

Entwicklung einer flexiblen bioinformatischen Plattform zur Analyse von Massenspektrometriedaten

Sebastian Gibb

Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)
Universität Leipzig

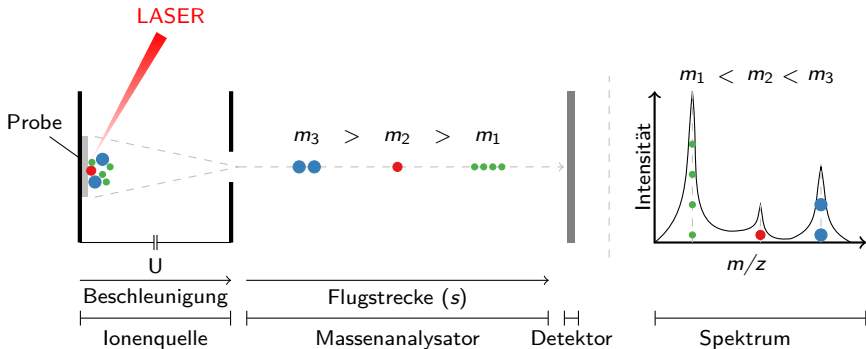
seit Dezember 2014: Klinik für Anästhesiologie, Universitätsmedizin Greifswald

22. Juli 2015

Früherkennung von Erkrankungen



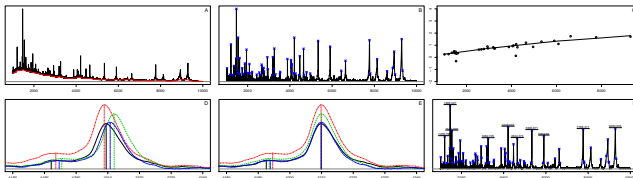
Funktionsweise MALDI/TOF MS



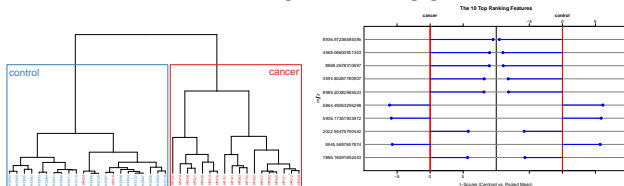
J. H. Gross. *Mass Spectrometry: A Textbook*. Springer, 2004. URL <http://www.springer.com/chemistry/analytical+chemistry/book/978-3-642-10709-2>

*omics Analyse

technologieabhängig



technologieunabhängig



Transparente Software

“The analysis and interpretation of the enormous volumes of proteomic data remains an unsolved challenge, . . .

*Therefore, the development of **transparent tools** for the analysis of proteomic data . . . is a **key challenge**.*

. . .

The development of such proteomics tools is still in its infancy.”

Ruedi Aebersold & Matthias Mann, Nature 2003

Open Source Software/Reproduzierbarkeit

- Reproduzierbarkeit.
- Transparenz/Dokumentation.
- Flexibilität und Erweiterbarkeit.
- Innovationen.
- Verständnis für Daten und Analyse.

Eigene Beiträge

- **MALDIquant**
 - “state-of-the-art” Methoden zur Analyse von 2D-MS Daten.
 - Abbildung individueller Arbeitsabläufe.
 - Behandlung von Spektren unterschiedlicher Auflösung und biologischer/technischer Replikate.
 - Ausführliche Dokumentation und Beispielanalysen.
 - Automatische Tests.
- **MALDIquantForeign**
 - Import von 11 verschiedenen Dateiformaten.
 - Export in 5 verschiedene Dateiformate.
 - `readBrukerFlexData` und `readMzXmlData`.

Funktionen

MALDIquant ist freie Software (GPLv3).

Funktionen in **MALDIquant** 1.12:

- Intensity Transformation.
- Intensity Smoothing.
- Baseline Correction.
- Peak Detection.
- Warping/Peak Alignment.
- Peak Binning.
- Peak Filtering.
- Calibration.
- Multiple plotting methods.
- Peak Labeling.
- Handling biological/technical replicates.
- Handling different resolutions.
- Merging mass spectra/peaks.
- Handling of MSI data.
- Fast (parallel support).
- Modular: easy to customize.

