong cross-modal support across the entire tissue (reconstr. score 87%), (**middle**) a pattern with partial cross-modal support (reconstr. score 81%, the location of good prediction is pinpointed via the absolute residuals image), and (**bottom**) a modality-specific pattern with little to no cross-modal support (reconstr. score 66%).



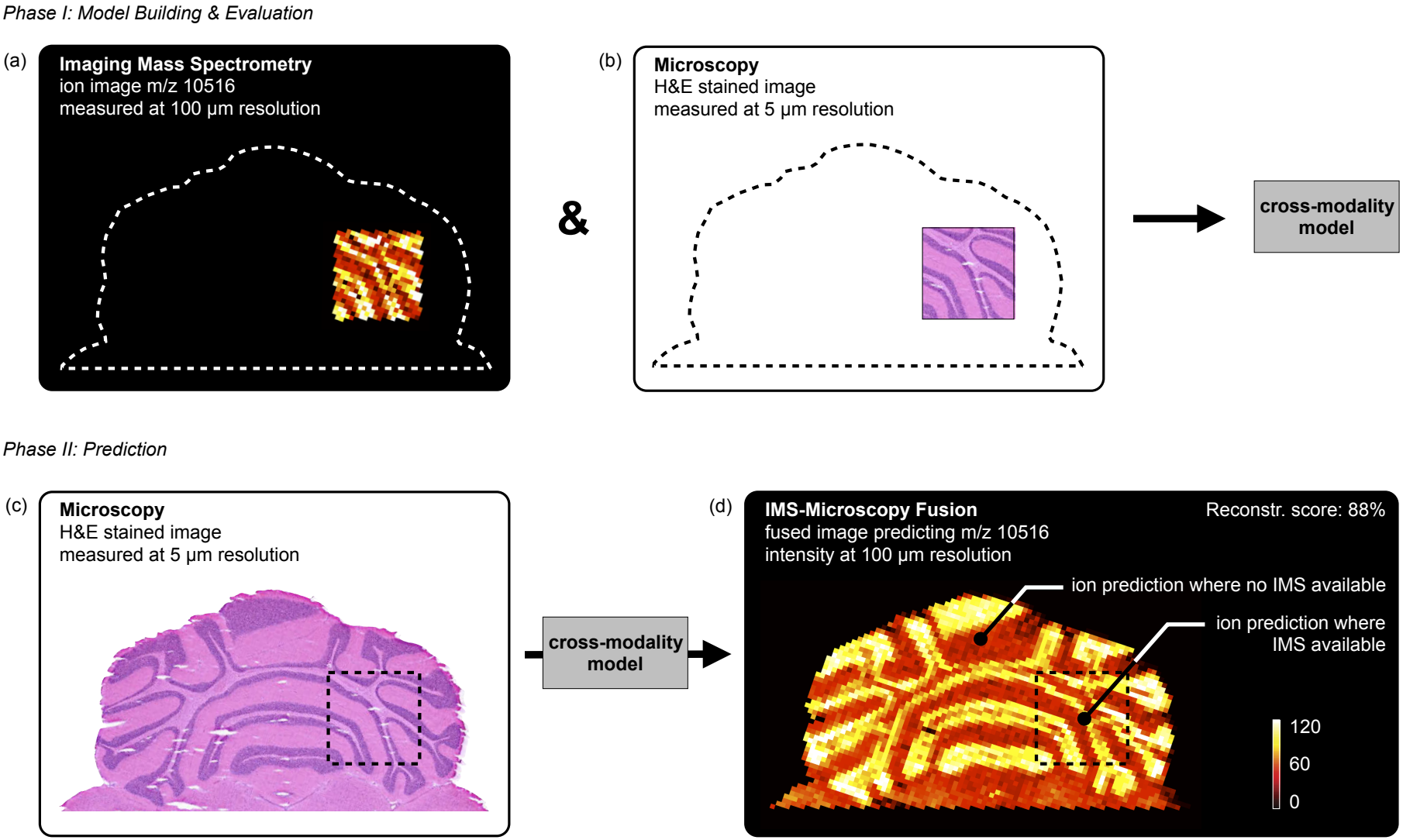
Supplementary Figure 18 Prediction of the ion distribution of *m/z* 12,166 in mouse brain at 5 µm resolution from 100 µm IMS and 5 µm microscopy measurements (sharpening). This example in mouse brain fuses a measured ion image for *m/z* 12,166 at 100 µm spatial resolution (**a**) with a measured H&E-stained microscopy image at 5 µm resolution (**b**), predicting the ion distribution of *m/z* 12,166 at 5 µm resolution (reconstr. score 89%) (**c**).



Supplementary Figure 19 Prediction of the ion distributions of *m/z* 7,060, 7,932, 11,336, and 14,198 in mouse brain at different target resolutions from 100 µm IMS and 5 µm microscopy measurements (sharpening). An IMS-microscopy model fuses information from ion images for *m/z* 7,060, 7,932, 11,336, and 14,198, measured at 100 µm spatial resolution (**a**), with that of an H&E-stained microscopy image measured at 5 µm resolution (**b**). Combined with the microscopy measurements, the fusion model is then used to predict the ion distribution of *m/z* 7,060, 7,932, 11,336, and 14,198 at 100, 50, and 5 µm resolution (**c**) (reconstr. score 90%, 90%, 89%, and 90% respectively).



Supplementary Figure 20 De-noising of the ion distribution of *m/z* 5,666 in mouse brain through round-trip prediction from 100 µm IMS to 100 µm fused resolution, using fusion to 5 µm microscopy measurements as a filter (de-noising). This example in mouse brain fuses a measured ion image for *m/z* 5,666 at 100 µm spatial resolution (**a**) with a measured H&E-stained microscopy image at 5 µm resolution (**b**), predicting the ion distribution of *m/z* 5,666 at 100 µm resolution (reconstr. score 89%) (**c**). No sharpening is pursued. The objective is to use fusion to employ microscopy measurements as a filter. This filtering application retains cross-modal (tissue) variation and removes modality-specific variation.



Supplementary Figure 21 Prediction of *m/z* 10,516 distribution at native IMS resolution (without sharpening) in mouse brain areas not measured by IMS. An IMS-microscopy model is built on a tissue sub-area for which IMS is available at 100 µm resolution (**a**) and H&E-stained microscopy is available at 5 µm resolution (**b**). The model is then used to predict the distribution of *m/z* 10,516 at 100 µm resolution in areas where no IMS was acquired and only microscopy is available (reconstr. score 88%) (**d**).

Supplementary Tables

Supplementary Table 1 Case study overview. Further details are provided in the Online Methods, section Case Study Materials and Methods.

Case study	Tissue type	Measured microscopy modality		Measured IMS modality		Predicted IMS-microscopy modality		Demonstrates	Figures
		type (staining)	pixel width (finest used) (µm)	type (focus)	pixel width (µm)	type (focus)	pixel width (µm)		
1	mouse brain (transversal)	H&E	5	MALDI-TOF (lipids) m/z 500 - 1000	100	MALDI-TOF (lipids) m/z 500 - 1000	10 (Fig. 2-3, S1-2, S4-5) 100 (Fig. S6-9) 100, 50, 5 (Fig. S10) 100, 75, 50, 25, 5 (Fig. S11)	modeling and prediction method workflow prediction vs. measurement prediction at different spatial resolutions	2-3, S1-2, S4-11
2	rat kidney	H&E	5	MALDI-TOF (proteins) m/z 3000 - 25000	100	MALDI-TOF (proteins) m/z 3000 - 25000	5 (Fig. 6, S12)	prediction in different tissue types prediction of different molecule types multi-modal enrichment	6, S12
3	mouse brain (transversal)	H&E	0.33	MALDI-TOF (lipids) m/z 500 - 1000	10	MALDI-TOF (lipids) m/z 500 - 1000	0.33 (Fig. 4)	prediction at sub-micron scales prediction beyond the capabilities of single modality	4
4	mouse brain (coronal)	H&E and Nissl	10	MALDI-TOF (lipids) m/z 400 - 1000	80	MALDI-TOF (lipids) m/z 400 - 1000	10 (Fig. S13-16)	prediction using different data sources	S13-16
5	synthetic	synthetic	5	synthetic variables '1' - '1000'	75	synthetic variables '1' - '1000'	5 (Fig. S17)	modeling algorithm check	S17
6	mouse brain (transversal)	H&E	5	MALDI-TOF (proteins) m/z 3000 - 20000	100	MALDI-TOF (proteins) m/z 3000 - 20000	5 (Fig. 1, S18) 100, 50, 5 (Fig. S19) 100 (Fig. S20)	multi-modal enrichment and de-noising	1, S18-20
7	mouse brain (transversal)	H&E	5 (same as case study 6)	MALDI-TOF (proteins) m/z 3000 - 20000	100 (rectangular sub-area of case study 6)	MALDI-TOF (proteins) m/z 3000 - 20000	5 (Fig. 5) 100 (Fig. S21)	prediction in non-IMS-measured areas	5, S21

Supplementary Notes

These Supplementary Notes describe the mathematical details of the image fusion framework. Supplementary Note 1 addresses the types of inputs the framework can accept, and discusses how these cover different imaging modalities. Supplementary Note 2 covers the model building and evaluation phase of the image fusion process, and provides details on the five different steps that make up this phase. Supplementary Note 3 addresses the prediction phase and its two sub-steps, while Supplementary Note 4 examines specific target applications and their underlying assumptions. The mathematical definitions first establish general framework implementations, independent of which image sources are being considered. These are then followed by IMS and microscopy-specific definitions that incorporate modality-specific considerations. The function definitions can be readily implemented using standard methods available in most algebraic environments such as MATLAB. Where a non-standard implementation is used, we provide pseudo-code.

Supplementary Note 1 – Inputs

The fusion procedure takes as inputs two image data sources and a function g that describes how these sources are spatially registered to each other. In the examples of the paper, these inputs consist of two 2-D images of different technological origin and an affine transformation matrix to define how they are registered to each other. However, the mathematical framework of the fusion method is generic and capable of handling any data source and/or any registration approach that adheres to the definitions below. This includes any 2-D or 3-D imaging modalities, and any rigid or non-rigid registration approaches.

Image data sources. Most imaging modalities natively deliver their measurements as a multidimensional array of values that can be accessed via n indices, a structure also known as an n-mode array or a tensor of order n. Such an nth order tensor of image data can be defined as $\mathcal{D} \in \mathbb{R}^{I_1 \times \ldots \times I_n}$, where $\mathbb{R}$ indicates the real numbers and each element in the image can be accessed as $d_{i_1 \ldots i_n}$ with $i_k \in \mathbb{N}$ an index along the kth mode with a value between one and I_k.

An imaging modality that records along n_s spatial dimensions and acquires an n_m-dimensional array of values at each measurement location, typically delivers its data as a tensor of order $n = n_\mathrm{s} + n_\mathrm{m}$. Depending on the nature of the imaging technology, values for n_s can vary from zero (single measurement; trivial case), to one (list of measurements; e.g. MS profiling), to two (2-D imaging; e.g. microscopy, IMS), to three (3-D imaging; e.g. MRI), and beyond. The number of measurement modes n_m, or $n - n_\mathrm{s}$, will typically range from zero (scalar value measured; e.g. gray-level image, profilometry), to one (vector of values measured; e.g. color image, MS), to two (matrix of values measured; e.g. ion mobility MS), and so on.

In order to keep the fusion method broadly applicable and independent of the particular technologies under study, the image data is translated into a tabular format, where each row represents a measurement at a particular spatial location and each column represents a particular feature recorded across all spatial locations. This re-organization of the image data from the native nth order tensor to a second order tensor or matrix can be accomplished by any 'flattening' or 'unfolding' procedure commonly available in algebra libraries (e.g. `reshape()` in MATLAB). The particular implementation of the flattening procedure is unimportant as long as long as it flattens each $(n - n_\mathrm{s})$-mode measurement to a row vector of features and it translates the n_s location indices (e.g. pixel/voxel locations) into a single index along a column vector (and this spatial translation can be reversed so fusion results can be shown as images afterwards). Note that this procedure does not change the content of the image data in any way,

but simply changes the way the content of the image is indexed and stored in memory. The re-organization not only makes the fusion method capable of accepting a broad variety of technologies, but also follows the objects-by-features table convention employed in most statistical texts, making the application of stochastic approaches more straightforward.

Definition 0.1. (image data source.) Let tensor $\mathcal{D}^{\text{mod}} \in \mathbb{R}^{I_1 \times \ldots \times I_{n_s} \times I_{n_s+1} \times \ldots \times I_n}$ denote an image with n_s spatial modes and $n - n_s$ measurement modes, obtained by employing imaging technology 'mod'. Let matrix $D^{\text{mod}} \in \mathbb{R}^{(\prod_{i=1}^{n_s} I_i) \times (\prod_{j=n_s+1}^{n} I_j)}$ denote an image data source, which is the tabular representation of the image $\mathcal{D}^{\text{mod}}$, obtained by flattening the n_s spatial modes to a single row index and the $n - n_s$ measurement modes to a single column index. If not all spatial locations that can be indexed by $I_1 \times \ldots \times I_{n_s}$ have been measured, only the measured rows are retained for analysis, making $D^{\text{mod}} \in \mathbb{R}^{P \times (\prod_{j=n_s+1}^{n} I_j)}$ with $P \leq (\prod_{i=1}^{n_s} I_i)$.

