R Fundamentals and Best Practices for Mass Spectrometry Data Analysis

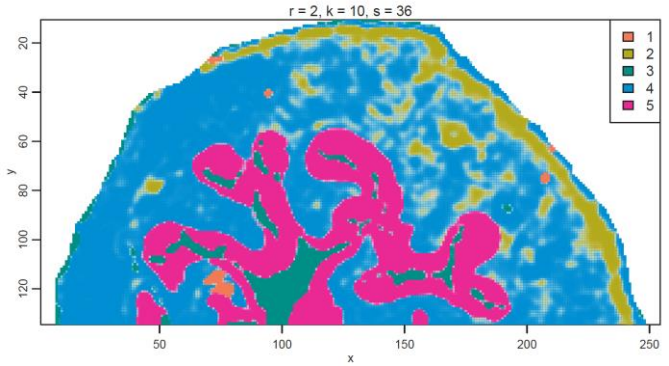
Sunday, November 15 (12:00-3:15pm Eastern)

Melanie Föll, University of Freiburg & Northeastern University

Module #8: Basic mass spectrometry imaging with Cardinal

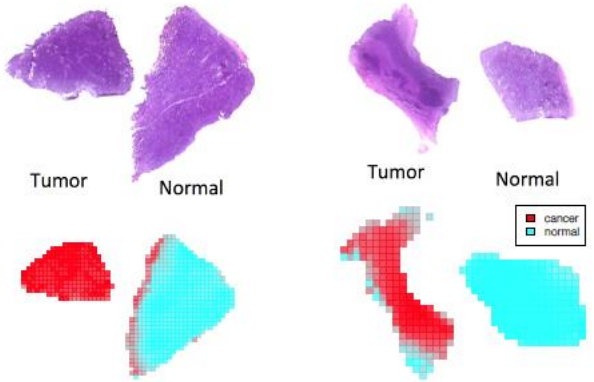
Mass spectrometry imaging (MSI)

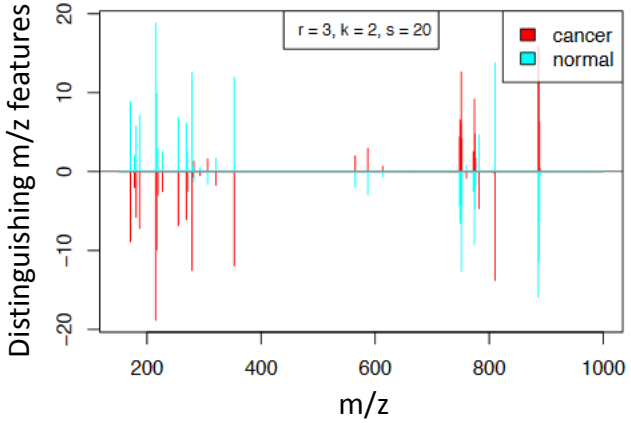
Tissue histology



Römpp 2010, Angewandte Chemie – international Edition

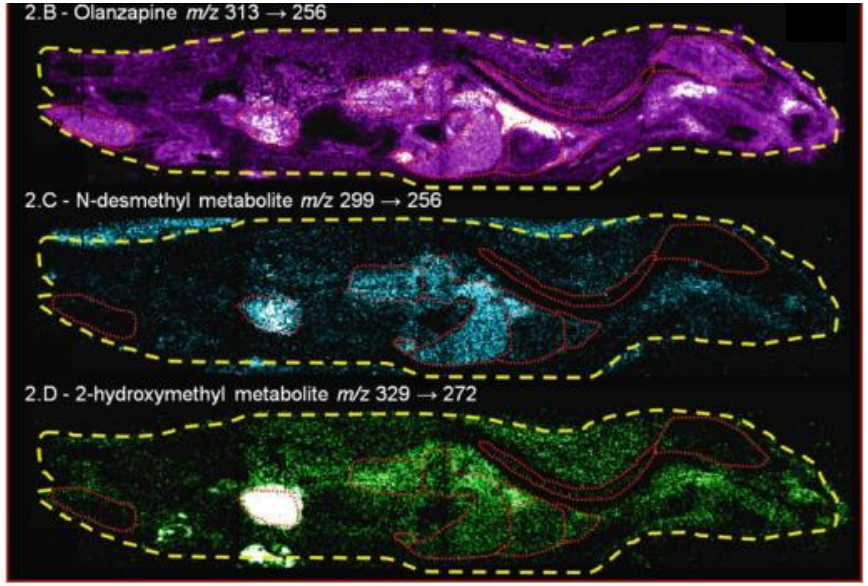
Biomarker discovery





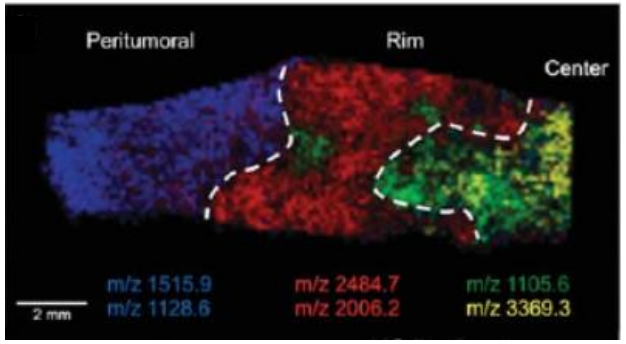
Bemis 2016, MCP

Drug imaging



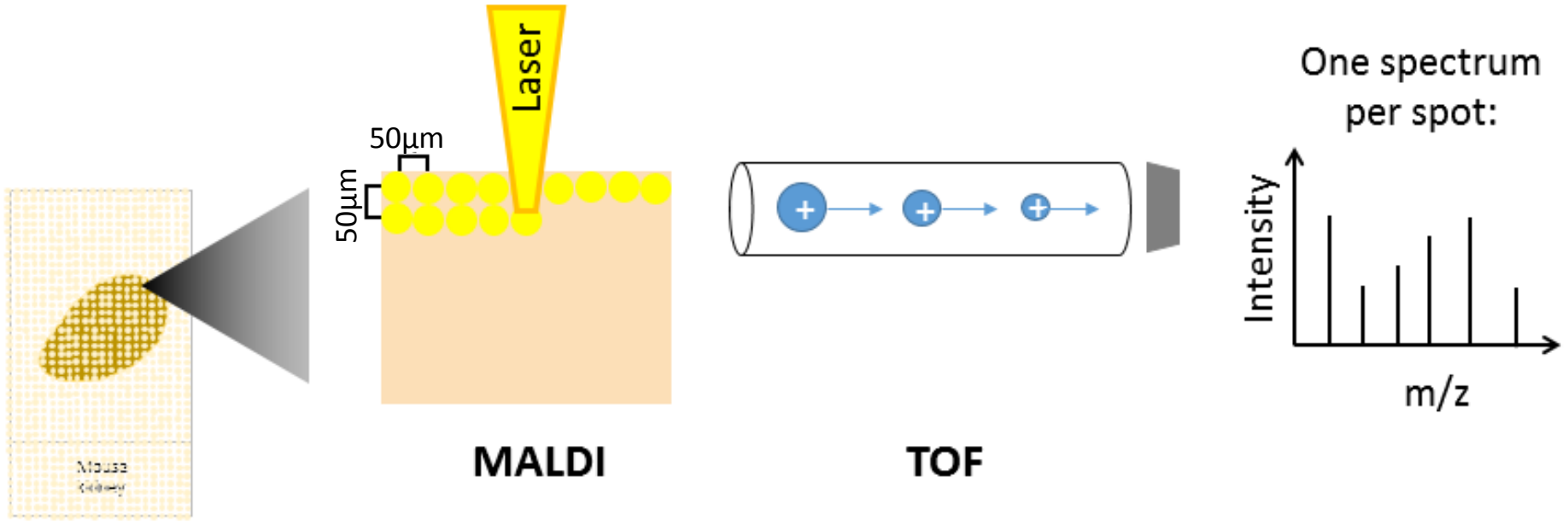
Khatib-Shahidi 2006, Anal. Chem.