Anwendungsbeispiel Fiedler et al. 2009

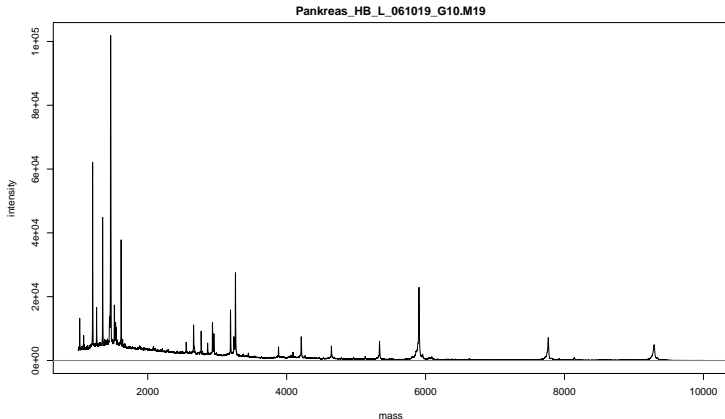
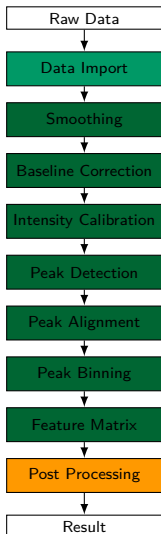
Serum Peptidome Profiling Revealed Platelet Factor 4 as a Potential Discriminating Peptide Associated with Pancreatic Cancer

G.M. Fiedler, A.B. Leichtle, J. Kase et al
Clin Cancer Res June 1, 2009 15:3812-3819

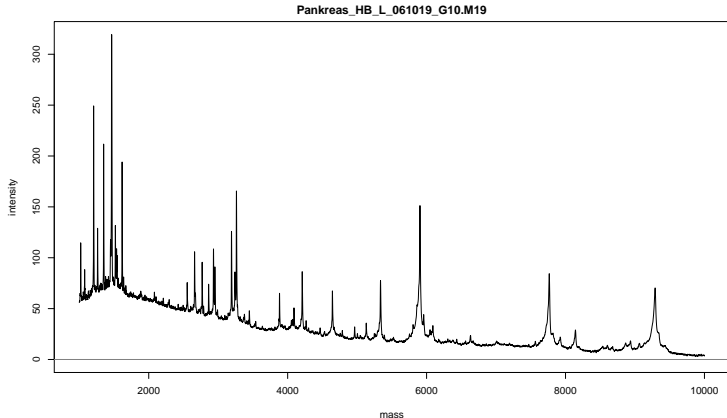
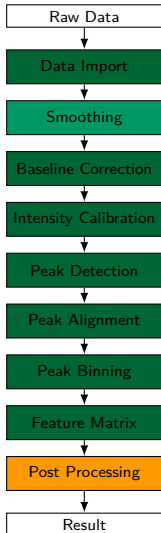
*“Two significant peaks (**m/z 3884**; 5959) achieved a sensitivity of 86.3% and a specificity of 97.6% for the discrimination of patients and healthy controls . . . ”*

*“MALDI-TOF MS-based serum peptidome profiling allowed the discovery and validation of platelet factor 4 [**m/z 3884, 7767**; S.G.] as a new discriminating marker in pancreatic cancer.”*

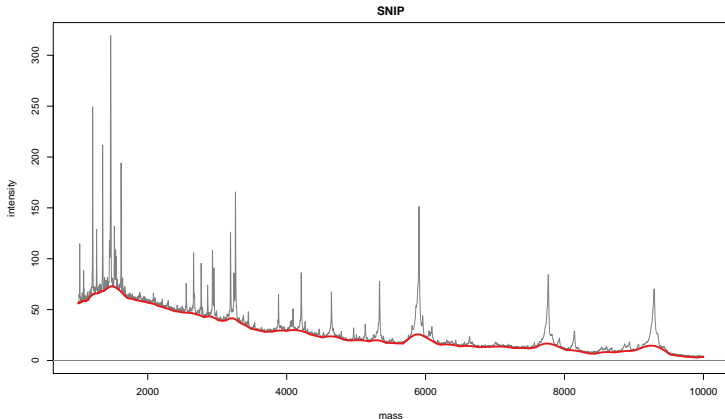
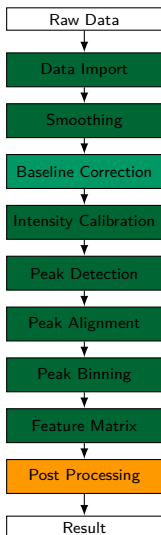
Import der Rohdaten