The image data sources used in the paper can now be specifically defined as follows.

Definition 0.2. (microscopy data source.) Let tensor $\mathcal{D}^{\text{micro}} \in \mathbb{R}^{R_{\text{micro}} \times C_{\text{micro}} \times U_{\text{micro}}}$ denote a microscopy image with R_{micro} rows and C_{micro} columns, and U_{micro} features measured per pixel. Let matrix $D^{\text{micro}} \in \mathbb{R}^{P_{\text{micro}} \times U_{\text{micro}}}$ with $P_{\text{micro}} \leq R_{\text{micro}} C_{\text{micro}}$ denote a microscopy data source obtained by flattening $\mathcal{D}^{\text{micro}}$ and retaining only the measured locations.

Definition 0.3. (IMS data source.) Let tensor $\mathcal{D}^{\text{ims}} \in \mathbb{R}^{R_{\text{ims}} \times C_{\text{ims}} \times U_{\text{ims}}}$ denote an IMS image with R_{ims} rows and C_{ims} columns, and U_{ims} features measured per pixel. Let matrix $D^{\text{ims}} \in \mathbb{R}^{P_{\text{ims}} \times U_{\text{ims}}}$ with $P_{\text{ims}} \leq R_{\text{ims}} C_{\text{ims}}$ denote an IMS data source obtained by flattening $\mathcal{D}^{\text{ims}}$ and retaining only the measured locations.

All microscopy images in the paper record a red, green, and blue intensity band, making $U_{\text{micro}} = 3$ in all case studies that feature microscopy. Since IMS measures a mass spectrum per pixel, the number of features U_{ims} in each IMS data source is equal to the number of *m/z* bins that were captured in the experiment. The microscopy data employed in this paper is natively always acquired at a spatial resolution of 0.33 μm, and it is subsequently down-sampled to the measurement resolution required for the case study at hand (as reported in **Supplementary Table 1**). Since this happens prior to any fusion process, the same equipment can thus be used to provide microscopy data sources (images) with different spatial resolutions. The spatial resolution of all IMS data sources is simply equal to the native measurement resolution at which the mass spectra were acquired.

Given that the ability to find cross-modality relationships will depend on the quality of the inputs, it is advisable to preprocess the image data sources before providing them to the fusion method. Any technology-specific noise that can be removed in a single-modality context prior to fusion will leave more bandwidth for cross-modality modeling. To this end, we applied standard MS preprocessing steps to the IMS data sources in this paper, going from raw IMS measurements in a data source $D_{\text{raw}}^{\text{ims}}$ to a preprocessed IMS data source $D_{\text{prepro}}^{\text{ims}}$. All IMS data sources in this paper have been baseline corrected (using `msbackadj()` from MATLAB), normalized[32], and aligned (using `mspalign()` from MATLAB). The microscopy data sources were used in their raw form, making $D_{\text{raw}}^{\text{micro}}$ and $D_{\text{prepro}}^{\text{micro}}$ equivalent, but also hinting at microscopy preprocessing as an area for further improvement.

As an example, in case study 6 (**Fig. 1**), microscopy data sources $D_{\text{raw}}^{\text{micro}}$ and $D_{\text{prepro}}^{\text{micro}}$ (natively acquired at 330 nm res., used at 5 μm res.) are equal and of size 4,716,671×3 with $R_{\text{micro}} = 3{,}230$, $C_{\text{micro}} = 2{,}485$, $P_{\text{micro}} = 4{,}716{,}671$, and $U_{\text{micro}} = 3$. IMS data source $D_{\text{raw}}^{\text{ims}}$ and its preprocessed version $D_{\text{prepro}}^{\text{ims}}$ (natively

at 100 μm res.) are of size $11{,}415\times13{,}728$ with $R_{\text{ims}} = 111$, $C_{\text{ims}} = 138$, $P_{\text{ims}} = 11{,}415$, and $U_{\text{ims}} = 13{,}728$.

Spatial registration of sources. Since fusion uses the spatial domain to map measurements from different technologies to each other, a registration function is an essential input to the fusion method. Fusion is independent of the implementation of the registration function as long as spatial coordinates can be translated back and forth between the two image data sources. As such the fusion method can accept any rigid or non-rigid registration methodology.

Definition 0.4. (registration function.) Let $g\colon \mathbb{R}^{n_S^A} \to \mathbb{R}^{n_S^B}$ denote a registration function from n_S^A-dimensional space to n_S^B-dimensional space, spatially registering image $\mathcal{D}^{\text{A}} \in \mathbb{R}^{I_1{}^A \times \ldots \times I_{n_S^A} \times I_{n_S^A+1} \times \ldots \times I_{n^A}}$ to image $\mathcal{D}^{\text{B}} \in \mathbb{R}^{I_1{}^B \times \ldots \times I_{n_S^B} \times I_{n_S^B+1} \times \ldots \times I_{n^B}}$.

Our case study samples show little tissue damage (e.g. due to freezing or cutting artifacts) and the morphology between the IMS and microscopy-imaged samples is identical when the same section is used, and very similar when neighboring sections are used. As a result, the examples do not require non-rigid registration approaches to 'warp' images to each other, and instead use a relatively simple affine registration to achieve good matching.

Definition 0.5. (affine IMS-to-microscopy registration.) Let $g_{\text{ims}\to\text{micro}}\colon \mathbb{R}^2 \to \mathbb{R}^2$ denote a registration function that maps a 2-D IMS image $\mathcal{D}^{\text{ims}}$ to a 2-D microscopy image $\mathcal{D}^{\text{micro}}$. Function $g_{\text{ims}\to\text{micro}}(\vec{x}_{\text{ims}}) = \vec{x}_{\text{micro}}$ transforms a vector of coordinates in IMS space, $\vec{x}_{\text{ims}}$, to coordinates $\vec{x}_{\text{micro}}$ in microscopy space, and is fully determined by the affine transformation matrix $G \in \mathbb{R}^{3\times3}$ with $G = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ 0 & 0 & 1 \end{bmatrix}$. Function $g_{\text{ims}\to\text{micro}}$ is implemented as $\begin{bmatrix} \vec{x}_{\text{ims}} \\ 1 \end{bmatrix} \mapsto G \begin{bmatrix} \vec{x}_{\text{ims}} \\ 1 \end{bmatrix}$.

In all examples, registration between $\mathcal{D}^{\text{ims}}$ and $\mathcal{D}^{\text{micro}}$ was determined by manual selection of corresponding fiducials in MATLAB, using the `cpselect()` function. The transformation matrix G was then obtained from the fiducial pairs by employing the `cp2tform()` function.

Supplementary Note 2 – Model Building and Evaluation Phase

Building a fusion model can be broken down into five steps: (step 1a) a technology-specific transformation of the variables provided by the first modality; (step 1b) a separate technology-specific transformation of the variables supplied by the second modality; (step 2) mapping both sets of transformed measurements to each other in the spatial domain, in order to build a multi-modal training set; (step 3) mining the training set for cross-modality relationships, attempting to model responses in one technology using observations from the other technology; and (step 4) evaluating the resulting model in both the chemical and spatial domains to ascertain for which variables (or ion peaks) and where in the tissue good and robust prediction is possible.

Each step can be considered a separate module with its own specific sub-problem to solve. The inputs and outputs of each step are specified by means of a mathematical function definition. The implementation of each step or function is either made mathematically explicit or the existing tools, software, or environments used to implement that function are specified.

The modularity of the model-building framework and the general definition through mathematical functions ensures that one can adjust or replace the implementation of one step without necessarily having to change the other steps. In the end, the quality of the cross-modality model and the prediction performance of the fusion application will be dependent on the combined parameters, constraints, and performance of each of these steps. **Supplementary Fig. 2** provides a detailed schematic of the model-building-and-evaluation phase and its five steps and **Supplementary Fig. 3** provides an overview of the inputs and outputs of each method step with mathematical details.

Step 1a and 1b – Transformation of source variables

The native form in which the source modalities deliver their measurements is not necessarily ideal for the efficient capture of cross-modality relationships. It is often preferable to transform the observations to a data space that brings out interesting patterns more clearly, while removing variables that add little information. If the source modality delivers few variables, the transformation can focus on increasing the number of relevant variables by mining the observations for additional insights (e.g. textural analysis of microscopy images; see **Supplementary Fig. 2**). If the source modality delivers many variables, sometimes with strong correlation among them, the transformation can entail some form of dimensionality reduction or feature extraction (e.g. peak picking for IMS; see **Supplementary Fig. 2**). Either way, step 1a and 1b are implemented as a feature transformation function.

Definition 1.1. (feature transformation function.) Let $f_{\text{trans}}: \mathbb{R}^{P\times U} \rightarrow \mathbb{R}^{P\times N}$ denote a feature transformation function that transforms the U feature bands of the input image data source $D \in \mathbb{R}^{P\times U}$ into N feature bands in the output image data source $D_{\text{tr}} \in \mathbb{R}^{P\times N}$, while leaving the number of measured pixels P unchanged.

The goal of the feature transformation function is to increase the relevant information content in the image data source, while removing uninformative feature bands and reducing the dimensionality of the pixel signatures without substantial information loss. It aims to make the overall fusion method more robust and to avoid resources being spent on uninformative features. Additionally, it aims to increase the number of viewpoints on the content of an image data source and expand the number of image aspects that cross-modality relationships can hook into. The implementation of f_{trans} is both modality-specific and task-specific, as it is the task that determines which features are informative and which are not. It is thus possible to give a modality-specific definition of f_{trans}, one focused on microscopy (step 1a) and one focused on IMS (step 1b).

Definition 1.2. (microscopy feature transformation function.) Let $f_{\text{trans}}^{\text{micro}}: \mathbb{R}^{P_{\text{micro}} \times U_{\text{micro}}} \rightarrow \mathbb{R}^{P_{\text{micro}} \times N_{\text{micro}}}$ denote a microscopy feature transformation function that transforms the U_{micro} feature bands of a microscopy data source $D^{\text{micro}} \in \mathbb{R}^{P_{\text{micro}} \times U_{\text{micro}}}$ into N_{micro} feature bands, leaves the number of measured pixels P_{micro} unchanged, and returns transformed microscopy data source $D_{\text{tr}}^{\text{micro}} \in \mathbb{R}^{P_{\text{micro}} \times N_{\text{micro}}}$.

Definition 1.3. (IMS feature transformation function.) Let $f_{\text{trans}}^{\text{ims}}: \mathbb{R}^{P_{\text{ims}} \times U_{\text{ims}}} \rightarrow \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$ denote an IMS feature transformation function that transforms the U_{ims} feature bands of an IMS data source $D^{\text{ims}} \in \mathbb{R}^{P_{\text{ims}} \times U_{\text{ims}}}$ into N_{ims} feature bands, leaves the number of measured pixels P_{ims} unchanged, and returns transformed IMS data source $D_{\text{tr}}^{\text{ims}} \in \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$.

Step 1a entails the expansion of the number of microscopy variables from the three that are natively provided (RGB - red, green, and blue) to additional variables that bring out different aspects encoded in the microscopy picture. These additional variables are for example obtained by converting the RGB values to different color spaces, providing additional bands encoding luminance, hue, saturation, and two-color difference bands. A second expansion technique is to calculate non-linear power law versions of the available bands so that more non-linear relationships between the technologies can be captured with a linear model. A third expansion method is to calculate the principal components of the available variables via principal component analysis[43] and to encode their spatial characterizations as image variables. A similar result can be obtained by applying non-negative matrix factorization[44] to the available bands. In order to reduce the pixel-specific variation in these variables somewhat, it is often helpful to calculate discrete wavelet de-noised versions of the available bands[45] and apply the preceding expansion methods to these de-noised bands as well. Finally, the available variables per pixel can be multiplied further by applying textural filters such as a range filter or an entropy filter to each individual band[45]. In each of our case studies, a combination of these techniques is used to expand the number of microscopy variables from three to several hundred.

Native IMS measurements return as many variables per pixel as there are *m/z* bins per spectrum, and for MALDI-TOF data this commonly runs into the tens of thousands. As a large amount of these variables will describe a non-peak part of the spectral profile, and others will correlate strongly due to describing the same ion peak, the task of step 1b is to reduce these expansive measurements to only variables of interest. This is accomplished by peak-picking the IMS data set[46] and only retaining ion images that describe distinct ion peaks.

Step 1a takes a microscopy image (**Supplementary Fig. 3**), encoded as an image data source matrix in $\mathbb{R}^{P_{\text{micro}} \times U_{\text{micro}}}$ with P_{micro} pixels and U_{micro} variables per pixel, and expands the number of variables per pixel to N_{micro}using the expansion methods discussed above. Step 1b reduces the number of variables per pixel from U_{ims} to N_{ims} through peak picking. In doing so, it reduces the size of the original IMS measurement matrix from $P_{\text{ims}} \times U_{\text{ims}}$ to $P_{\text{ims}} \times N_{\text{ims}}$ with P_{ims} being the number of pixels in the ion images.

