Zentrum für Technomathematik

Cross-normalization for MALDI TOF MSI

Improving inter-lab comparability and multi-center studies

Tobias Boskamp
Head, Bioinformatics Group
Center for Industrial Mathematics
University of Bremen, Germany

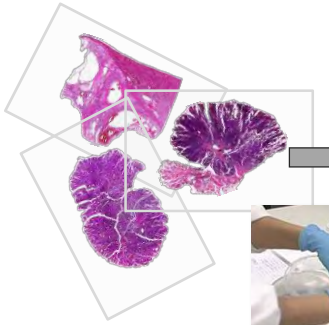
CoI disclosure: TB is consultant with SCiLS (Bremen, Germany)

SCiLS

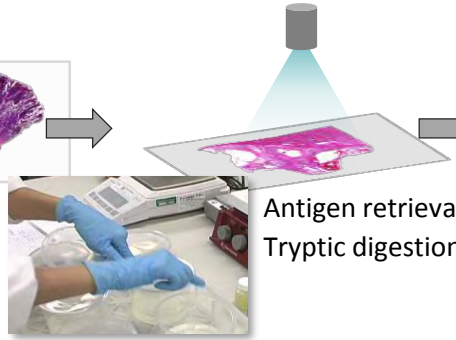
Universität Bremen*

The Role of Reproducibility in MALDI Imaging Research

FFPE tissue samples



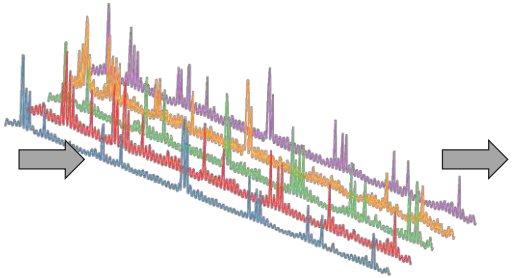
Sample preparation



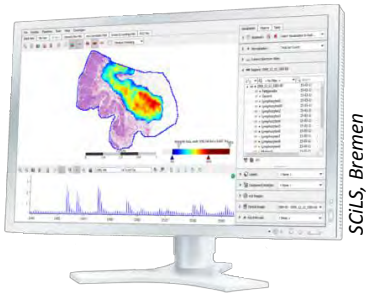
Antigen retrieval
Tryptic digestion

MALDI data acquisition

Bruker Daltonik, Bremen



Data analysis



SCiLS, Bremen

Process complexity

Operator dependence

Technical variability

Reproducibility and comparability …

- *across measurements*
- *over time*
- *across instruments*
- *across laboratories*

Large cohort studies

Longitudinal studies

Multi-center studies

Biomarker discovery

Pathway analysis

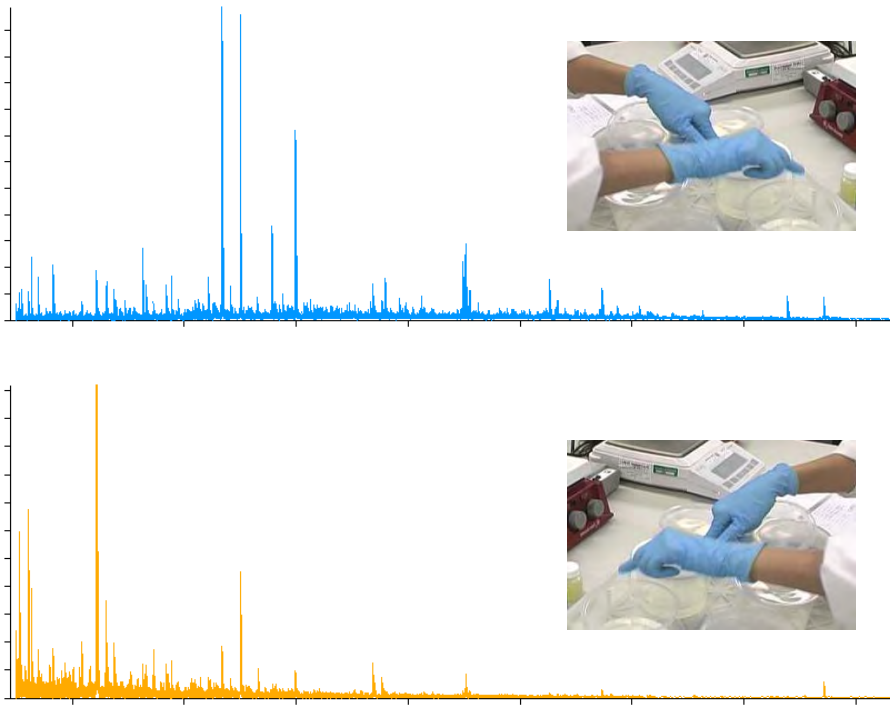
Prognostic factors

Drug efficacy

Technical Variation in MALDI TOF MSI

- High sensitivity to process variations
- Differences between measurements often larger than between phenotypes
- Variability effects include
  - ***Ionization / ion suppression***
  - ***Delocalization***
  - ***Noise***
  - ***Intensity / sensitivity variations***
  - ***Mass distortions***
  - …

All data is dirty!