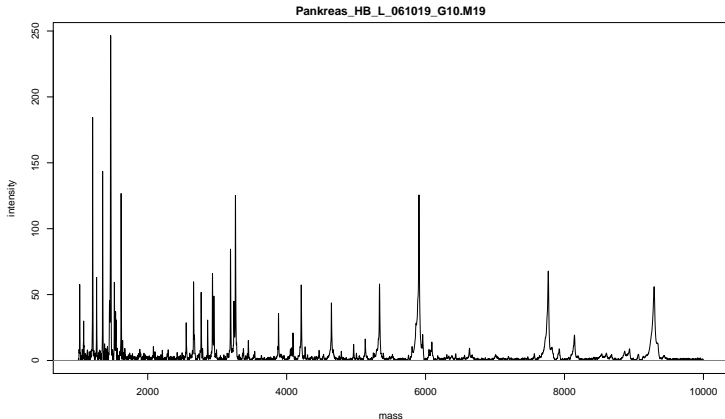
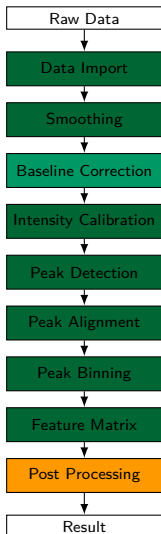
Transformation der Intensitäten



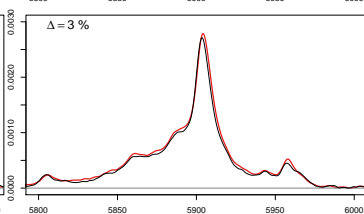
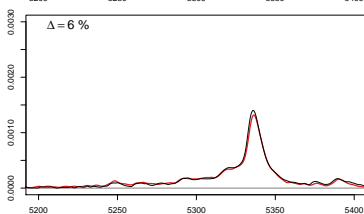
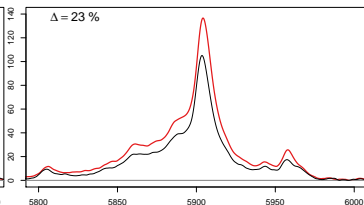
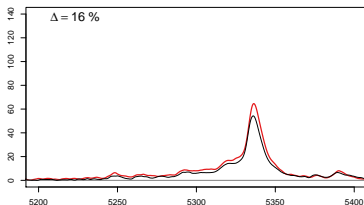
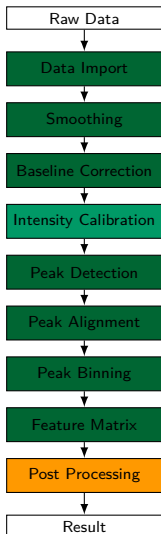
Korrektur der Grundlinie



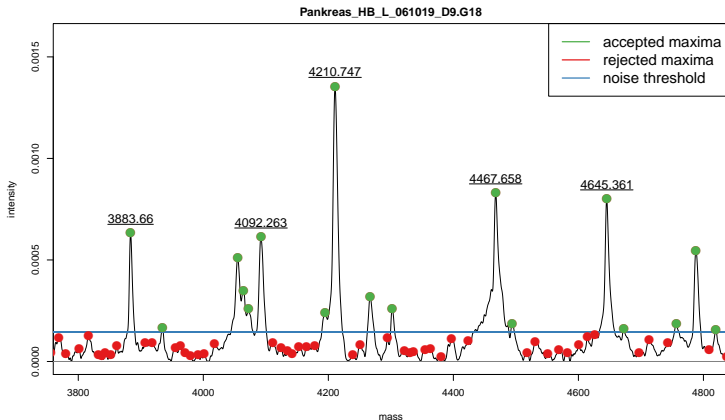
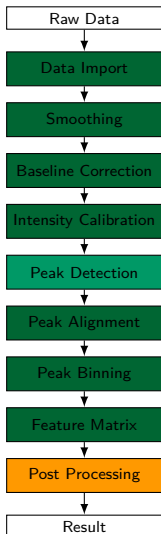
Korrektur der Grundlinie