Tissue heterogeneity

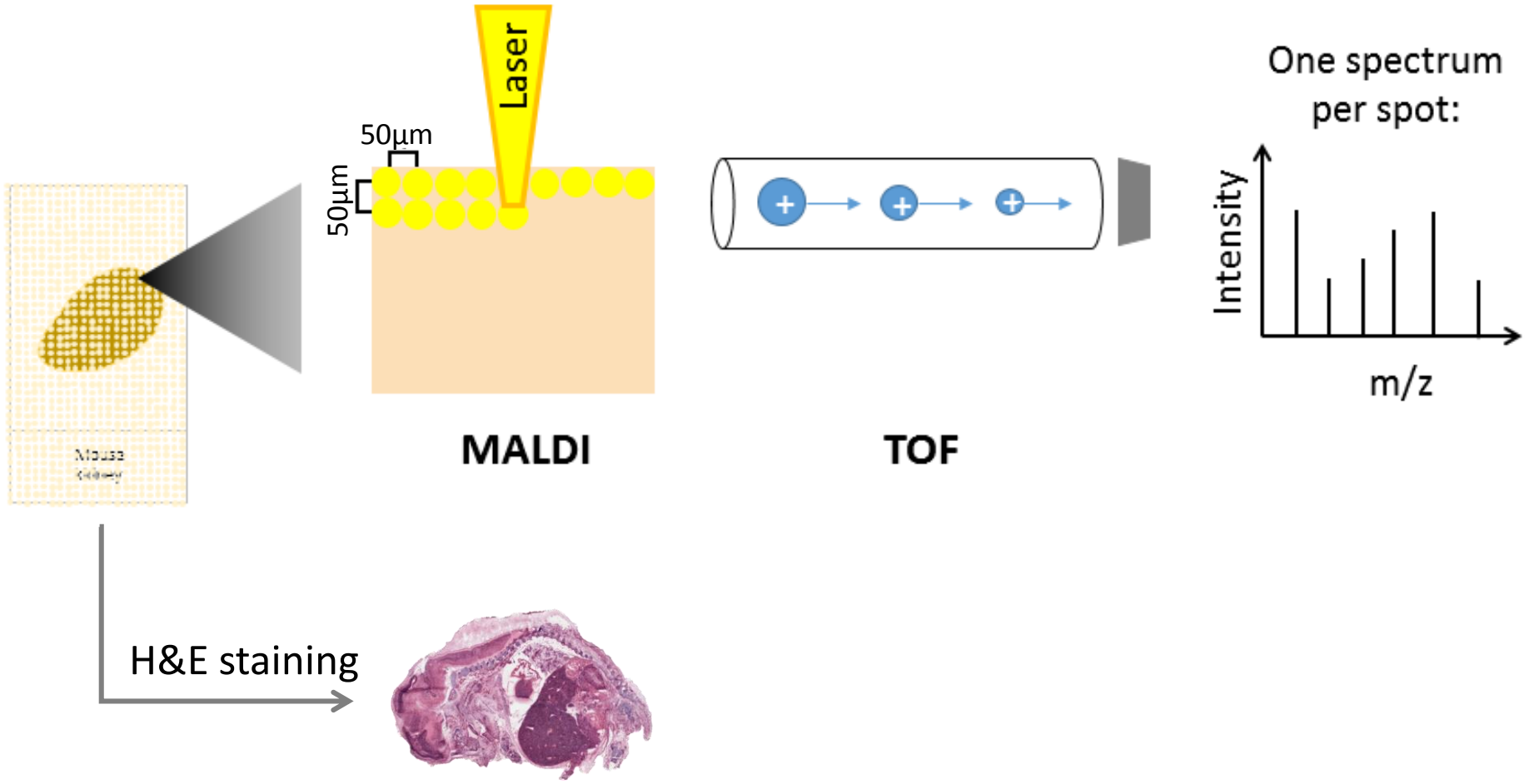


Turtoi 2014, Hepatology

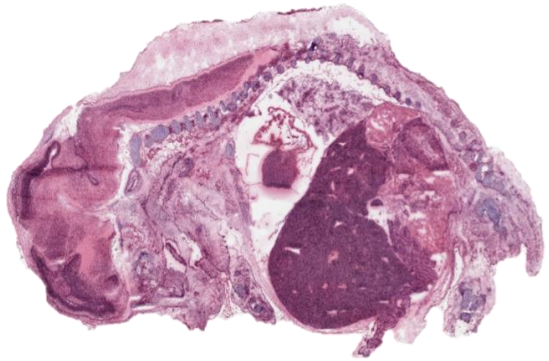
Data acquisition

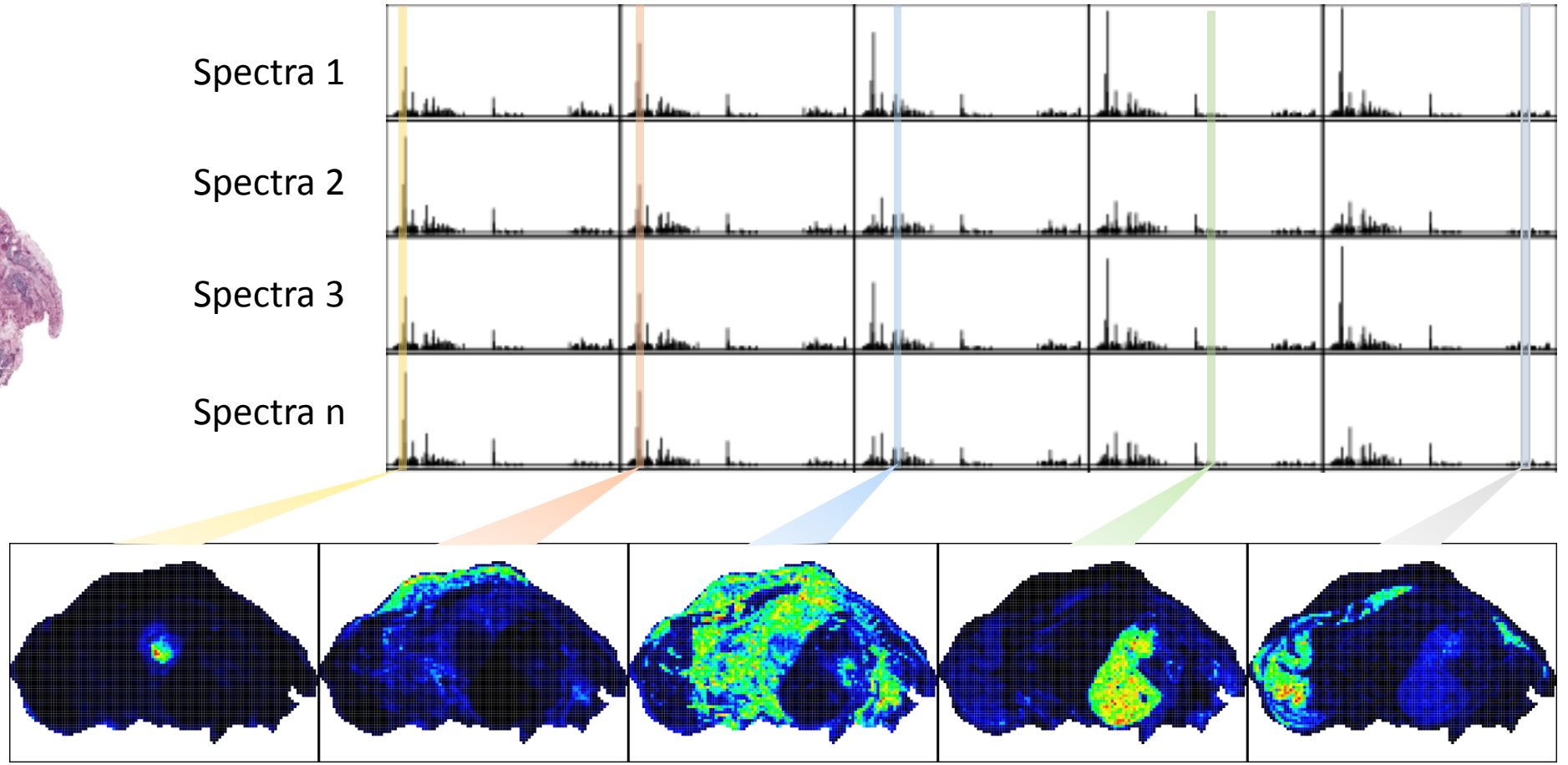


Data acquisition



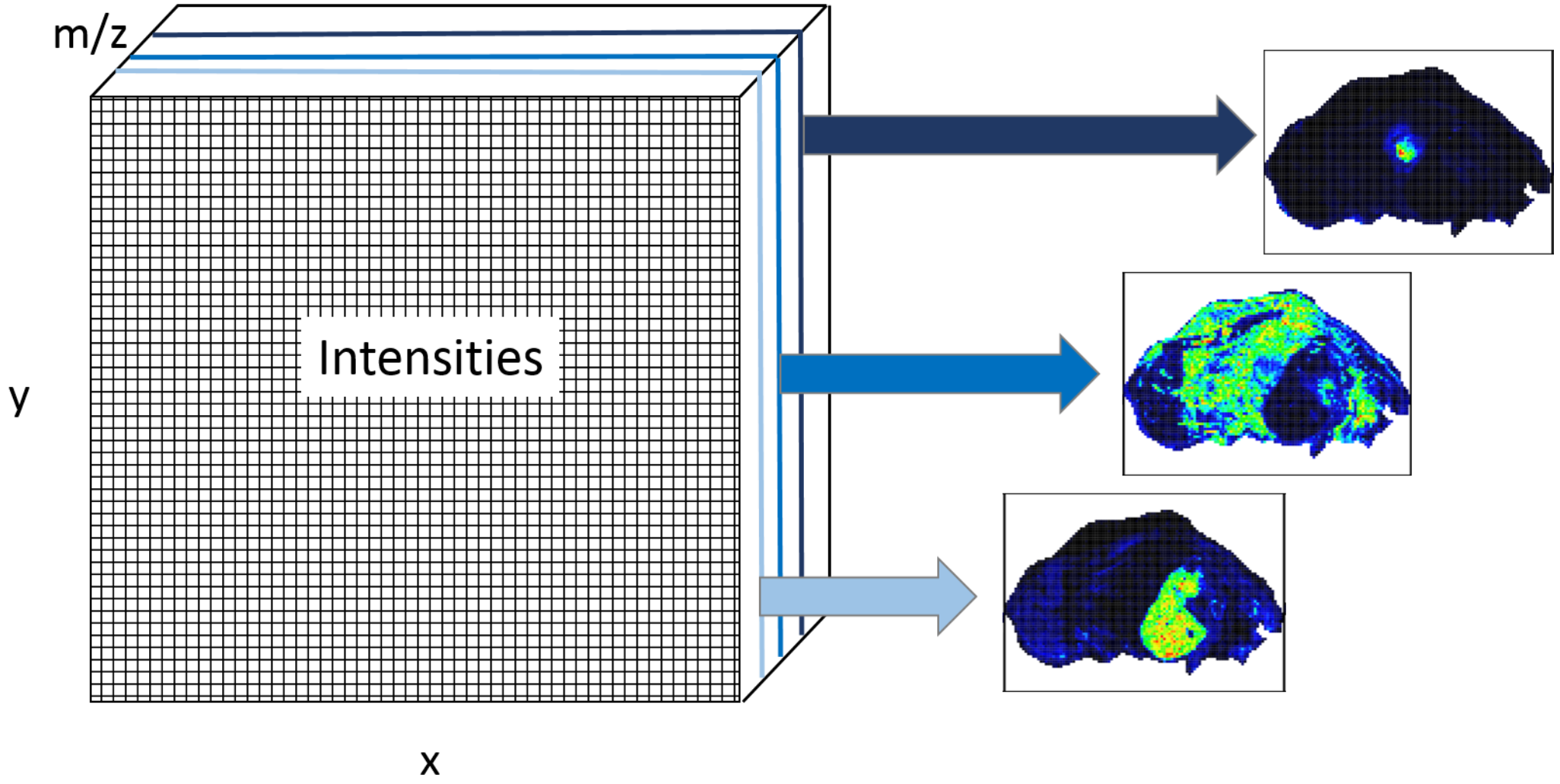
From spectra to distribution profiles





Data provided by R. Graham Cooks and lab

From spectra to distribution profiles



Cardinal: open-source statistical software for MSI

www.cardinalmsi.org

- Free, open-source
- R-based
- Rich methods portfolio
- Compatible with larger than memory datasets
- Extensive documentation
- Active mailing list