These transformations are specific to the modality in question, and can be customized for each new sensor type submitted to the fusion methodology. The transformation steps are the ideal place to put code for the removal of technology-specific noise, feature expansion, dimensionality reduction, or the selection of which variables to let participate in the fusion process. Due to the modularity of the framework, other fusion steps are largely independent of the particularities of the source modalities used, and they will typically require little to no adjustment when a new sensor type is introduced.

The applications pursued in the paper are exploratory and no prior information is available as to which feature bands in the sources will be relevant to the task at hand. A broader vocabulary of microscopy-derived patterns provides more degrees of freedom to describe the ion distributions with and is thus preferable. However, more microscopy-derived variables also incur higher computational costs in memory and processing time. Therefore, the features in the microscopy data source are expanded to fill

the computational resources available in order to practically maximize the chance of finding cross-modality relationships. The number of microscopy-derived variables is thus determined by the specifics of the case study at hand (e.g. the number of microscopy pixels) and the computational resources available to hold these variables. We employ the `rgb2ntsc()`, `rgb2ycbcr()`, `princomp()`, `nnmf()`, `rangefilt()`, `entropyfilt()`, `wavedec2()`, and `wrcoef2()` MATLAB functions to calculate the additional bands in $f_{\text{trans}}^{\text{micro}}$. The features of the IMS data source are reduced by removing non-peak *m/z* bins. This is accomplished by running the `mspeaks()` function of MATLAB on the average mass spectrum of D^{ims} and retaining only the bins that are closest to the center mass of the robustly detected ion peaks. In case study 1, the number of features per microscopy pixel in D^{micro} was expanded from $U_{\text{micro}} = 3$ to $N_{\text{micro}} = 204$, providing a transformed microscopy data source $D_{\text{tr}}^{\text{micro}} \in \mathbb{R}^{99,458,571\times 204}$. The number of features per IMS pixel in D^{ims} was reduced from $U_{\text{ims}} = 24{,}576$ to $N_{\text{ims}} = 218$, providing a transformed IMS data source $D_{\text{tr}}^{\text{ims}} \in \mathbb{R}^{11,857\times 218}$.

Step 2 –Mapping transformed measurement sets to each other

Once the transformations are complete, the two sets of transformed observations need to be linked to each other. Since the spatial domain is common to both data sources, this link is established by spatially mapping pixels in one technology to pixels in the other technology (**Supplementary Fig. 2**). The goal is to generate a set of examples to train the cross-modality model on. Each of these examples needs to consist of a transformed observation on the IMS side and a corresponding transformed observation on the microscopy side.

To establish a model for observations in modality *B* on the basis of observations in modality *A*, a training set of examples in both modalities is required. Constructing a training set is not straightforward when modality *A* and *B* use different spatial resolutions. The multi-resolution problem in multi-modal studies entails that when modalities obtain measurements at different spatial resolutions, it is often not possible to establish a one-to-one mapping between measurements of different technical origin. We handle this problem by not imposing a one-to-one mapping at all. Instead, our mapping step embraces a one-to-many mapping, letting all potential measurement couples participate in the building of a model, and not just the 'right' couples (which are often unknown). That way, a model can be left to generalize across a set of examples that is admittedly noisy, but at the same time massive in number. The approach uses the large number of training examples to mitigate the noise introduced by measurement couples that were wrongly linked to each other. The approach effectively uses computational resources to avoid having to make assumptions about how modalities might be related. We also specify a weighted version of the mapping step, which uses prior knowledge on the spatial sampling behavior of a modality to reduce the amount of noise examples in the training set and help improve the fusion results further.

Mapping. The mapping step aims to (spatially) map observations in one modality to observations in the other modality, in order to establish a training set of corresponding signatures in both modalities. Spatial mapping entails that a measurement in modality *B* is 'linked' to the measurement(s) in modality *A* that describe the same space (e.g. pixel surface area). If for the space characterized by an observation in modality *B*, there are no observations in modality *A* (e.g. if only part of the measured areas overlap), no link will be established and this observation will not be part of the training set. If there are observations in modality *A* that do characterize this same space, all of them will be linked to the observation in modality *B*. If the spatial resolution of modality *A* is more fine-grained than that of modality *B*, several observations in *A* are linked to a single observation in *B*. For example, if modality *A* has a spatial resolution such that one of its measurements characterizes a pixel surface area of 10×10 μm, and a measurement in modality *B* characterizes a pixel area of 50×50 μm, there will be 25 measurements in *A* linked to each measurement in *B*. The repetitive aspect of the same measurement in *B* being linked to

different partner measurements in A is a key characteristic of our fusion method. By allowing the training set to contain false positive examples, there can be a 'voting' effect across all of the observations in the (often massive) training set. This avoids the need for human interaction to define which out of the 25 measurements in A is the right one to be linked to the one measurement in B. Instead, not just the 'right' connection but all possible connections are represented in the training set and get a chance to influence the model. The model building step can then generalize across a noisy but large database of examples, and figure out in an unsupervised way which (or which part) of the 25 measurements in A are responsible for the observations in the one measurement in B.

Definition 2.1. (mapping function.) Let $f_{\text{map}}: \mathbb{R}^{P_A \times N_A} \times \mathbb{R}^{P_B \times N_B} \times g_{B \to A} \to \mathbb{R}^{M \times N_A} \times \mathbb{R}^{M \times N_B}$ denote a mapping function. Function f_{map} uses the registration function $g_{B \to A}$ to establish a binary or numeric link between each of the P_A observations from image data source $D^A \in \mathbb{R}^{P_A \times N_A}$ and zero, one, or more of the P_B observations from image data source $D^B \in \mathbb{R}^{P_B \times N_B}$. The non-zero links are used to establish two data sets $S^A \in \mathbb{R}^{M \times N_A}$ and $S^B \in \mathbb{R}^{M \times N_B}$, in which each i_mth row in S^A maps to the i_mth row in S^B, with $i_m \in \mathbb{N}$ between 1 and M.

Together, S^A and S^B make up a training set $[\, S^A \;\; S^B \,] \in \mathbb{R}^{M \times (N_A + N_B)}$ with M training examples, one in each row. The training set delivers M times a $1 \times N_A$-sized observation in modality A that corresponds to a $1 \times N_B$-sized observation in modality B.

Permutation mapping. A mapping function is not required to build its training set using only observations that are natively encountered in the image data sources, and could include, for example, observations that are some form of combination of the original observations in D^A and D^B. For example, in the case described above, a mapping function could create a training example by linking the one measurement in B to the average of some or all of the 25 corresponding measurements in A. However, unless there is specific information available on how measurements can be combined into a consensus measurement without jeopardizing the task at hand, combining observations runs the risk of creating training examples that might never be encountered in real measurements, rendering the model useless for prediction on new measurements. To avoid this risk, we introduce a more narrow type of mapping, permutation mapping, where the outputs S^A and S^B are allowed to contain only observations that were provided in D^A and D^B (with repetition). This ensures the training set consists only of observations that were actually measured.

Definition 2.2. (permutation mapping function.) Let $f_{\text{pmap}}: \mathbb{R}^{P_A \times N_A} \times \mathbb{R}^{P_B \times N_B} \times g_{B \to A} \to \mathbb{R}^{M \times N_A} \times \mathbb{R}^{M \times N_B}$ denote a permutation mapping function. Function f_{pmap} is a mapping function where $S^A \in \mathbb{R}^{M \times N_A}$ is a permutation of M elements with repetition of the rows of $D^A \in \mathbb{R}^{P_A \times N_A}$, and $S^B \in \mathbb{R}^{M \times N_B}$ is a permutation of M elements with repetition of the rows of $D^B \in \mathbb{R}^{P_B \times N_B}$.

Weighted permutation mapping. Most modalities assume that a measurement characterizes an entire (often square) pixel area. However, the measurement principle behind the modality might really only be sampling a subarea within that pixel. If there is information available on which subarea of the pixel is really being sampled and in what amount parts of that subarea contribute to the acquired signal, it is beneficial to incorporate that information into the mapping function to further improve the quality of the training set.

In the example above, each of the 5×5 observations in A was linked with equal weight to the one observation in B, resulting in 25 training examples. If we know that modality B physically only samples

the center 3×3 grid rather than the full 5×5 locations in A, we can remove the training examples that stem from the outer ring of the 5×5 grid from the training set, essentially eliminating 16 false examples per measurement in B. If we also have access to quantitative information that specifies that the center pixel in the 3×3 grid delivers twice as much signal or material as the surrounding pixels, such information can be reflected in the training set as well. We could for example give the training example that stems from the center pixel more prominence in the training set by repeating it proportional to its importance.

If such sampling information is available for a modality, either by theoretical or empirical means, we suggest including it into the mapping function in the form of link weights encoded by a space-to-measurement function. Every connection between an observation in A and an observation in B on the basis of the registration function $g_{B\to A}$ can then receive a weight according to the space-to-measurement function of the modality with the coarsest spatial resolution. One way to translate these weights into training set prominence is by repeating the training example proportional to its weight. This has the effect that examples that receive lots of confidence from the weighting scheme (which encodes prior knowledge on the modality) will occur many times in the training set, while those with little weight will occur rarely and those with zero weights will be removed from the training set altogether. This weighting scheme avoids that resources are spent on training examples that are probably noise, and it improves the quality of the training set and model, thus improving the quality of the fusion results overall.

Definition 2.3. (space-to-measurement function.) Let $h_{B\text{-in-}A}\colon \mathbb{R}^{n_s^A} \to \mathbb{N}^{I_1^h\times\ldots\times I_{n_s^A}^h}$ denote a space-to-measurement function for an observation in modality B, defined at the spatial resolution of modality A. Function $h_{B\text{-in-}A}$ takes a location $\vec{x} \in \mathbb{R}^{n_s^A}$ in the space of modality A and delivers a weighted representation of how a measurement in modality B at that location would sample the surrounding space in A. The space-to-measurement relationship is provided as an n_s^Ath order tensor $\mathcal{H} \in \mathbb{N}^{I_1^h\times\ldots\times I_{n_s^A}^h}$ of natural numbers where the center element represents the given location $\vec{x}$, and the surrounding tensor elements represent the locations in the space of A that surround $\vec{x}$. Each tensor element encodes a weight assigned to the location in A it represents. The weights of locations in A that lie beyond those indexed by $\mathcal{H}$ are assumed zero.

To be able to incorporate prior information on the sampling behavior of a modality into the mapping step, we generalize our definition of a permutation mapping function to a weighted permutation mapping function, which takes one additional input in the form of a space-to-measurement function.

Definition 2.4. (weighted permutation mapping function.) Let $f_{\text{wpmap}}\colon \mathbb{R}^{P_A\times N_A} \times \mathbb{R}^{P_B\times N_B} \times g_{B\to A} \times h_{B\text{-in-}A} \to \mathbb{R}^{M\times N_A} \times \mathbb{R}^{M\times N_B}$ denote a weighted permutation mapping function. Function f_{wpmap} is a permutation mapping function where the link between an observation from $D^A \in \mathbb{R}^{P_A\times N_A}$ and an observation from $D^B \in \mathbb{R}^{P_B\times N_B}$ is encoded as a $1\times(N_A+N_B)$ row vector and that row is repeated in $[\,S^A\ S^B\,] \in \mathbb{R}^{M\times(N_A+N_B)}$ proportional to the weight assigned to that link by the space-to-measurement function $h_{B\text{-in-}A}$.

Let us now apply these mapping concepts to the IMS and microscopy case. In our examples, the spatial resolution of IMS is coarser than that of microscopy. It is thus valuable to encode any prior information we have on the spatial sampling behavior of IMS into the mapping step using an IMS space-to-measurement function.

Definition 2.5. (IMS space-to-measurement function.) Let $h_{\text{ims-in-micro}}: \mathbb{R}^2 \rightarrow \mathbb{N}^{I^h_{\text{row}} \times I^h_{\text{col}}}$ denote an IMS space-to-measurement function for an observation in the IMS modality, defined at the spatial resolution of the microscopy modality. Function $h_{\text{ims-in-micro}}$ takes a location $\vec{x}_{\text{ims-in-micro}} \in \mathbb{R}^2$ in the microscopy space and delivers a weighted representation of how an IMS measurement at that location would sample the surrounding space in terms of microscopy pixels. The space-to-measurement relationship is provided as a matrix $H \in \mathbb{N}^{I^h_{\text{row}} \times I^h_{\text{col}}}$ in which a grid of I^h_{row} rows and I^h_{col} columns represents the microscopy pixels around the given location $\vec{x}_{\text{ims-in-micro}}$ with the center element of H representing $\vec{x}_{\text{ims-in-micro}}$. Each element in H contains a weight value that describes how strongly an IMS measurement at $\vec{x}_{\text{ims-in-micro}}$ is related to the microscopy pixel area encoded by that element. The weights of microscopy pixels that lie beyond those indexed by H are assumed zero.

