Tissue Based Proteomics and Biomarker Discovery

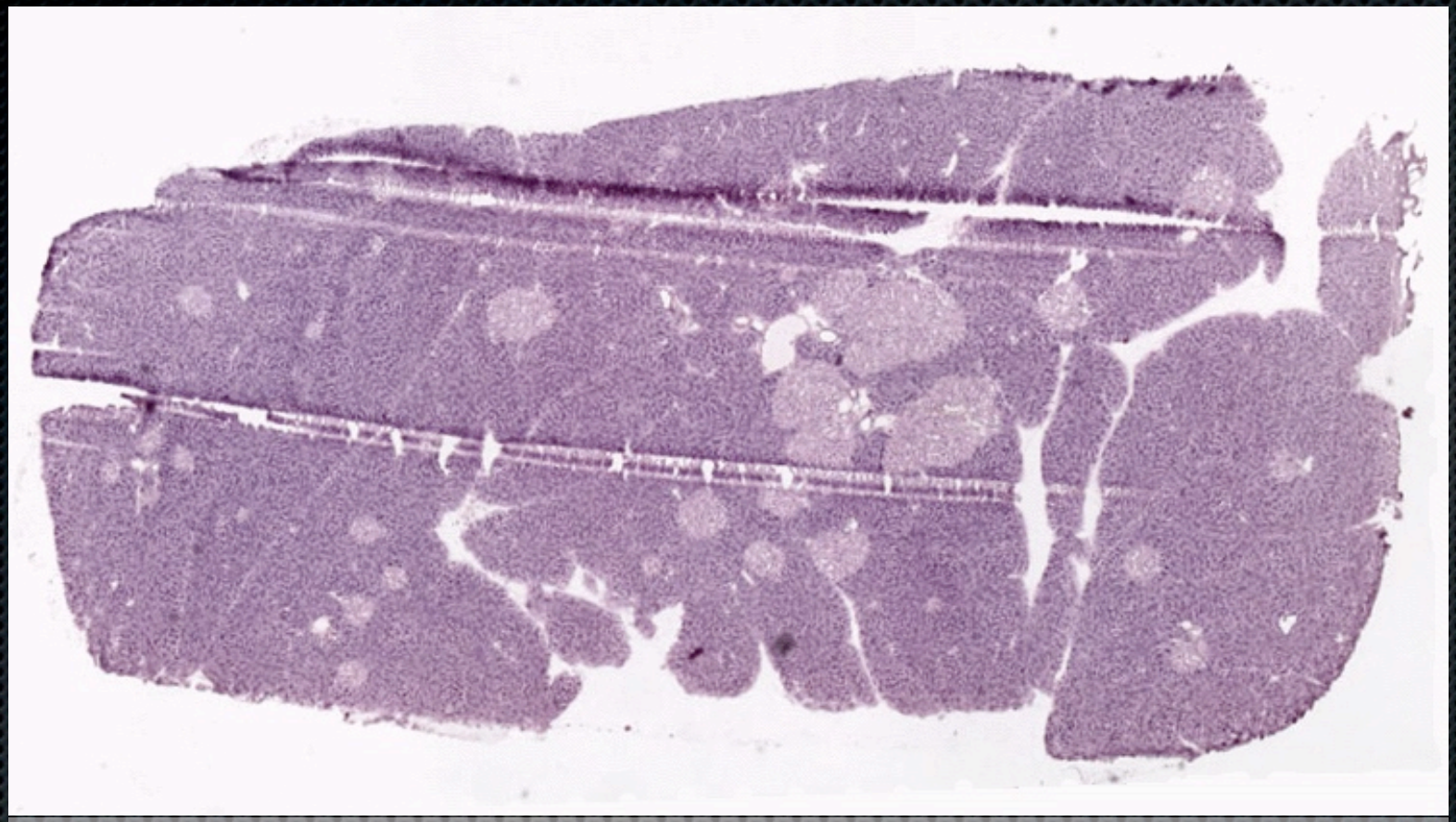
Multivariate Data Mining Strategies for Mass Spectral Imaging

Raf Van de Plas

Promotor
Prof. dr. ir. Bart De Moor

Co-Promotor
Prof. dr. Etienne Waelkens

Katholieke Universiteit Leuven



premise *- The direct analysis of tissue is important for understanding disease.*

problem *- Mass spectral imaging is a promising technology for direct tissue analysis, but is currently hampered by lack of computational techniques.*

goal of this work *- To answer that challenge and develop multivariate data mining techniques to unlock MSI's full potential.*

- Mass Spectral Imaging
- Computational Analysis of MSI Data

Mass Spectral Imaging

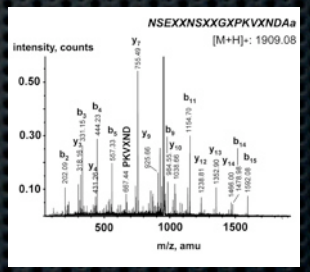
Mass Spectral Imaging

Objective

To link together two aspects:

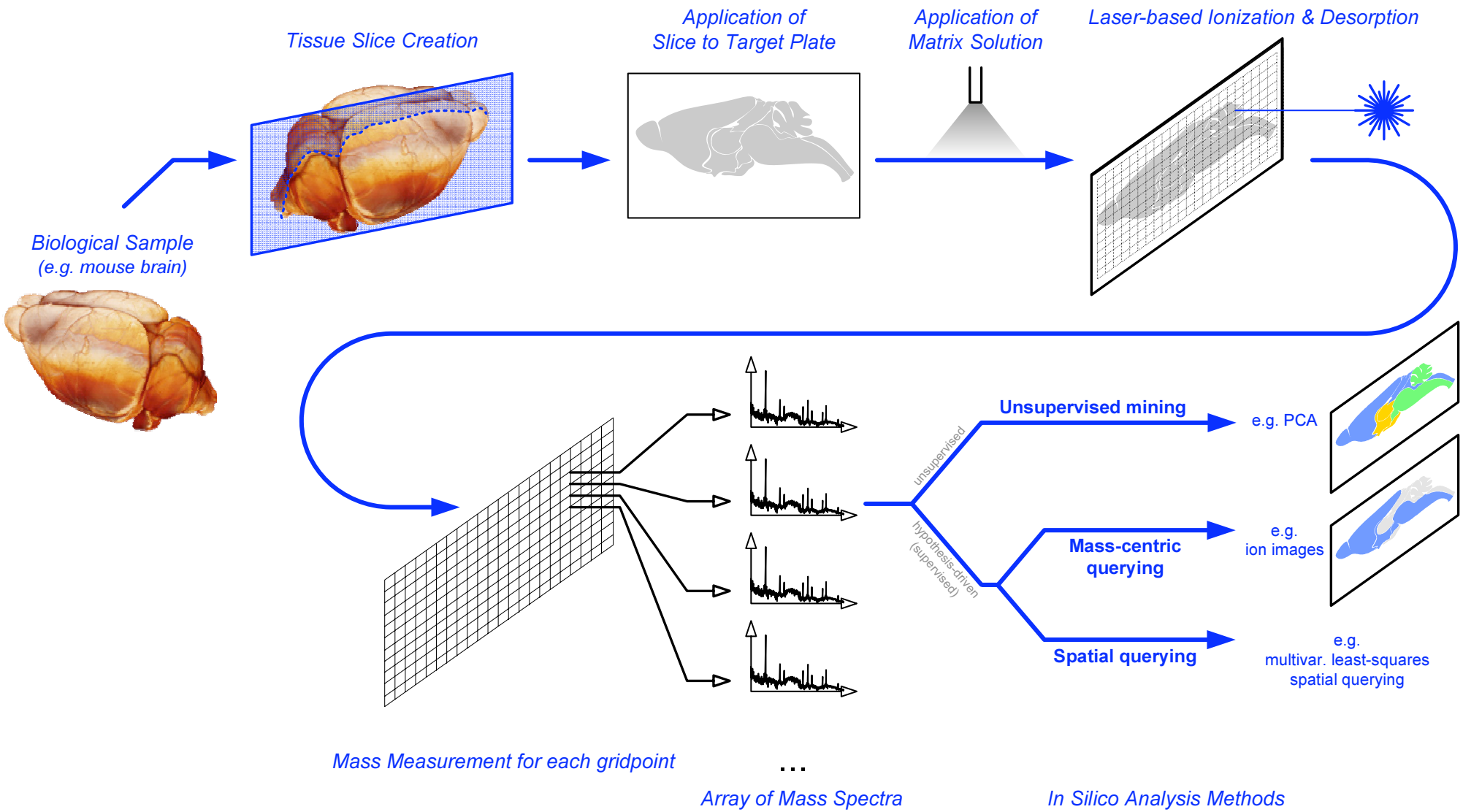
Biochemical characterization

MS

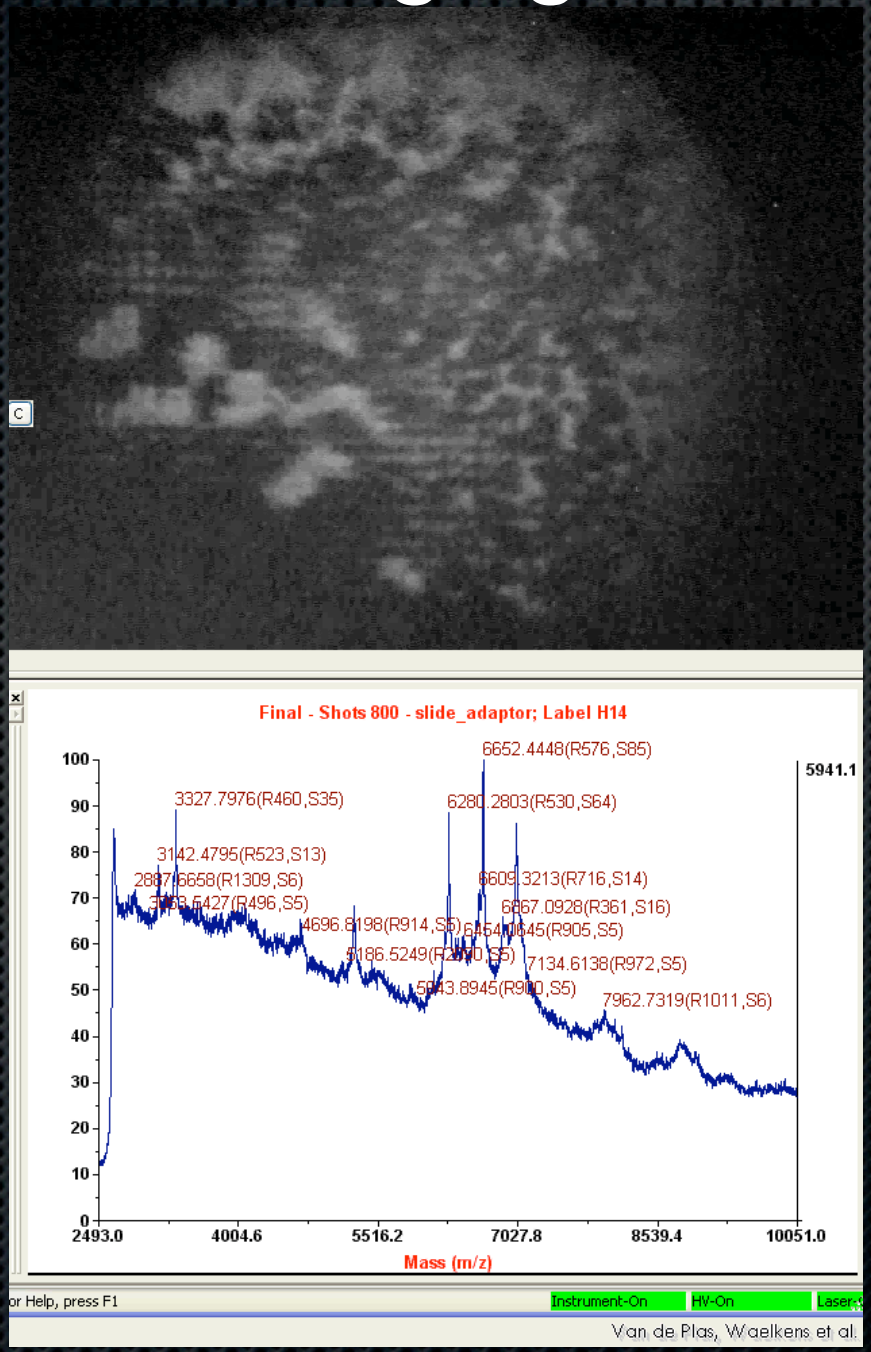


Spatial information

Mass Spectral Imaging



Mass Spectral Imaging



Mass Spectral Imaging

Pros and cons

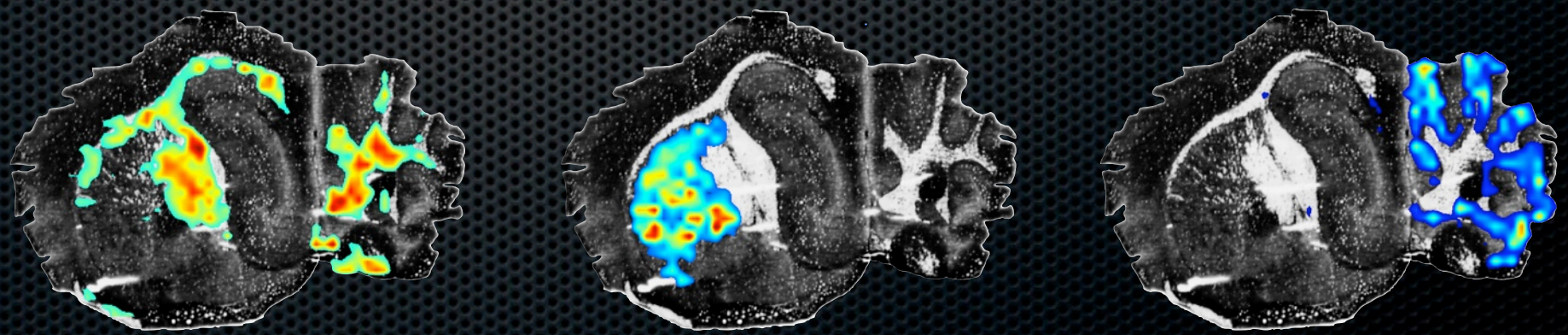
Pros:

- **Requires no chemical labeling of sample**
- **Requires no prior hypothesis** - Excellent exploratory merit
- **Very high specificity** - Better than most immunobodies
- **Massively multiplexed** - Hundreds of ions spatially followed in one experiment
- **Insight into intermolecule relationships** - Inherent multivariate nature

Cons:

- **Relatively low spatial resolution**
- **Sensitivity is tissue-dependent**

Complementary to traditional molecular imaging modalities:



Mass Spectral Imaging

Starting up…

- ***Proper sample preparation and extraction***

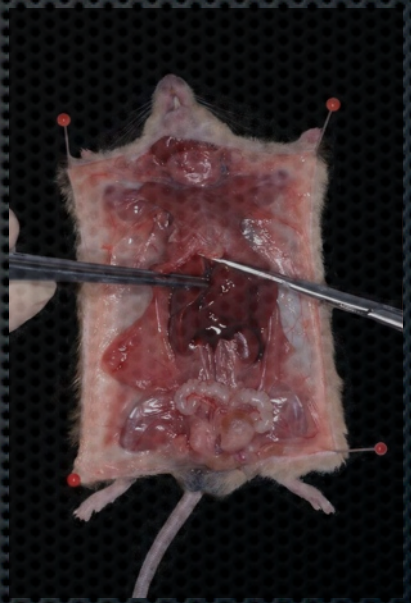
  tissue extraction, freezing, washing, cutting, collaboration with pathologists,…

- ***Experimentation with matrix deposition***

  experimentation with matrix solutions, heterogenous deposition operation (self-made, Labcyte, Shimadzu)

- ***Fusing heterogenous equipment***

  custom target plate (ABI), ITO conductive glass slides (Bruker), conductive tape (Shimadzu), different MS platforms,…



Conclusion:

Proper MS-related computational analysis requires a lot of physics, biochemistry, and MS-specific context and experience. 'Standardized' computational approaches rarely fit MS or MSI data directly.

Mass Spectral Imaging

Starting up...

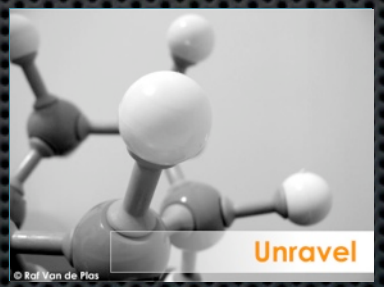
Heterogeneous instruments & formats

Applied Biosystems Inc.
4800 MALDI-TOF/TOF Analyzer

Bruker Daltonics
UltraFlex II w/ SmartBeam Laser

Shimadzu AXIMA TOF^2

Others...