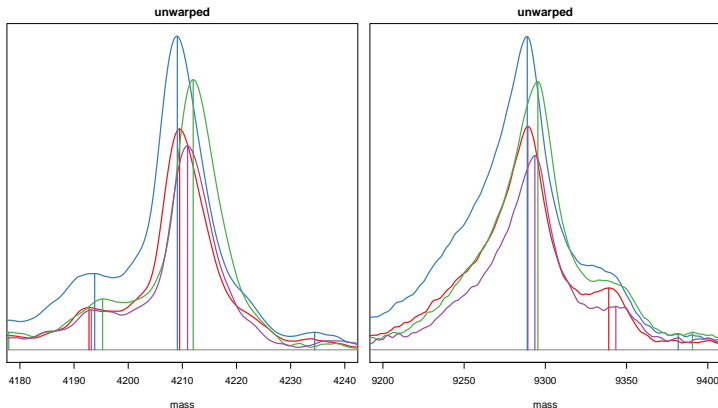
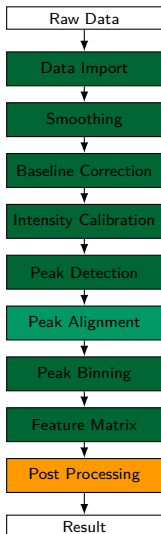
Kalibrierung der Intensitäten



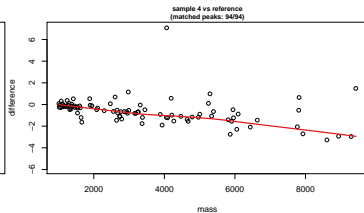
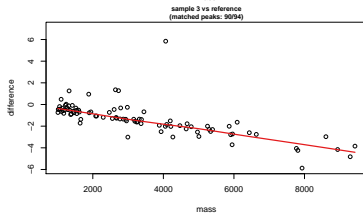
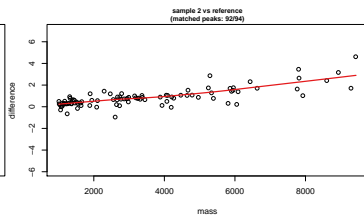
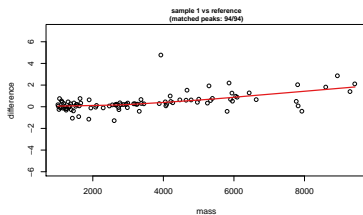
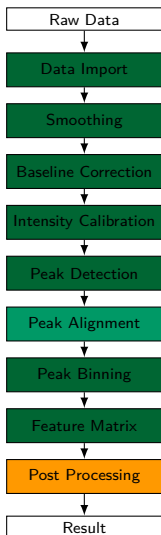
Identifizierung von Merkmalen



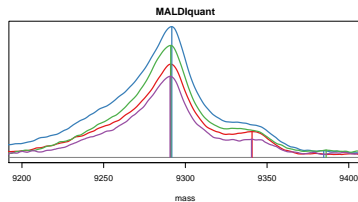
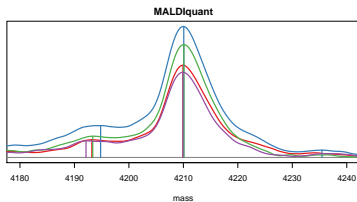
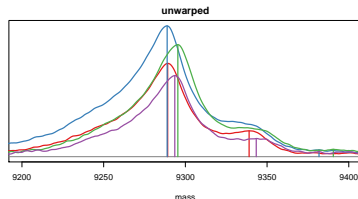
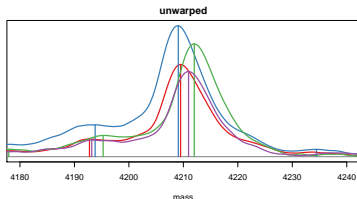
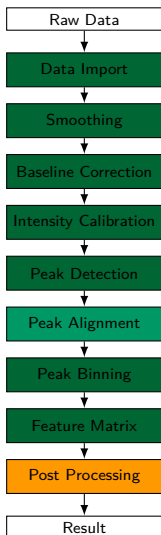
Kalibrierung der m/z -Werte