When MALDI-based IMS acquires a mass spectrum at a particular tissue location, it rarely ablates all material within the rectangular tissue area that corresponds to an IMS pixel. Instead, the mass spectrum will typically report on a subarea of the pixel, determined by the laser footprint and whether the laser was allowed to 'walk' through the pixel area. Additionally, the subarea that is ablated is often not sampled homogeneously, digging off more material in certain areas and less so in other areas. The sampling behavior of an IMS instrument can be determined theoretically on the basis of instrumental knowledge, or can be established empirically by, for example, examining ablation craters. Either way, such sampling information can be encoded into matrix H as an integer grid of sampling weights.

In all case studies, we employed a 2-D Gaussian bell curve centered on the IMS pixel as an approximation for how the IMS laser samples an IMS pixel, and the matrix H consists of projecting the bell curve to natural integer weights. This IMS sampling model gives more importance to microscopy pixels that are found closer to the center of the IMS pixel, and less importance to microscopy pixels that are found at the edges of the IMS pixel. The bell curve approximation gives a more robust training set than if all microscopy pixels within an IMS pixel would be considered equally valid, yet it does not give all importance to only the microscopy pixel at the absolute center of the IMS pixel, which helps us handle registration inaccuracies and morphological variation when images from neighboring tissue sections are fused. It is clear that the 2-D bell curve is but an approximation, and that the training set could be improved further by encoding in H an empirically determined ablation profile for that particular experiment and instrument.

Once we have an IMS image data source $D^{\text{ims}}_{\text{tr}}$, a microscopy image data source $D^{\text{micro}}_{\text{tr}}$, a registration function $g_{\text{ims}\rightarrow\text{micro}}$ that ties them spatially together (implemented as affine transformation matrix G), and an IMS space-to-measurement function $h_{\text{ims-in-micro}}$ (implemented as weight matrix H), the mapping step can be completed using an IMS-to-microscopy weighted permutation mapping function.

Definition 2.6. (IMS-to-microscopy weighted permutation mapping function.) Let $f^{\text{ims}\rightarrow\text{micro}}_{\text{wpmap}}: \mathbb{R}^{P_{\text{micro}} \times N_{\text{micro}}} \times \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}} \times g_{\text{ims}\rightarrow\text{micro}} \times h_{\text{ims-in-micro}} \rightarrow \mathbb{R}^{M \times N_{\text{micro}}} \times \mathbb{R}^{M \times N_{\text{ims}}}$ denote an IMS-to-microscopy weighted permutation mapping function. Function $f^{\text{ims}\rightarrow\text{micro}}_{\text{wpmap}}$ is a weighted permutation mapping function where the link between a microscopy observation from $D^{\text{micro}}_{\text{tr}} \in \mathbb{R}^{P_{\text{micro}} \times N_{\text{micro}}}$ and an IMS observation from $D^{\text{ims}}_{\text{tr}} \in \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$ is encoded as a $1 \times (N_{\text{micro}} + N_{\text{ims}})$ row vector and that row is repeated in $[\, S^{\text{micro}} \;\; S^{\text{ims}} \,] \in \mathbb{R}^{M \times (N_{\text{micro}} + N_{\text{ims}})}$ proportional to the weight assigned to that link by the IMS space-to-measurement function $h_{\text{ims-in-micro}}$.

The implementation of the $f_{\text{wpmap}}^{\text{ims}\rightarrow\text{micro}}$ function is shown in the following pseudo-code:

```
SET S^micro to an empty array of 0×N_micro values
SET S^ims to an empty array of 0×N_ims values

FOR each observation p_ims in IMS image data source D_tr^ims
    SET x⃗_ims^{p_ims} to coordinates of p_ims in IMS space
    COMPUTE coordinates of p_ims in microscopy space x⃗_micro^{p_ims} (by calculating x⃗_micro^{p_ims} = G x⃗_ims^{p_ims})
    COMPUTE surface area of IMS pixel p_ims in microscopy space
    DETERMINE K microscopy pixels {p_micro,1, …, p_micro,K} overlapping with surface area of IMS pixel p_ims in microscopy space
    DETERMINE K weights {w_1, …, w_K} for overlapping microscopy pixels (by retrieving weight mask H = h_ims-in-micro(x⃗_micro^{p_ims}) and centering it on microscopy pixel closest to x⃗_micro^{p_ims})

    FOR each microscopy pixel p_micro,k from {p_micro,1, …, p_micro,K}
        SET weight to w_k
        SET count to zero
        WHILE count < weight
            APPEND observation p_micro,k to S^micro
            APPEND observation p_ims to S^ims
            INCREMENT count by one
        END WHILE
    END FOR
END FOR
```

The algorithm uses the `imtransform()` function in MATLAB to calculate the projections from IMS to microscopy space.

For case study 1, step 2 resulted in a training set of $M = 777{,}760$ training examples with $S^{\text{micro}} \in \mathbb{R}^{777{,}760\times 204}$ and $S^{\text{ims}} \in \mathbb{R}^{777{,}760\times 218}$. This was the result of combining the spatial mapping as depicted in figure files spat_mapping.png and spat_mapping_zoom.png with a bell curve implementation of $h_{\text{ims-in-micro}}$ as depicted in files h_ims_in_micro.pdf and h_ims_in_micro_surf.pdf, and the integer weights derived from them as shown in h_ims_in_micro_weights.pdf and h_ims_in_micro_weights_surf.pdf (all files provided in a zip-file in Supplementary Information).

A key aspect of the mapping step is that in building a training set, it allows this set of training examples to be noisy. This means that a training set not only contains examples that link signatures together that have a genuine correspondence between them, but that it also contains false positive examples linking signatures that have little to do with each other. The reason false training examples are tolerated in the training set is that it is relatively straightforward to spatially link signatures together across technologies, but much more difficult to find out which ones are truly linked content-wise. Any registration can tell us which observations in one modality lie within a spatial sphere of influence of an observation in another modality. However, it is much harder to take this collection of linked signatures that could potentially correspond, and find out which ones actually do. Since little prior information is available to aid in this effort, the mapping step avoids culling the training set prematurely on the basis of assumptions, and instead lets the model-building step (Step 3) generalize across a noisy training set. The disadvantage of noisy training data is that it can cause reduced model quality, but in practice this is rarely a problem for the fusion scenarios discussed here. The cross-modality relationships tend to be captured

well from noisy training data as long as the true positive examples dominate the false positive examples, and the number of training examples is high (in our case studies, often close to a million). The advantage of this approach is that it allows fusion to handle the problem of different spatial resolutions in a graceful way instead of forcing measurements to be down-sampled into a single spatial resolution for all modalities. For example, in a multi-resolution scenario, many pixels in one modality might contribute to a single pixel in the other modality, potentially as a mixture. The noisy training approach gives all components of the mixture a chance to influence the model. The noisy training data also makes the fusion procedure remarkably robust against registration errors and outlier pixels. For example, any registration procedure will introduce some registration errors, spatially linking pixels to the wrong partners in the other technology. This can be due to fiducial inaccuracies, but also due to the use of neighboring tissue sections for example (as in **Fig. 4**). As long as the false mappings do not dominate the training set and the majority of training examples still connect observations together that correspond to each other, the model will generalize out these false examples as outliers, and the predictions undergo little effect from a registration that is somewhat off.

Step 3 – Build a cross-modality model

The third sub-task is to mine the training set for any detectable relationships between observations in IMS and observations in microscopy (**Supplementary Fig. 2**). If these connections can be captured as mathematical descriptions in a model, they can later on be used to predict IMS observations when only microscopy observations are available. The main goal of the third step is therefore to build a cross-modality modeling function on the basis of the training set.

Definition 3.1. (cross-modality modeling function.) Let $q_{B/A}: \mathbb{R}^{N_A} \rightarrow \mathbb{R}^{N_B}$ denote a cross-modality modeling function that takes an observation of size $1 \times N_A$ in modality A, and predicts a corresponding observation of size $1 \times N_B$ in modality B.

The building of this modeling function $q_{B/A}$ is taken care of by a model building function f_{model} that implements step 3.

Definition 3.2. (model building function.) Let $f_{\text{model}}: \mathbb{R}^{M \times N_A} \times \mathbb{R}^{M \times N_B} \rightarrow q_{B/A}$ denote a model building function that builds a cross-modality modeling function $q_{B/A}$ from M training examples.

For IMS-microscopy fusion, the goal is thus to establish an IMS-microscopy modeling function $q_{\text{ims/micro}}$ capable of predicting IMS on the basis of microscopy.

Definition 3.3. (IMS-microscopy modeling function.) Let $q_{\text{ims/micro}}: \mathbb{R}^{N_{\text{micro}}} \rightarrow \mathbb{R}^{N_{\text{ims}}}$ denote an IMS-microscopy modeling function that takes a microscopy observation of size $1 \times N_{\text{micro}}$, and predicts a corresponding IMS observation of size $1 \times N_{\text{ims}}$.

In our method, the building of a model that implements this IMS-microscopy modeling function is approached as a regression task. The field of regression analysis focuses on estimating relationships between variables. In our setting, regression analysis is used to understand how the ion intensity of a particular ion species changes when the values of one or more microscopy variables change. The analysis provides for each IMS variable or ion image a regression function, which describes the intensity and distribution of that ion as a function of the intensity distributions of a subset of the microscopy variables. The final cross-modality modeling function consists of the combined regression functions, and enables concurrent prediction for all ion images (or IMS variables) when presented with a microscopy

observation. Regression analysis comprises a wide assortment of methods to choose from, tailored towards various data types and constraints (e.g. linear vs. non-linear, parametric vs. non-parametric, etc.). In the proof-of-principle examples demonstrated in this paper, IMS-microscopy fusion is achieved by implementing the function $q_{\text{ims/micro}}$ as a linear regression model.

Definition 3.4. (linear IMS-microscopy model.) Let

$$Z_{\text{ims}} = Y_{\text{micro}}\, Q_{\text{ims/micro}}$$

with $Y_{\text{micro}} \in \mathbb{R}^{V\times(1+N_{\text{micro}})}$ as V observations in microscopy to predict for;
$Q_{\text{ims/micro}} \in \mathbb{R}^{(1+N_{\text{micro}})\times N_{\text{ims}}}$ as the coefficient matrix; and
$Z_{\text{ims}} \in \mathbb{R}^{V\times N_{\text{ims}}}$ as V predicted observations in IMS;

denote a linear IMS-microscopy model. This linear model structure implements an IMS-microscopy modeling function $q_{\text{ims/micro}}$, such that a predicted IMS observation $v'_{\text{ims}} \in \mathbb{R}^{1\times N_{\text{ims}}}$ can be obtained from a measured microscopy observation $v_{\text{micro}} \in \mathbb{R}^{1\times N_{\text{micro}}}$ by calculating $v'_{\text{ims}} = [1 \quad v_{\text{micro}}]\, Q_{\text{ims/micro}}$.

The linear IMS-microscopy model actually contains N_{ims} predictive sub-models, one for each IMS variable separately. Since in step 1b, each IMS variable was set to report on the measured ion peak intensity at a certain *m/z* value, each sub-model here predicts the ion peak intensity at that *m/z* value, using all microscopy variables as its guide.

Definition 3.5. (linear IMS-microscopy sub-model.) Let each column in the transformation matrix $Q_{\text{ims/micro}}$ of a linear IMS-microscopy model, establish a linear IMS-microscopy sub-model. This sub-model predicts a single IMS variable as a linear combination of all N_{micro} microscopy variables plus one intercept value:

$$z_{ij} = \alpha_j + \beta_{kj} y_{ik}$$

with i between 1 and V, indicating the ith observation in Y_{micro} and Z_{ims};
j between 1 and N_{ims}, indicating the jth IMS variable;
k between 1 and N_{micro}, indicating the kth microscopy variable;
y_{ik}, the measured value for microscopy variable k in observation i;
z_{ij}, the predicted value for IMS variable j in observation i;
α_j, the intercept value for IMS variable j; and
β_{kj}, the slope value that connects microscopy variable k and IMS variable j.

Definition 3.6. (coefficient matrix.) Let $Q_{\text{ims/micro}} \in \mathbb{R}^{(1+N_{\text{micro}})\times N_{\text{ims}}}$ denote the coefficient matrix of a linear IMS-microscopy model. The coefficient matrix $Q_{\text{ims/micro}}$ is the concatenation of a vector of intercept values, $\alpha \in \mathbb{R}^{1\times N_{\text{ims}}}$, and a matrix of slope values, $\beta \in \mathbb{R}^{N_{\text{micro}}\times N_{\text{ims}}}$, such that

$$Q_{\text{ims/micro}} = \begin{bmatrix}\alpha \\ \beta\end{bmatrix}$$

The vector α contains an intercept value α_j for each IMS variable j, and the matrix β contains a slope value β_{kj} for each combination of microscopy variable k and IMS variable j.

By concatenating α and β into a single transformation matrix $Q_{\text{ims/micro}}$, and by having the matrix Y_{micro} precede each microscopy measurement by a one in the first column, the predictions can be calculated in a straightforward way using a single matrix multiplication. The concatenated matrix formulation of the prediction step not only makes the calculation of predictions efficient (e.g. no for-loops in an algebra-capable environment such as MATLAB), it also generalizes prediction from a single observation to a large number of observations in a straightforward way that can be easily parallelized.