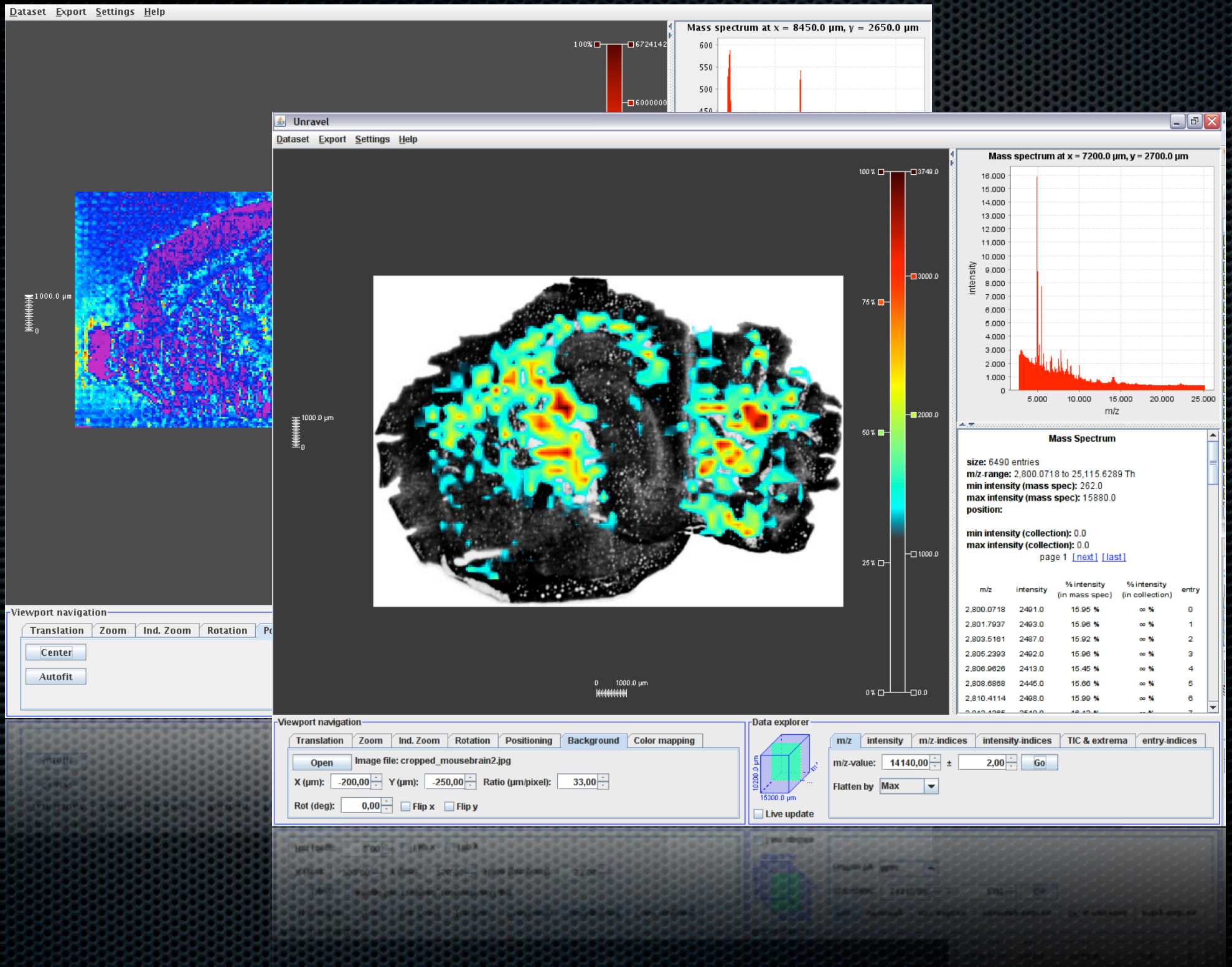


Unravel software

Java-based
20.000 lines of code

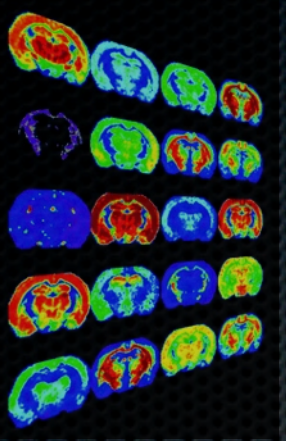
Uniform analysis & exploration

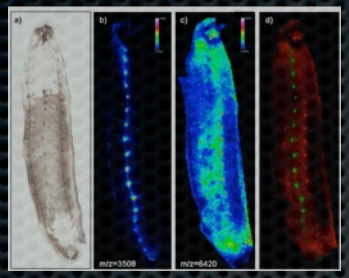
- Data export
  - ↳ Computational & statistical analysis
- Image export
- Data exploration
- Real-scale representation
- Ion image generation
- Overlay with different modalities (e.g. microscopy)

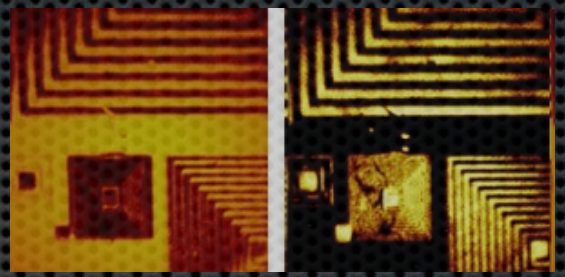

Dataset Export Settings Help
100% 6724142 6000000
Mass spectrum at x = 8450.0 µm, y = 2650.0 µm
600
550
500
1000.0 µm
0
Unravel
Dataset Export Settings Help
100 % 3749.0
75 % 3000.0
50 % 2000.0
25 % 1000.0
0 % 0.0
1000.0 µm
0 1000.0 µm
Mass spectrum at x = 7200.0 µm, y = 2700.0 µm
intensity
m/z
5.000 10.000 15.000 20.000 25.000
Mass Spectrum
size: 6490 entries
m/z-range: 2,800.0718 to 25,115.6289 Th
min intensity (mass spec): 262.0
max intensity (mass spec): 15880.0
position:
min intensity (collection): 0.0
max intensity (collection): 0.0
page 1 [next] [last]
m/z intensity % intensity (in mass spec) % intensity (in collection) entry
2,800.0718 2491.0 15.95 % ∞ % 0
2,801.7937 2493.0 15.96 % ∞ % 1
2,803.5161 2487.0 15.92 % ∞ % 2
2,805.2393 2492.0 15.96 % ∞ % 3
2,806.9626 2413.0 15.45 % ∞ % 4
2,808.6868 2445.0 15.66 % ∞ % 5
2,810.4114 2498.0 15.99 % ∞ % 6
Viewport navigation
Translation Zoom Ind. Zoom Rotation
Center
Autofit
Viewport navigation
Translation Zoom Ind. Zoom Rotation Positioning Background Color mapping
Open
Image file: cropped_mousebrain2.jpg
X (µm): -200,00 Y (µm): -250,00 Ratio (µm/pixel): 33,00
Rot (deg): 0,00 Flip x Flip y
Data explorer
10200.0 µm
15300.0 µm
Live update
m/z intensity m/z-indices intensity-indices TIC & extrema entry-indices
m/z-value: 14140,00 ± 2,00 Go
Flatten by Max

Mass Spectral Imaging

Experiencing rapid growth...







MALDI

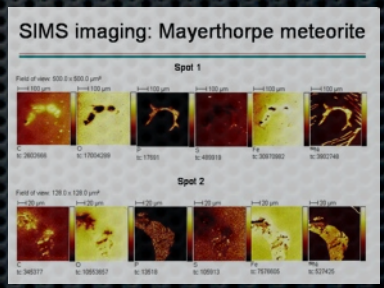
organic, biomacromolecules

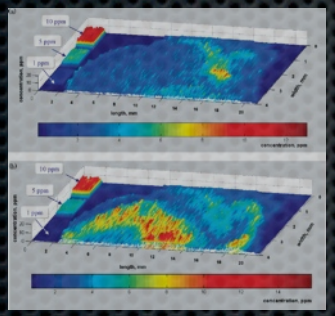
Caprioli, Vanderbilt University
McDonnell, Leiden University
Fournier & Salzet, University of Lille
K.U.Leuven

SIMS

anorganic, small molecules

Heeren, AMOLF Amsterdam
Vickerman, University of Manchester
Winograd, Penn State University





LA-ICP

anorganic, material science

Becker, Juelich Research Centre

LA-ESI

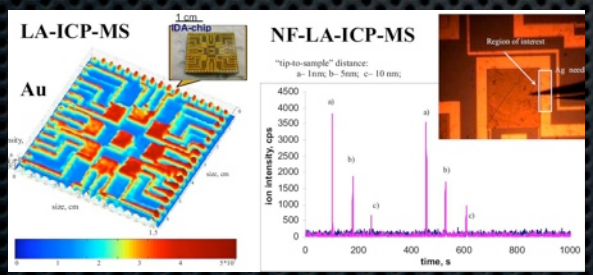
organic, ambient pressure

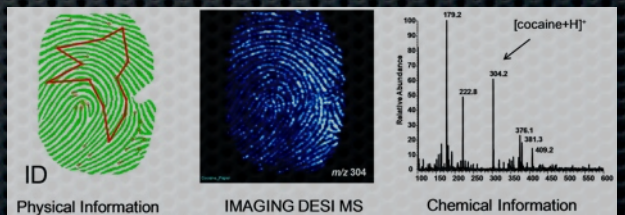
Vertes, George Washington University

DESI

forensic apps, latent fingerprints,...