- Over 7,000 unique downloads since public release in 2015
- 2015 John M. Chambers Statistical Software Award

Bioconductor

platforms all | downloads top 20%
build ok | commits 1.00

Software	K. Bemis et al. Bioinformatics, 2015.
Statistical methods	K. Bemis et al. Molecular and Cellular Proteomics, 2016.
Large-than memory data	K. Bemis et al. Bioinformatics, 2017.

Cardinal: Methods portfolio

Import

- imzML
  - MSI data
- Analyze 7.5
  - MSI data
- Text files
- Image files
- RData
- Rscript
- Markdown files

Visualization

- Ion Images
  - Singe m/z
    - Contrast enhancement
    - Image smoothing
    - Scaling
    - color palette
  - Multiple m/z overlay
  - Summary statistics (TIC, median, …)
- Mass spectra
  - Single spectrum
  - Multiple spectra overlay
  - Summary statistics (Mean, sum, …)
- Statistical results

Preparation

- Annotate ROIs
  - Select ROI
  - Annotate spectra
- Subset & Filter
  - m/z and mass range
  - Spectra and pixels
  - Intensities
- Combine
  - Samples
  - Spectra
- Transformation
  - Log-transform
  - Replace values (NA, negative, …)

Pre-processing

- Normalization
  - Total ion current
  - Root-mean-square
  - Reference m/z
- Baseline removal
  - Median
  - Locmin
- Smoothing
  - Gaussian
  - Moving average
  - Savitzky-golay
- Spectral alignment
  - Mean
  - Reference
- Re-calibration
- Peak detection
  - Simple
  - Adaptive
  - MAD
- Peak alignment
  - Mean
  - Reference
- Peak integration
  - Peak height
  - Peak area
- Mass binning

Statistics & AI

- Colocalization
  - Reference m/z
  - ROI
- Classification
  - PLS-DA
  - OPLS-DA
  - Spatial Shrunken Centroids
    - Gaussian
    - Adaptive
- Segmentation
  - Single ions
    - Spatial-DGMM
  - Multivariate
    - PCA
    - Spatial k-means
      - Gaussian
      - Adaptive
    - Spatial Shrunken Centroids
      - Gaussian
      - Adaptive
- Class comparison
- Summary statistics

Export

- imzML
  - MSI data
- Analyze 7.5
  - MSI data
- Text files
  - Intensity matrix
  - m/z features
  - Mass spectra
  - Pixel coordinates
  - Experimental metadata
- Image files
  - Ion images
  - Mass spectra
  - Statistical results
- RData
  - Stastistical models
- Rscript
  - R analysis
- Markdown files
  - R analysis

Software

Parallel CPU computation

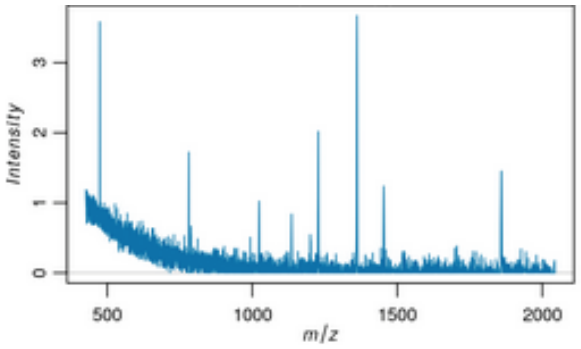
Out-of-memory computation

Visualizations

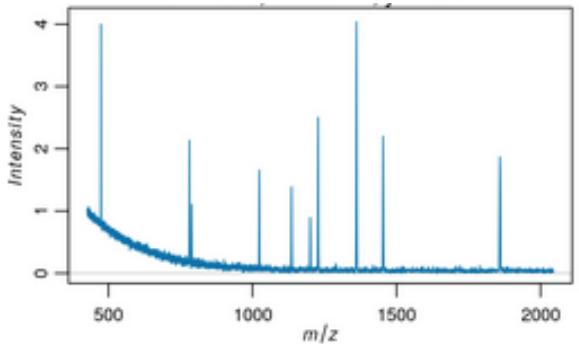
Mass spectra

plot(mse, …)

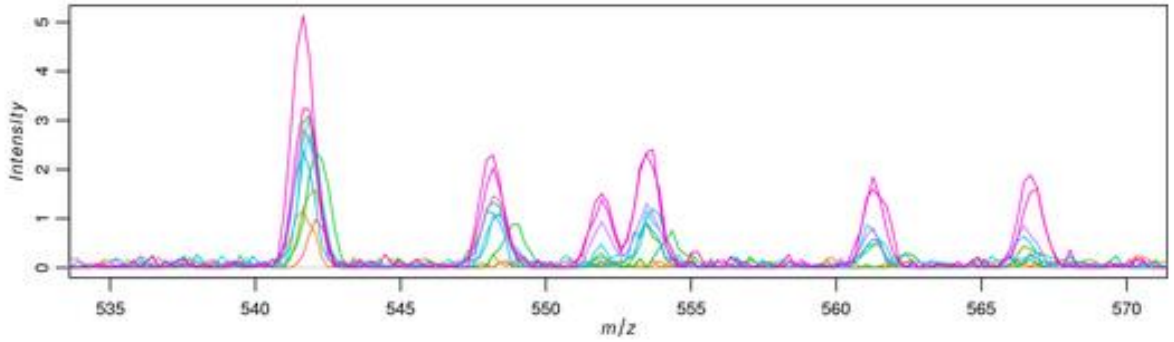
Single mass spectra



Average mass spectra



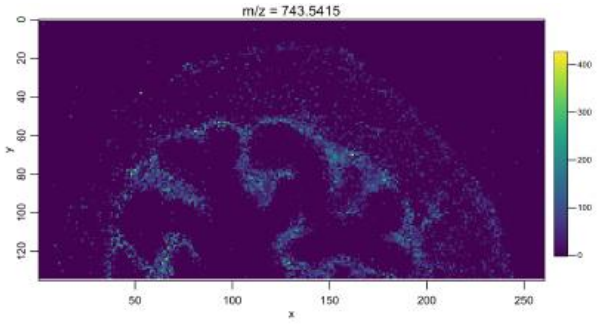
Zoomed in & overlaid mass spectra



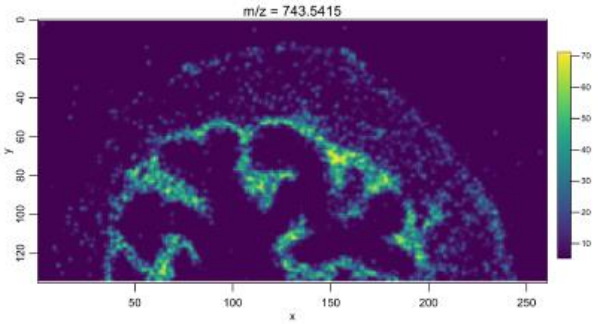
Ion images

image(mse, …)

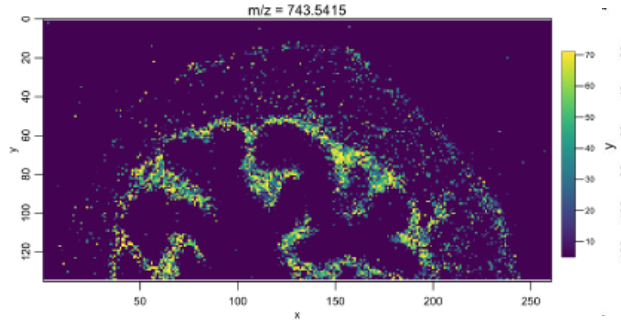
Ion image



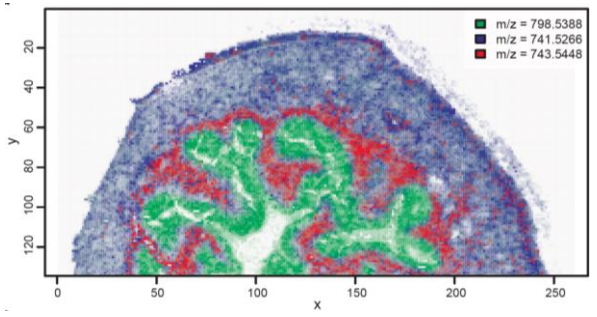
Ion image (smoothed)



Ion image (contrast enhanced)