The choice of regression method, method parameters, and model type will be important towards the quality of the predictions that are possible with a particular fusion approach. The choice of model type will determine the type of cross-modality relationships that can be captured by it. For example, if the distribution and intensity of a variable in one modality shows a quadratic or logarithmic relationship to the expression of a variable in another modality, a linear modeling approach will have a hard time capturing that relationship adequately and will often result in only partial approximation over a sub-window of the range of these variables. For this paper, we focus on a linear regression model to implement $q_{\text{ims/micro}}$, and as such our predictions can only be based on linear cross-modality relationships. However, the compromise seems reasonable, as the calculation load of linear methods is typically smaller than nonlinear methods, keeping this step within practical resource boundaries. Simultaneously, the case studies demonstrate that linear (multivariate) modeling is sufficiently complex to outperform univariate correlation-based results, and that it can already capture a lot of structure hidden in the variation across different modalities. An additional advantage of linear models is that once the model has been built, it allows for straightforward interpretation of the relationships between variables using clear concepts also used elsewhere in analytical chemistry (e.g. line intercepts and slopes in calibration). In more intricate non-linear models this will often be much less straightforward.

Once a linear model structure is set, the challenge is to set the coefficients $Q_{\text{ims/micro}}$ to concrete values, in such a way that they capture the linear relationships that can be observed between IMS and microscopy variables in the training examples. The model coefficients are estimated from the examples in the training set, for which both the model input (microscopy measurements) and the desired model output (IMS measurements) are available. If that estimation can be completed with reasonable success (note: model performance will be IMS variable-specific), the same model (both structure and coefficients) can be used to predict IMS observations when only microscopy measurements are available. The goal of the regression method is therefore to determine $Q_{\text{ims/micro}}$ by plugging the two sides of the training set, S^{micro} and S^{ims}, into the formula $Z_{\text{ims}} = Y_{\text{micro}}\, Q_{\text{ims/micro}}$, such that the difference between the measurement-based S^{ims} and the prediction-based $S'^{\text{ims}} = [1 \quad S^{\text{micro}}]\, Q_{\text{ims/micro}}$ is minimal. In other words, a single ion distribution and a vocabulary of microscopy-derived patterns are given per sub-model. The sub-model coefficients are calculated to be the values that best approximate the measured ion distribution using the vocabulary of microscopy-derived patterns. The determination of the model coefficients can be considered an optimization problem over all training set examples. There are many different regression algorithms available to drive this estimation of $Q_{\text{ims/micro}}$. For our fusion purposes, we use a well-studied method called PLS regression to implement the regression engine of our cross-modality modeling effort. PLS regression stands for Partial Least Squares regression or Projection to Latent Structures regression[40]. This type of regression projects its observed and predicted variables to a new data space where it finds a linear regression model. The projection to latent variables by PLS regression is related to principal component analysis and enables robust regression results even when there is a lot of correlation between the measured variables. This characteristic is often present in our multi-modal data sets since the transformation steps 1a and 1b do not specifically pursue orthogonal variables and, for example, many microscopy variables are often texture-filtered versions of others. Although any regression technique could potentially be used, we found projection to latent variables methods to be more robust for our

purposes. The choice for PLS regression demands specification of the number of latent components used. In order not to introduce an analytical bottleneck, this parameter is set to its maximum value.

Step 4 – Evaluate the cross-modality model

The principle of integration via image fusion is that measurements made in one technology can be related to measurements acquired via another technology. However, the same microscopy variables that are very telling towards certain IMS variables will not necessarily be relevant to other IMS variables. As a result, fusion-driven prediction is variable-specific and the quality of prediction will need to be assessed for each ion peak individually. It is hard to determine beforehand for which ion species cross-modality relationships will be available and for which the modeling assumptions will hold as well. Our method therefore takes an empirical approach, training a sub-model for each IMS variable (step 3) and assessing prediction performance afterwards (this step). The goal of the fourth step is thus to provide a variable-specific measure of predictive potential so that fusion applications can be constrained to those variables for which good prediction is possible given these data sources and model type.

The benchmarking of predictive models is a well-studied problem in statistics and machine learning, and a plethora of approaches has been formulated[47]. However, some imaging-specific characteristics make direct application of common evaluation procedures difficult. One such characteristic is the potential for spatial autocorrelation between pixels, which invalidates the independent and identical distribution (i.i.d.) assumption made by many statistical approaches. Other causes for difficulty are the difference in spatial resolution between the data sources and the fact that spatially sampled data are not technical replicates of each other. For example, a traditional evaluation approach would be to separate the available IMS-microscopy measurement pairs into a training set for model building and a test set for model evaluation. However, withholding a sizable amount of measurements from the model training step can seriously hurt model performance as the model will not see any examples from a substantial amount of tissue surface area, particularly on the side of the imaging modality with the coarsest spatial resolution. This would further exacerbate the problem of smaller tissue areas having only limited representation within the training set, and will sometimes remove representation of these areas from the model entirely. Although such issues can be circumvented using approaches such as bootstrapping[41,42], the task of calculating thousands of bootstrap models is often impractical given the size and multivariate nature of the data sets. It is clear that the spatial nature of the fusion-based prediction task makes traditional statistical approaches only partially applicable, and that in the long run more spatially-aware techniques from spatial statistics[47] will need to be incorporated to improve reliability and performance.

Although more advanced evaluation procedures will be pursued in the future, this paper describes five measures that provide the user guidance by highlighting different aspects of prediction performance. The different viewpoints make it possible for the user to assess whether the fusion model can be useful for the study at hand, given what is specifically important to that study. Together these measures cover two types of assessment: (i) reporting variable-specific prediction performance across the entire tissue section, as captured by a global evaluation function, and (ii) reporting variable-specific prediction performance at a particular location in the tissue, as captured by a local evaluation function. Evaluation measures of the first type give us a broad idea of how well a particular ion can be predicted in general using the model. Evaluation measures of the second type tell us in which sub-areas of the tissue the ion can be predicted well and in which tissue areas the predictions for a particular ion are less robust.

Definition 4.1. (global evaluation function.) Let $f_{\mathrm{ge}}^{q_{B/A}}: \mathbb{R}^{L\times N_{\mathrm{A}}} \times q_{B/A} \times \mathbb{R}^{L\times N_{\mathrm{B}}} \to \mathbb{R}^{N_B}$ denote a global evaluation function that summarizes the performance of cross-modality modeling function $q_{B/A}$ across all provided observations L as a single value per variable in target modality B. A global evaluation function $f_{\mathrm{ge}}^{q_{B/A}}$ takes L measured observations of size $1\times N_A$ in modality A, designated as matrix $S_{\mathrm{test}}^{A} \in \mathbb{R}^{L\times N_{\mathrm{A}}}$, predicts L corresponding observations of size $1\times N_B$ in modality B using cross-modality modeling function $q_{B/A}$, designated as matrix $S'^{B}_{\mathrm{test}} \in \mathbb{R}^{L\times N_{\mathrm{B}}}$, compares these predictions to L corresponding measured observations of size $1\times N_B$ in modality B, designated as matrix $S_{\mathrm{test}}^{B} \in \mathbb{R}^{L\times N_{\mathrm{B}}}$, and returns a scalar prediction performance value per variable in modality B, encoding variable-specific global model performance as a vector of size $1\times N_B$.

Definition 4.2. (local evaluation function.) Let $f_{\mathrm{le}}^{q_{B/A}}: \mathbb{R}^{L\times N_{\mathrm{A}}} \times q_{B/A} \times \mathbb{R}^{L\times N_{\mathrm{B}}} \to \mathbb{R}^{L\times N_B}$ denote a local evaluation function that summarizes the performance of cross-modality modeling function $q_{B/A}$ as a value per provided observation and per variable in target modality B. A local evaluation function $f_{\mathrm{le}}^{q_{B/A}}$ takes L measured observations of size $1\times N_A$ in modality A, designated as matrix $S_{\mathrm{test}}^{A} \in \mathbb{R}^{L\times N_{\mathrm{A}}}$, predicts L corresponding observations of size $1\times N_B$ in modality B using cross-modality modeling function $q_{B/A}$, designated as matrix $S'^{B}_{\mathrm{test}} \in \mathbb{R}^{L\times N_{\mathrm{B}}}$, compares these predictions to L corresponding measured observations of size $1\times N_B$ in modality B, designated as matrix $S_{\mathrm{test}}^{B} \in \mathbb{R}^{L\times N_{\mathrm{B}}}$, and returns a scalar prediction performance value per variable in modality B and per provided observation, encoding variable-specific and observation/location-specific local model performance as a matrix of size $L\times N_B$.

Both types of evaluation function require a test set $[S_{\mathrm{test}}^{A} \quad S_{\mathrm{test}}^{B}] \in \mathbb{R}^{L\times(N_{\mathrm{A}}+N_{\mathrm{B}})}$ of L measurements in modality A that are matched to L measurements in modality B. In most multi-modal scenarios the only measurements available are image data sources $D_{\mathrm{tr}}^{\mathrm{A}} \in \mathbb{R}^{P_{\mathrm{A}}\times N_{\mathrm{A}}}$ and $D_{\mathrm{tr}}^{\mathrm{B}} \in \mathbb{R}^{P_{\mathrm{B}}\times N_{\mathrm{B}}}$, so the test set will have will have to be derived from these observation sets. Since we usually do not have prior information on how the image data source with the coarsest spatial resolution, $D_{\mathrm{tr}}^{\mathrm{B}}$, can be up-sampled from P_{B} observations to P_{A} observations with $P_{\mathrm{A}} > P_{\mathrm{B}}$, and this in fact makes up part of the modeling task in for example a sharpening application of fusion, it is more reliable to perform the model evaluation at the coarsest spatial resolution by down-sampling the image data source with the finer spatial resolution, $D_{\mathrm{tr}}^{\mathrm{A}}$, from P_{A} observations to P_{B} observations. The result is that in most multi-modal scenarios, and in all examples of this paper, the test set for model evaluation consists of $L = P_{\mathrm{B}}$ observations in modality A and B, such that S_{test}^{B} is equal to transformed image data source $D_{\mathrm{tr}}^{\mathrm{B}} \in \mathbb{R}^{P_{\mathrm{B}}\times N_{\mathrm{B}}}$ and S_{test}^{A} is equal to the down-sampled transformed image data source $\breve{D}_{\mathrm{tr}}^{\mathrm{A}} \in \mathbb{R}^{P_{\mathrm{B}}\times N_{\mathrm{A}}}$ obtained by down-sampling $D_{\mathrm{tr}}^{\mathrm{A}} \in \mathbb{R}^{P_{\mathrm{A}}\times N_{\mathrm{A}}}$. In the evaluation of our IMS-microscopy models, where microscopy takes the role of modality A and IMS takes the role of modality B, the test set consists of the P_{ims} measurements from transformed IMS data source $D_{\mathrm{tr}}^{\mathrm{ims}} \in \mathbb{R}^{P_{\mathrm{ims}}\times N_{\mathrm{ims}}}$ and a corresponding set of P_{ims} measurements from down-sampled microscopy data source $\breve{D}_{\mathrm{tr}}^{\mathrm{micro}} \in \mathbb{R}^{P_{\mathrm{ims}}\times N_{\mathrm{micro}}}$. The data source $\breve{D}_{\mathrm{tr}}^{\mathrm{micro}}$ is obtained by taking $\mathcal{D}_{\mathrm{tr}}^{\mathrm{micro}} \in \mathbb{R}^{R_{\mathrm{micro}}\times C_{\mathrm{micro}}\times N_{\mathrm{micro}}}$, the image tensor representation of transformed microscopy data source $D_{\mathrm{tr}}^{\mathrm{micro}} \in \mathbb{R}^{P_{\mathrm{micro}}\times N_{\mathrm{micro}}}$, and spatially down-sampling it to image tensor $\breve{\mathcal{D}}_{\mathrm{tr}}^{\mathrm{micro}} \in \mathbb{R}^{R_{\mathrm{ims}}\times C_{\mathrm{ims}}\times N_{\mathrm{micro}}}$, which encodes microscopy measurements at the IMS resolution. The image data source $\breve{D}_{\mathrm{tr}}^{\mathrm{micro}} \in \mathbb{R}^{P_{\mathrm{ims}}\times N_{\mathrm{micro}}}$ is the flattened representation of down-sampled image $\breve{\mathcal{D}}_{\mathrm{tr}}^{\mathrm{micro}}$. The down-sampling from multiple to a single microscopy signature is implemented as taking the mean of each of the N_{micro} bands. In case study 1, the test set contains $L = P_{\mathrm{ims}} = 11{,}857$ IMS signatures of $N_{\mathrm{ims}} = 218$ variables per IMS pixel and the same number of microscopy signatures of $N_{\mathrm{micro}} = 204$ variables per IMS pixel.