Cooks, Purdue University



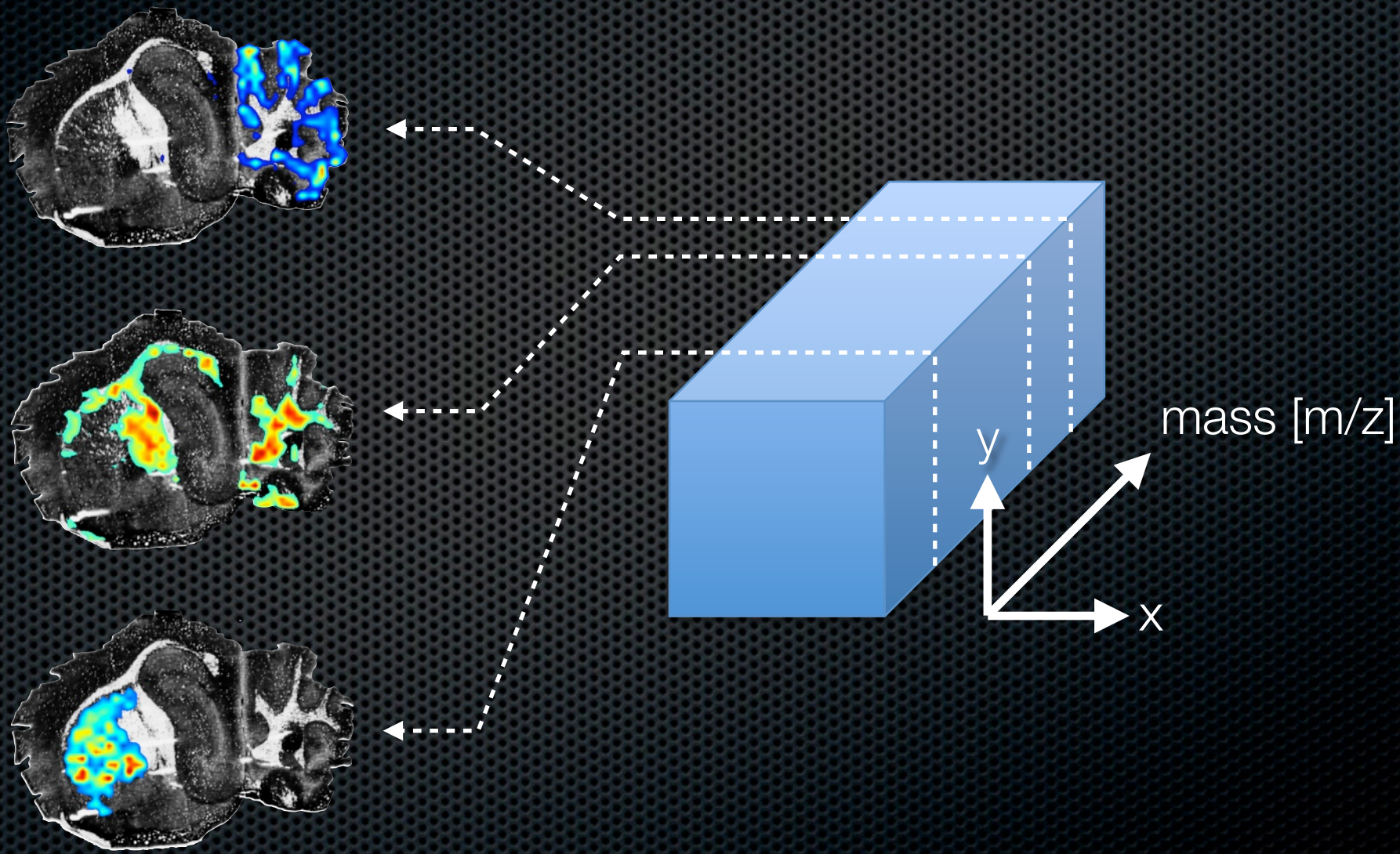


Computational Analysis of MSI Data

Contributions of this thesis

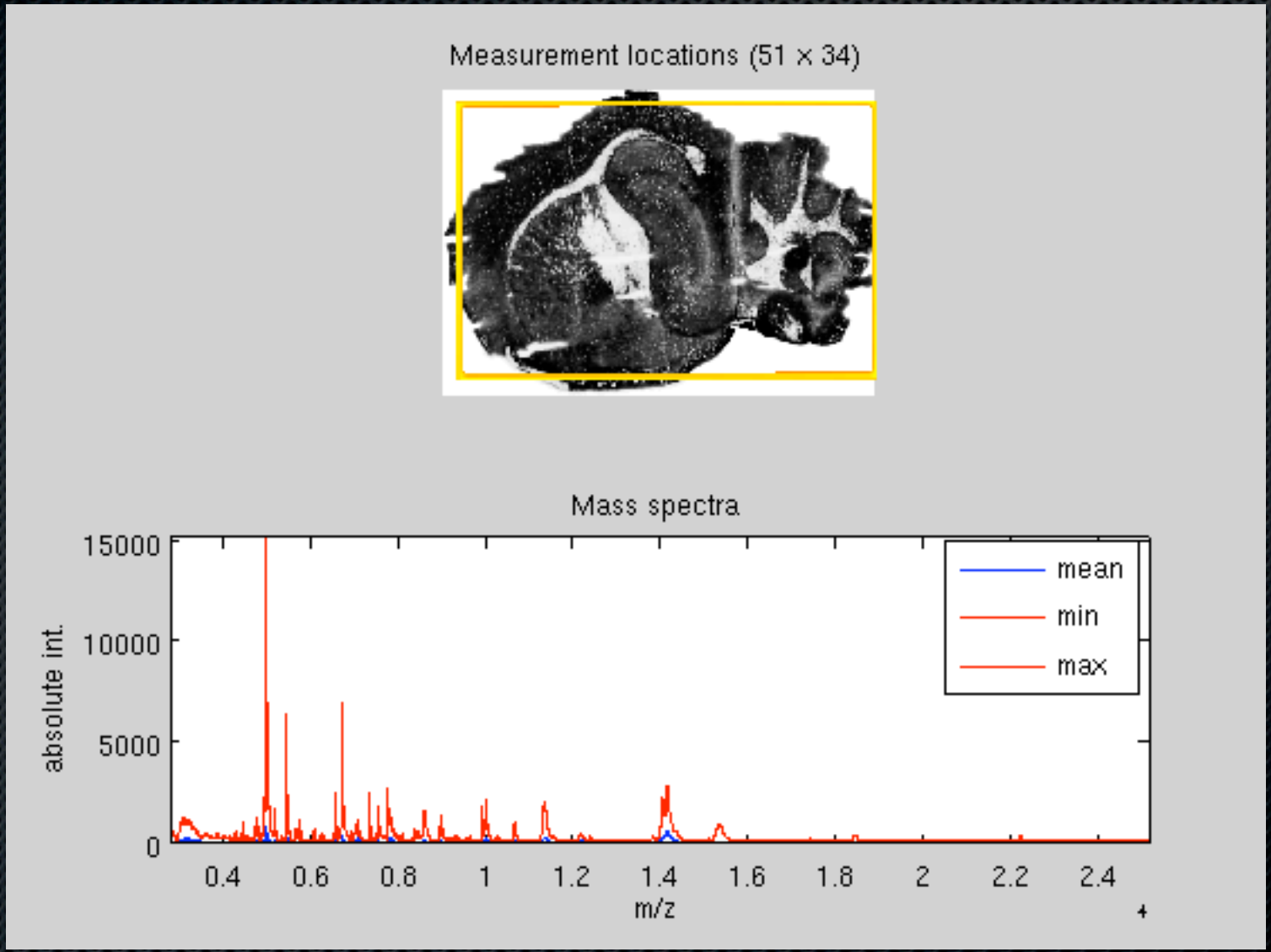
Mass Spectral Imaging

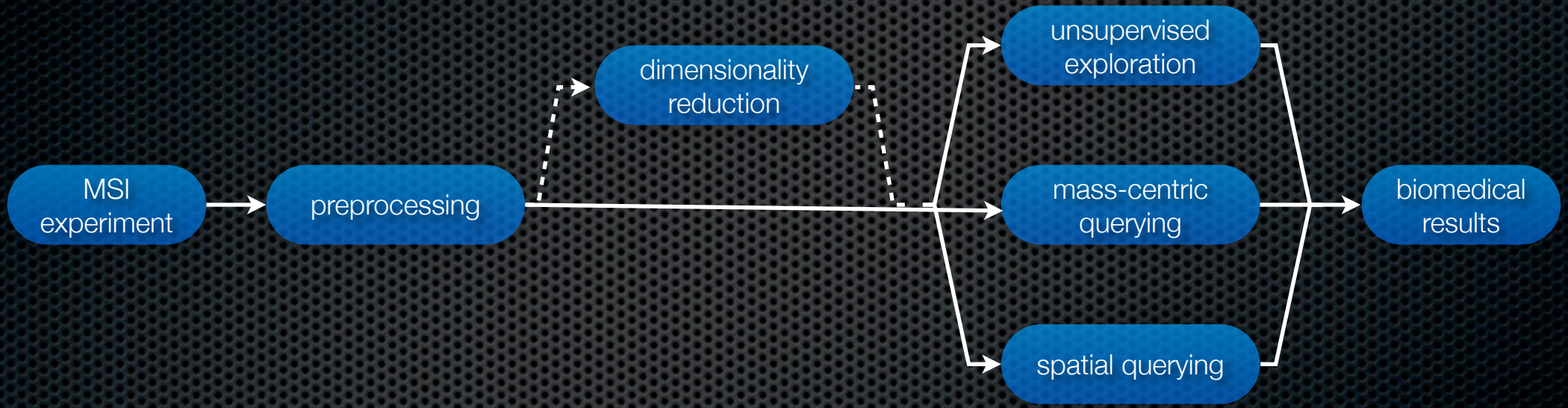
MSI data

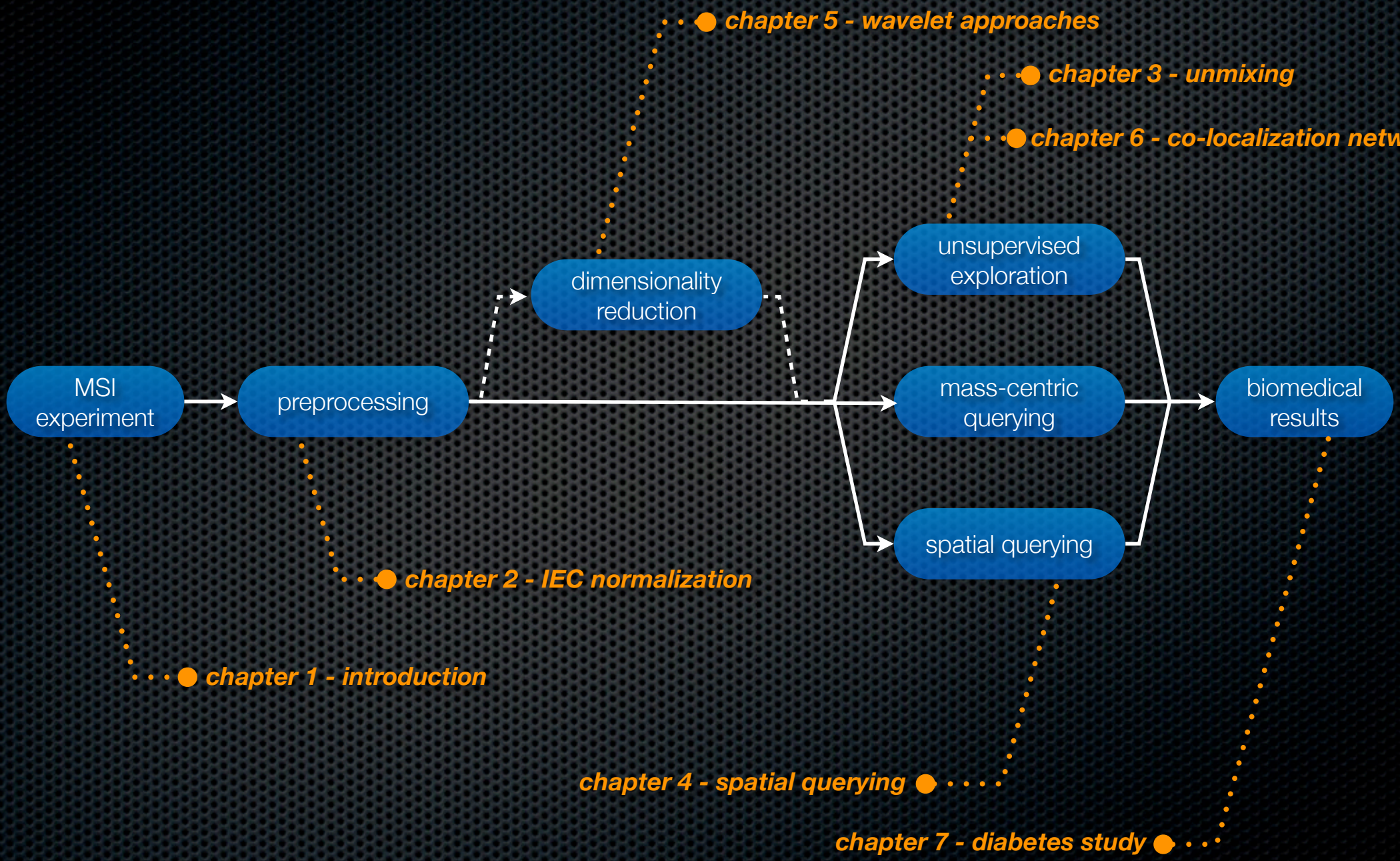


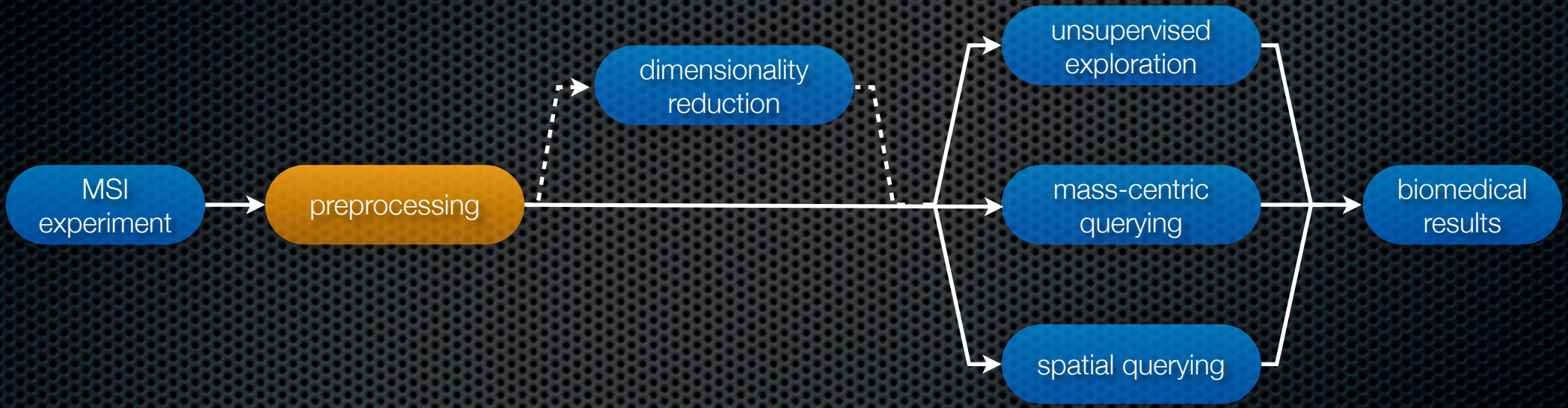
Mass Spectral Imaging

MSI data




MSI experiment
preprocessing
dimensionality reduction
unsupervised exploration
mass-centric querying
spatial querying
biomedical results


chapter 5 - wavelet approaches
chapter 3 - unmixing
chapter 6 - co-localization netwo
dimensionality reduction
unsupervised exploration
MSI experiment
preprocessing
mass-centric querying
biomedical results
spatial querying
chapter 2 - IEC normalization
chapter 1 - introduction
chapter 4 - spatial querying
chapter 7 - diabetes study


MSI experiment
preprocessing
dimensionality reduction
unsupervised exploration
mass-centric querying
spatial querying
biomedical results

Preprocessing - Normalization

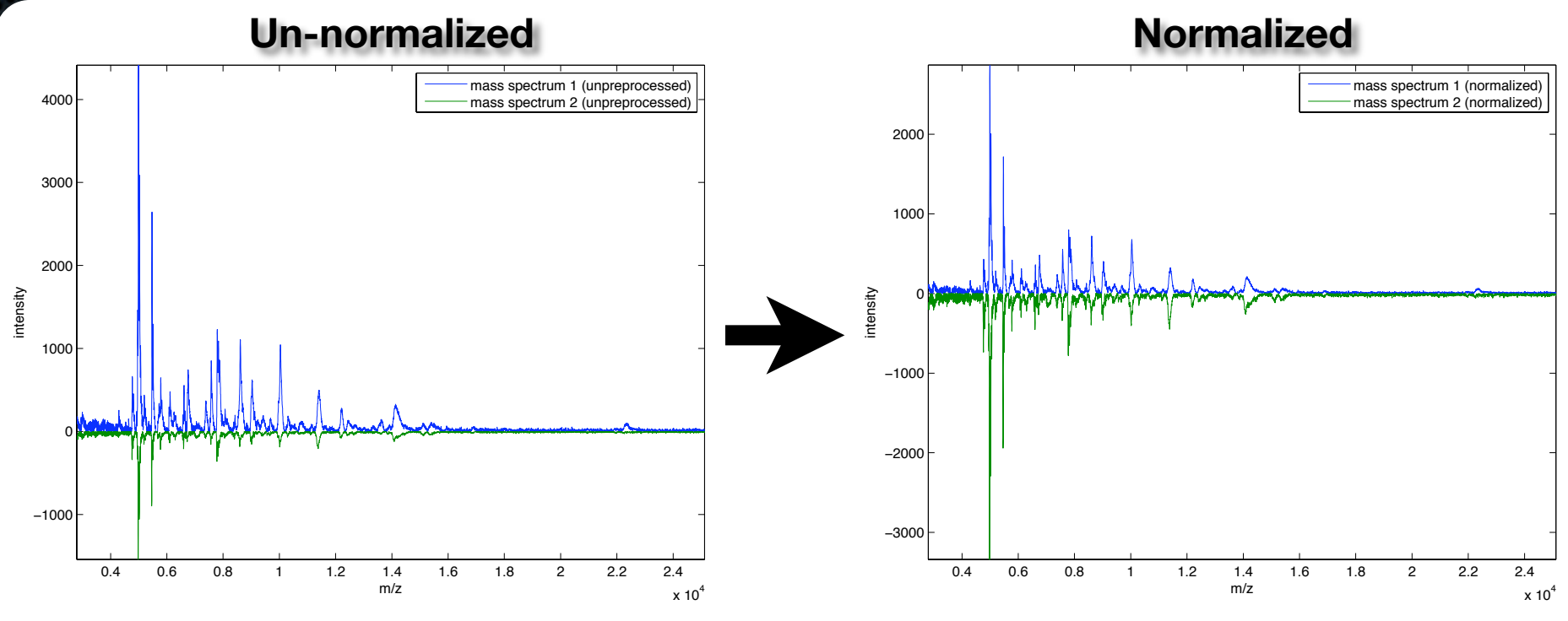
Problem definition

Problem

SNR differs from spectrum to spectrum due to wet lab sample prep and instrumental noise. ➜ Non-biological variation

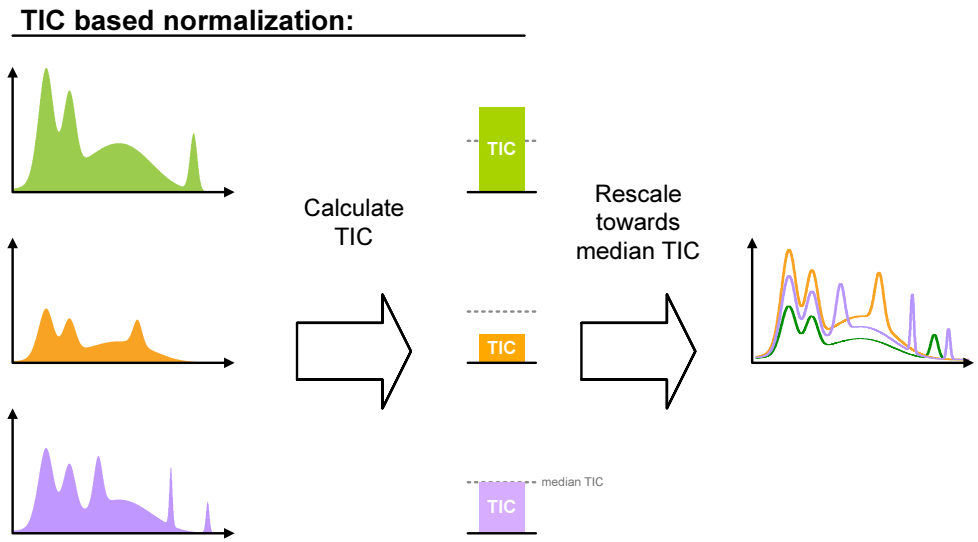
Goal

Project peak heights from several spectra onto common intensity scale.



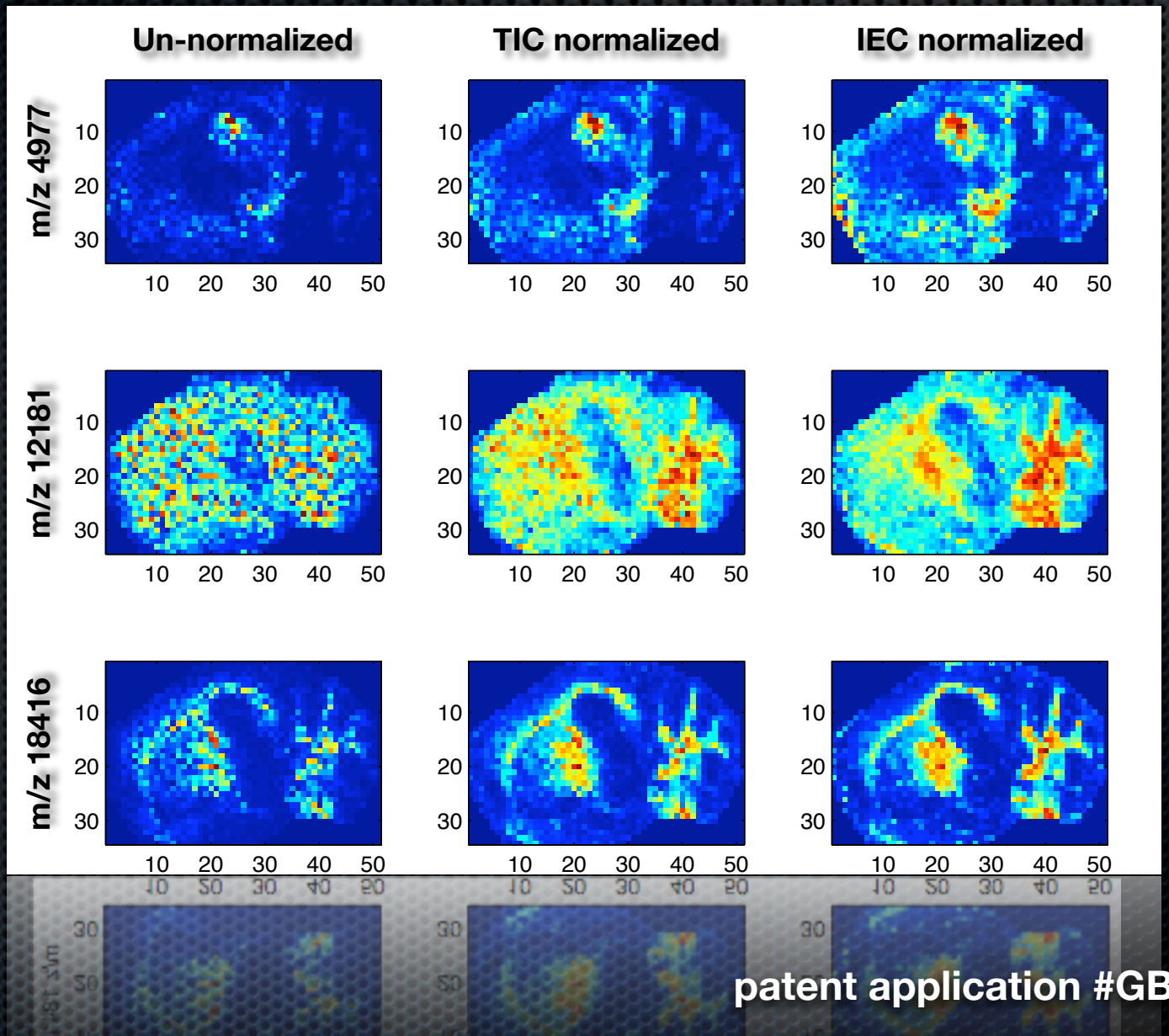
Preprocessing - Normalization

Contribution: Ionization Efficiency Correction



Preprocessing - Normalization

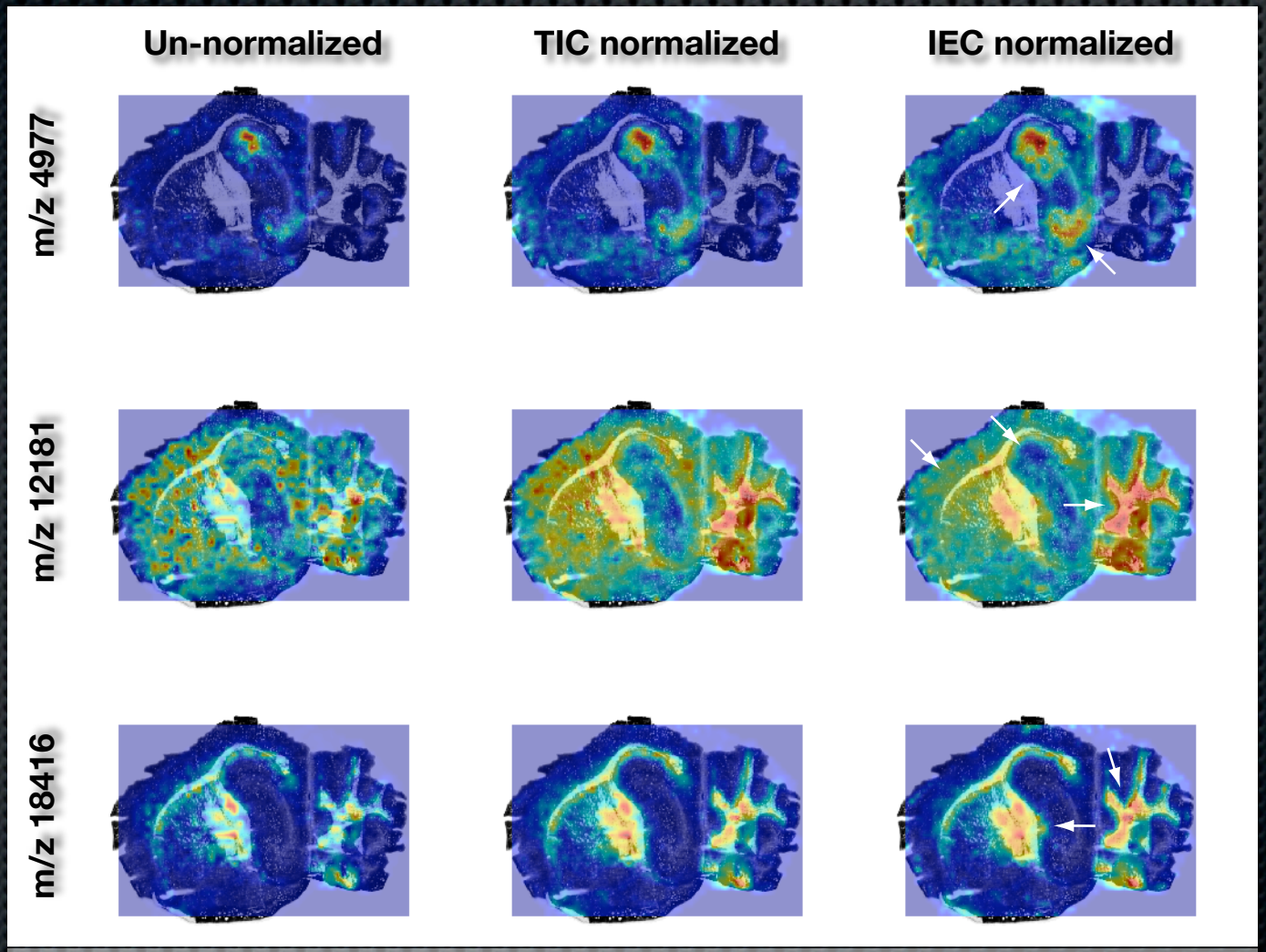
Contribution: Ionization Efficiency Correction