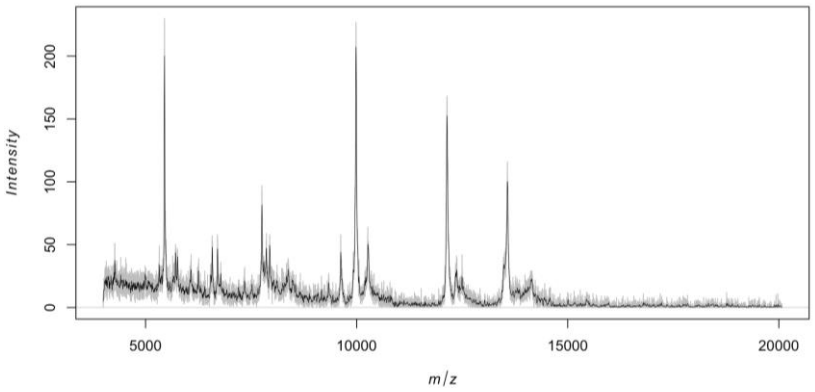


Ion image overlay of three m/z



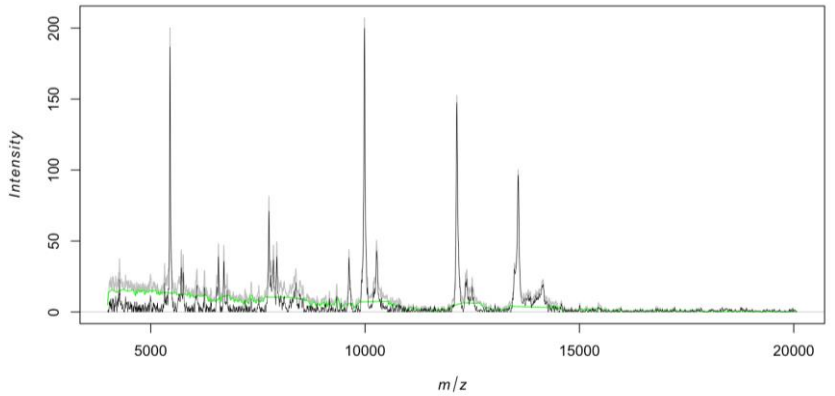
Pre-processing

Smoothing



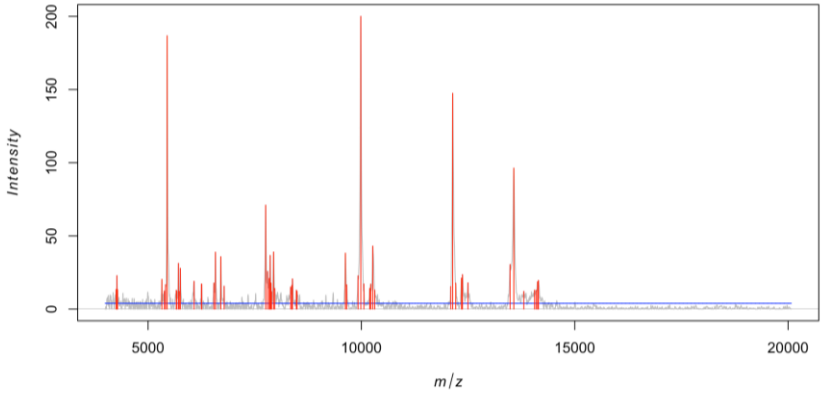
smoothSignal(mse, …)

Baseline removal



reduceBaseline(mse, …)

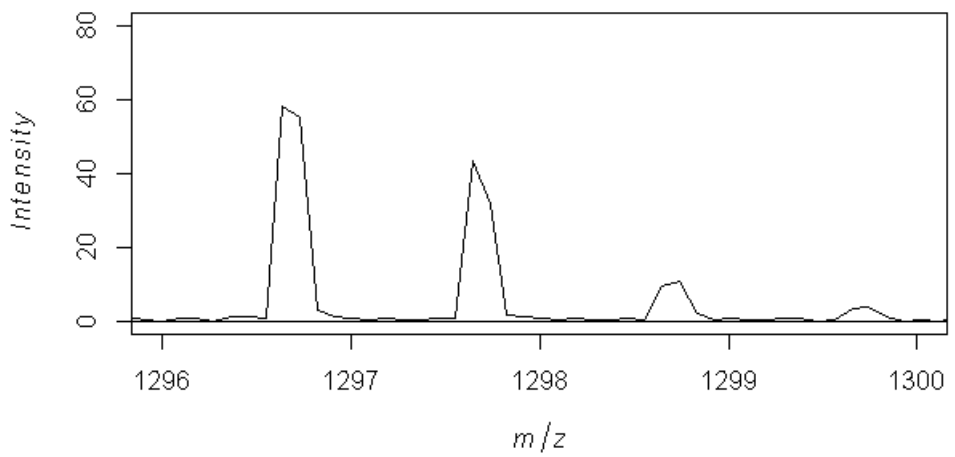
Peak detection



peakPick(mse, …)

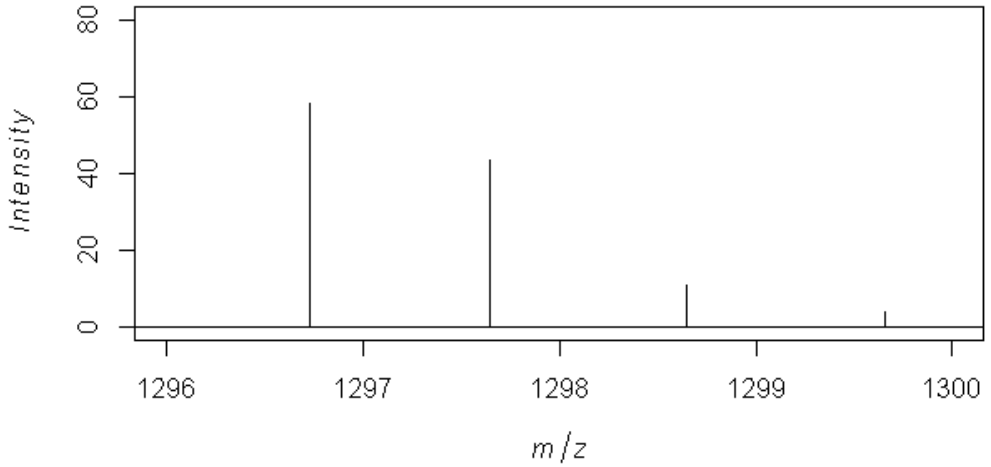
Pre-processing

Profile spectrum



centroided(mse) = FALSE

Centroided spectrum

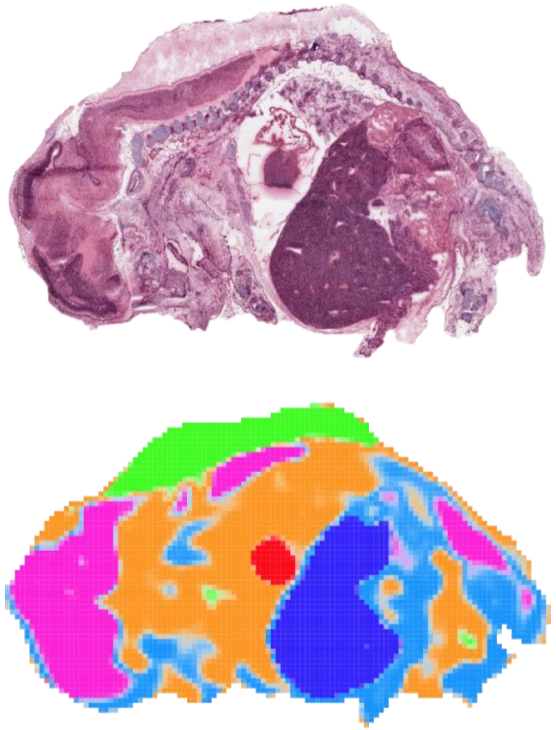


centroided(mse) = TRUE

Statistics & AI

1. Class discovery

(segmentation)



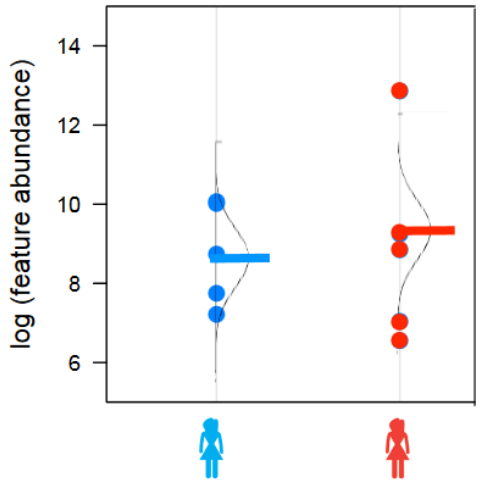
spatialShrunkenCentroids(mse, …)

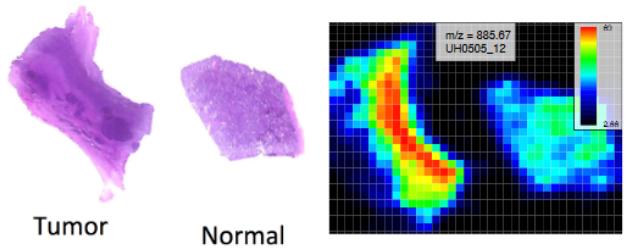
spatialDGMM(mse, …)

….