The role of global evaluation in this step is to provide an elementary score-based idea of how well a particular IMS variable can be predicted overall using this model, regardless of where in the tissue we want to predict. The objective is to determine how close a microscopy-based predicted distribution for an

ion peak approximates the measured distribution for that ion peak. As we only have IMS measurements at the native IMS resolution, the evaluation step essentially performs a round-trip prediction at the IMS resolution without performing any up-sampling. The challenge in comparing the predictions against this 'gold standard' is that the difference between the predicted ion pattern and the measured ion image is not only caused by the model. In fact, the measured ion images cannot be considered a true gold standard as they are themselves corrupted by noise. This means that a common evaluation measure such as root-mean-square error, summarizing the amplitude of all differences between prediction and measurement, would underestimate true model performance. To mitigate this issue, the global evaluation measures we define do not directly report on the 'difference image' of an ion (the measured minus the predicted ion image), but use a combination of sub-measures to gauge different aspects of that difference image. This provides a means of preventing certain noise sources from substantially impacting a global evaluation measure. We define three different measures of global evaluation, each aimed at assessing a different aspect of the quality of prediction. We strive to have these global evaluation measures cover a range from zero to one, where values close to one report excellent microscopy-based reconstruction of the pattern observed in the ion image. An advantage of such a score-based approach is that it can be reported as a percentage value alongside the fusion result. This allows for the human observer to assess whether the predicted ion distribution is reliable enough for her or his purposes.

The first global measure is the intensity reconstruction score, which assesses how well the absolute peak intensities are approximated tissue-wide. If the task at hand requires correctly predicted peak height, regardless of how well the relative spatial distribution pattern is approximated, the intensity reconstruction score can serve as a guide.

Definition 4.3. (intensity reconstruction scoring function.) Let $f_{\text{irs}}^{q_{\text{ims/micro}}}: \mathbb{R}^{P_{\text{ims}} \times N_{\text{micro}}} \times q_{\text{ims/micro}} \times \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}} \to \mathbb{R}^{N_{\text{ims}}}$ denote an intensity reconstruction scoring function, which is a global evaluation function that summarizes the prediction performance of the IMS-microscopy model implementing $q_{\text{ims/micro}}$ as a single value per IMS variable, emphasizing tissue-wide absolute peak intensity approximation. The intensity reconstruction score reports on the difference matrix $\Delta_{\text{test}}^{\text{ims}} \in \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$, obtained by

$$\Delta_{\text{test}}^{\text{ims}} = S_{\text{test}}^{\text{ims}} - S'^{\text{ims}}_{\text{test}}$$

with $S_{\text{test}}^{\text{ims}} \in \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$, the measured IMS observations in the test set; and
$S'^{\text{ims}}_{\text{test}} \in \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$, the predicted IMS observations in the test set,
calculated as $S'^{\text{ims}}_{\text{test}} = q_{\text{ims/micro}}(S_{\text{test}}^{\text{micro}})$.

The intensity reconstruction score vector $\varphi_{\text{irs}} \in [0,1]^{N_{\text{ims}}}$ provides a scalar percentage score for each IMS variable, and is calculated per column from $\Delta_{\text{test}}^{\text{ims}}$. Two sub-scores, the difference score (**definitions 4.3.1**) and the spatial autocorrelation score (**definitions 4.3.2**), are combined (**definitions 4.3.3**) to gauge the severity of the differences between measurements and predictions.

Definition 4.3.1. (difference score) Let the $\varphi_{\text{diff}_j} \in [0,1]$ denote a difference score for IMS variable j, such that the difference score vector $\varphi_{\text{diff}} \in [0,1]^{N_{\text{ims}}}$ reports per column the importance of the prediction differences relative to the measured signal. The difference score φ_{diff_j} for IMS variable j is calculated as the ratio of the average of absolute intensity differences to the average of absolute measured intensities:

$$\varphi_{\text{diff}_j} = \frac{\sum_{i=1}^{P_{\text{ims}}} abs(\delta_{ij})/P_{\text{ims}}}{\sum_{i=1}^{P_{\text{ims}}} abs(\sigma_{ij})/P_{\text{ims}}}$$

with δ_{ij} as the intensity difference for observation i and variable j in $\Delta_{\text{test}}^{\text{ims}}$, and σ_{ij} as the measured intensity for observation i and variable j in $S_{\text{test}}^{\text{ims}}$. To avoid extreme values caused by outliers, a capped version $\hat{\varphi}_{\text{diff}_j}$ of the difference score φ_{diff_j} is used in further calculations, such that

$$\hat{\varphi}_{\text{diff}_j} = \begin{cases} 1 & \text{if } \varphi_{\text{diff}_j} > 1 \\ \varphi_{\text{diff}_j} & \text{if } 0 < \varphi_{\text{diff}_j} < 1 \\ 0 & \text{if } \varphi_{\text{diff}_j} < 0 \end{cases}$$

.

When $\hat{\varphi}_{\text{diff}_j}$ is large, across the tissue the absolute difference in prediction versus measurement for IMS variable j is large compared to the signal that was measured for this variable. When $\hat{\varphi}_{\text{diff}_j}$ is small, the prediction differences are relatively small compared to the measured intensity pattern.

Definition 4.3.2. (spatial autocorrelation score) Let the $\varphi_{\text{sac}_j} \in [0,1]$ denote a spatial autocorrelation score for IMS variable j, such that the spatial autocorrelation score vector $\varphi_{\text{sac}} \in [0,1]^{N_{\text{ims}}}$ reports per column the amount of spatial structure found in the prediction differences, or how far the spatial distribution of the prediction differences is removed from a random spatial distribution. The spatial autocorrelation score φ_{sac_j} for IMS variable j is calculated using Geary's C as a measure for spatial autocorrelation[47]:

$$\varphi_{\text{sac}_j} = abs(C_j - 1)$$

with C_j as Geary's contiguity ratio for variable j, calculated from the image representation of the differences in column j of $\Delta_{\text{test}}^{\text{ims}}$.

Geary's $C \in [0,2]$ with $\begin{cases} C > 1 & \text{negative spatial autocorrelation} \\ C = 1 & \text{no spatial autocorrelation} \\ C < 1 & \text{positive spatial autocorrelation} \end{cases}$.

When φ_{sac_j} is large, the prediction differences seem to have spatial structure for this variable, indicating that a potential tissue or biological structure was not well predicted. When φ_{sac_j} is small, the prediction differences seem randomly distributed across the tissue and the probability of the differences reporting genuine 'missed' tissue structure goes down.

Definition 4.3.3. (intensity reconstruction score) Let the $\varphi_{\text{irs}_j} \in [0,1]$ denote an intensity reconstruction score for IMS variable j, such that the intensity reconstruction score vector $\varphi_{\text{irs}} \in [0,1]^{N_{\text{ims}}}$ reports per column how well the absolute intensity predictions approximate the measurements, taking into account the relative severity of the differences through difference score vector $\hat{\varphi}_{\text{diff}}$ and the spatial randomness of the differences through spatial autocorrelation score vector φ_{sac}. The intensity reconstruction score φ_{irs_j} for IMS variable j is calculated as:

$$\varphi_{\text{irs}_j} = 1 - \hat{\varphi}_{\text{diff}_j}\varphi_{\text{sac}_j}$$

with φ_{sac_j} reducing the influence of $\hat{\varphi}_{\text{diff}_j}$ proportional to how randomly distributed the differences are.

When φ_{irs_j} is large, the absolute intensities across the tissue are well approximated through prediction. When φ_{irs_j} is small, the absolute intensities across the tissue are not well approximated through prediction, or the differences describe a structured sub-area within the tissue. Since the measured ion images are prone to contain notable amounts of so-called 'salt-and-pepper' noise[45], a result of their Poisson-like signal nature, and such noise can considerably overinflate the difference score, it is necessary to selectively attenuate difference scores based on randomly distributed salt-and-pepper noise, compared to difference scores reporting structural differences. We accomplish this by multiplying the difference score with a spatial autocorrelation score before using it to obtain the intensity reconstruction score. If the difference image for a variable primarily reports structural patterns missing from the prediction, such as differences in homogeneous regions of pixels, this score will be close to one and multiplication with the difference score will make little difference. However, if the difference image consists primarily of randomly distributed noise variation, the prediction should be considered of good quality. In that case the spatial autocorrelation score will be close to zero, largely eliminating by multiplication the difference score from the intensity reconstruction score.

The second global measure is the pattern reconstruction score, which assesses how well the relative spatial distribution of intensities matches that of the measured ion image. This score is less concerned with whether a predicted ion peak reaches the right absolute height, and instead grades the prediction on putting an ion peak at the right location in the tissue. If the task is to predict relative presence in tissue and the goal is not necessarily quantitative, this score can act as a guide in selecting for which ions fusion can bring something to the table.

Definition 4.4. (pattern reconstruction scoring function.) Let $f_{\text{prs}}^{q_{\text{ims/micro}}}: \mathbb{R}^{P_{\text{ims}} \times N_{\text{micro}}} \times q_{\text{ims/micro}} \times \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}} \rightarrow \mathbb{R}^{N_{\text{ims}}}$ denote a pattern reconstruction scoring function, which is a global evaluation function that summarizes the prediction performance of the IMS-microscopy model implementing $q_{\text{ims/micro}}$ as a single value per IMS variable, emphasizing approximation of the relative spatial distribution of the intensities. The pattern reconstruction score vector $\varphi_{\text{prs}} \in [0,1]^{N_{\text{ims}}}$ provides a scalar percentage score for each IMS variable, and is calculated per column j from $S_{\text{test}}^{\text{ims}}$ and $S'^{\text{ims}}_{\text{test}}$. The elements in pattern reconstruction score vector φ_{prs} report per column how well the relative intensity pattern of the predictions approximates that of the measurements, taking care to remove localized intensity spikes from the analysis by using a 2-D median filter on the image representation of the measurements and predictions. The pattern reconstruction score φ_{prs_j} for IMS variable j is calculated as:

$$\varphi_{\text{prs}_j} = \text{corr}(\text{medfilt}(\sigma_{\cdot j}), \text{medfilt}(\sigma'_{\cdot j}))$$

with $\sigma_{\cdot j}$ as the jth column vector from $S_{\text{test}}^{\text{ims}}$ containing the measured intensities for all observations of variable j, and $\sigma'_{\cdot j}$ as the jth column vector from $S'^{\text{ims}}_{\text{test}}$ containing the predicted intensities for all observations of variable j. The medfilt function takes a column vector, converts it to its unflattened image representation, performs 2-D median filtering (with a default 3×3 window) on the image to remove outlier intensities, and flattens the resulting image back into a column vector. The corr function takes two column vectors and calculates Pearson's linear correlation coefficient to report their relative pattern similarity. If the correlation coefficient reports a negative correlation value, $\varphi_{\text{prs}_j} = 0$.

When φ_{prs_j} is large, the relative intensity distribution pattern across the tissue for IMS variable j is well approximated through prediction. When φ_{prs_j} is small, the relative intensity distribution pattern of the measurements is not well approximated. Applying a 2-D median filter on the measured and predicted patterns before calculating the correlation measure substantially diminishes perturbation of the correlation measure by noise.

The third global measure is the overall reconstruction score, simply indicated as 'reconstruction score' in the images, and it is a combination of the intensity and pattern reconstruction scores, giving both prediction aspects equal weight. Since we have no reason to prefer one aspect over the other for the applications examined in the paper, all examples assess prediction performance for an IMS variable on the basis of the overall reconstruction score.

Definition 4.5. (overall reconstruction scoring function.) Let $f_{\text{rs}}^{q_{\text{ims/micro}}}: \mathbb{R}^{P_{\text{ims}} \times N_{\text{micro}}} \times q_{\text{ims/micro}} \times \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}} \rightarrow \mathbb{R}^{N_{\text{ims}}}$ denote an overall reconstruction scoring function, which is a global evaluation function that summarizes the prediction performance of the IMS-microscopy model implementing $q_{\text{ims/micro}}$ as a single value per IMS variable, emphasizing both the tissue-wide absolute peak intensity approximation and the approximation of the relative spatial distribution of the intensities. The overall reconstruction score vector $\varphi_{\text{rs}} \in [0,1]^{N_{\text{ims}}}$ provides a scalar percentage score for each IMS variable, and is calculated per column j from $S_{\text{test}}^{\text{ims}}$, $S'^{\text{ims}}_{\text{test}}$, and $\Delta_{\text{test}}^{\text{ims}}$. The overall reconstruction score φ_{rs_j} for IMS variable j is calculated as:

$$\varphi_{\text{rs}_j} = \frac{\varphi_{\text{irs}_j} + \varphi_{\text{prs}_j}}{2}$$

with equal weight given to absolute intensity prediction and relative pattern prediction.

When φ_{rs_j} is large, the measurements for IMS variable j are well approximated through prediction. When φ_{rs_j} is small, the predictions are not good approximations of the measured values. In the examples presented here, it is the vector of reconstruction scores that is used later on in prediction scenarios to provide an indicator of prediction reliability.