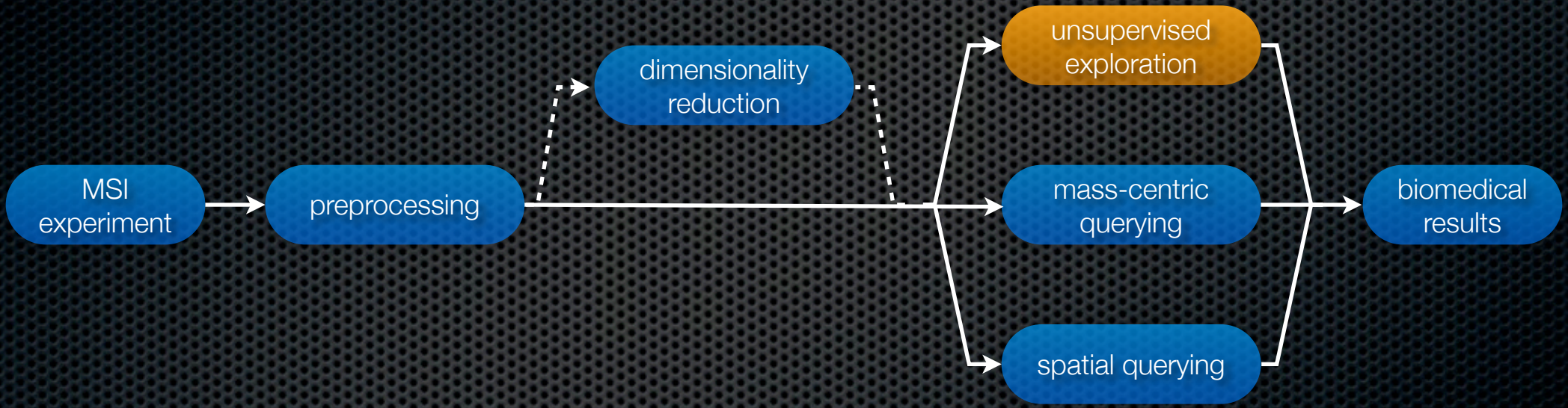
patent application #GB1005959.0

Preprocessing - Normalization

Contribution: Ionization Efficiency Correction



patent application #GB1005959.0


MSI experiment
preprocessing
dimensionality reduction
unsupervised exploration
mass-centric querying
spatial querying
biomedical results

Unmixing

Problem definition

Problem

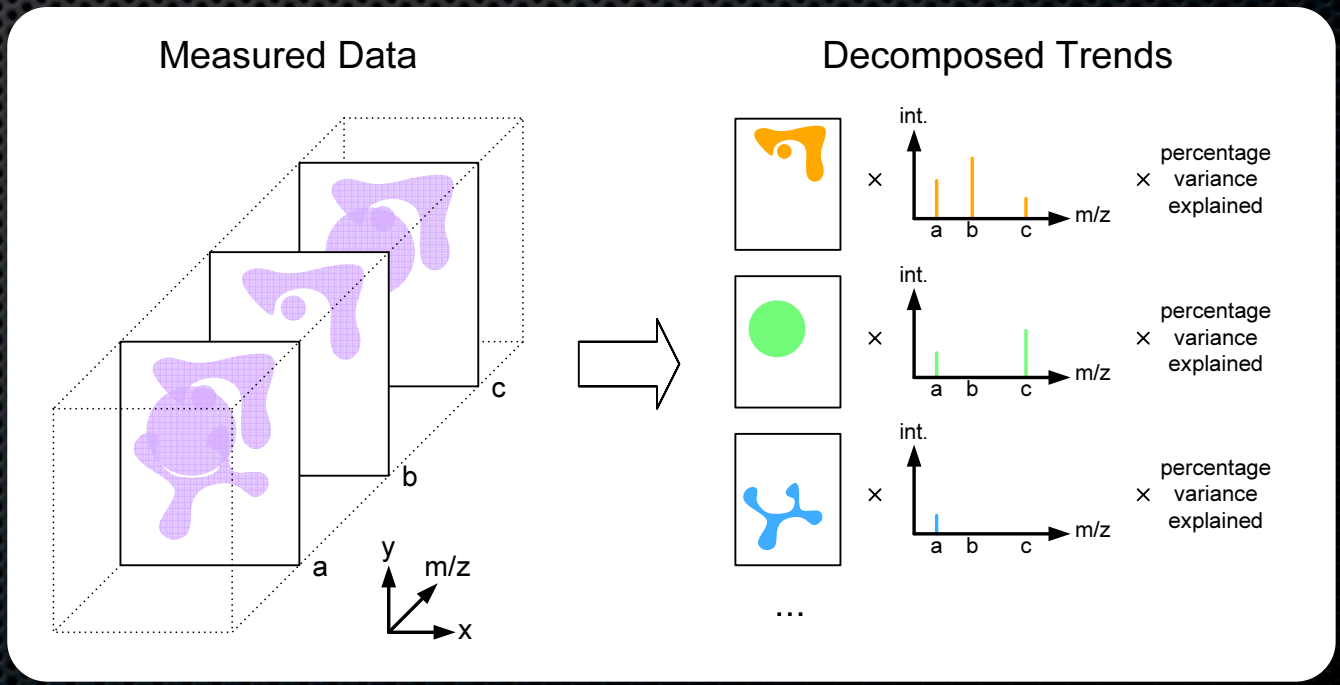
Organic tissue is a complex mixture of chemical content signatures due to biological, wet lab, and instrumental effects.

Goal

To reverse the mixing process and disassemble the measurements into underlying components.

Applications

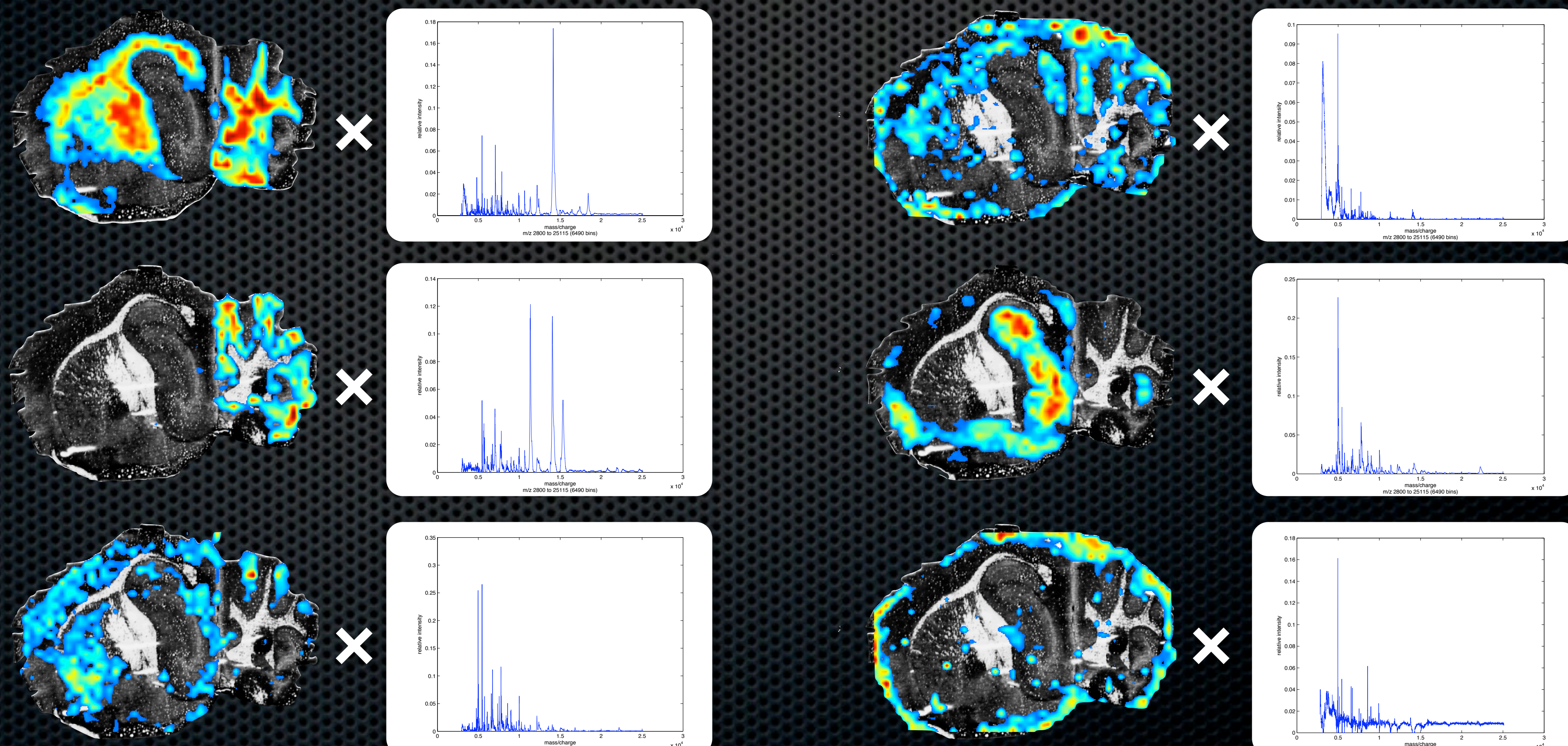
- Aid interpretation
- Filtering
- Dimensionality reduction
- Denoising



Unmixing
Contributions

- Principal Component Analysis (PCA)
- Peak Intensity Weighted PCA (PIW-PCA)
- Non-negative matrix factorization (NMF)

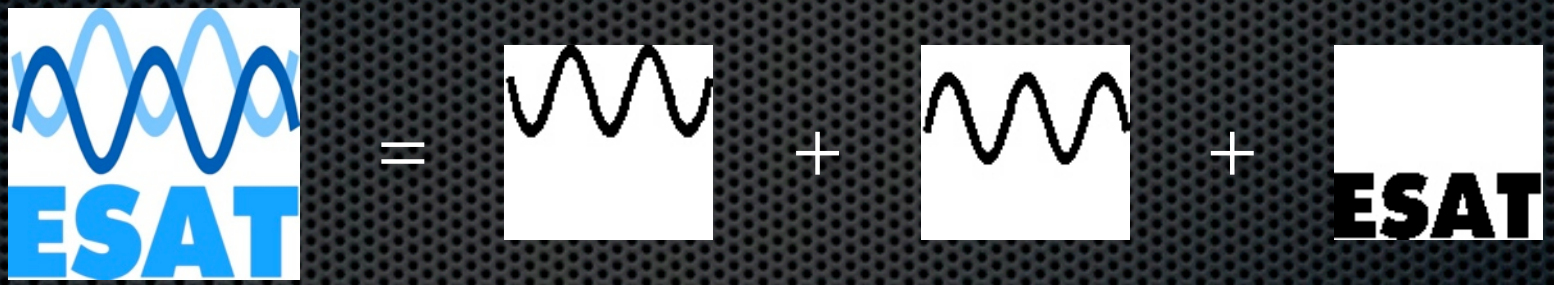
$$X \approx \sum_{a=1}^{r} w_a h_a^T = W'H'^T$$



Unmixing

Algorithm comparison

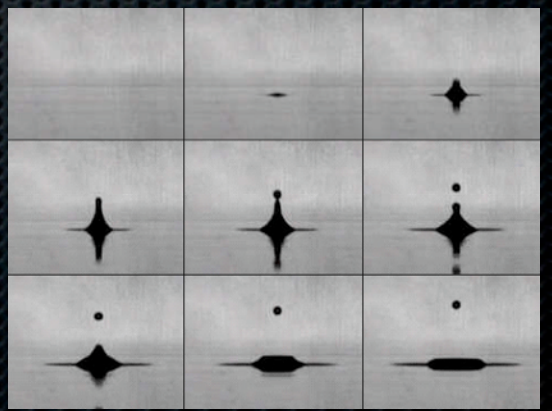
The ESAT logo at micrometer scale
divided in 3 distinct subzones



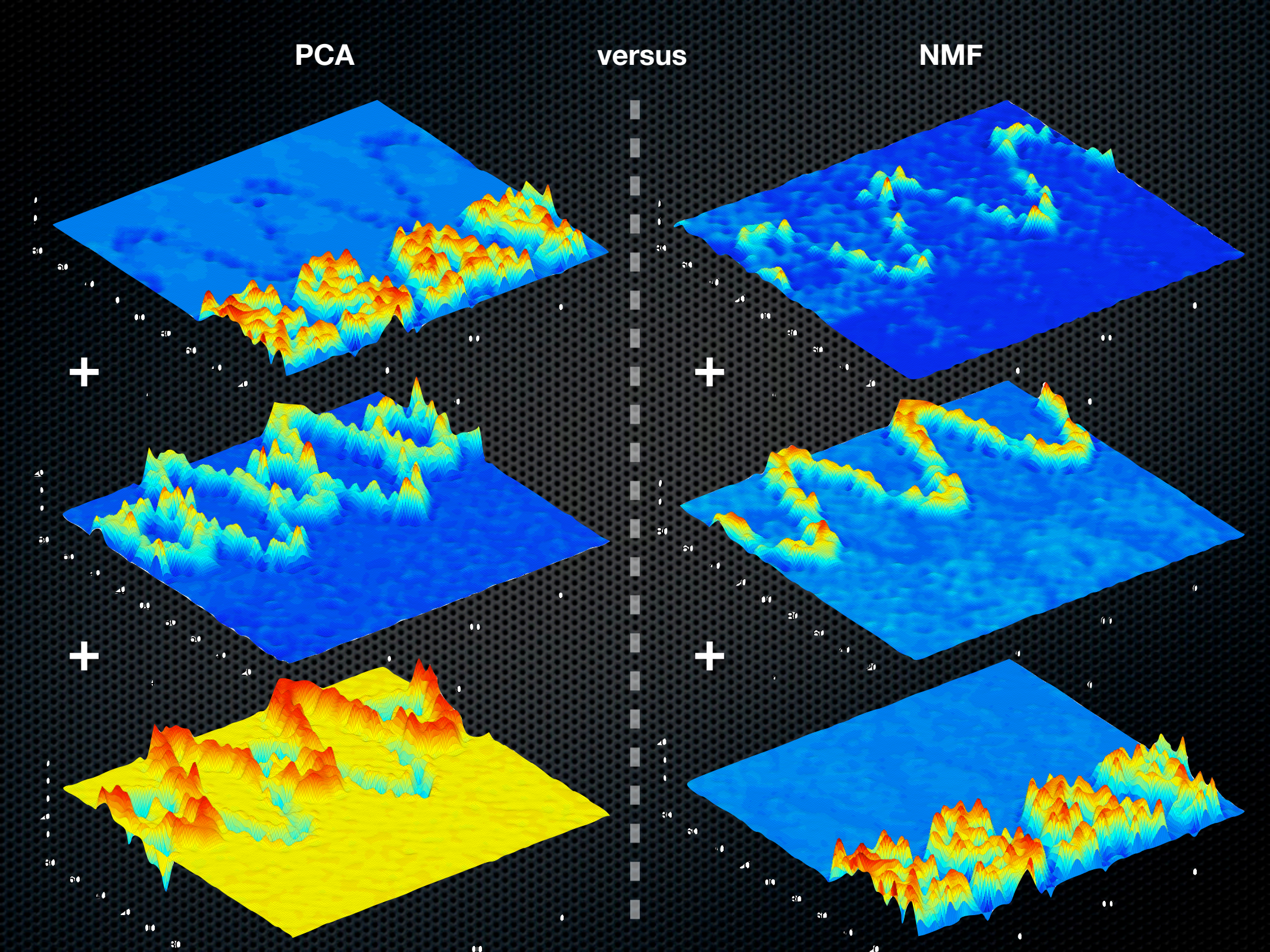
How?