2. Class comparison

(linear models)





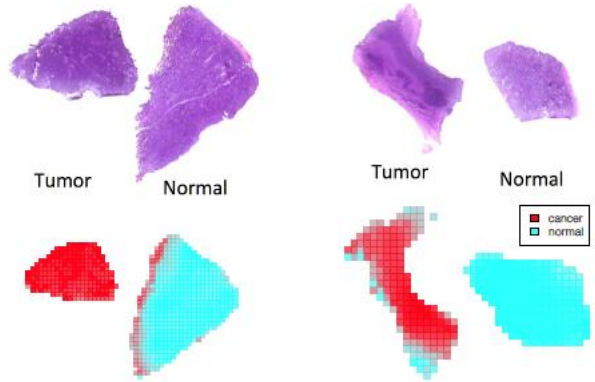
meansTest(mse, …)

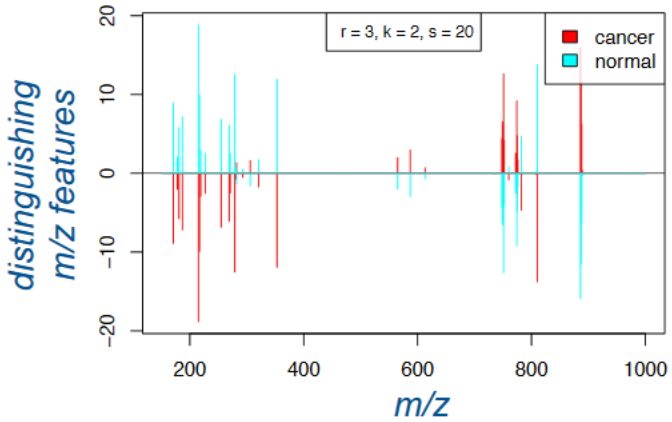
segmentationTest(mse, …)

….

3. Class prediction

(classification)





spatialShrunkenCentroids(mse, …)

….

MSI data import & export

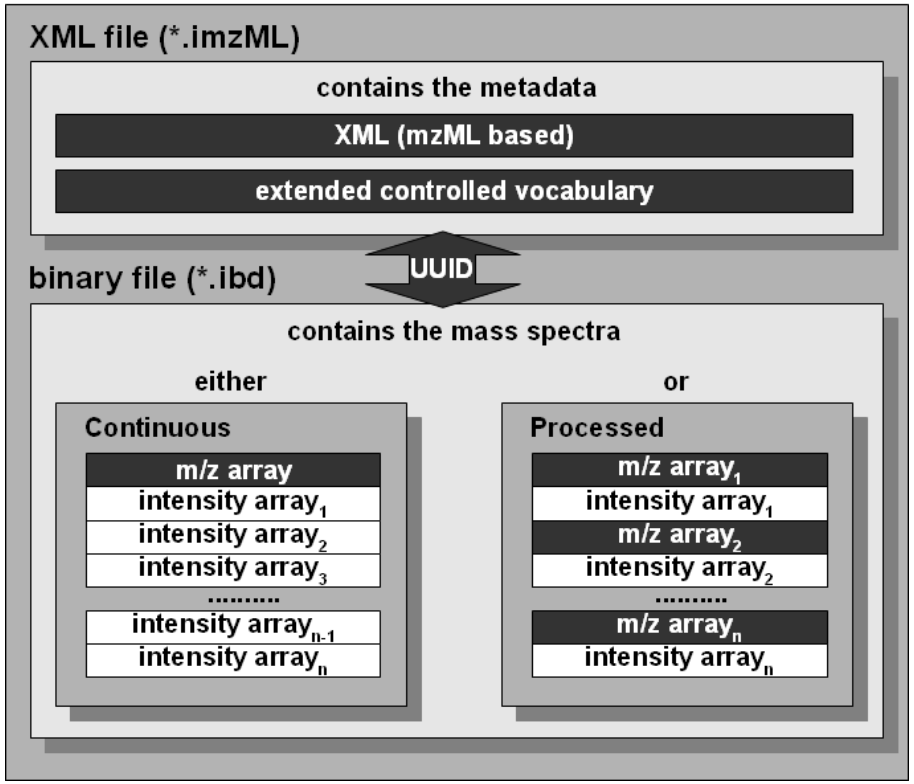
readImzML(mse, …)

writeImzML(mse, …)

imzML: open standard file format for MSI data

File1: .imzML

File2: .ibd



MSContinuous ImagingExperiment

MSProcessed ImagingExperiment

imzML converter:
https://www.cs.bham.ac.uk/~ibs/imzMLConverter

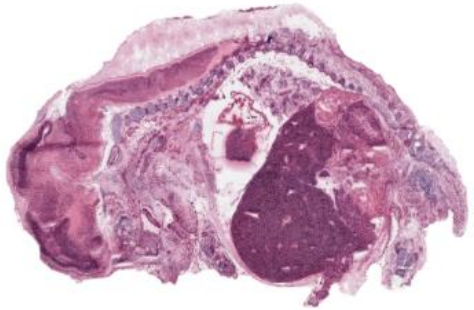
Race, Alan M., Iain B. Styles, and Josephine Bunch. "Inclusive sharing of mass spectrometry imaging data requires a converter for all." Journal of proteomics 75.16 (2012): 5111-5112.

imzML validator:
https://gitlab.com/imzML/imzMLValidator

Race, Alan M., and Andreas Römpp. "Error-Free Data Visualization and Processing through imzML and mzML Validation." *Analytical chemistry* 90.22 (2018): 13378-13384..

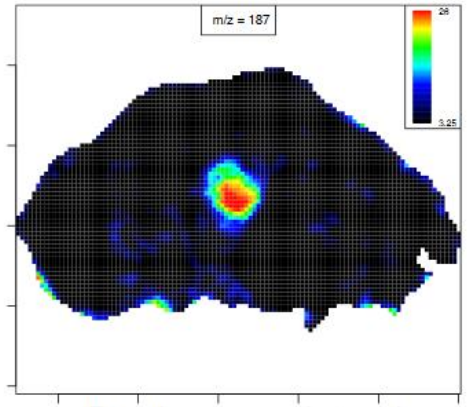
Schramm, Thorsten, et al. "imzML—a common data format for the flexible exchange and processing of mass spectrometry imaging data." *Journal of proteomics* 75.16 (2012): 5106-5110.

Today‘s dataset & analysis

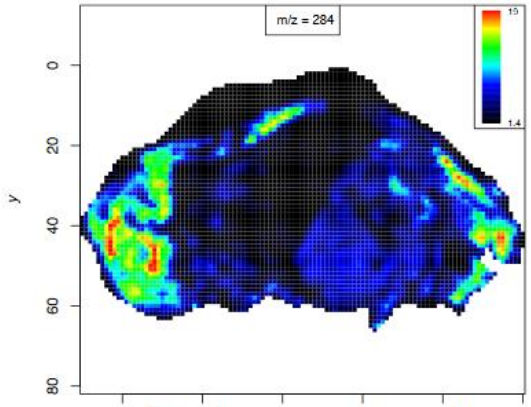


Optical image

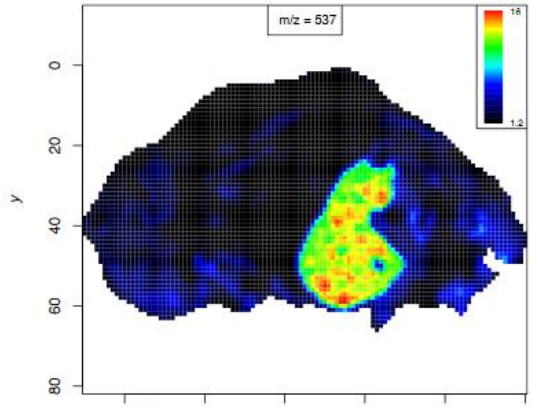
- Pig fetus section
- DESI LTQ mass spectrometer
- m/z range 150-1000



Peak associated with heart



Peak associated with brain



Peak associated with liver

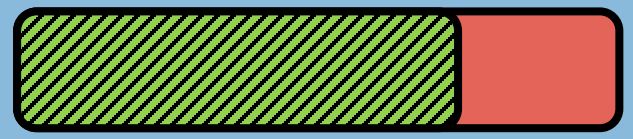
Data provided by R. Graham Cooks and lab

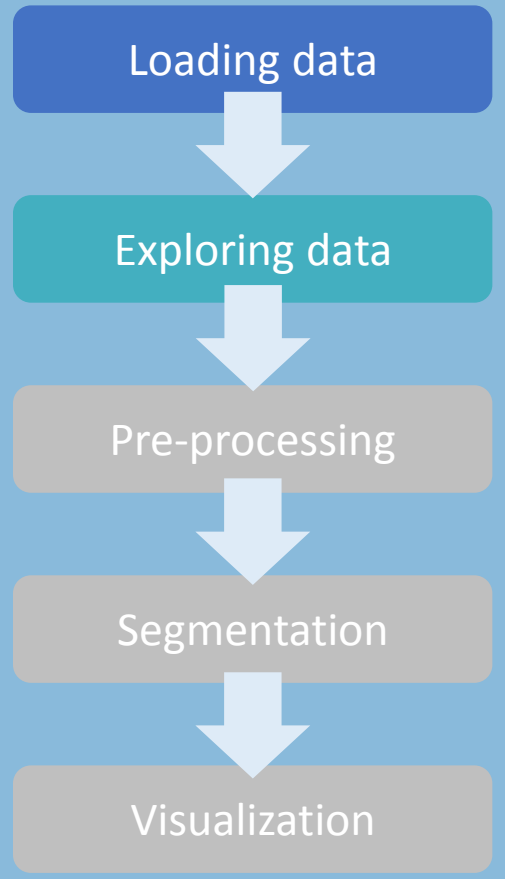
- Loading data
- Exploring data
- Pre-processing
- Segmentation
- Visualization

Vignette:
http://bioconductor.org/packages/release/data/experiment/vignettes/CardinalWorkflows/inst/doc/MSI-segmentation.html