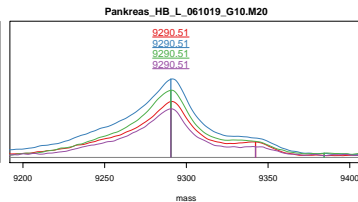
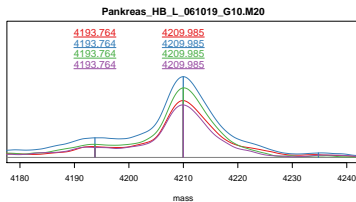
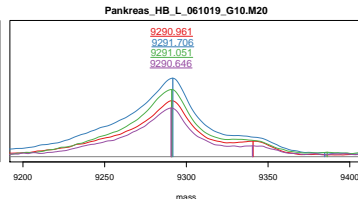
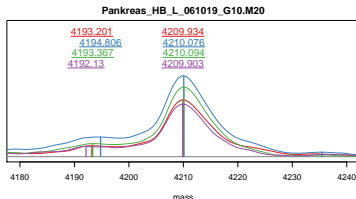
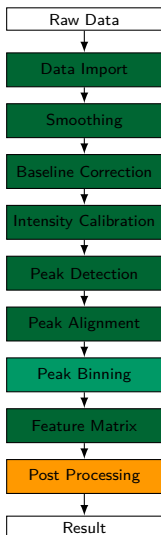
Kalibrierung der m/z -Werte



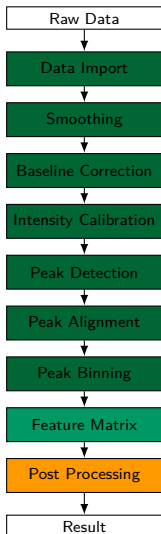
Kalibrierung der m/z -Werte



Kalibrierung der m/z -Werte



Feature Matrix



	1011.95683040433	1020.93287697588	
HC056	0.0001662288	0.0010911416	...
HC001	0.0001865509	0.0007400781	...
HC002	0.0001690267	0.0006584539	...
HC008	0.0001516142	0.0002370617	...
HC011	0.0001514336	0.0004122820	...
HC033	0.0001457631	0.0006348643	...
HC049	0.0001615118	0.0004522258	...
HC050	0.0001637069	0.0011229157	...
HC054	0.0002076213	0.0005488659	...
HC055	0.0001860750	0.0005867420	...
HC122	0.0001261214	0.0005200141	...
HC057	0.0001386024	0.0003449330	...
...

Arbeitsablauf in R

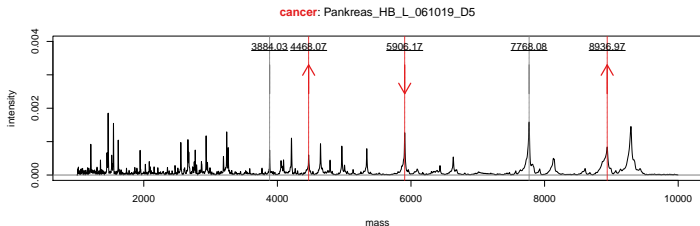
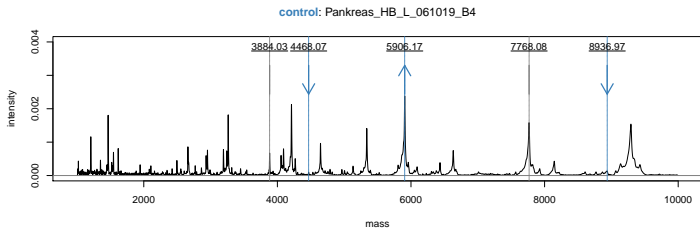
```
1 library("MALDIquant")
2 library("MALDIquantForeign")
3
4 spectra <- import("fiedler2009spectra.tar.gz")
5
6 spectra <- transformIntensity(spectra, method="sqrt")
7 spectra <- smoothIntensity(spectra, method="SavitzkyGolay")
8 spectra <- removeBaseline(spectra, method="SNIP")
9 spectra <- calibrateIntensity(spectra, method="TIC")
10
11 peaks <- detectPeaks(spectra)
12
13 warpingFunctions <- determineWarpingFunctions(peaks)
14 spectra <- warpMassSpectra(spectra, warpingFunctions)
15 peaks <- warpMassPeaks(peaks, warpingFunctions)
16 peaks <- binPeaks(peaks)
17
18 featureMatrix <- intensityMatrix(peaks, spectra)
```

