Supplementary information

for

Image fusion of mass spectrometry and microscopy: a new multi-modality paradigm for molecular mapping of tissue

Raf Van de Plas[1], Junhai Yang[1], Jeffrey Spraggins[1], and Richard M. Caprioli[1,2]

[1] Mass Spectrometry Research Center and Department of Biochemistry, Vanderbilt University, Nashville, Tennessee, U.S.A.
[2] Departments of Chemistry, Pharmacology, and Medicine, Vanderbilt University, Nashville, Tennessee, U.S.A.

Correspondence should be addressed to R.M.C. (richard.m.caprioli@vanderbilt.edu).

Supplementary Results

The results shown in Figure **3** demonstrate the potential for predictive signal improvement achieved with multi-modal fusion, without the need for instrument modification (Fig. **3**c vs. **3**a). When the assumptions are met, prediction can come close to actual measurement at the same resolution (Fig. **3**c vs. **3**d), providing a means of circumventing physical acquisition when not reasonably attainable. Finally, the advantage of multi-modality measurements over same-modality data processing is illustrated in Supplementary Figure 4, where *in silico* up-sampling via fusion is clearly superior to that of interpolation (Supplementary Fig. 4c vs. 4e).

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 shown in Supplementary Figure 9 is an example from the same IMS experiment as in Figure **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 in Supplementary Figure 10, the ion m/z 778.5, 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. Supplementary Figure 11 further extends these results for m/z 778.5 with 75 and 25 μm predictions, and for comparison includes measured ion images 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, Supplementary Figure 12 uses protein images of rat kidney measured between m/z 3,000 and 20,000. The example takes an IMS measurement of a renal cross-section at 100 μm resolution (a), with ion distributions localizing to the kidney cortex, medulla, and pelvis, and fuses it with an H&E stained microscopy image of the same tissue section at 5 μm resolution (b), measured after IMS analysis. The fusion result predicts protein ion abundance in the kidney up to the native microscopy resolution of 5 μm (c).

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. Supplementary Figures 13-16 show 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. Supplementary Figures 13-16 illustrate that the developed fusion method is applicable across different image sources, and 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. Supplementary Figure 17 shows the method behavior and fusion result for three of the embedded patterns, each with differing amounts of cross-modal support. 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 residual image, giving the researcher the information to assess whether this area of reduced prediction confidence overlaps with a tissue area of interest. Finally, the third and last example of Supplementary Figure 17 (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 of Supplementary Figure 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 we develop here), the pattern is picked up, connected to the other modality, and reports an excellent score making 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 Tables

Supplementary Table 1 Case study overview and details.

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 - 21000	100	MALDI-TOF (proteins) m/z 3000 - 21000	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 500 - 1000	80	MALDI-TOF (lipids) m/z 500 - 1000	10 (Fig. S13-16)	prediction using different data sources	S13-16
5	synthetic	synthetic	5	synthetic	75	synthetic	5 (Fig. S17)	modeling algorithm check	S17
6	mouse brain (transversal)	H&E	5	MALDI-TOF (proteins) m/z 3000 - 21000	100	MALDI-TOF (proteins) m/z 3000 - 21000	5 (Fig. 1, S18) 100, 50, 5 (Fig. S19) 100 (Fig. S20)	multi-modal enrichment and denoising	1, S18-20
7	mouse brain (transversal)	H&E	5 (same as case study 6)	MALDI-TOF (proteins) m/z 3000 - 21000	100 (rectangular sub-area of case study 6)	MALDI-TOF (proteins) m/z 3000 - 21000	5 (Fig. 5) 100 (Fig. S21)	prediction in non-IMS-measured areas	5, S21