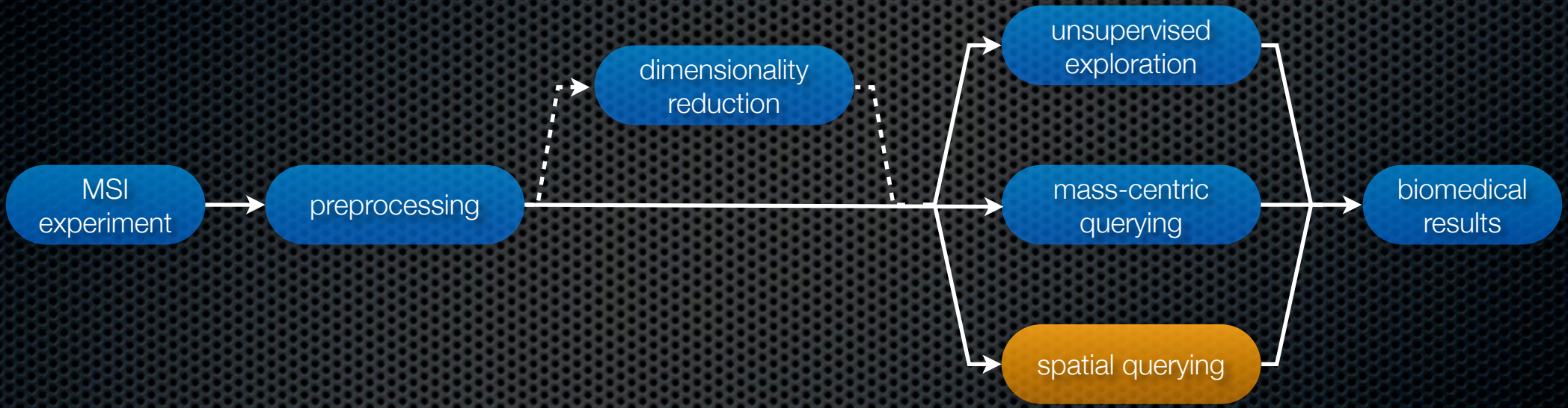
Using the Labcyte Portrait
P630 chemical spotter

Principle: Acoustic Droplet
Ejection for "Touchless"
Liquid Transfer



We kindly acknowledge Siobhan Pickett, Jovica Pavlović, and Elaine Heron from Labcyte Inc.


PCA
versus
NMF


MSI experiment
preprocessing
dimensionality reduction
unsupervised exploration
mass-centric querying
spatial querying
biomedical results

Spatial Querying

Problem definition

Problem

Need method to interrogate MSI data for molecules that exhibit a spatial expression pattern of interest. Many disease cases have spatial prior information available.

Goal

The ability to ask a biological question using a spatial expression pattern of interest (= spatial query). A spatial query

- is a hypothesis-testing question **formulated in the spatial domain**.
- arises from scientific questions that **focus on a particular area in the tissue**.
- retrieves the **chemical signatures and relationships specific to that area**.

A method for spatial querying of MSI data is currently lacking, although there are many opportunities:

- Parkinson's disease - amygdala and putamen in the brain
- Huntington's disease - striatum in the brain
- Amyotrophic Lateral Sclerosis - motor neurons in spinal cord
- ...

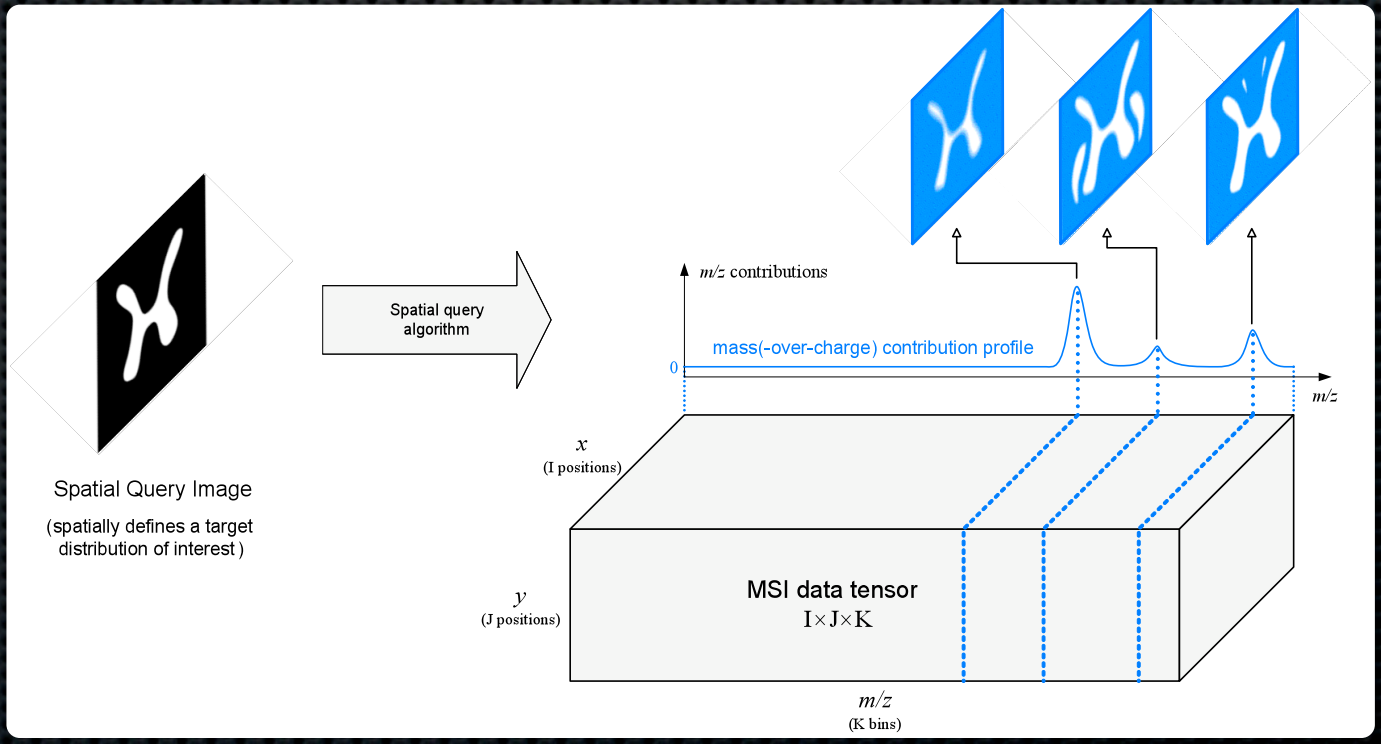
Spatial Querying

Contribution: non-negative least squares approach

The query q is considered a weighted sum of all features (ion images, m/z bins).

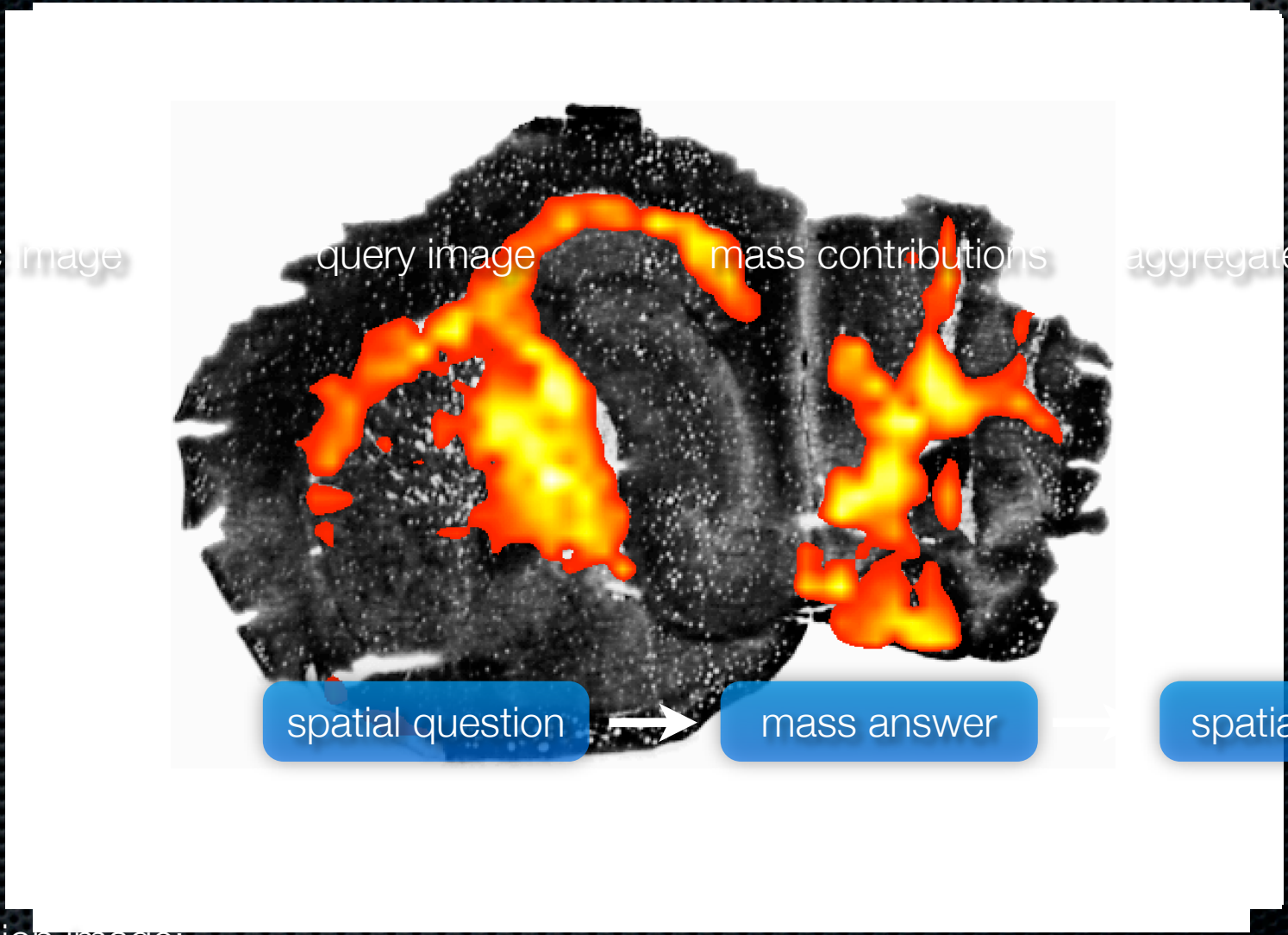
A linear model is adopted: $$q_k \equiv \sum_{m=1}^{M} \phi_k^m p_m + \epsilon_k, \quad \forall k = 1, \ldots, K$$

Goal - To find mass contribution coefficients: $$p = (p_1, \ldots, p_M)^T$$

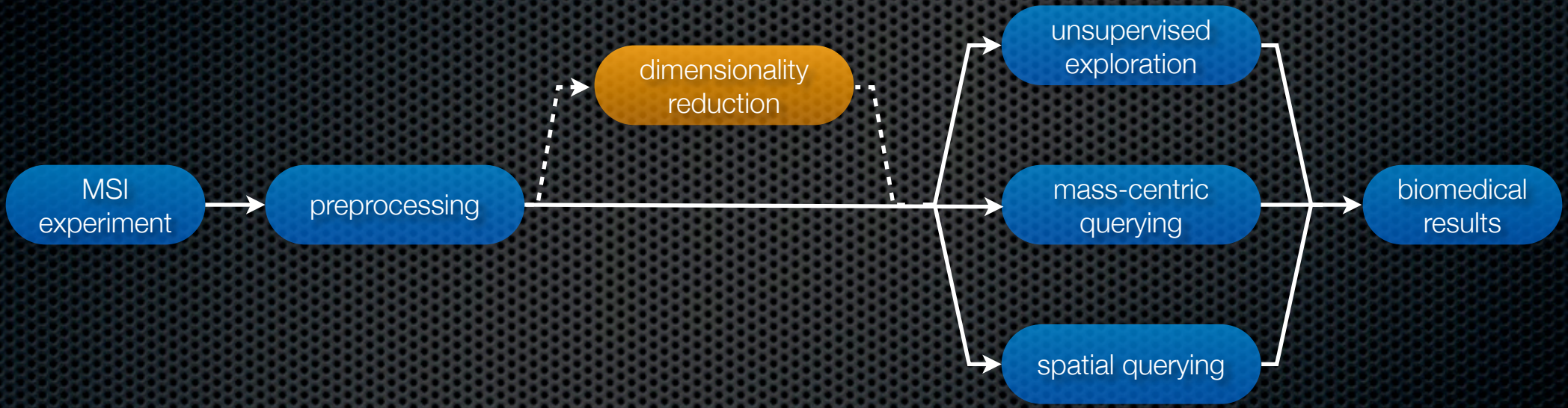


Spatial Querying

Contribution: non-negative least squares approach



Aggregated ion image:
Pulls in other areas in the tissue that exhibit a similar chemical signature to the queried area.


MSI experiment
preprocessing
dimensionality reduction
unsupervised exploration
mass-centric querying
spatial querying
biomedical results

Wavelet approaches

Problem definition

Problem

MSI commonly become very large, making computational analysis from a computational and memory standpoint increasingly difficult. Size is influenced by:

- covered tissue surface area
- spatial resolution
- extent of mass range
- mass resolution

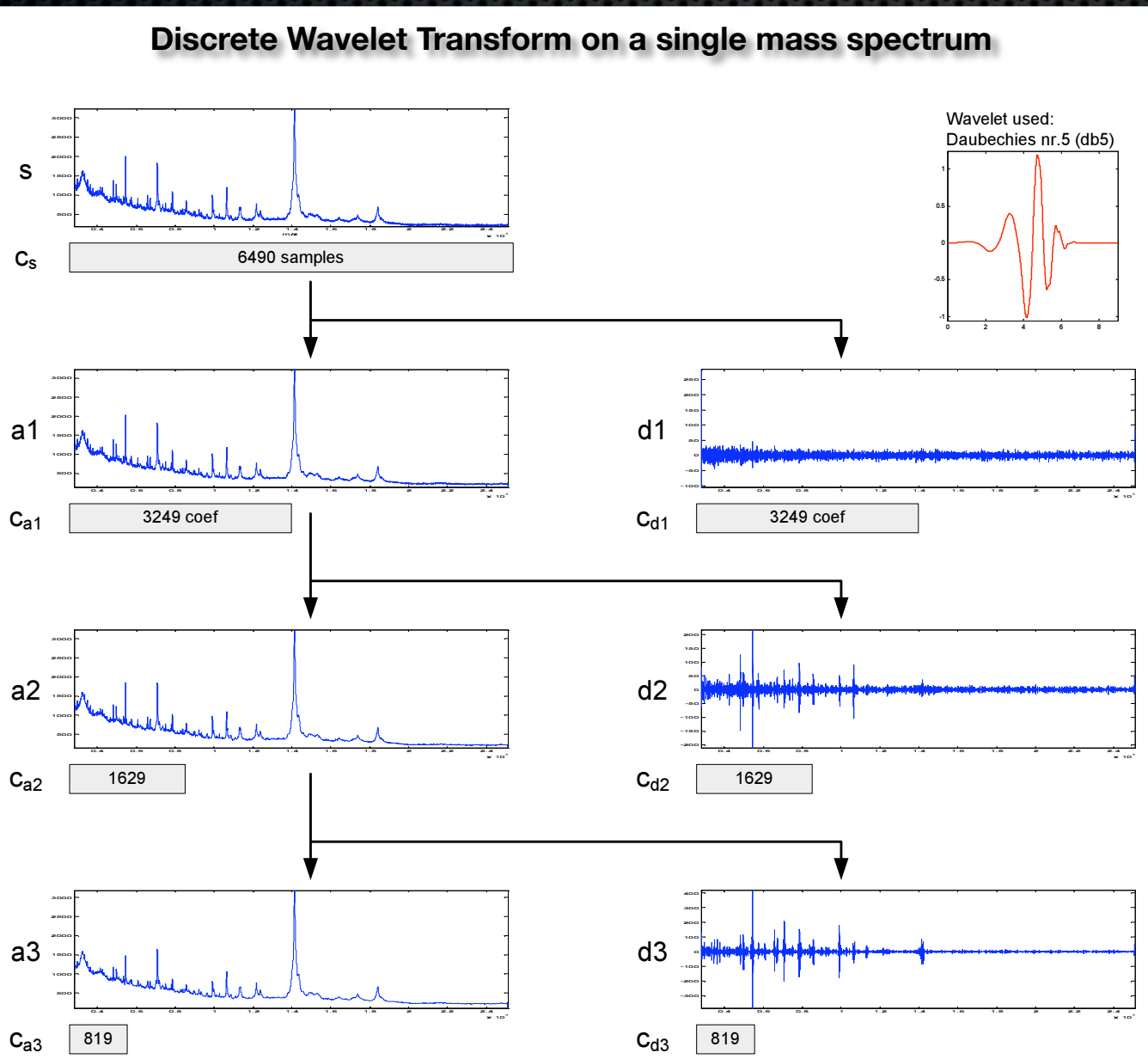
Goal

Provide a strategy for those situations where the multivariate analysis of a MSI data set becomes either

- infeasible due to memory constraints;
- impractical due to exorbitant calculation times or scalability of the algorithm;
- unreliable due to the ‘curse of dimensionality’.