Part I: Loading & exploring MSI data

Loading: 75%

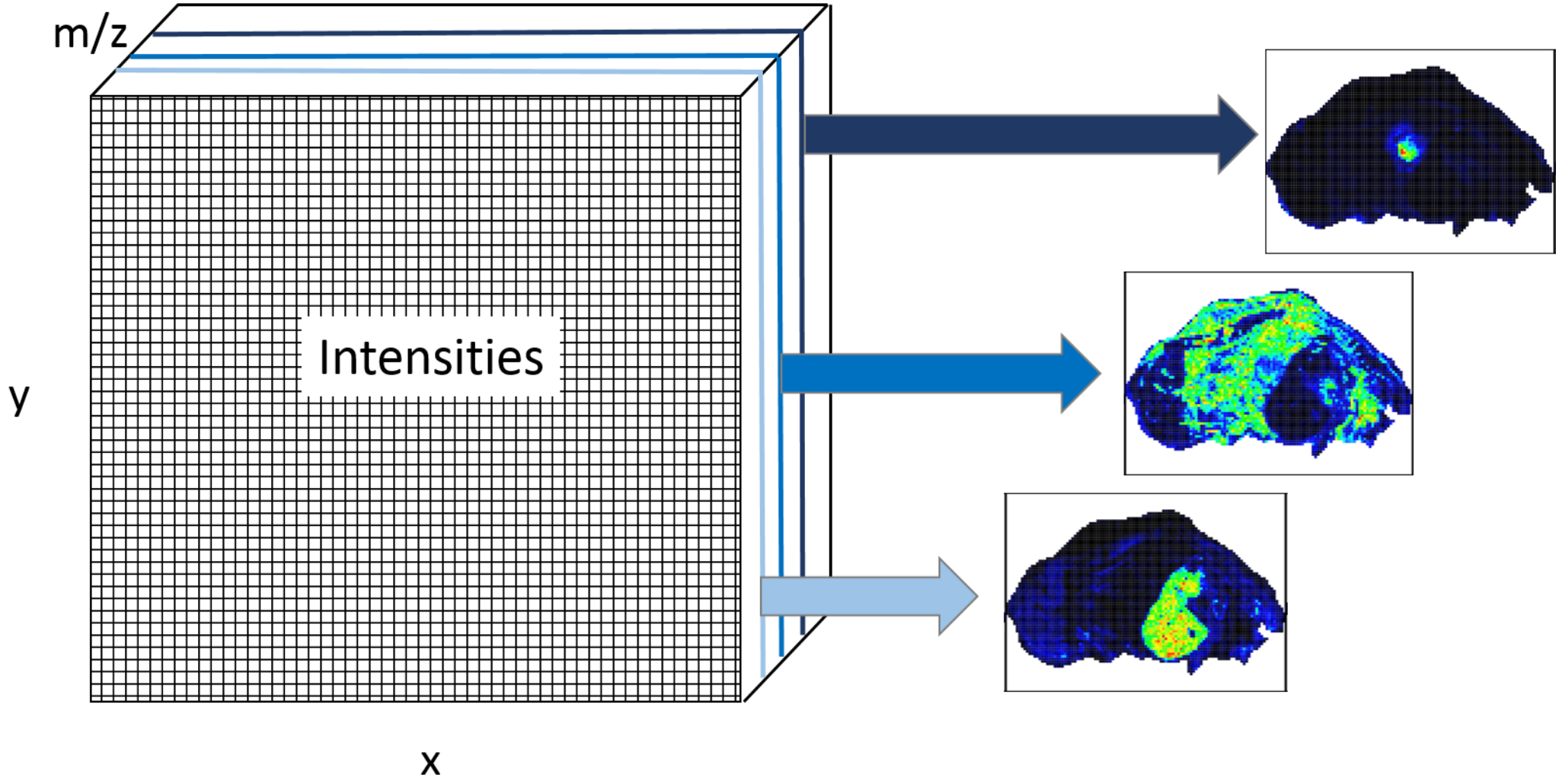




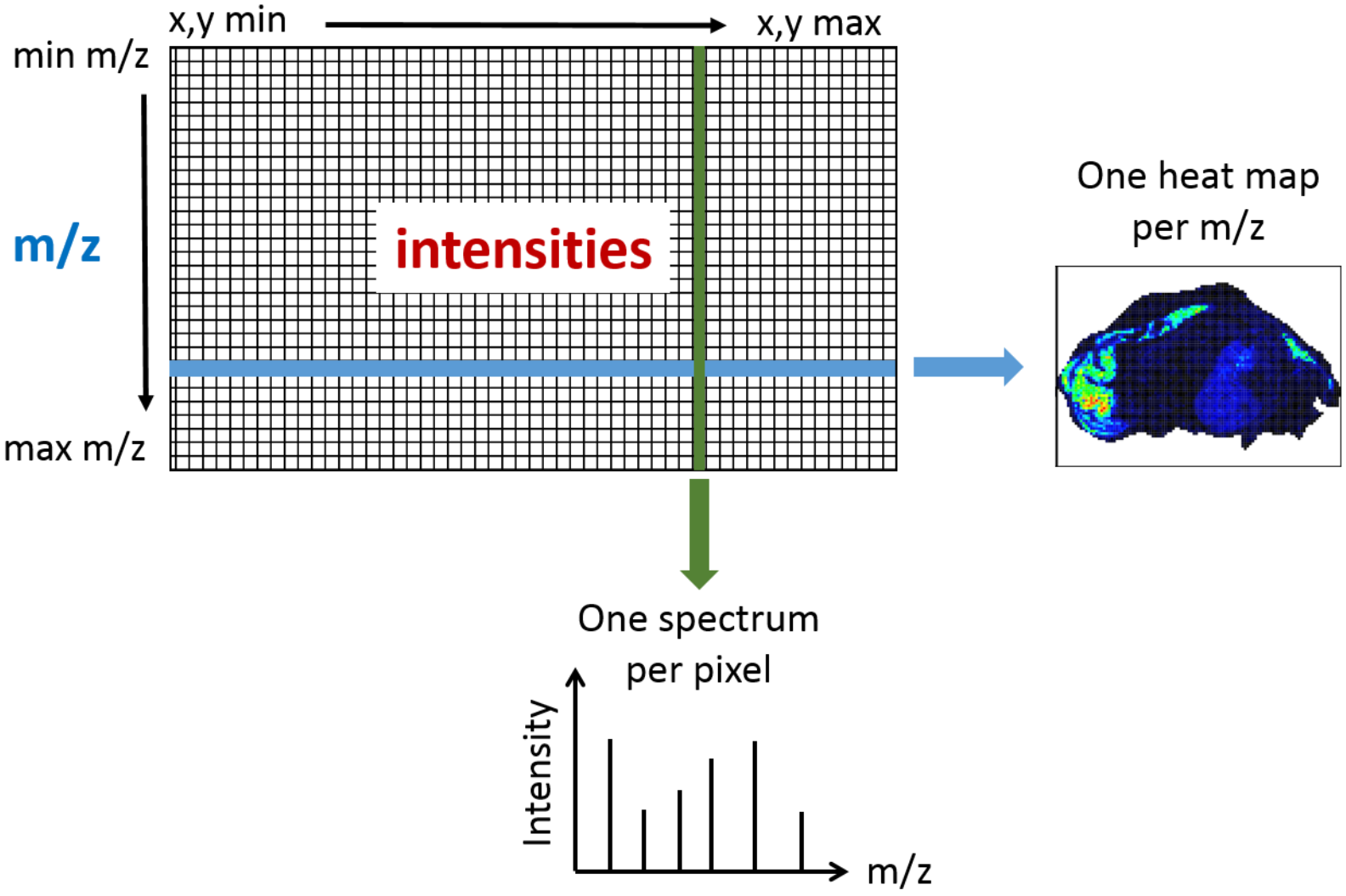
Importing MSI data

- readMSIdata(), readimzML()
- From CardinalWorkflows package
  - Object type: MSImageSet (old format)
  - Coercion to MSImagingExperiment (current format)