The role of local evaluation is to provide a location-specific assessment of how well a particular IMS variable can be predicted using this model at that location in the tissue. Where a reconstruction score summarizes prediction performance across the entire tissue section as a single number, which is a good method to obtain initial insight into the overall fusion potential of a particular IMS variable, it is important to note that prediction performance of a fusion model is not only a function of the output variable, but also of location. In some sub-areas of the tissue, a model can give excellent predictions for a certain ion species, while in other tissue areas its prediction for that same ion species is off (e.g. due to other underlying cell types giving less telling microscopy clues). It is also important to realize that in

most cases, researchers are interested in a particular tissue region of interest, which means that as long as a fusion model delivers good performance in that area its predictions might have value regardless of how it performs elsewhere. In order to obtain some insight into where in the tissue a fusion model delivers good performance for a particular IMS variable, we employ the same round-trip prediction used to obtain the various reconstruction scores to calculate two spatially aware prediction performance measures. The first is the absolute residuals image, which is obtained by taking the absolute values of the difference image.

Definition 4.6. (absolute residuals imaging function.) Let $f_{\text{ari}}^{q_{\text{ims/micro}}}: \mathbb{R}^{P_{\text{ims}} \times N_{\text{micro}}} \times q_{\text{ims/micro}} \times \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}} \rightarrow \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$ denote an absolute residuals imaging function, which is a local evaluation function that reports localized prediction performance of the IMS-microscopy model implementing $q_{\text{ims/micro}}$ as an image per IMS variable, making local deviations between the measurements and predictions visible. The absolute residuals image of IMS variable j shows the spatial distribution throughout the tissue of the absolute peak intensity differences between the measured ion image and the predicted ion image. The absolute residuals image matrix $\theta_{\text{ari}} \in \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$ provides for each IMS variable j a column vector $\theta_{\text{ari}.j} \in \mathbb{R}^{P_{\text{ims}} \times 1}$, which can be unflattened into an absolute residuals image of size $R_{\text{ims}} \times C_{\text{ims}}$. The absolute residuals image column vector for IMS variable j is calculated as

$$\theta_{\text{ari}.j} = \text{abs}(\Delta_{\text{test}.j}^{\text{ims}})$$

.

In an absolute residuals image, areas with low values indicate good predictive approximation of the measured ion image, while areas that light up indicate tissue regions where the fusion-based prediction is off.

A second spatially aware performance measure is the 95% confidence interval image, which is obtained by calculating in addition to the main fusion model a set of bootstrap models[41,42].

Definition 4.7. (95% confidence interval imaging function.) Let $f_{\text{ari}}^{q_{\text{ims/micro}}}: \mathbb{R}^{P_{\text{ims}} \times N_{\text{micro}}} \times q_{\text{ims/micro}} \times \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}} \rightarrow \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$ denote a 95% confidence interval (CI) imaging function, which is a local evaluation function that reports localized prediction performance of the IMS-microscopy model implementing $q_{\text{ims/micro}}$ as an image per IMS variable, making local model robustness visible. The CI image of IMS variable j shows the spatial distribution throughout the tissue of the dispersion in predictions by a set of different bootstrap models, trained in addition to the main fusion model on perturbed versions of the training set. The CI image matrix $\theta_{\text{cii}} \in \mathbb{R}^{P_{\text{ims}} \times N_{\text{ims}}}$ provides for each IMS variable j a column vector $\theta_{\text{cii}.j} \in \mathbb{R}^{P_{\text{ims}} \times 1}$, which can be unflattened into a CI image of size $R_{\text{ims}} \times C_{\text{ims}}$. The CI image column vector for IMS variable j is calculated as

$$\theta_{\text{cii}.j} = \text{confint}(\{S'^{\text{ims},b}_{\text{test}\ .j}\}_{b=1}^{\mathcal{B}})$$

with $S'^{\text{ims},b}_{\text{test}\ .j}$ as the predictions for IMS variable j made by bootstrap model b, and $\mathcal{B}$ being the total number of bootstrap models that were calculated in addition to the main fusion model. The confint function calculates row-wise 95% confidence intervals over the matrix that concatenates the set of column vectors $S'^{\text{ims},b}_{\text{test}\ .j}$ it is passed, and returns a column vector of size $P_{\text{ims}} \times 1$.

The CI-image reports for each native IMS pixel the dispersion of the predictions provided by different bootstrap models. In the areas where this dispersion is large, prediction will be less robust. Due to the computational cost involved with bootstrapping, the examples we show have been obtained on a set of only ten bootstrap models. Given the multivariate nature of the data, this number is too low to deliver any strong statistical conclusions on prediction confidence. However, even a limited number of bootstrap models provide an indication of which tissue areas are more robustly predicted than others.

One could argue that the fusion model will be over-fitted to the tissue area on which it is trained, and that these evaluation scores will usually underestimate the true prediction error. In that respect it is important to note that although the training and test sets are derived from the same image data sources, they are not copies of each other. The round-trip prediction on which the evaluation is based is done on a simple down-sampled version of the microscopy and does not contain the weighting aspects introduced into the training set in step 2. Also, the issue of over-fitting to a certain tissue area is less of a problem in fusion applications such as sharpening, where the prediction typically needs to happen within that same tissue area. For fusion applications that predict outside the IMS-measured area, it will be important to keep the microscopy measurements as similar to those in the modeled area as possible. However, even in cases where this is difficult and prediction performance is therefore degraded, out-of-sample prediction applications are often still useful to obtain some exploratory insight, particularly when measured alternatives are lacking.

The fusion methodology described here runs regressions against a vast number of variables in one technology versus a vast number of variables in the other technology, and it is not unexpected that some cross-modality relationships are found. Some of these relationships will turn out to be generic and will hold up across any measurements made with these source technologies. Other relationships will be specific to the instruments, tissue type, sample preparation, or other case-specific parameters. Some relationships might even be coincidental or spurious, although that would require them to hold up consistently across millions of pixel measurements in the training set, which is unlikely. The methodology described in this paper does not discriminate sub-types of cross-modality relationships. Instead, it focuses solely on finding cross-modality relationships of any nature that are supported by substantial evidence, and on exploiting these relationships for predictive applications. However, evaluating what the nature is of the cross-modality relationship that drives a prediction is an interesting future research challenge. Another area of future extension is in the statistical treatment of fusion. The current method explores the multi-modal data in descriptive terms, but does not make a statement on significance using inferential statistics. We believe this work can be a stepping stone towards that challenge by revealing the unexplored potential that is locked behind cross-modality relationships and multi-modal studies, and by providing a framework to extract such information regardless of the underlying source technologies.

Supplementary Note 3 – Prediction Phase

The prediction phase of the fusion process is application-specific, but most scenarios consist of combining the model learned in the previous phase and the measurements acquired in one of the modalities, to obtain predictions in the other modality. This phase entails a prediction step (step 5), which provides predictions for all variables in the target modality, and a selection step (step 6), which prunes target variables for which prediction performance is insufficiently reliable.

Step 5 - Prediction

The prediction step is straightforward in that it consists of plugging a microscopy-derived signature into the modeling function $q_{\text{ims/micro}}$, which returns a predicted IMS signature as output. In the case of a linear IMS-microscopy model, the IMS predictions for a set of microscopy measurements $\tilde{Y}_{\text{micro}}$ can be calculated as

$$\tilde{Z}_{\text{ims}} = \tilde{Y}_{\text{micro}}\, Q_{\text{ims/micro}}$$

with $\tilde{Y}_{\text{micro}} \in \mathbb{R}^{\tilde{V}\times(1+N_{\text{micro}})}$ as $\tilde{V}$ observations in microscopy to predict for;
$Q_{\text{ims/micro}} \in \mathbb{R}^{(1+N_{\text{micro}})\times N_{\text{ims}}}$ as the coefficient matrix; and
$\tilde{Z}_{\text{ims}} \in \mathbb{R}^{\tilde{V}\times N_{\text{ims}}}$ as $\tilde{V}$ predicted observations in IMS;

In a sharpening scenario, $\tilde{Y}_{\text{micro}}$ will typically contain all microscopy pixels within the IMS-measured area at the fine-grained native microscopy resolution (or a down-sampled set if another target resolution is pursued). For an out-of-sample prediction outside the IMS-measured area, $\tilde{Y}_{\text{micro}}$ will contain microscopy pixels that were not part of the modeling phase to begin with.

Step 6 - Selection

Although the prediction step returns intensity predictions for each ion peak in the IMS data source, it is important to realize that the model only gives good results for a subset of the N_{ims} variables and that fusion results should only be considered for that subset of ion peaks. The quality of the prediction for an ion peak depends on whether there are cross-modality relationships relevant to that ion peak present between the provided data sources. It also depends on how well the assumptions made during the model-building-and-evaluation phase are met by the provided data. These assumptions include both structural parameters (e.g. the chosen model structure) as well as practical parameters (e.g. the number of weight levels employed in the mapping step). The reconstruction scores, calculated in the step 4, can serve as a heuristic for the extent to which these assumptions are met for each IMS variable and whether there is potential for good prediction. For this reason, step 5, which calculates predictions for each of the N_{ims} variables, will typically be followed by a selection step retaining only the IMS variables that exhibit good prediction according to their reconstruction score.

Definition 6.1. (selection function.) Let $f_{\text{sel}}: \mathbb{R}^{N_B}\times f_{\text{ge}}^{q_{B/A}}\times[0,1] \rightarrow \mathbb{R}^{\Psi}$ denote a selection function that takes a predicted observation of size $1\times N_B$ in modality B, a global evaluation function $f_{\text{ge}}^{q_{B/A}}$, a threshold value $\psi \in [0,1]$, and returns a vector of size $1\times\Psi$ that is a subset of the predicted observation, retaining only the variables whose prediction performance, as determined by $f_{\text{ge}}^{q_{B/A}}$, exceeds ψ.

In our IMS-microscopy setting, the global evaluation function $f_{\text{ge}}^{q_{B/A}}$ is implemented by the overall reconstruction score φ_{rs} and the threshold value $\psi = 0.75$. In essence, step 6 reduces a predicted IMS

signature vector in $\mathbb{R}^{N_{\text{ims}}}$ to a vector of predictions that meet a threshold of reliability as reported by the reconstruction score. This final set of predictions is a vector in $\mathbb{R}^{\Psi}$, where Ψ is the number of IMS variables that give good fusion-driven prediction. In all case studies described here, this threshold was set to 75% overall reconstruction score. Below that value, the fusion-based prediction was ignored.

Supplementary Note 4 – Application-specific assumptions

Besides the general assumptions tied to fusion and cross-modality modeling, there are also application-specific facets to consider. These include any assumption that needs to be met to be able to use the cross-modality model in a particular prediction scenario. For sharpening, the assumption is that the relationships that are modeled hold across multiple scales, at least down to the target spatial resolution. For prediction outside the IMS-measured area, we assume that the modeled relationships are sufficiently general such that they hold for the non-IMS tissue area or section as well, and that the microscopy provided for those sections is similar to the microscopy measurements in the modeled area. However, even with moderate violation of these assumptions we have found prediction results to be quite robust and commonly in line with independent verification via measurement.

Although it is necessary to keep the imposed constraints and assumptions in mind when interpreting the fusion results, the promising insights already delivered by relatively basic step implementations demonstrate the potential of heterogeneous fusion for multi-modal imaging. At the same time, the approaches we describe show ample opportunity for further improvement and hint at a lot of un-mined potential still hidden behind cross-modality relationships that were not sufficiently captured by our current implementations.

Supplementary Results

The results demonstrate the potential for predictive signal improvement through multi-modal fusion and this without the need for instrument modification (**Fig. 3c** versus **3a**). When the assumptions are met, prediction can come close to actual measurement at the same spatial resolution (**Fig. 3c** versus **3d**), and this provides a means of circumventing physical acquisition when not reasonably attainable. Additionally, we use the same data set to illustrate the advantage of multi-modality measurements over same-modality data processing (**Supplementary Fig. 4**). We show an example of *in silico* up-sampling via fusion (using two modalities) clearly outperforming classical bilinear interpolation[45] (using a single modality) (**Supplementary Fig. 4c** versus **4e**) when it comes to estimating the true tissue content (**Supplementary Fig. 4d**).

Although the fusion-predicted ion distribution at 10 μm (**Fig. 3c**) approximates the measured ion distribution at 10 μm (**Fig. 3d**) well, some differences in ion patterning are still visible between prediction and measurement. There are two reasons for this. (a) The fusion-based prediction is not perfect. This is reported to the user by a reconstruction score of 82% rather than 100% for this ion species, and indicators such as the absolute residuals and CI images further characterize this spatially as well. (b) The IMS measurement is not perfect either and therefore not necessarily a true gold standard to compare against. A measurement acquired at 10 μm often has reduced sensitivity due to the reduced amount of ablated material. It will therefore report only the most abundant of biological patterns and is more sensitive to the presence of measurement noise (e.g. matrix inhomogeneities). Considering the coarseness of the native IMS measurement at 100 μm (**Fig. 3a**) that the prediction is based on and the lackluster results that are obtained in the absence of other-modality information (**Supplementary Fig. 4e**), the fusion-based prediction comes remarkably close to the measurement at 10 μm resolution. Whether the prediction is close enough for a particular purpose depends on the case at hand and on whether qualitative prediction is sufficient or whether quantitative accuracy is also required (and if so, what sensitivity level would answer the question at hand).