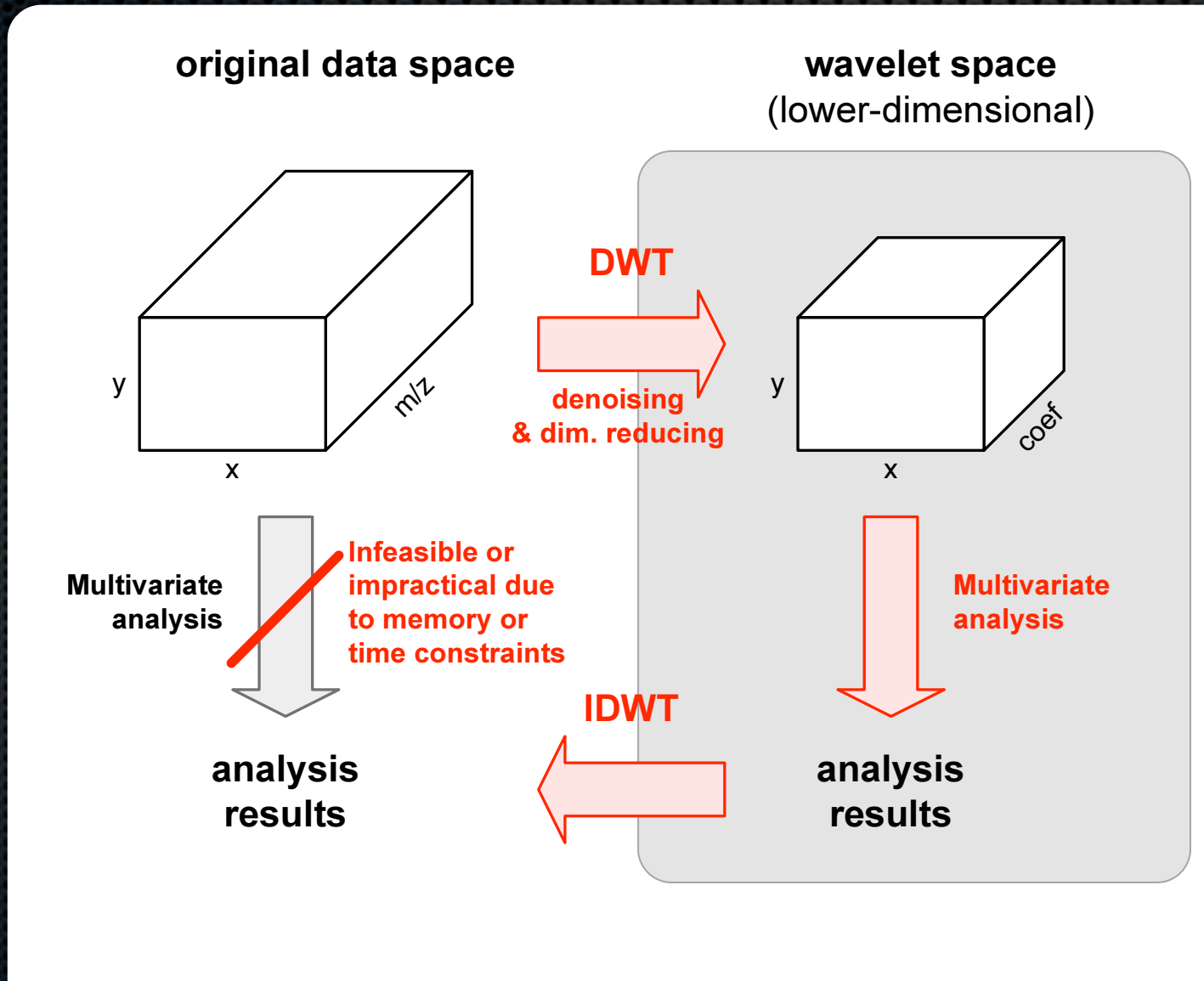
Wavelet approaches

Contribution: multivariate analysis in DWT space



Wavelet approaches

Contribution: multivariate analysis in DWT space

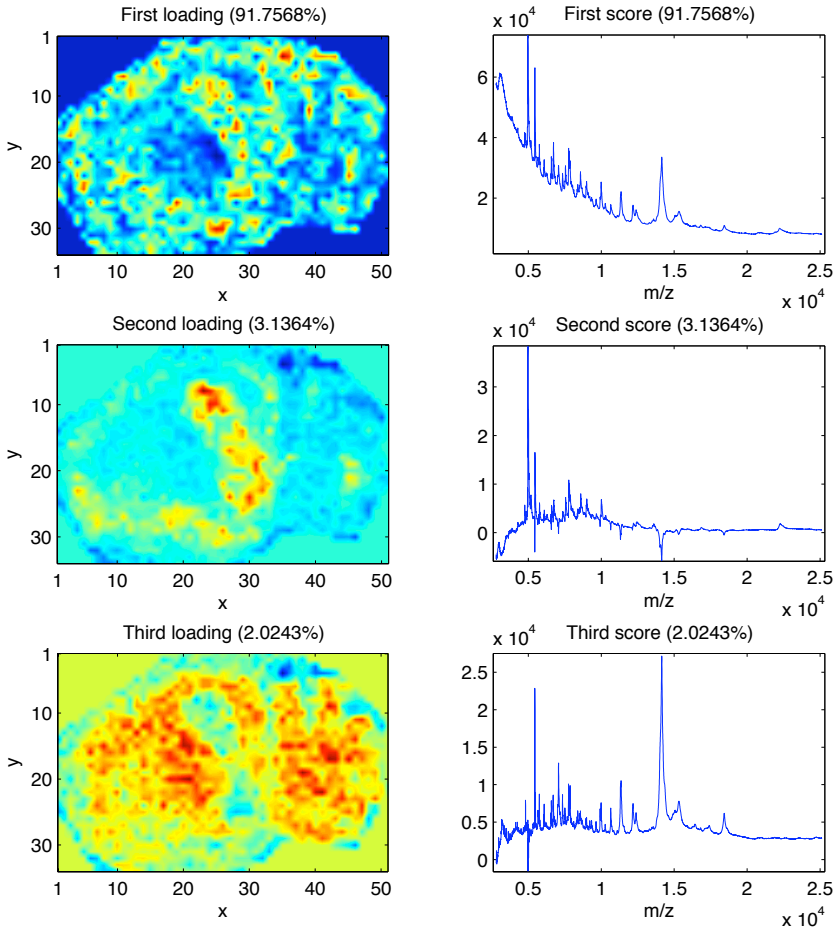


Wavelet approaches

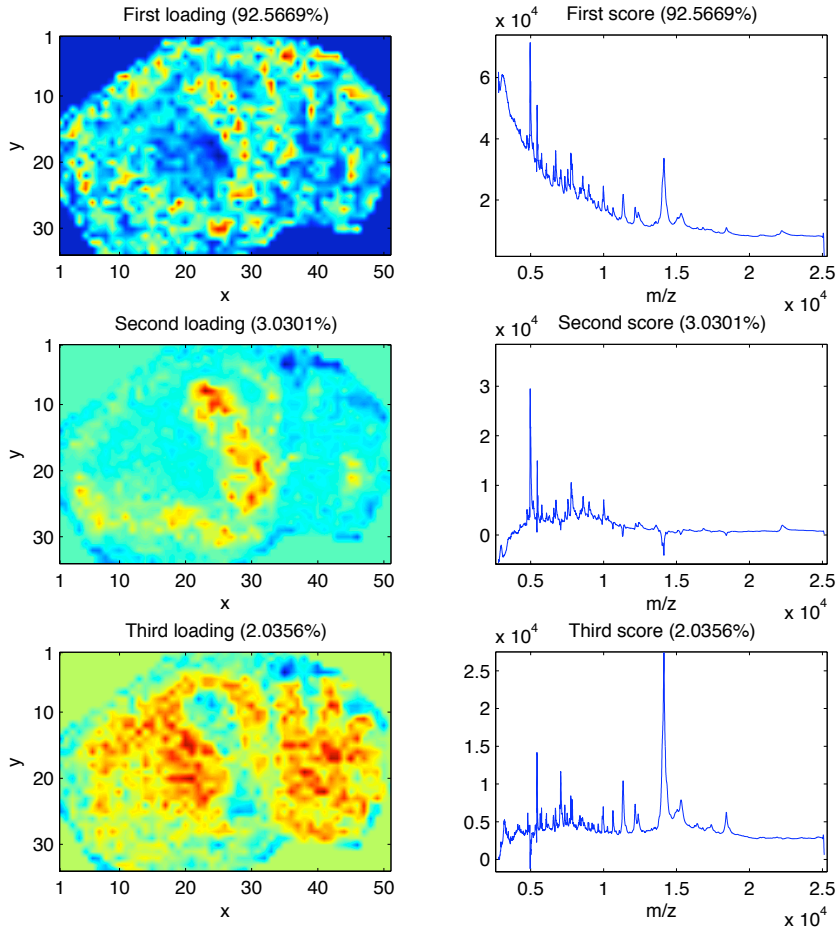
Contribution: multivariate analysis in DWT space

visually indiscernible results

PCA on raw data



PCA on in wavelet space



Wavelet approaches

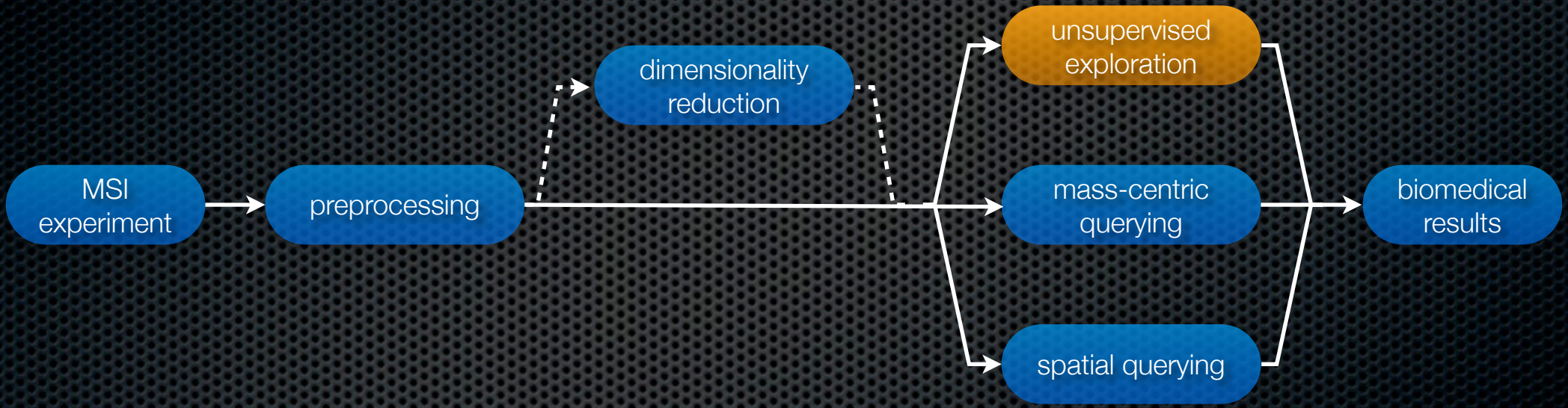
Contribution: multivariate analysis in DWT space

	Non-DWT PCA	DWT PCA
size cov. matrix	m/z bins × m/z bins	coef × coef
required memory	6490 × 6490 (337MB)	819 × 819 (5.3MB)
contrib./cov. entry	# pixels	# pixels

63-fold reduction in memory requirements

	Non-DWT PCA	DWT PCA
size cov. matrix	pixels × pixels	pixels × pixels
required memory	1734 × 1734 (24MB)	1734 × 1734 (24MB)
contrib./cov. entry	# m/z bins	# coefficients
required calc. time	158.3 s	13.2+75.6 s = 88.8 s
	(PCA)	(DWT) + (PCA)

44% reduction in calculation time


MSI experiment
preprocessing
dimensionality reduction
unsupervised exploration
mass-centric querying
spatial querying
biomedical results

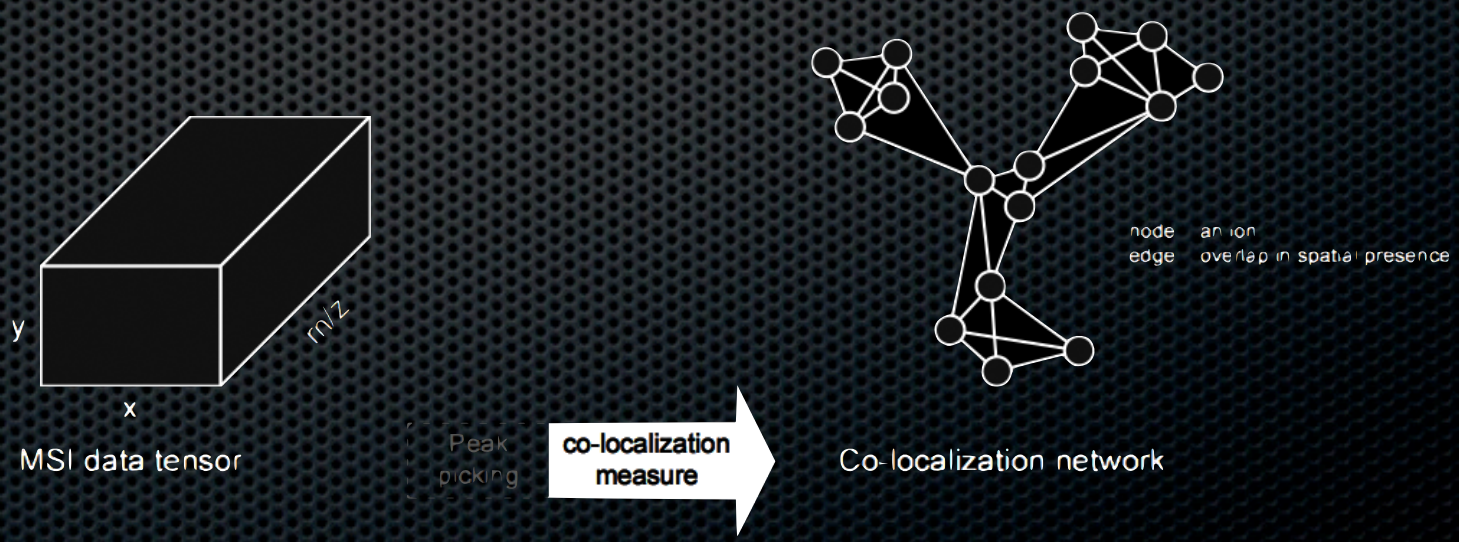
Co-localization networks

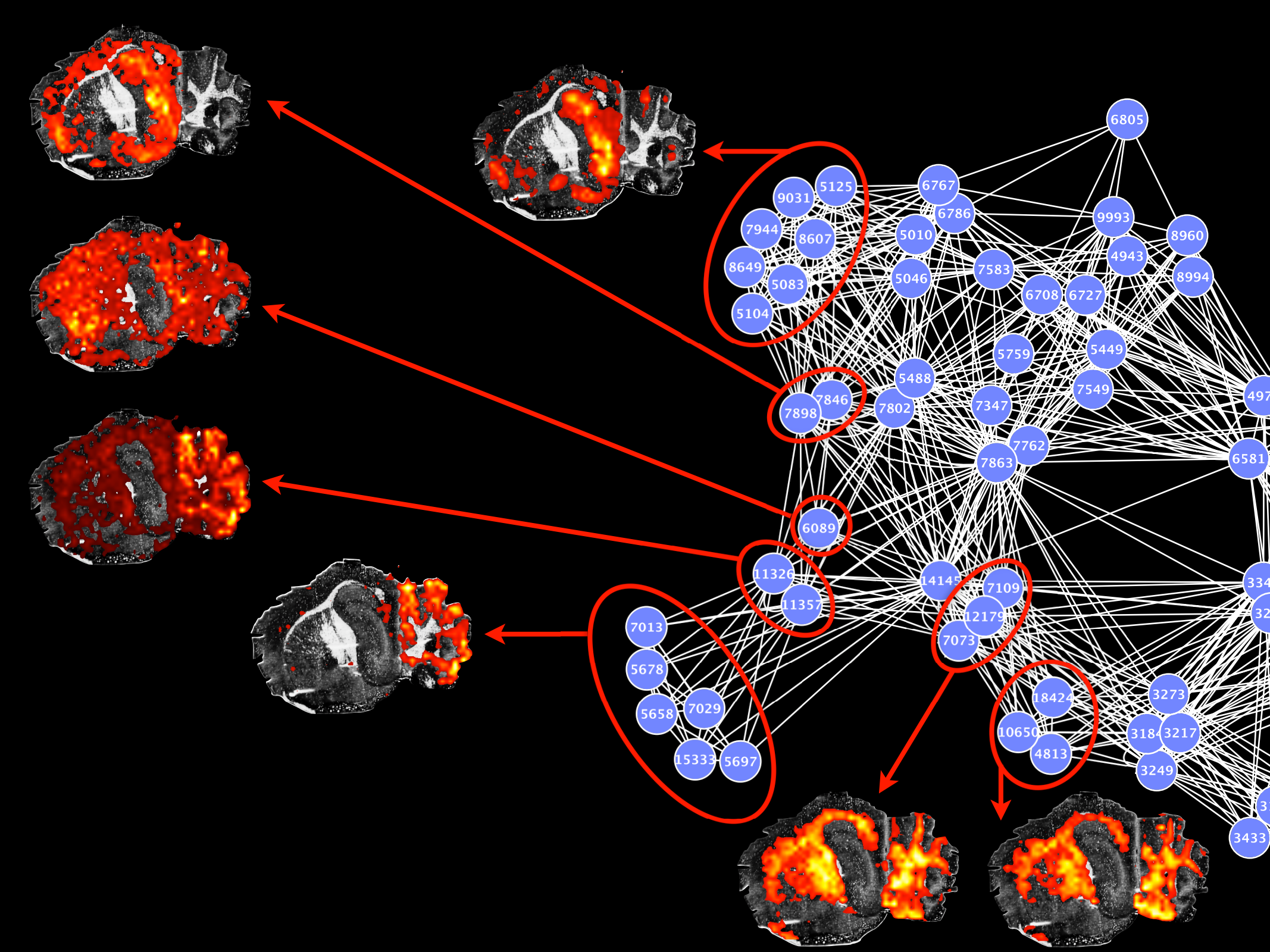
Problem definition

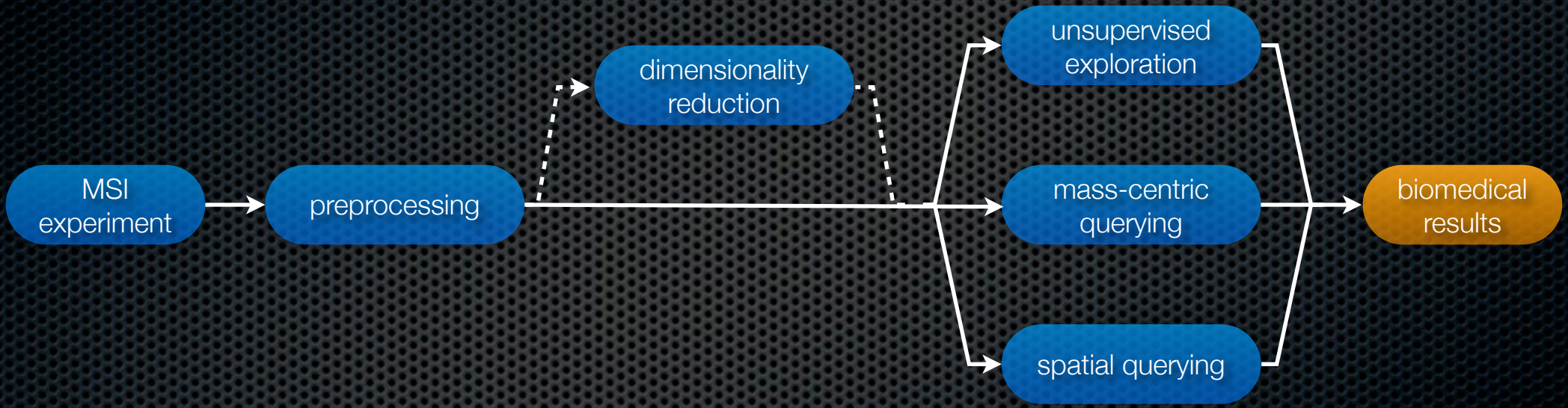
Problem
Most MSI analysis has focused on the imaging aspect of the data. MSI, however, has a dual spatio-biochemical nature due to its strong mass resolution.