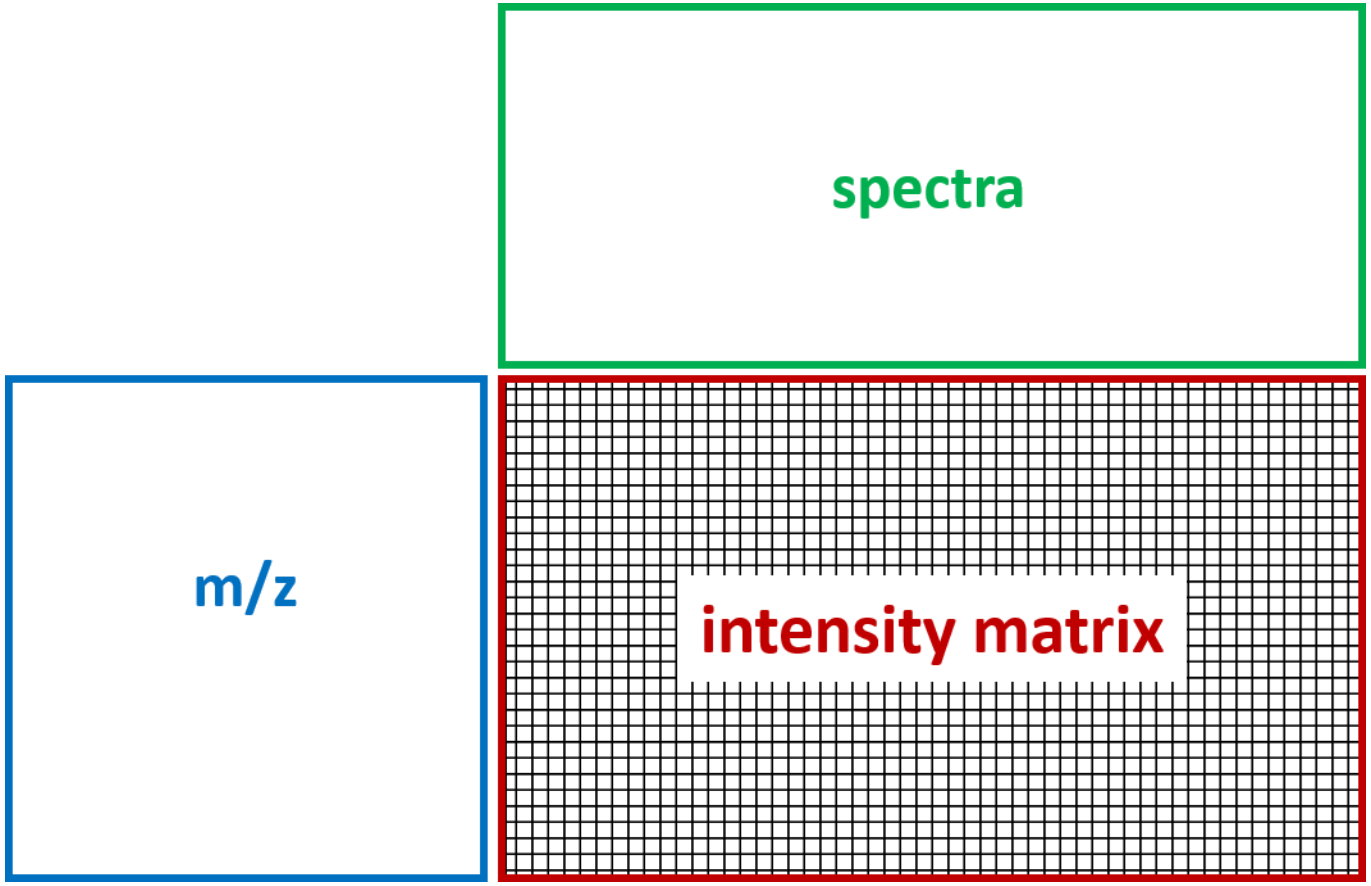
MSI data structure – 3D presentation



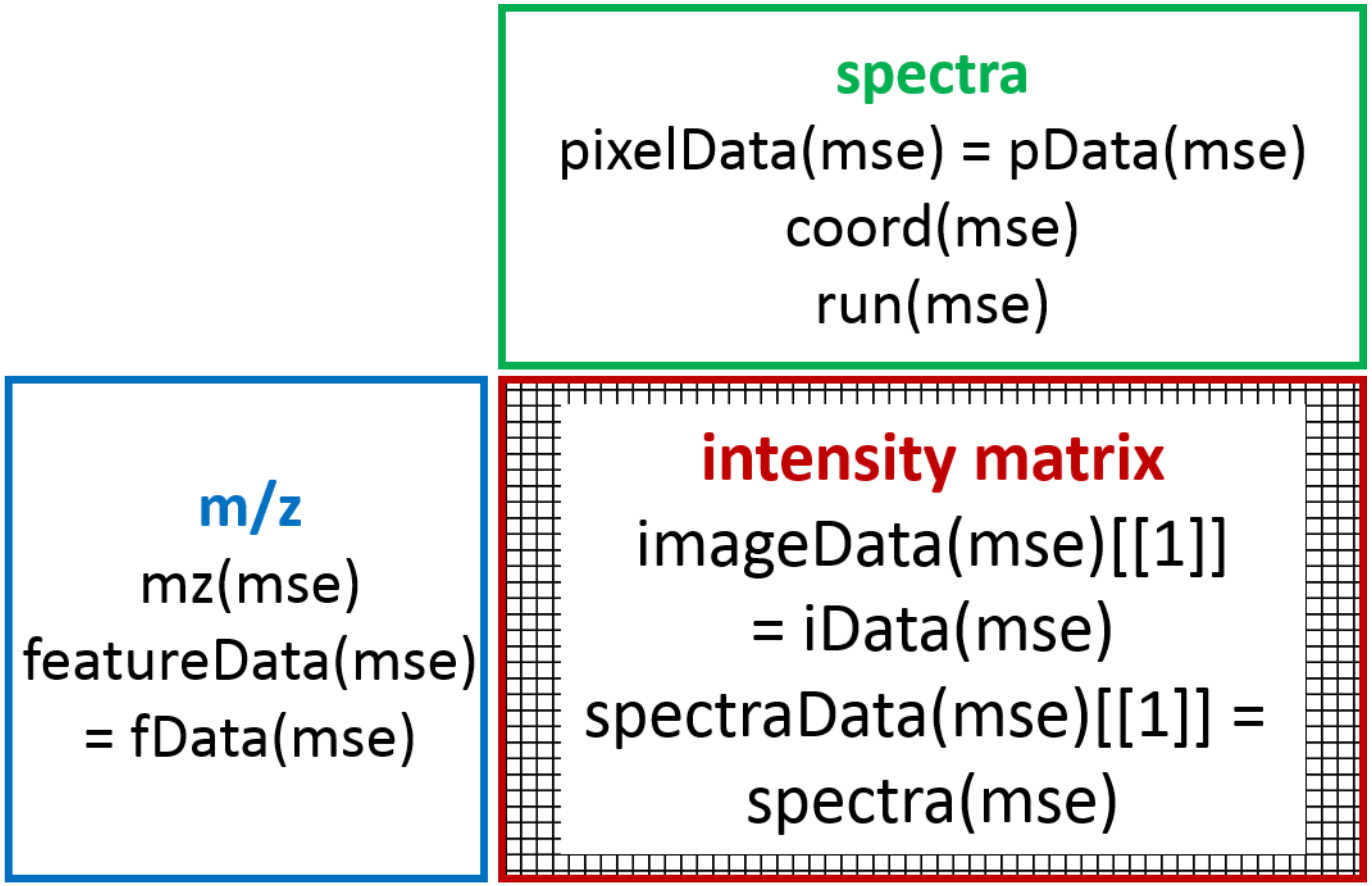
MSI data structure – 2D presentation



MSImagingExperiment – basic data structure



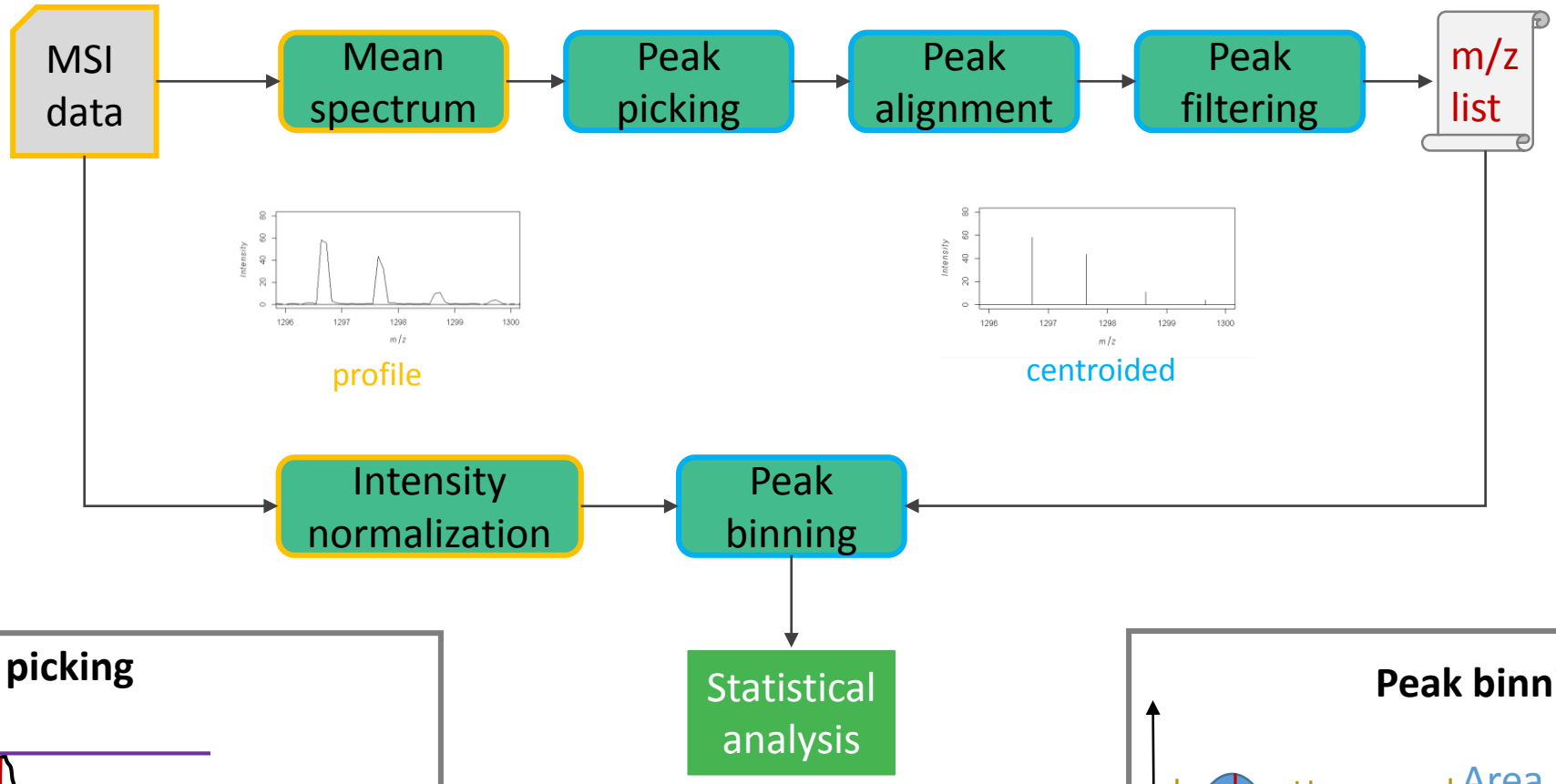
MSImagingExperiment – basic data structure