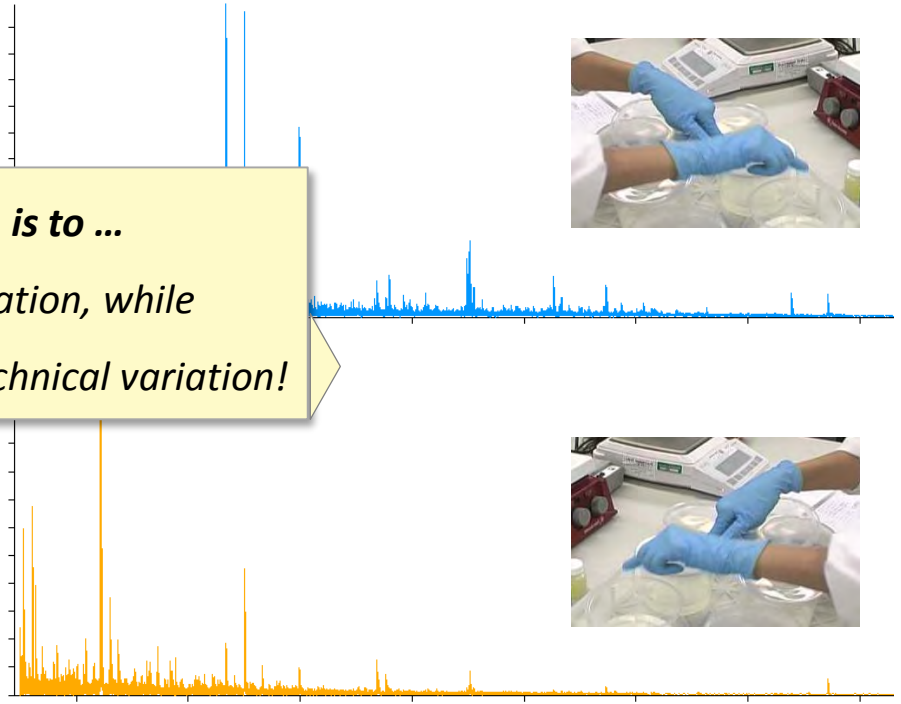


Phenotypes vs. Technical Variation

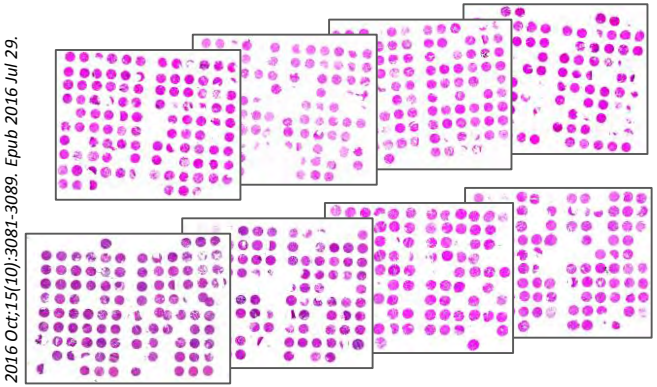
The challenge is to …

preserve biological variation, while

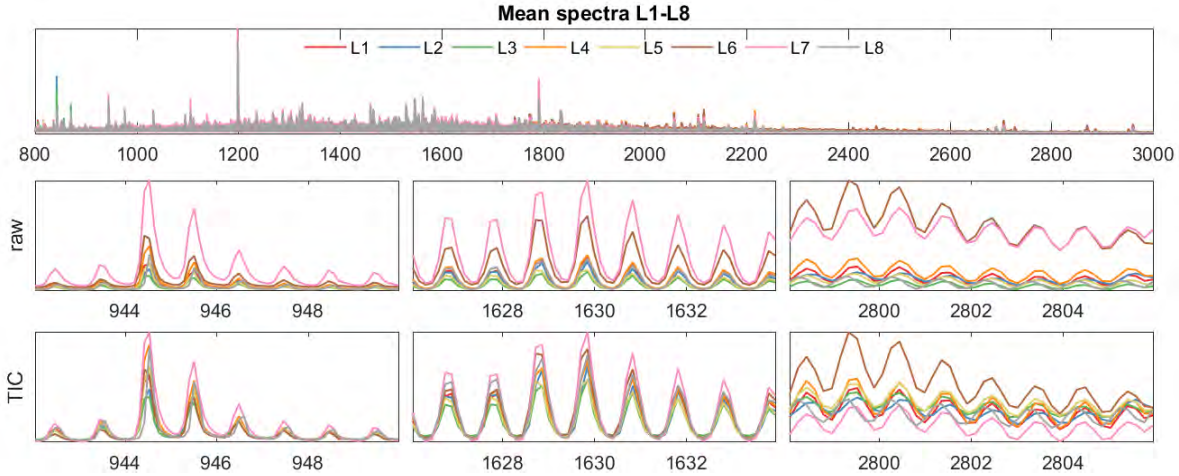
reducing technical variation!



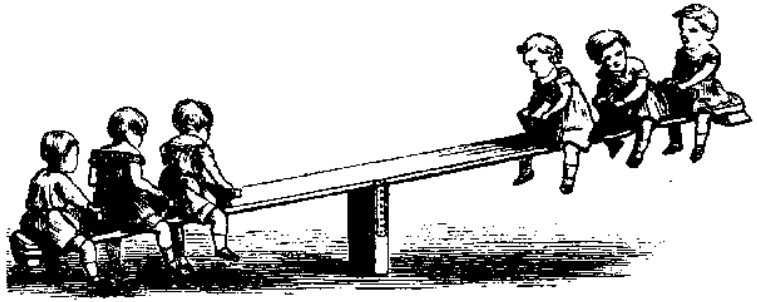
Example: Intensity Variations



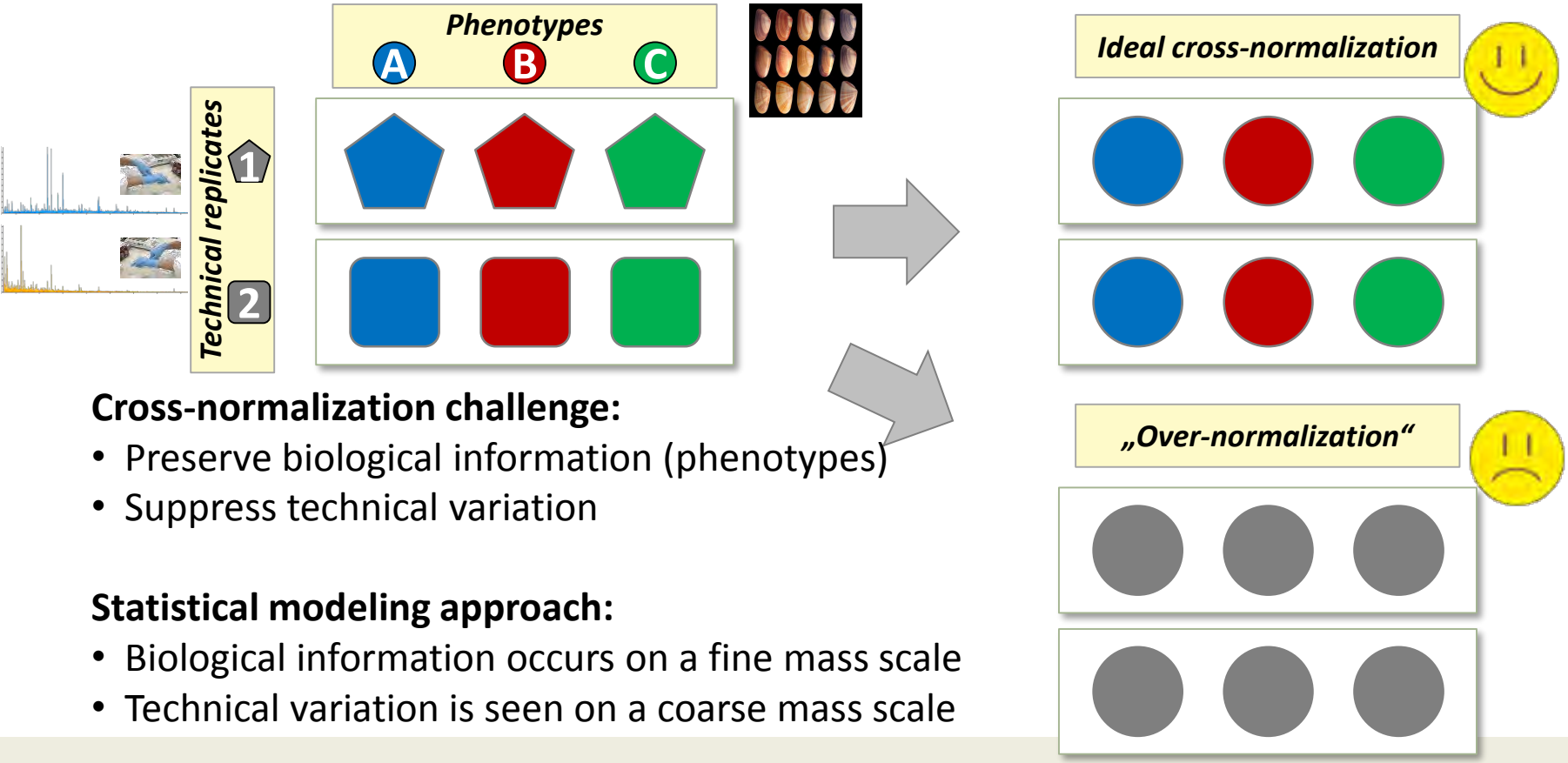
MSI Data: *Kriegsmann et.al. Mol Cell Proteomics. 2016 Oct;15(10):3081-3089. Epub 2016 Jul 29.*