Goal
Construct a network representation of the inter-molecule relationships present in a MSI experiment, such as co-localization and correlated expression.

Such a co-localization network can provide **insight into both biological and instrumentally induced relationships between ions**.




6805
9031
5125
6767
6786
9993
8960
7944
8607
5010
4943
8649
5083
5046
7583
8994
6708
6727
5104
5759
5449
5488
7549
7846
7898
7802
7347
7762
7863
6581
6089
11326
11357
14145
7109
12179
7073
7013
5678
5658
7029
15333
5697
18424
3273
10650
4813
3217
3249
3433


MSI experiment
preprocessing
dimensionality reduction
unsupervised exploration
mass-centric querying
spatial querying
biomedical results

Diabetes Case Study

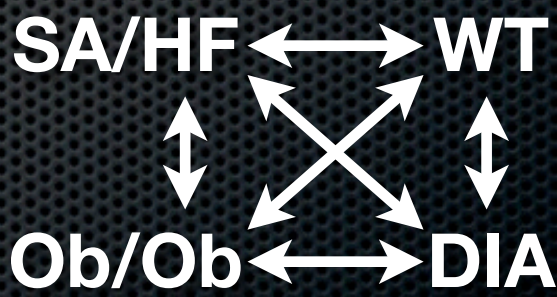
Problem Definition

Problem

Is it possible to detect where diseased metabolic pathways start deviating from normal metabolic function straight from tissue?

Goal

To construct a **metabolic screening method straight from tissue** by combination of MSI and computational multivariate analysis techniques.



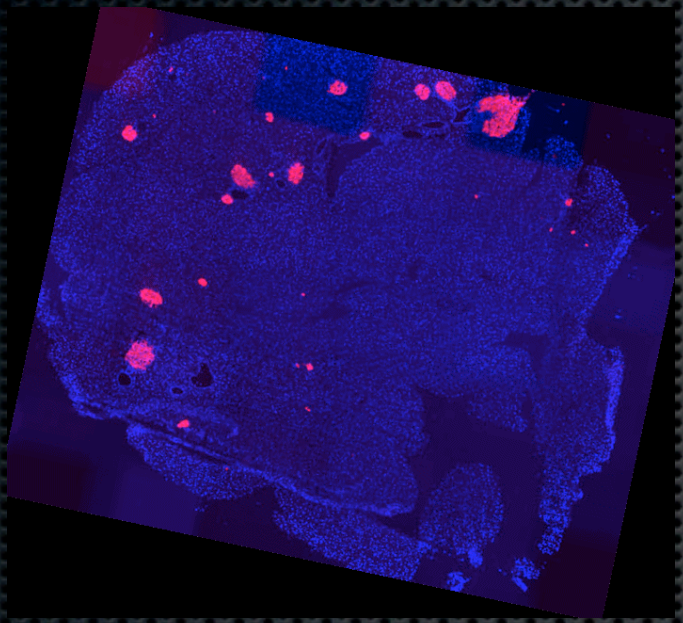
Sub-goals

1. Label-free *in situ* imaging of insulin in tissue;
2. Relative quantification of insulin;
3. Extract multi-molecular content signature for the islets of Langerhans;
4. Differential comparison of islet content;
5. Mapping to metabolic pathways.

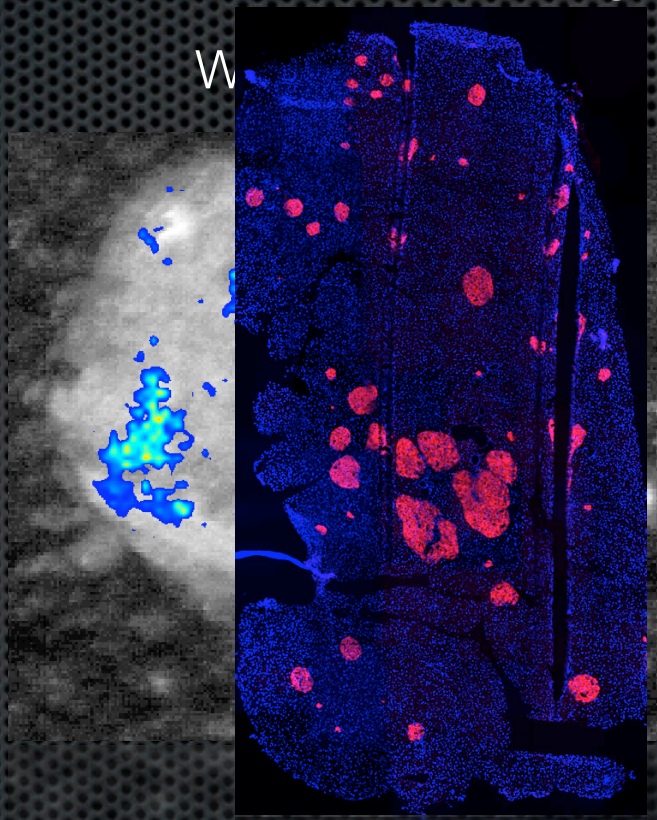
Diabetes Case Study

Metabolic screening straight from tissue

WT immuno-staining

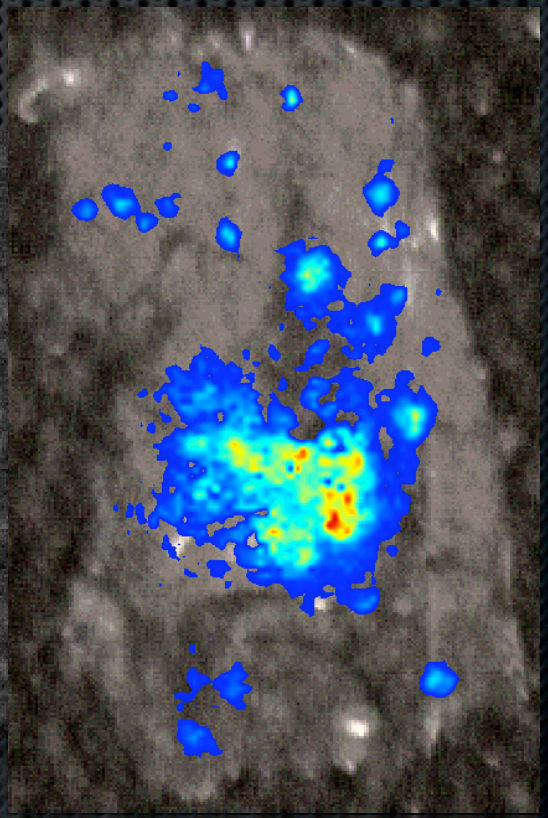


W



Ob/Ob
immuno-staining

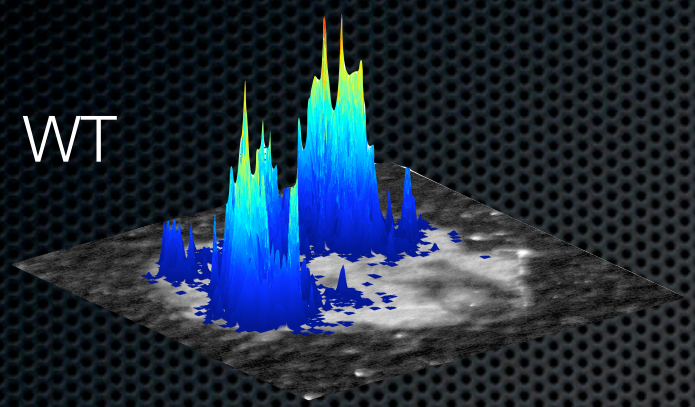
Ob/Ob
ion image

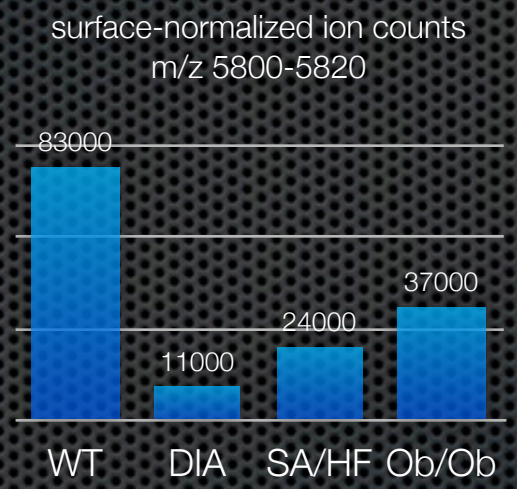


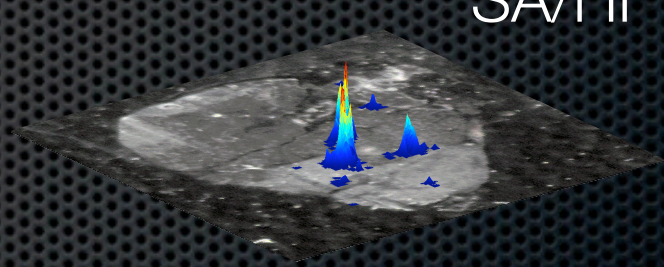
1. Label-free *in situ* imaging of insulin in tissue

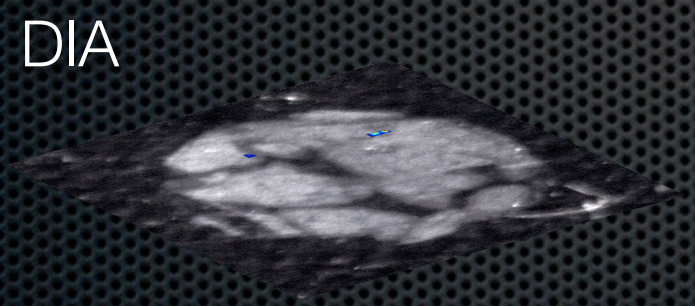
Diabetes Case Study

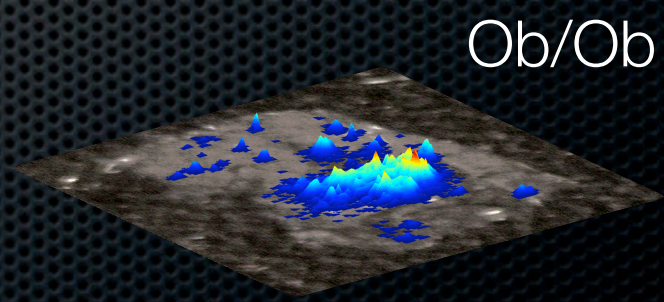
Metabolic screening straight from tissue







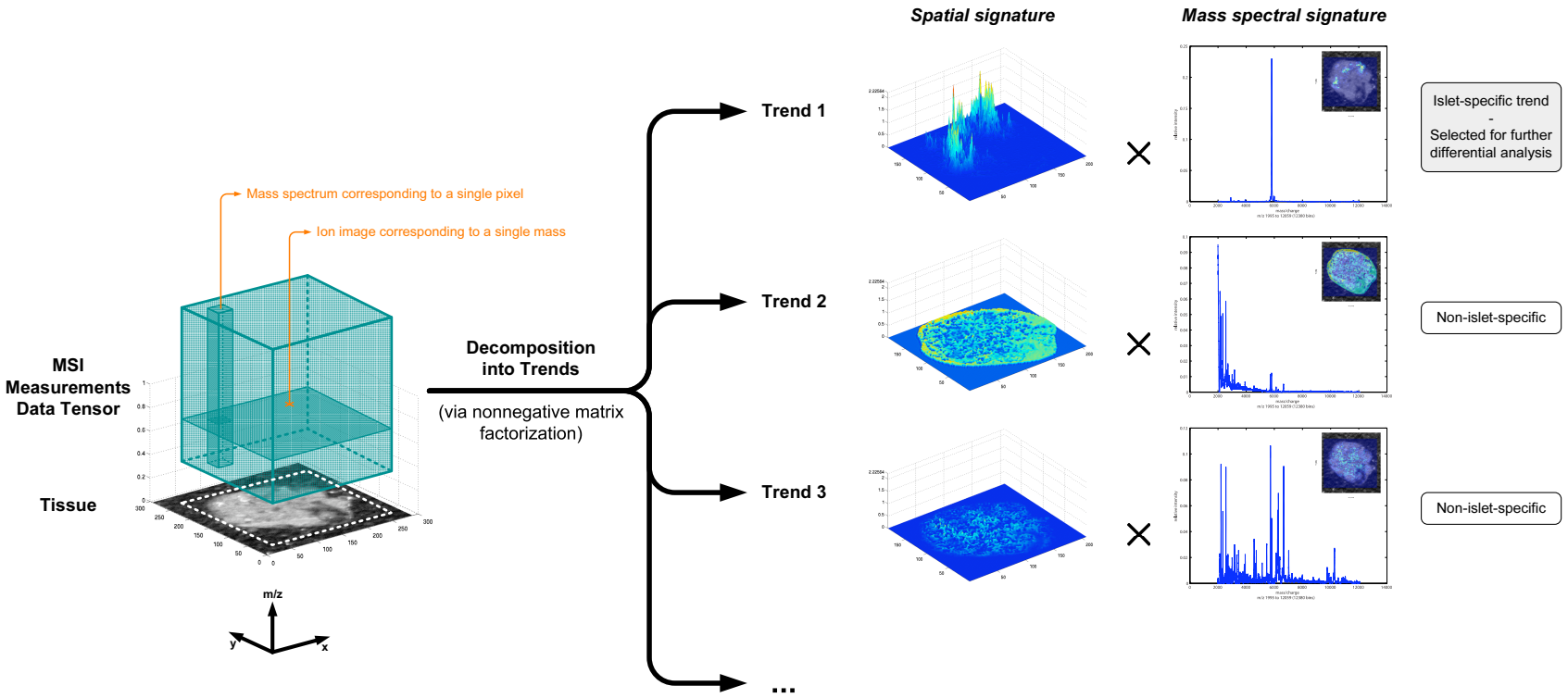




2. Relative quantification of insulin

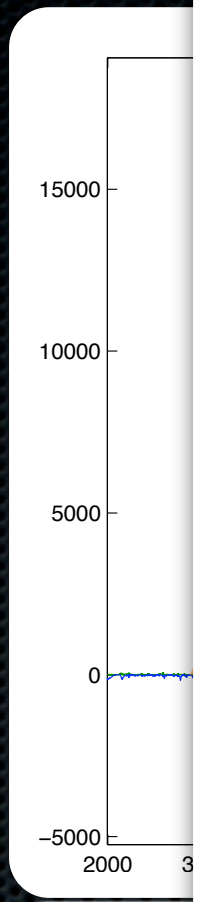
Diabetes Case Study

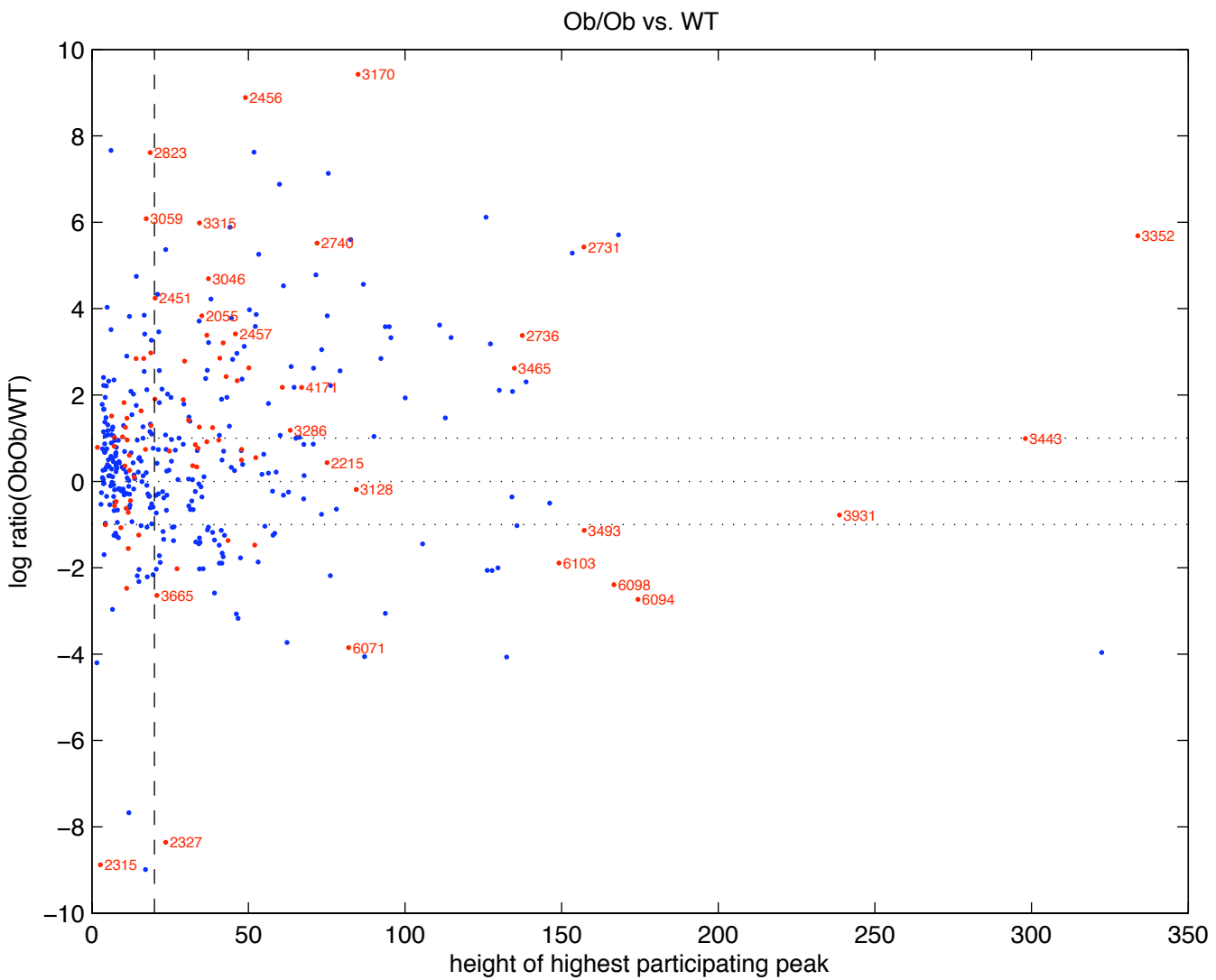
Metabolic screening straight from tissue



