Supporting Information:

MSI Warp: A General Approach to Mass Alignment in Mass Spectrometry Imaging

Jonatan O. Eriksson,† Alejandro Sánchez Brotons,‡ Melinda Rezeli,† Frank Suits,¶ György Markó-Varga,† and Peter Horvatovich*,†,‡,†

†Department of Biomedical Engineering, Lund University, 221 00 Lund, Sweden
‡Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV, Groningen, the Netherlands
¶IBM Research - Australia. 60 City Rd., Southbank VIC 3006, Australia

E-mail: p.l.horvatovich@rug.nl
Data sets

TOF kidney

This data set was downloaded from ProteomeXchange (identifier PXD013860). It was generated from two mouse kidney tissue sections (from two different animals). The spectra were collected in negative ion mode with a raster size of 100 µm. Some peaks of interest were subsequently identified with MALDI-LIFT (MS/MS). The tissue sections were sprayed with a binary matrix that was a mixture of 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (CHCA) (with a concentration of 7 mg/mL each). Baseline removal and smoothing was performed with the SCiLS software (version 2018b Core).

TOF spheroids

This data set was downloaded from ProteomeXchange (identifier PXD013069). The spectra were collected in positive ion mode with a raster size of 50 µm. A solution of 7 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA, Bruker Daltonics, Bremen, Germany) in 50% acetonitrile (Merck, Darmstadt, Germany) and 0.2% trifluoroacetic acid (Merck, Darmstadt, Germany) was used as matrix. Baseline removal, TIC (total ion current) normalization, and smoothing was performed with the SCiLS software (version 2016a, GmbH, Bremen, Germany).
**Orbitrap liver**

We previously generated this data set from a rat liver tissue section for a different study. A number of drug mixtures were spotted on the tissue section prior to MSI analysis. It is available at ProteomeXchange with the identifier PXD016146. The spectra were collected in positive ion mode with automatic gain control switched off and a raster size of 50 µm. The matrix was a 5 mg/mL solution of α-cyano-4-hydroxycinnamic acid (CHCA, Sigma-Aldrich) dissolved in 50% MeOH containing 0.1% trifluoroacetic acid (TFA, Sigma-Aldrich, Steinheim, Germany).

**Orbitrap DESI**

This data set was downloaded from MetaboLights (identifier MTBLS289). This data set contains MSI analyses of a large number of ex vivo colorectal adenocarcinoma samples, and we downloaded and aligned the sample named ”A67 CT S4-centroid.imzML”. The MSI analysis was performed using a home-built motorized DESI ion source and an LTQ XL Orbitrap Discovery mass spectrometer (Thermo Fisher Scientific Inc., Bremen, Germany). The spectra were collected in negative ion mode. Although the MetaboLights description states that the samples were analyzed over the $m/z$ range 600-1000 we found that the spectra in this data set were analyzed over the $m/z$ range 200-1000.
Table S1: Median mass dispersion around mean spectrum peaks in the TOF kidney data set for different scalings of modelled peak width, $\sigma$, and the peak matching threshold, $\epsilon$. The spectrum with index 200 was used as reference instead of that with the highest TIC. The base value of $\sigma$ was set to 150 ppm and that of $\epsilon$ was 2 FWHM (approx. 750 ppm).

|      | 0.2\(\sigma\) | 0.5\(\sigma\) | 1.0\(\sigma\) | 2.0\(\sigma\) | 5.0\(\sigma\) |
|------|---------------|---------------|---------------|---------------|---------------|
| 0.2\(\epsilon\) | 53.19         | 59.64         | 66.03         | 65.28         | 75.89         |
| 0.5\(\epsilon\) | 37.83         | 11.20         | 11.89         | 14.57         | 30.58         |
| 1.0\(\epsilon\) | 37.83         | 11.34         | 11.92         | 16.58         | 59.12         |
| 2.0\(\epsilon\) | 37.83         | 11.34         | 11.93         | 25.19         | 95.78         |
| 5.0\(\epsilon\) | 37.83         | 11.34         | 11.93         | 25.19         | 104.55        |

Table S2: Median mass dispersion around mean spectrum peaks in the Orbitrap liver data set for different scalings of modelled peak width, $\sigma$, and the peak matching threshold, $\epsilon$. The base value of $\sigma$ was set to 6 ppm at 400 m/z and that of $\epsilon$ was 2 FWHM (approx. 30 ppm at 400 m/z).

|      | 0.2\(\sigma\) | 0.5\(\sigma\) | 1.0\(\sigma\) | 2.0\(\sigma\) | 5.0\(\sigma\) |
|------|---------------|---------------|---------------|---------------|---------------|
| 0.2\(\epsilon\) | 0.269         | 0.191         | 0.178         | 0.185         | 0.2233        |
| 0.5\(\epsilon\) | 0.269         | 0.191         | 0.178         | 0.186         | 0.2233        |
| 1.0\(\epsilon\) | 0.269         | 0.191         | 0.178         | 0.186         | 0.2242        |
| 2.0\(\epsilon\) | 0.269         | 0.191         | 0.178         | 0.187         | 0.2989        |
| 5.0\(\epsilon\) | 0.269         | 0.191         | 0.178         | 0.187         | 0.3103        |
Figure S1: TIC images for the four data sets.
Figure S2: RANSAC results for three example spectra from the TOF kidney data set. True and spurious matches are marked with circles and squares, respectively. The orange crosses mark the sampled matches (2 from each segment). Note that one of the sampled matches is spurious in both (b) and (c), but is ignored in the warping function.
Figure S3: TOF kidney mean spectrum peak scatters 1-35.
Figure S4: TOF kidney mean spectrum peak scatters 36-70.
Figure S5: TOF kidney mean spectrum peak scatters 71-100.
Figure S6: TOF spheroids mean spectrum peak scatters 1-30.
Figure S7: TOF spheroids mean spectrum peak scatters 31-50.
Figure S8: Orbitrap liver mean spectrum peak scatters 1-35.
Figure S9: Orbitrap liver mean spectrum peak scatters 36-70.
Figure S10: Orbitrap liver mean spectrum peak scatters 71-100.
Figure S11: Orbitrap DESI mean spectrum peak scatters 1-35.
Figure S12: Orbitrap DESI mean spectrum peak scatters 36-70.
Figure S13: Orbitrap DESI mean spectrum peak scatters 71-100.