- MALDI MSI of 8 FFPE lung cancer biopsy TMAs (L1 – L8), identical protocol (incl. trypsin digestion), constant acquisition conditions
- Strong intensity shifts between L1 – L8 mean spectra, varying across m/z range
- Standard normalization (TIC, median, RMS, …) cannot avoid ***"seesaw effect"***



Cross-Normalization



Cross-normalization challenge:

- Preserve biological information (phenotypes)
- Suppress technical variation

Statistical modeling approach:

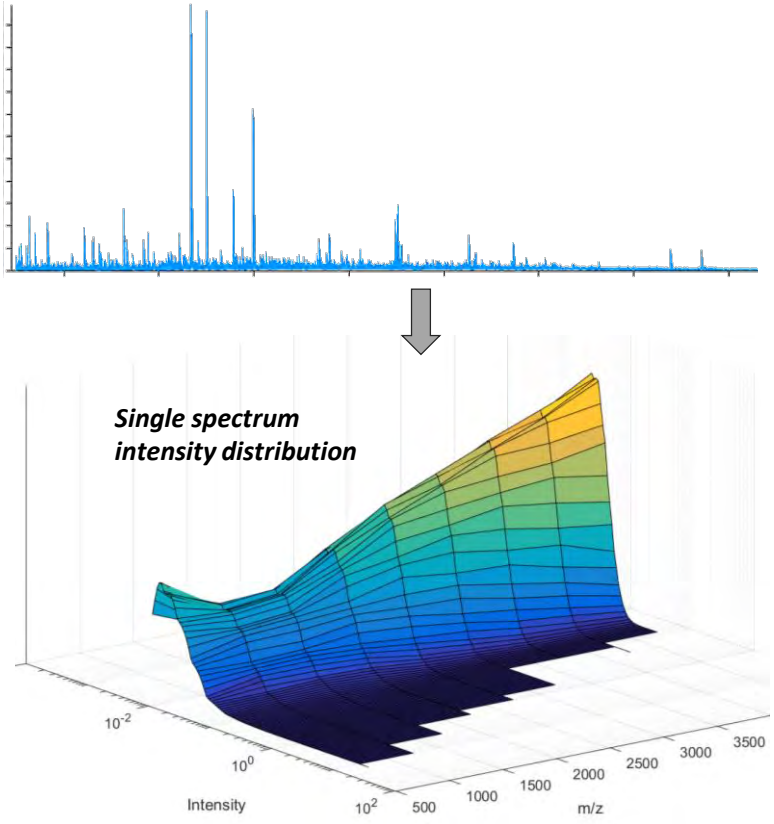
- Biological information occurs on a fine mass scale
- Technical variation is seen on a coarse mass scale

Intensity Cross-Normalization

- Compute ***intensity profiles*** per spectrum
  - *m/z* range split in *K* intervals
  - F_k: cumulative intensity distribution per interval

$$Q = \left(\frac{d}{dp} \log F_k^{-1}(p) \right)_{k=1\ldots K}$$

- Average profile used as ***reference***
- ***Transform*** individual spectra to match reference profile

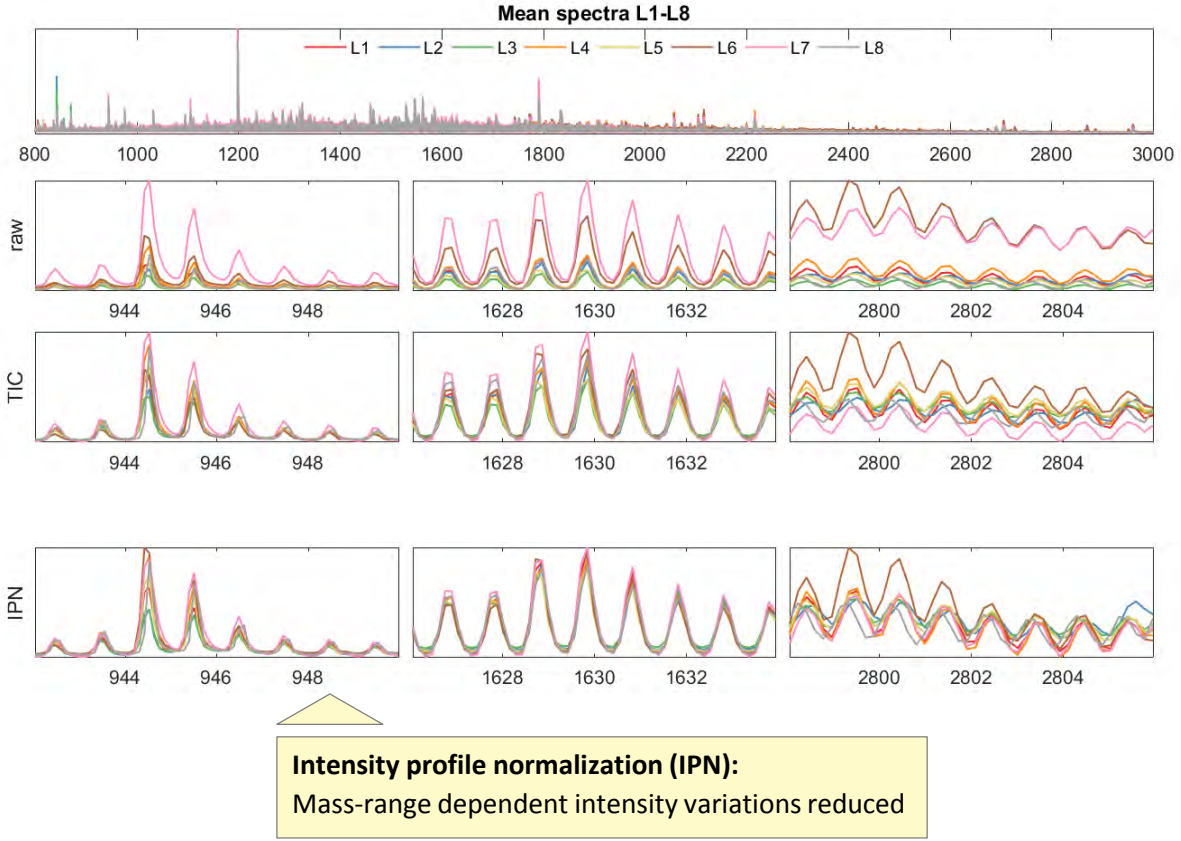


Intensity Cross-Normalization

- Compute ***intensity profiles*** per spectrum
  - *m/z* range split in *K* intervals
  - F_k: cumulative intensity distribution per interval