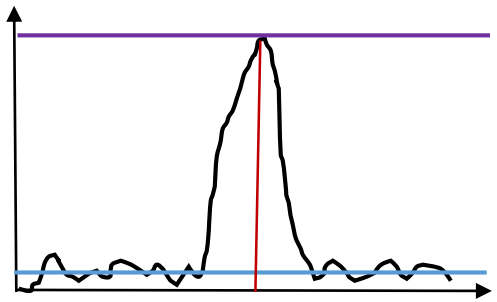
Part II: Pre-processing

- Loading data
- Exploring data
- **Pre-processing**
- Segmentation
- Visualization

Pre-processing workflow

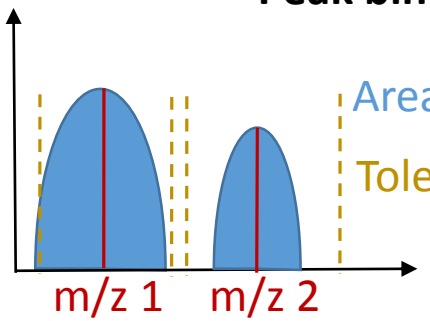


Peak picking

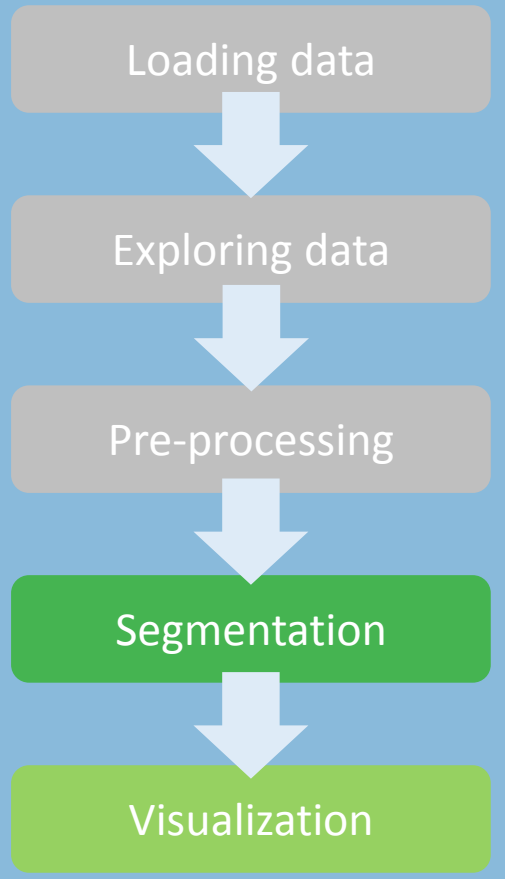


SNR:
Signal to noise ratio

Peak binning

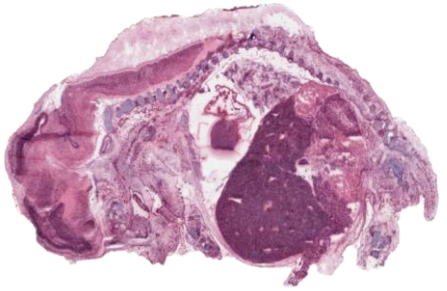


Part III: Segmentation & visualization



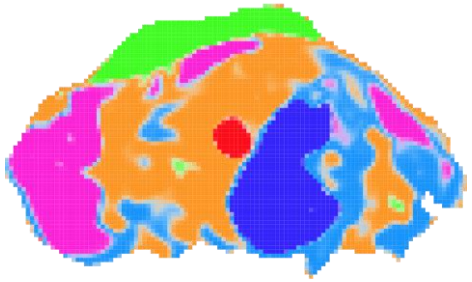
Unsupervised segmentation

Spatial shrunken centroids:



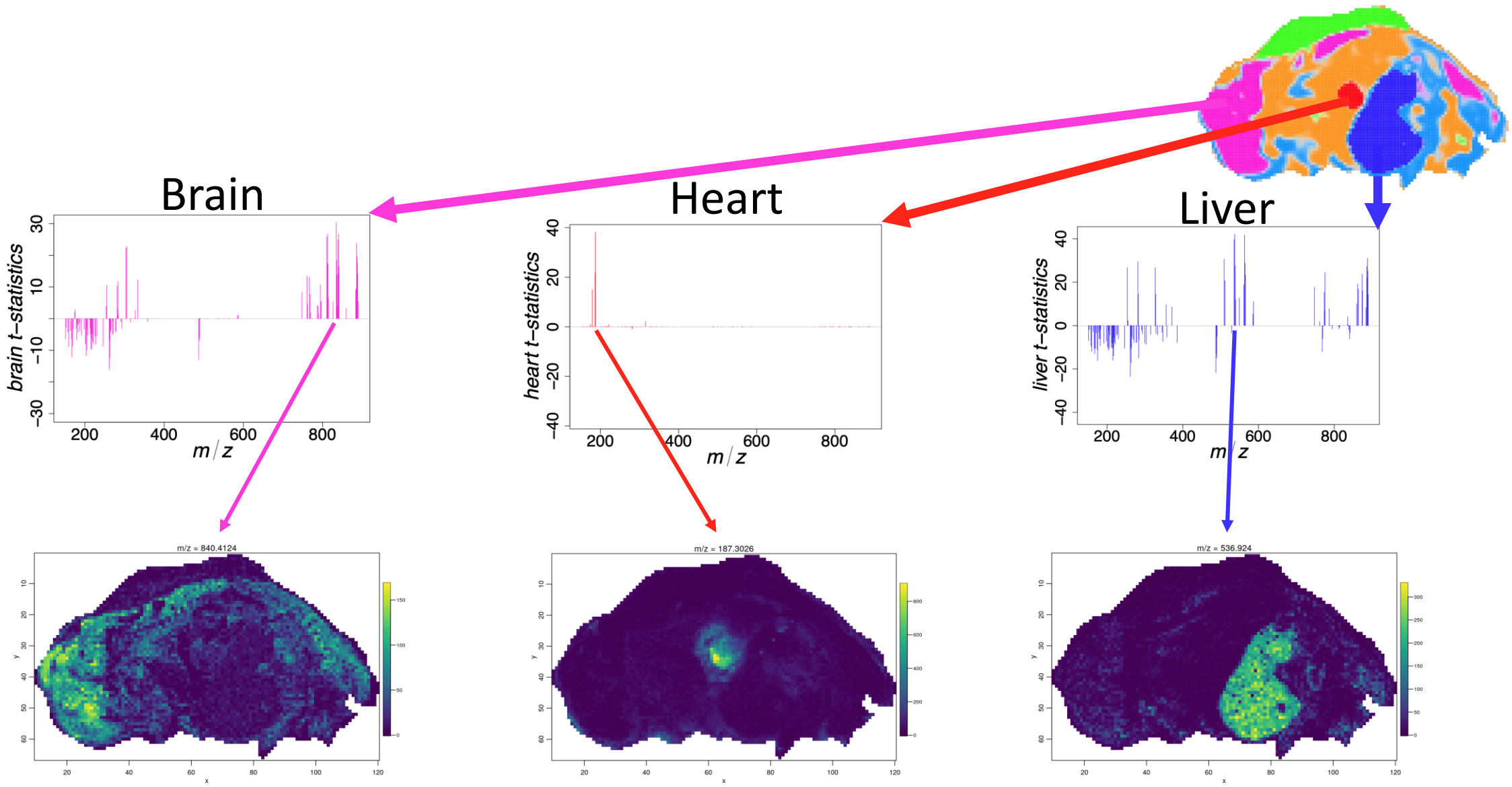
spatially-aware segmentation with feature selection

- Data-driven selection of appropriate number of segments
- Selects most important m/z for distinguishing segments
- Probablity model characterizes uncertainty



K. Bemis, et al. "Probabilistic segmentation of mass spectrometry (MS) images helps select important ions and characterize confidence in the resulting segments." *Molecular & Cellular Proteomics* 15.5 (2016): 1761-1772.

Visualization of most important m/z



Cardinal: resources & thank you

Kylie Bemis

Olga Vitek

Dan Guo

Sai Lakkimsetty

User guides/vignettes	http://cardinalmsi.org/workflows/
Videos May Institute	https://www.youtube.com/c/MayInstituteNEU/playlists
Help forum	http://groups.google.com/d/forum/CardinalMSI

Thank you to the instructors and to the teaching assistants!

Ryan Benz
Meena Choi
Niyati Chopra
Miguel Cosenza
Matthias Fahrner
Amanda Figueroa-Navedo
Melanie Foell
Omkar Reddy Gojala
Dan Guo
Shubhanshu Gupta
Ting Huang
Maanasa Kaza
Smit Anish Kiri
Devon Kohler
Sai Srikanth Lakkimsetty
Danielle LaMay
Ajeya Makanahalli Kempegowda
Yogesh Nizzer
Harish Ramani
Ruthvik Ravindra
Abdul Rehman
Sai Divya Sangeetha Bhagavatula
Siddarth Sathyanarayanan
Gopalika Shama
Rishabh Rajesh Shanbhag
Sagar Singh
Mateusz Staniak
Sara Taheri
Anuska Tak
Derrie Susan Varghese
Amrutha Vempati