Good prediction is possible even for ion peaks reporting panels of ions, and it does not necessarily require uniquely resolved ion species. The *m/z* 747.5 image (**Supplementary Fig. 9**) is provided by the same IMS experiment that delivered the *m/z* 762.5 example discussed earlier (**Fig. 3**). Although the nominal *m/z* peak is composed of multiple ion species (^{13}C PE-NME$_2$(16:0/18:0), ^{13}C PE(P-16:0/22:6), and PA(18:0/22:6)), unresolved by the MS analyzer in this experiment, its overall reconstruction score of 86% indicates that for this IMS variable strong modeling and prediction is possible. For a multiple-species variable such as this, the measured distribution is a superposition of the distributions of the individual species. As the fusion procedure has no information on the individual species and only has access to the combined peak distribution, its prediction for this variable will pertain to the combined panel.

The fusion model also has the capability to predict at any spatial resolution between the low (native IMS) resolution and the high (microscopy) resolution. This is made possible by training the model on data that characterizes the tissue at different resolutions and letting the model generalize relationships that span the scales between the source resolutions. For example, the ion *m/z* 778.5 (**Supplementary Fig. 10**), which has been identified as PE(P-40:4) and gives a reconstruction score to H&E stained microscopy of 76%, is predicted at spatial resolutions of 100, 50, and 5 μm. These results are further extended (**Supplementary Fig. 11**) with predictions for *m/z* 778.5 at 75 and 25 μm, and with measured ion images for comparison, acquired via time-of-flight (TOF) and Fourier transform ion cyclotron resonance (FTICR) instruments.

The fusion process is not exclusive to a particular tissue or molecule type. Although many examples in this study focus on the lipid mass range and on mouse brain samples, fusion capabilities are also demonstrated on protein images of rat kidney measured between *m/z* 3,000 and 20,000 (**Supplementary Fig. 12**). These examples take an IMS measurement of a renal cross-section at 100 μm

resolution (**Supplementary Fig. 12a**), with ion distributions localizing to the kidney cortex, medulla, and pelvis, and fuses them with an H&E stained microscopy image of the same tissue section at 5 μm resolution (**Supplementary Fig. 12b**), measured after IMS analysis. The fusion result predicts protein ion abundances in the kidney up to the native microscopy resolution of 5 μm (**Supplementary Fig. 12c**).

Furthermore, the variety of native IMS resolutions presented by the different case studies (100, 80, 75, and 10 μm, see **Supplementary Table 1**), demonstrate that fusion is not tied to a particular absolute laser footprint. As long as the data sources provide a strong signal, and there are relationships to be found at the resolutions in question, fusion applications can be considered.

The fusion method does not require multi-modal relationships to be defined prior to operation, but instead searches for them itself and evaluates whether they are sufficiently strong to drive prediction applications. As a result, the fusion method is not tied to any particular imaging technology, and will function with other modalities than IMS and H&E stained microscopy. The method can mine cross-modality relationships between any image types that share a common spatial basis, and use these modeled links for fusion-driven prediction. The ability to conduct the fusion process on different data sources is demonstrated (**Supplementary Figs. 13-16**) using examples from an experiment where an IMS measurement of a coronal mouse brain section, acquired at 80 μm spatial resolution in the lipid mass range, is fused with an H&E stained microscopy image on the one hand and a Nissl stained microscopy image on the other hand. Both microscopy sources are measured at 10 μm resolution and acquired from neighboring tissue sections. The difference in stain type between the two fusion runs reveals different structures and tissue patterns in their respective microscopy sources, and thus influences the cross-modality connections that can be made to IMS variables. These examples illustrate that the fusion method we have developed can be applied across different image sources, and they also demonstrate that prediction performance is dependent on the content and particular combination of source modalities.

The ability of the fusion method to capture cross-modality relationships (and the evaluation step to accurately score them) is hard to assess on real-world biological measurements, as these data sets do not provide a gold standard to compare against. For this purpose, we created a synthetic multi-modal data set that mimics IMS and microscopy characteristics (e.g. spatial resolution, number of variables per pixel, etc.). We embedded into the data set known cross-modal and modality-specific patterns for the algorithm to find and use. In order to better approximate real measurement conditions, we also added a mixture of Gaussian and Poisson noise on top of these patterns to mimic measurement uncertainty and detector noise. The fusion task consists of integrating an IMS-like modality at 75 μm spatial resolution with a microscopy-like modality acquired at 5 μm, and to sharpen the IMS-like patterns to 5 μm. The method behavior and fusion results are shown for three of the embedded patterns, each with differing amounts of cross-modal support (**Supplementary Fig. 17**). The first example (top) focuses on sharpening a pattern with strong cross-modal support across the entire tissue, and demonstrates excellent prediction. This indicates that the fusion method is able to detect such relationships even with substantial IMS and microscopy noise present in the measurements. The predictive power for this variable is also accurately captured by the reconstruction score, which reports a value of 87%. A second example (middle) shows method behavior in the case of a pattern that is only partially supported across modalities, with some tissue subareas providing good support and other subareas providing little to none. Also in this case the cross-modal prediction is excellent in areas that have a connection to the microscopy, but there is a serious prediction error in areas that do not have such a cross-modal connection for this variable. Since the reconstruction score is meant to summarize performance across all tissue, the presence of subareas with reduced fusion-driven prediction performance or IMS-specific features is reflected in the reduced reconstruction score of 81%. Additionally, the method pinpoints the tissue location of the modality-specific feature through the absolute residuals image, giving the researcher the information to assess whether this area of reduced prediction confidence overlaps with a tissue area of interest. The third and last example (bottom) reports method behavior for a modality-specific pattern that does not have cross-

modal support. Although the method tries to approximate the IMS pattern as best it can, using the vocabulary of microscopy-derived patterns available to it (the eight native patterns shown in **Supplementary Figure 17** plus the patterns derived through textural filters etc.), good prediction is never really achieved and the low 66% reconstruction score accurately reports this to the user. In addition to assessing the behavior of the fusion method in various cross-modal support situations, the synthetic data set also highlights the necessity for multivariate fusion models rather than univariate measures such as correlation between modalities. A good example of this is the pattern at the top (**Supplementary Fig. 17**). This pattern has great cross-modal support in a multivariate sense, since it can be approximated well by a combination of multiple microscopy-derived patterns. However, it does not have good cross-modal support in a univariate sense, since none of the microscope variables alone can provide a good approximation of the IMS pattern. Hence a correlation measure, which assesses univariate cross-modal support, would have reported a low value for this pattern and an opportunity to reveal and utilize cross-modal information would have gone unnoticed. Instead, when fusion is approached in a multivariate sense that allows pattern combinations (such as the linear models developed here), the pattern is picked up, connected to the other modality, and reports an excellent score that makes fusion applications possible. A key observation is that most modalities will not provide patterns that directly correlate with patterns measured by another technology, and thus multivariate mixing-capable models are essential to making fusion and finding cross-modality information possible in the majority of cases.

Supplementary Discussion.

Prediction beyond the IMS measurement resolution

Predicting at resolutions that exceed the measurement resolution might seem counterintuitive at first, but it is an established approach throughout science. A simple one-dimensional example is the standard curve, where a least squares line is fit to a set of measured points. When that line is used to predict concentration for a previously unseen instrument response, we predict points in between the original measured points, effectively predicting at a resolution more fine-grained than the measurement resolution. Standard curve-predicted values are commonly used to establish quantitative measurements, which further demonstrates the analytical value of predictions. Where the standard curve predicts values between measured points using only a model assumption, image fusion employs a model together with actual measurements (from another modality) to drive its estimations, giving it the potential for more reliable and sample-driven prediction. We demonstrate this (**Supplementary Fig. 4**) with a fusion-based prediction of the distribution of *m/z* 762.5 at 10 µm (**Supplementary Fig. 4c**). This fusion-driven prediction is much closer to the measured distribution (**Supplementary 4d**) than the prediction obtained via classical (**Supplementary Fig. 4e**), which uses only a model assumption and no other-modality information.

It is useful to note that we predict beyond the measurement resolution of the IMS data, but not beyond the measurement resolution of the microscopy data. By not predicting beyond the resolution of the spatially most fine-grained data source, we ensure that predictions can be driven by actual measurements, albeit from another modality, rather than that they need to be driven by assumptions of finer resolution behavior. This measurement-based aspect of fusion-driven prediction translates into substantially increased prediction reliability for variables that exhibit strong relationships to another modality.

Reliability of fusion-driven prediction

The reliability of fusion-based prediction is subject to the same considerations as any other modeling effort. In addition to application-specific assumptions, the validity of any cross-modality model will depend, for example, on whether the measurements that make up the training set are representative for the measurements on which prediction will be based; on whether the model type, structure, and parameters provide the necessary degrees of freedom to accurately describe the cross-modality relationships without over-fitting on the training examples; and on whether fusion parameters such as the chosen mapping function hold true for the modalities in question. To provide guidance, each fusion-based prediction is accompanied by indicators that assess its reliability both along the spectral (reconstruction scores) as well as the spatial domain (absolute residuals and CI images). These indicators allow the user to determine whether the prediction is reliable for a particular ion species of interest and/or a tissue sub-area of interest. Say, we want to predict the distribution of an ion species in a tissue area not measured via IMS (a scenario similar to **Fig. 5**), and that area contains an anomalous structure such as a tumor. If the modeled area contains tumor traces, the training set contains tumor signature examples and the model can predict for tumor and healthy tissue areas alike. If the modeled area does not contain tumor tissue, the training set does not provide any tumor examples and the model will have no material on which to base its prediction for tumor areas. In the latter case, the training area is not fully representative for the prediction area, and predictions for unrepresented (tumor) areas will be less reliable. However, the user is alerted to a reduced reliability in tumor areas by indicators such as the CI-image, which will spatially outline the anomalous areas with reduced prediction confidence. Further research on integrating classical validation approaches into the fusion workflow, and on how the non-independent nature of spatial measurements influences the modeling process are promising areas for future extension of image fusion.

Empirical nature of the image fusion method

Since the relationships between measurement principles do not need to be specified beforehand, but instead are empirically inferred from a set of measurements, our image fusion method is generic and can be applied to a wide variety of different imaging modality combinations. This empirical discovery aspect also brings some nuance to the type of correspondences that exist between different sensor types, which is information often absent from data integration efforts that follow connections defined prior. Some cross-modality relationships are highly reliable and give strong confident predictions, while others give only rudimentary clues about the tissue content at particular location. Some relationships hold across an entire tissue sample, while others only hold in a particular tissue sub-type. Some relationships report a well-understood biological connection between the observations in two different image types (e.g. same cell type reported in different modalities). Other relationships report a previously unknown connection between sensors, giving good predictions without a full understanding of the underlying mechanism that drives this connection (e.g. a microscopy texture type that consistently corresponds to the presence of a particular molecule).

Contribution to microscopy

Although most examples in this paper discuss IMS-microscopy fusion from the perspective of what microscopy can bring to IMS-based analysis of tissue, it is important to note that image fusion between modalities is a two-way street. Just as microscopy brings spatially fine-grained information to IMS, the same fusion model allows IMS to bring chemically fine-grained information to microscopy in the form of mass spectrometry-grade molecular specificity. The pixel color in a fusion result is not just any re-coloring of the microscopy image, but rather a chemically very specific re-coloring that predicts the spatial presence and abundance of a particular molecule species. This specificity is biologically very important, yet it is unavailable from the broad dye coloring natively provided by stained microscopy, and exceeds the molecular specificity that can be achieved through tag-based approaches such as fluorescence microscopy and immunohistochemistry (e.g. regarding post-translational modifications, families of similar molecule species, etc.). For example, in pathology the added chemical specificity of a fusion result could allow a diseased tissue area to be differentiated from healthy tissue, where from the microscopy alone one might not be able to tell. Furthermore, microscopy benefits from any fusion application that utilizes cross-modality relationships to gain deeper insight into what is biological signal and what is noise. Enrichment specifically is an interesting application for microscopy. We gave an example of tissue features that were barely visible in microscopy, if not below the limit of detection of the human eye (**Fig. 6**, annotations *a*-*c*). The fusion process and the cross-modality relationships with IMS allowed us to enrich these features for microscopy users to a point where they can be clearly observed and used for tissue interpretation. By using multi-modality information we are able to pull more information from a microscopy image than would be possible or practical if the microscopy image would be examined as part of a single-modality study.

Supplementary References

41. Wehrens, R. & Linden, W. v. d. Bootstrapping principal component regression models. *Journal of Chemometrics* **11,** 157-171 (1997).

42. Wehrens, R., Putter, H., & Buydens, L. The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems* **54,** 35-52 (2000).

43. Jolliffe, I. T. *Principal component analysis.* Springer, New York, 2002.

44. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401,** 788-791 (1999).

45. Gonzalez, R. C. & Woods, R. E. *Digital image processing.* Prentice Hall, Upper Saddle River, NJ, 2002.

46. Coombes, K. R. *et al.* Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics* **5,** 4107-4117 (2005).

47. Schabenberger, O. & Gotway, C. A. *Statistical methods for spatial data analysis.* Chapman & Hall/CRC, Boca Raton, 2005.