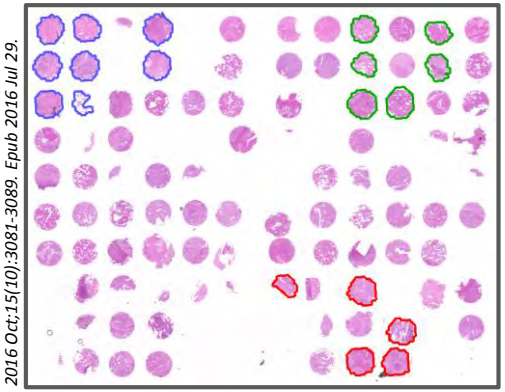
$$Q = \left(\frac{d}{dp} \log F_k^{-1}(p) \right)_{k=1\ldots K}$$

- Average profile used as ***reference***
- ***Transform*** individual spectra to match reference profile

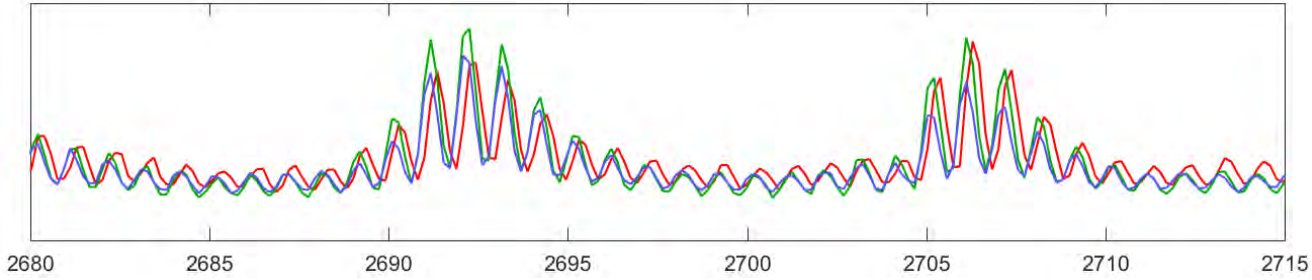


Intensity profile normalization (IPN):
Mass-range dependent intensity variations reduced

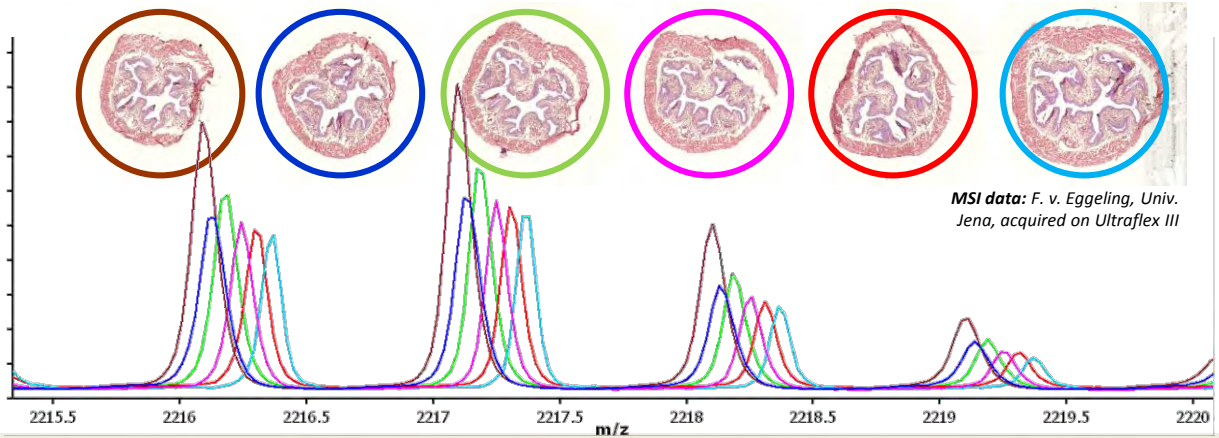
Example: Mass Shift Variation



MSI data: *Kriegsmann et.al. Mol Cell Proteomics. 2016 Oct;15(10):3081-3089. Epub 2016 Jul 29.*



Lung cancer TMA L8: Regional mass and intensity shifts within individual measurements



MSI data: *F. v. Eggeling, Univ. Jena, acquired on Ultraflex III*

Mouse bladder sections: Heavy mass shifts due to non-optimal acquisition conditions

Mass Shift Cross-Normalization

- Compute ***mass shift profiles*** per spectrum
  - Fourier integrals of spectrum *S* over *m/z* intervals I_k

$$R = \left(\int_{I_k} S(t) e^{i\omega t} dt \right)_{k=1\ldots K}, \omega = \frac{2\pi}{1+\delta}$$

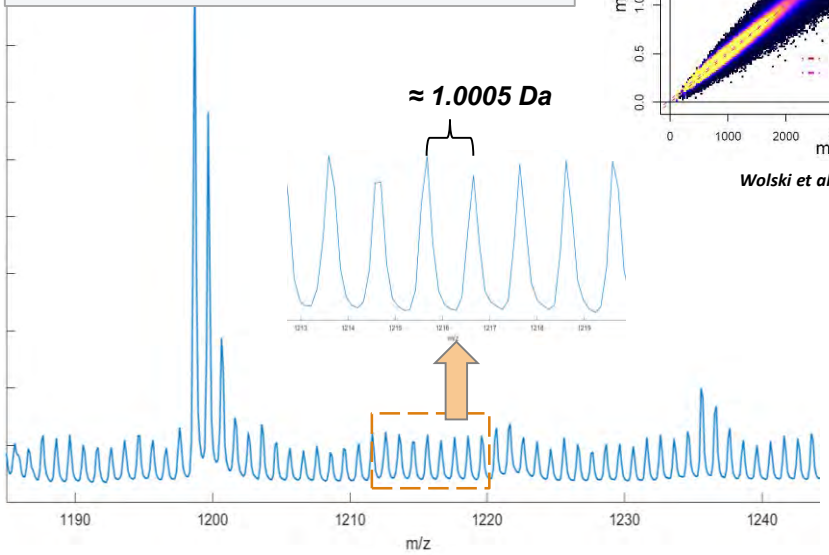
  - Equivalent to circular moments of peptide scale defect Δ
- Average or null profile used as ***reference***
- ***Transform*** individual spectra to match reference profile

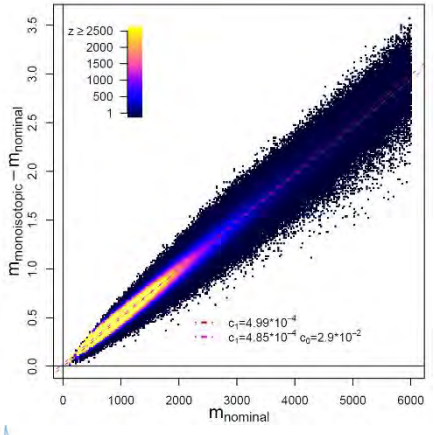
Peptide mass rule:

$$\bar{m} = (1+\delta) m_N, \delta \approx 4.95 \times 10^{-4}$$

Peptide scale mass defect:

$$\Delta = \mathrm{frac}\left(\frac{m}{1+\delta} + 0.5\right) - 0.5$$





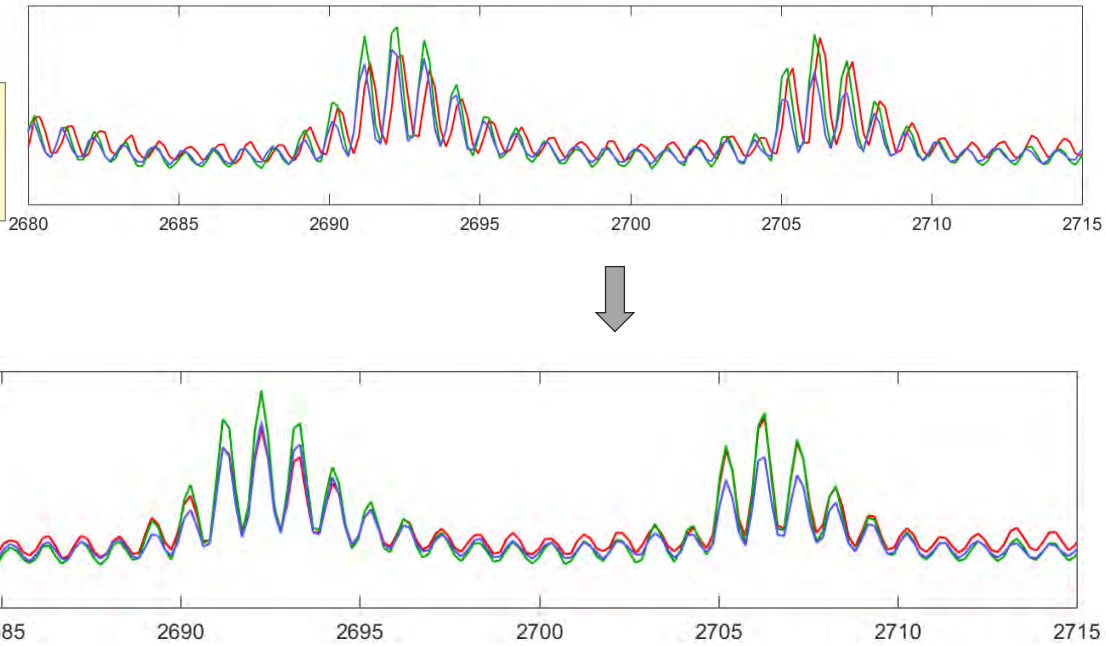
Wolski et al., Proteome Science, 2006

Mass Shift Cross-Normalization

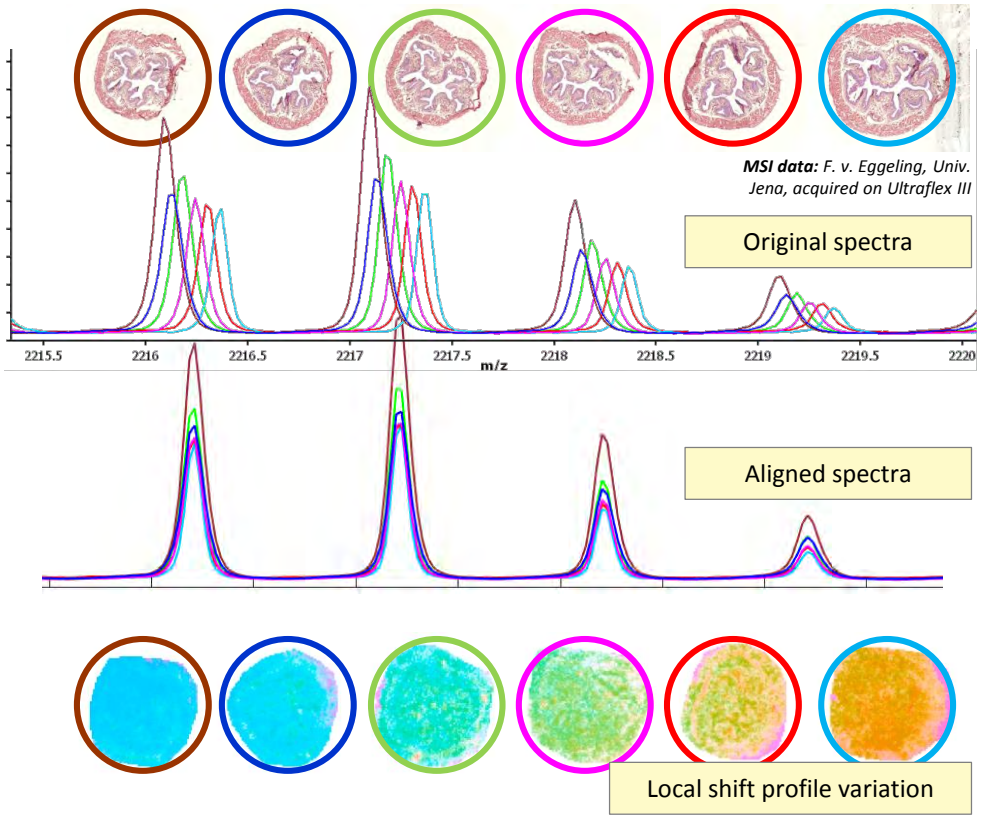
- Compute ***mass shift profiles*** per spectrum
  - Fourier integrals of spectrum *S* over *m/z* intervals I_k

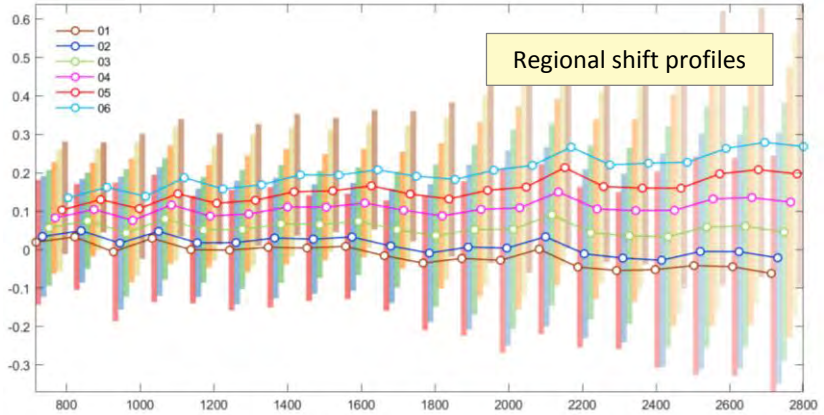
$$R = \left(\int_{I_k} S(t) e^{i\omega t} dt \right)_{k=1\ldots K}, \omega = \frac{2\pi}{1+\delta}$$

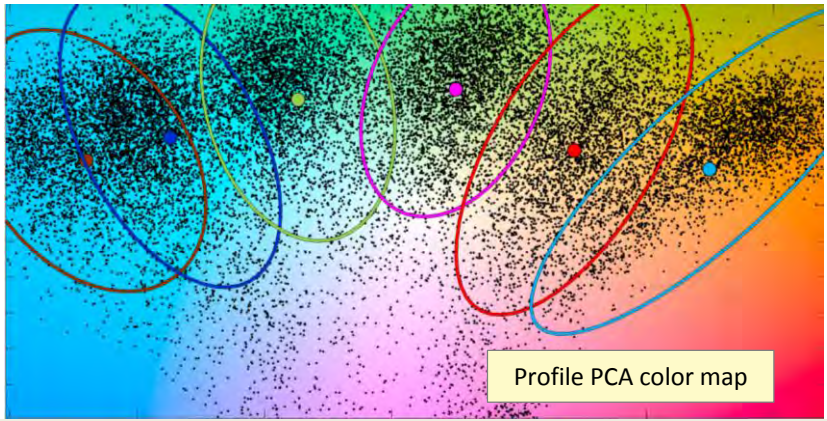
  - Equivalent to circular moments of peptide scale defect Δ
- Average or null profile used as ***reference***
- ***Transform*** individual spectra to match reference profile