3. Extract multi-molecular content signature for the islets of Langerhans

Diabetes Case
Meta

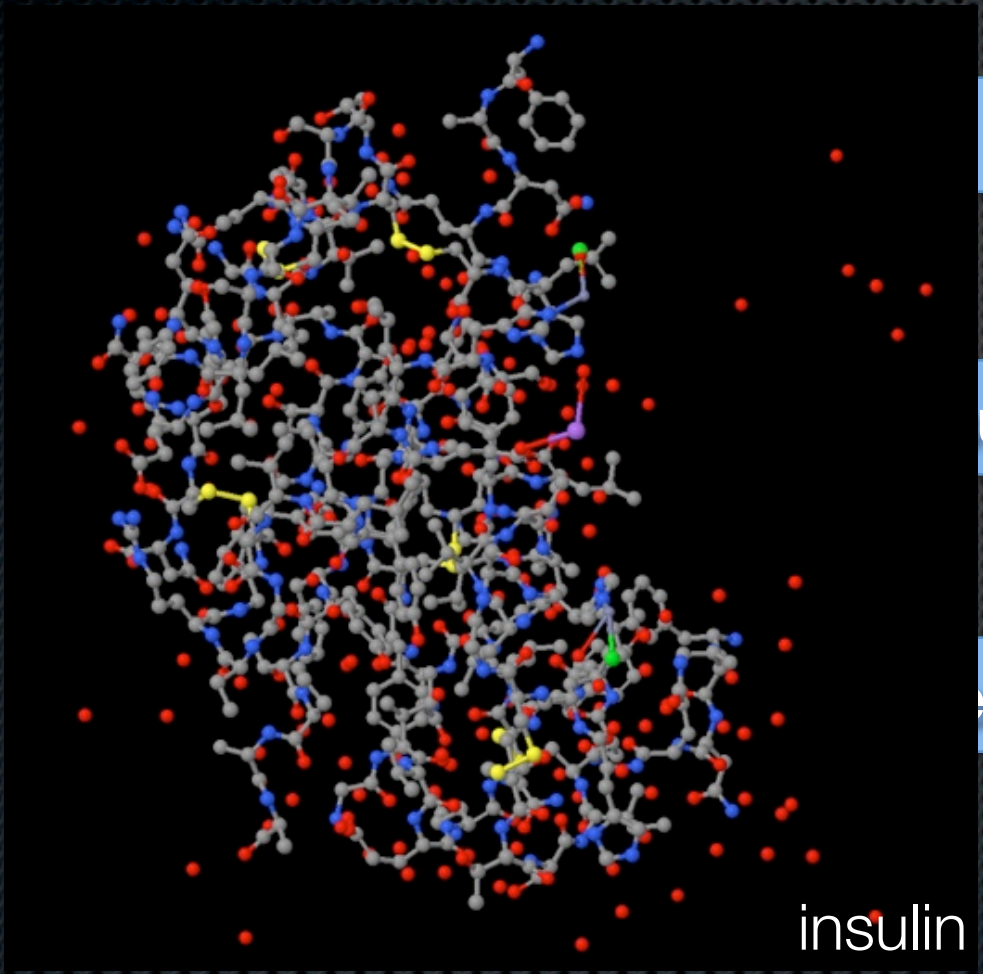


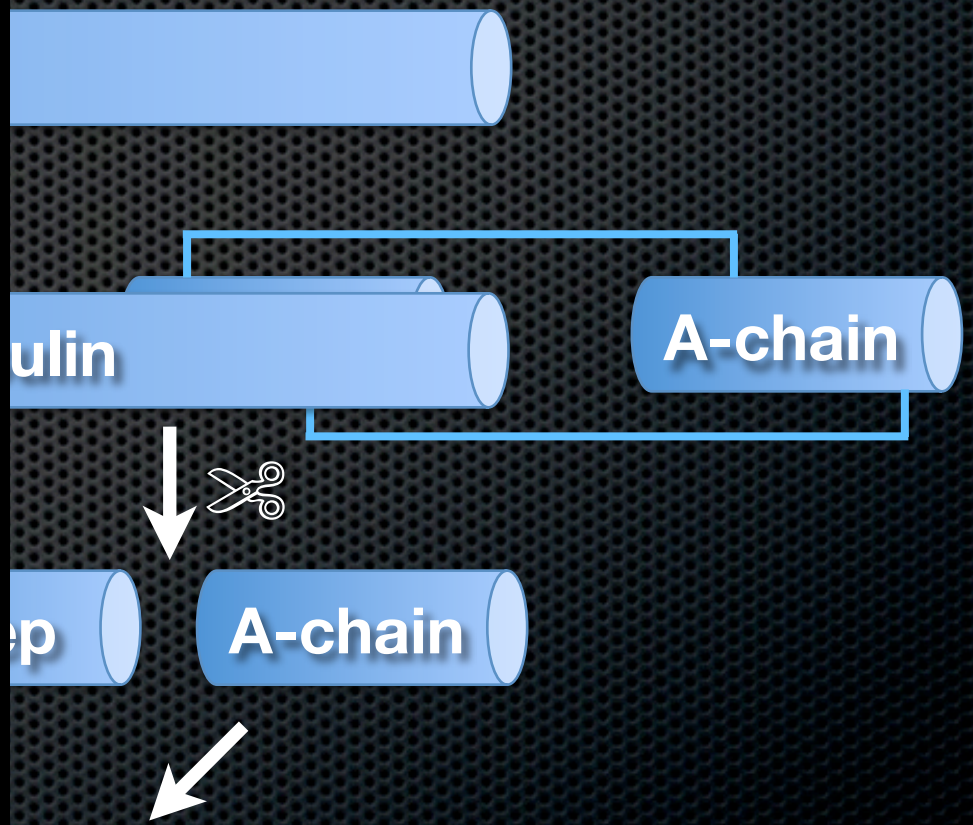


4. Differential comparison of islet content

Diabetes Case Study

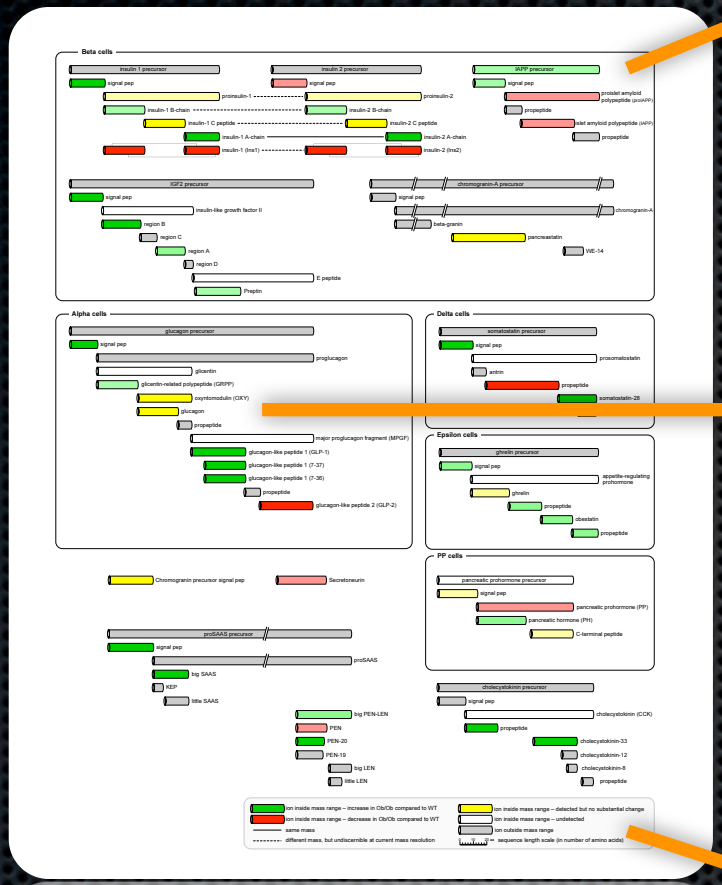
Metabolic screening straight from tissue

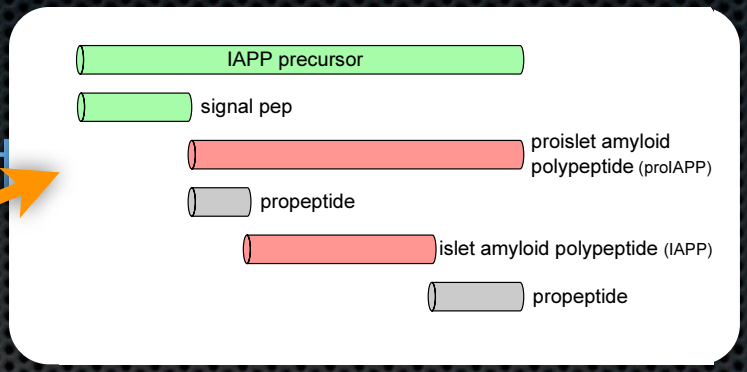


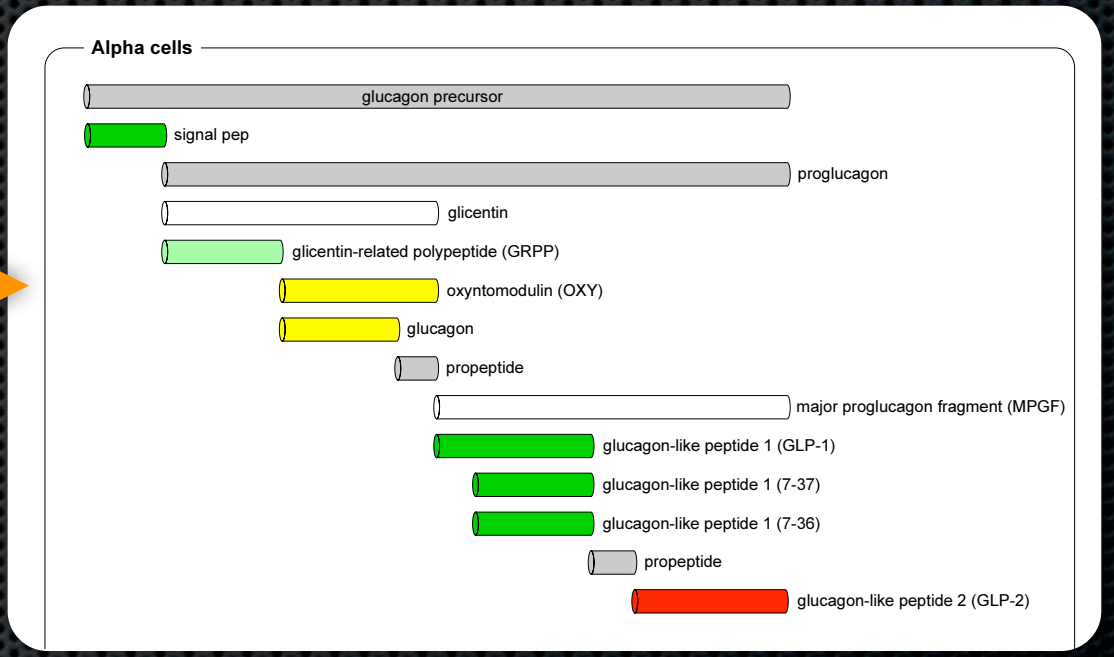


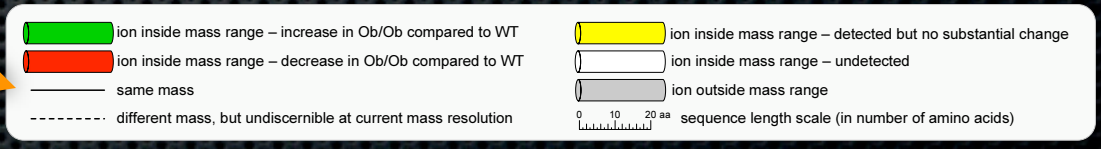
Diabetes Case Study

Metabolic screening straight

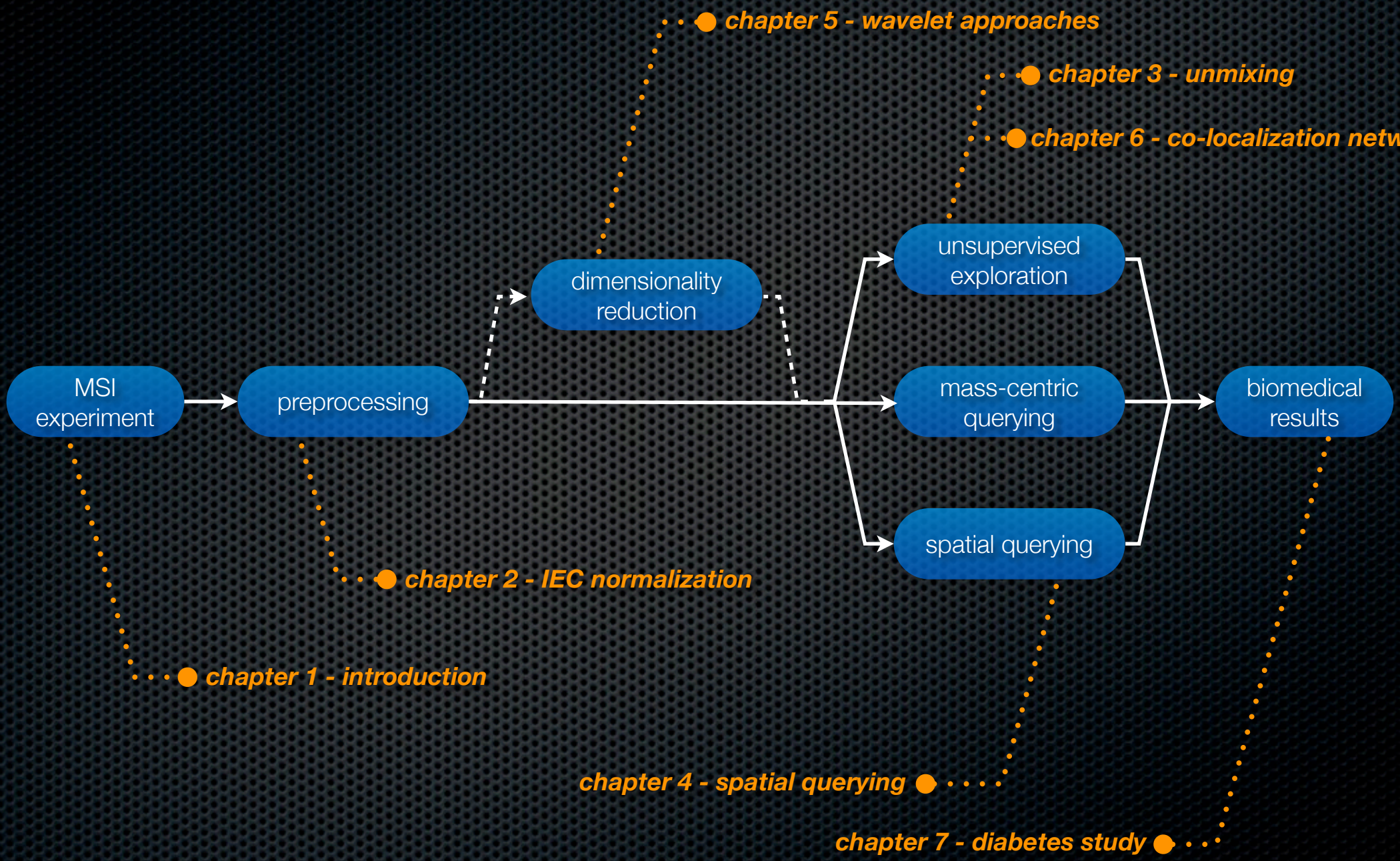








5. Mapping to metabolic pathways


chapter 5 - wavelet approaches
chapter 3 - unmixing
chapter 6 - co-localization netwo
dimensionality reduction
unsupervised exploration
MSI experiment
preprocessing
mass-centric querying
biomedical results
spatial querying
chapter 2 - IEC normalization
chapter 1 - introduction
chapter 4 - spatial querying
chapter 7 - diabetes study

Acknowledgements

K.U.Leuven, Department of Electrical Engineering

Bart De Moor • Johan Suykens • Fabian Ojeda • Kristiaan Pelckmans • Olivier Gevaert • Anneleen Daemen • Sabine Van Huffel • Jan Luts

K.U.Leuven, Department of Molecular Cell Biology

Etienne Waelkens • Rita Derua • Justin Vijay Louis Frans Schuit • Dirk Vander Mierde • Kathleen Lemaire

K.U.Leuven & University Hospitals Leuven, Department Woman and Child

Ignace Vergote • Thomas D'Hooghe • Dirk Timmerman • Isabelle Cadron • Toon Van Gorp • Cleophas Kyama • Attila Mihalyi • Amelie Fassbender

K.U.Leuven & University Hospitals Leuven, Department of Experimental Neurology

Wim Robberecht • Ludo Van Den Bosch • Thomas Philips • Maarten Dewil

Vrije Universiteit Brussel, Department of Pathology

Peter in't Veld

Vlaams Supercomputer Centrum

Kurt Lust

ProMeta

www.prometa.kuleuven.be

•

BioMacS

biomacs.kuleuven.be