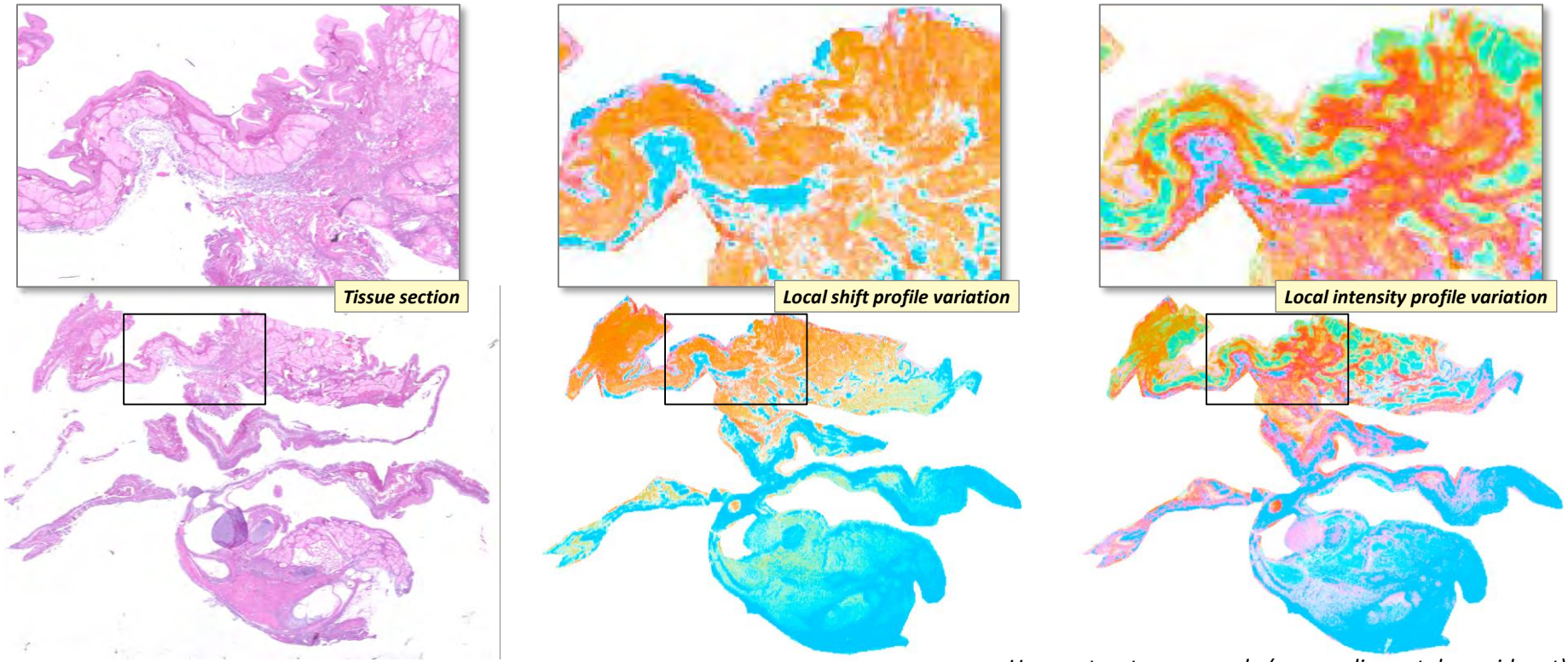
Local Mass Shift Variation





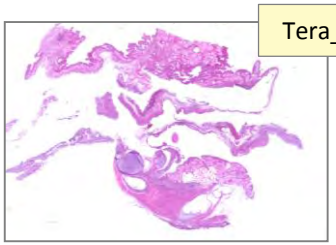


Local Profile Variation – Correlation to Anatomy



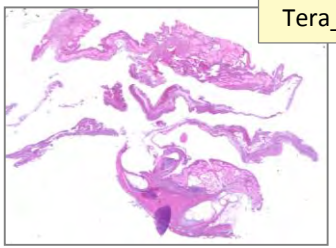
Human teratoma sample (non-malignant dermoid cyst).
Data acquired on rapiflex, 50 µm lat. res., 600–3200 m/z.

Normalization Across Instruments



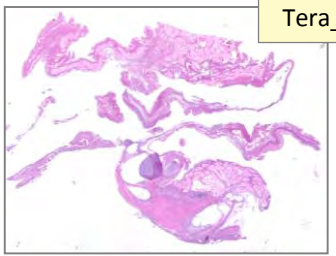
Tera_12

ImagePrep + Autoflex



Tera_13

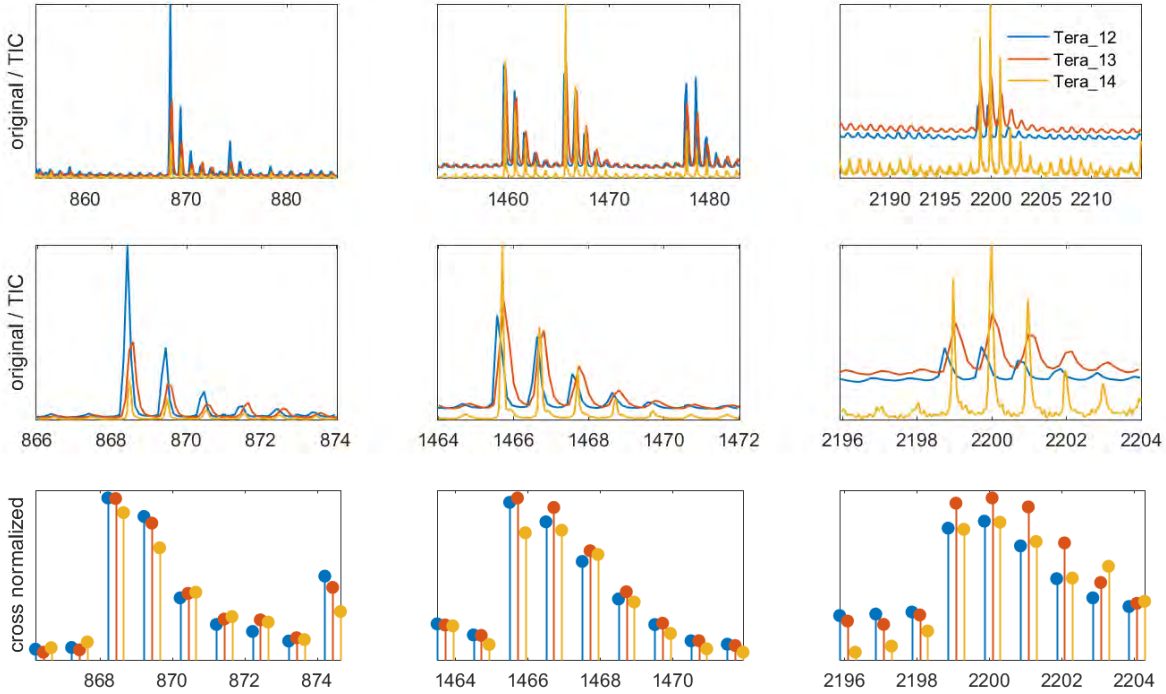
TM Sprayer + Autoflex



Tera_14

TM Sprayer + Rapiflex

Three serial sections of human teratoma prepared and measured using different instrumentation and protocols

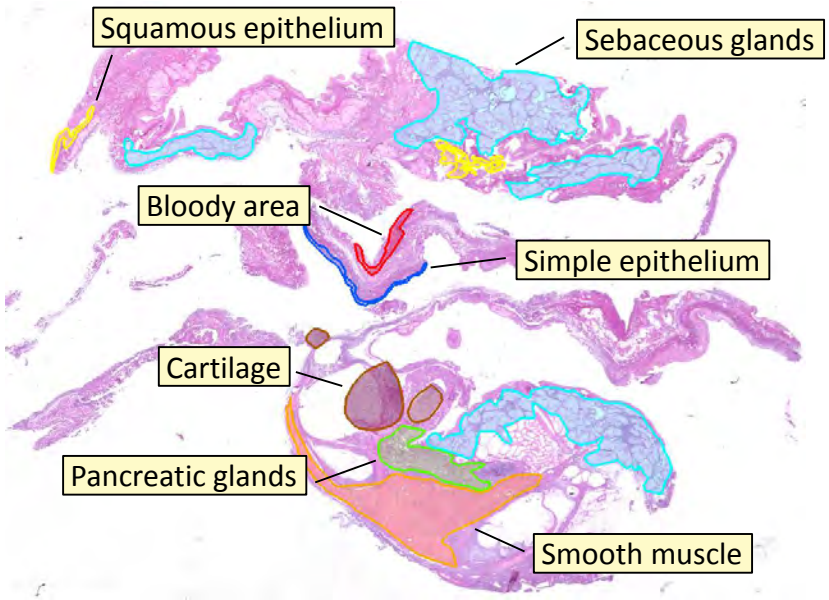


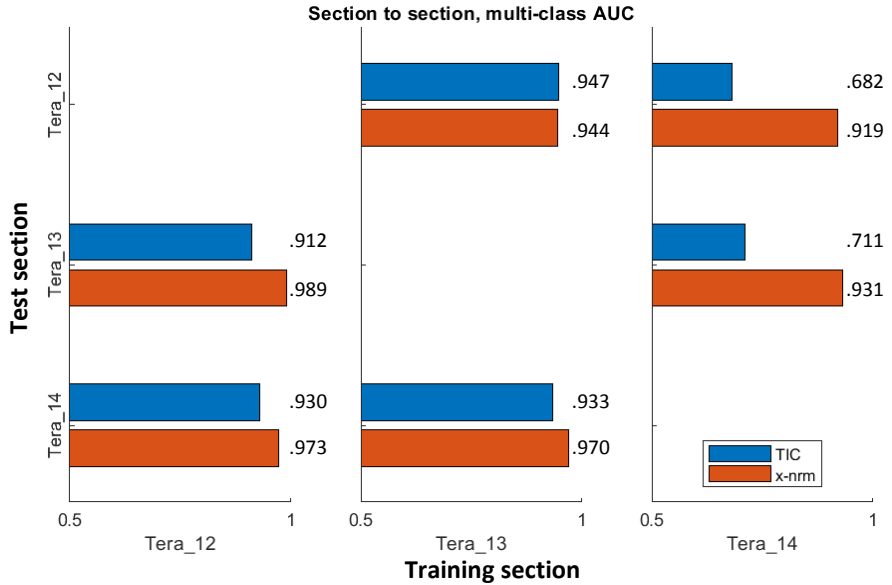
Cross-normalization allows combined analysis of all three datasets

Section-to-section Tissue Typing

- Pathologist annotations for seven tissue types
- Corresponding regions annotated on all three sections
- LDA classification models on 20 ROC selected peaks, trained seperately for each section

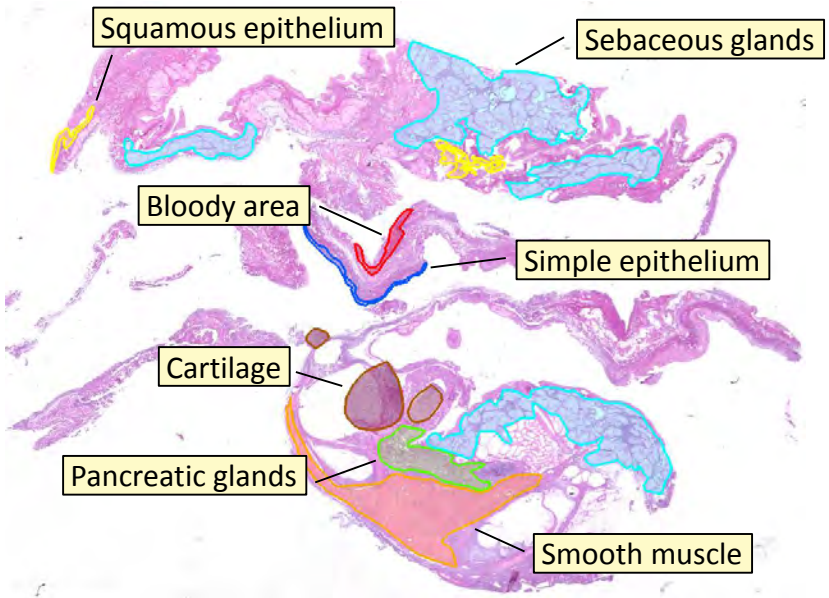
- AUC performance with cross-normalization (x-nrm) better than with standard TIC normalization
- Largest gain in rapiflex-to-autoflex scenario





Section-to-Section Tissue Typing

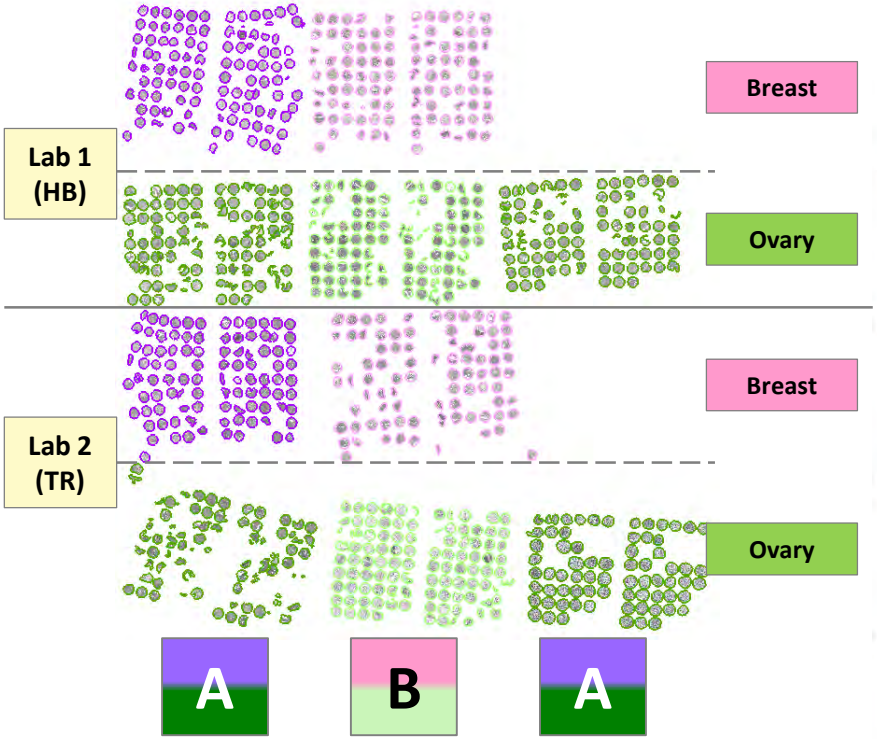
- 21 discriminative m/z values extracted
- 9 m/z values possibly attributable to tissue specific proteins



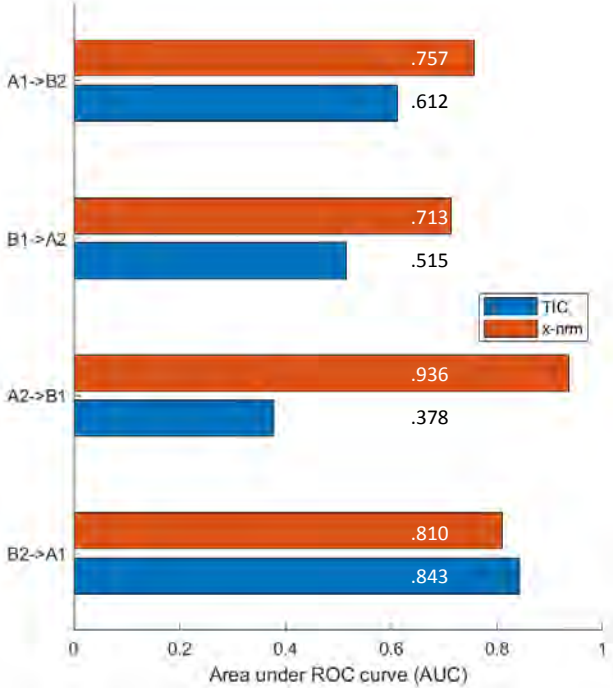
m/z	Tissue	Protein candidate	
1275.63	Bloody area	HBM	Hemoglobin
1315.65	Bloody area	HBG1/2	Hemoglobin
964.48	Bloody area		
912.45	Cartilage		
913.45	Cartilage		
883.44	Cartilage		
788.39	Pancreatic glands	MUC2/12	Mucin
982.49	Pancreatic glands	AMY1/2B/P	Alpha-amylase
1731.86	Pancreatic glands		
921.46	Sebaceous glands		
1080.53	Simple epithelium		
882.44	Simple epithelium	MUC4/5B	Mucin
1130.56	Smooth muscle	MYH2	Myosin type II
1055.52	Smooth muscle		
1209.60	Smooth muscle	MYH2	Myosin type II
1056.52	Smooth muscle		
1659.82	Squamous epithelium		
1043.52	Squamous epithelium	K1C15	Cytokeratin-15
1363.67	Squamous epithelium	K1C15	Cytokeratin-15
706.35	Squamous epithelium		
1535.76	Squamous epithelium		

Inter-Lab Normalization and Classification

- Five TMAs of breast and ovary cancer samples (*N* = 238 patients)
- Serial sections prepared and measured at two labs
- Identical protocols using ImagePrep, data acquired on Autoflex, 100 µm lat. res., 700–4000 m/z
- TMAs / patients divided into subsets A, B
- LDA classification models on 80 ROC selected peaks
- Cross-validation across labs and TMA sets A, B

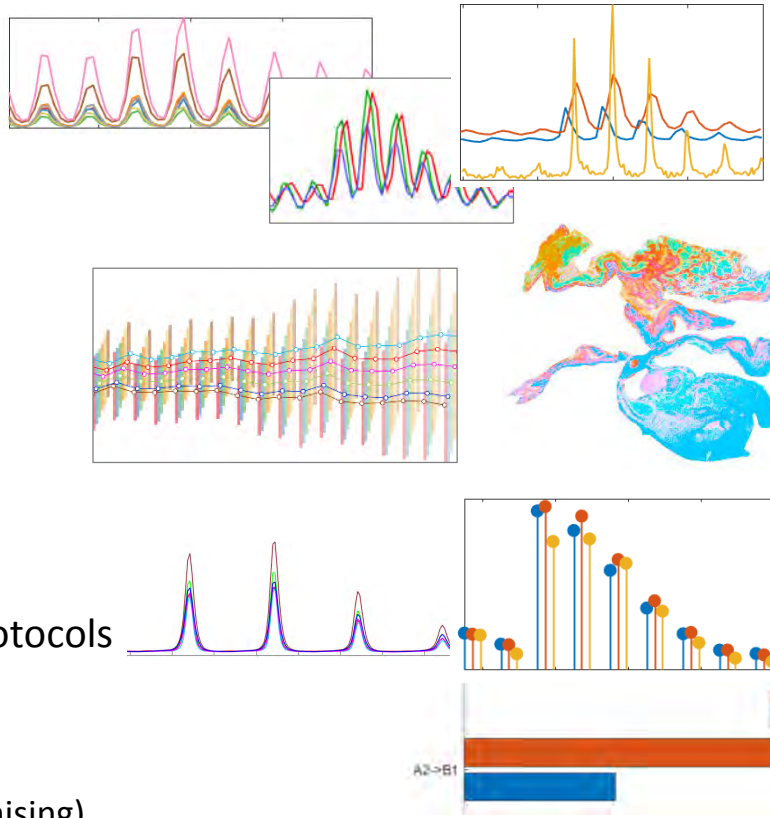


Cross-normalization (x-nrm) improves classification accuracy



Conclusion

- ***Technical variability in MALDI MSI …***
  - … obscures biological signals
  - … impedes routine applications
- ***Local mass shift / intensity profiles …***
  - … characterize technical variation
  - … show correlations to tissue anatomy
- ***Cross-normalization …***
  - … reduces mass shifts and intensity variations
  - … improves data comparability
  - … facilitates analysis across labs, instruments, protocols
- ***Future work***
  - More extensive evaluation
  - Transfer to metabolites and lipids (first results promising)



Thank You!

Center for Industrial Mathematics

- Jens Behrmann
- Yovany Cordero
- Christian Etmann
- Pascal Fernsel
- Lena Hauberg-Lotte
- Delf Lachmund
- Jonathan v. Schröder
- Max Westphal
- Peter Maaß

MALDI Imaging Lab

- Janina Oetjen
- Annette Peter

SCiLS

- Jan-Hendrik Kobarg
- Orlando Galashan
- Dennis Trede

Proteopath, Trier

- Rita Casadonte
- Petra Wandernoth
- Jörg Kriegsmann

Bruker

- Sören Deininger

More tissue and data contributed by

- Mark Kriegsmann, UK Heidelberg
- Ferdinand v. Eggeling, UK Jena

SPONSORED BY THE

Federal Ministry of Education and Research

DFG Deutsche Forschungsgemeinschaft