Ergebnisse/Klassifizierung

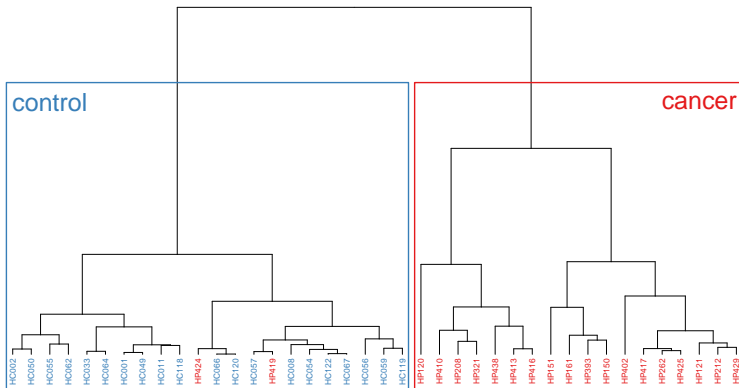


M. Ahdesmäki and K. Strimmer. Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. *The Annals of Applied Statistics*, 4 (1):503–519, Mar 2010. doi: 10.1214/09-AOAS277. URL <http://dx.doi.org/10.1214/09-AOAS277>

Ergebnisse/Spektrenvergleich



Ergebnisse/Clustering

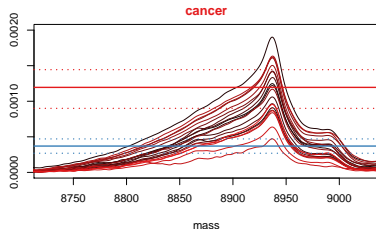
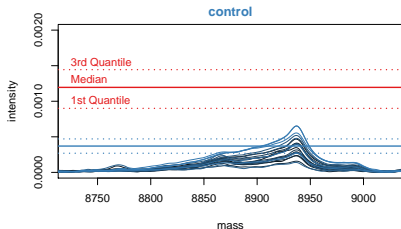


Ergebnisse/Biologische Relevanz

- Complement C3 (CO3_HUMAN)
- Pancreatic Progenitor Cell Differentiation and Proliferation Factor-Like Protein (PDPFL_HUMAN)

Komplette Analyse unter:

<http://strimmerlab.org/software/malDIquant/>



Publikation

MALDIquant: a versatile R package for the analysis of mass spectrometry data

S. Gibb, K. Strimmer - *Bioinformatics*, 2012

> 50 Publikationen:

Antibiotikaresistenzen von Bakterien, Speziesbestimmung (Bakterien, Insekten), MSI, Profiling/Früherkennung von Krankheiten, Aktivierung des Immunsystems, Spiegelbestimmung von Medikamenten, ...

Software: MSnbase, MSI.R, Mass-Up

Verfügbarkeit: CRAN, RforProteomics, MASSyPup, Debian, Ubuntu

Zusammenfassung

MALDIquant

- Freie, transparente Software für 2D-MS Daten.
- Reproduzierbare Analysen.
- Ausführliche Dokumentation und Beispielanalysen.
- Flexible Anwendungsmöglichkeiten.
- Vielfältige Verwendung.

Kontakt:

mail@sebastiangibb.de

MALDIquant Software:

<http://strimmerlab.org/software/malDIquant/>



Danksagung

G. M. Fiedler und A. B. Leichtle: Diskussionen, Beispieldaten

(Universitätsinstitut für Klinische Chemie, Universitätsspital Bern)

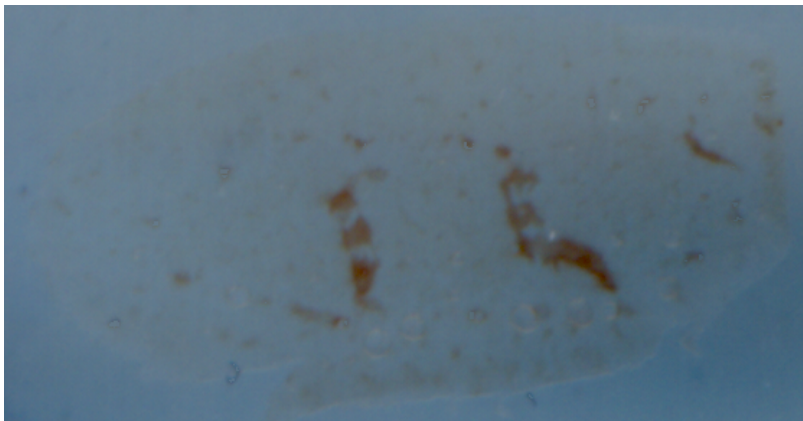
Katrin Uhlmann und Ralph Feltens: Diskussionen, Beispieldaten

(Department Proteomik, Helmholtz-Zentrum für Umweltforschung (UFZ) Leipzig)

Korbinian Strimmer: Betreuung der Disseration

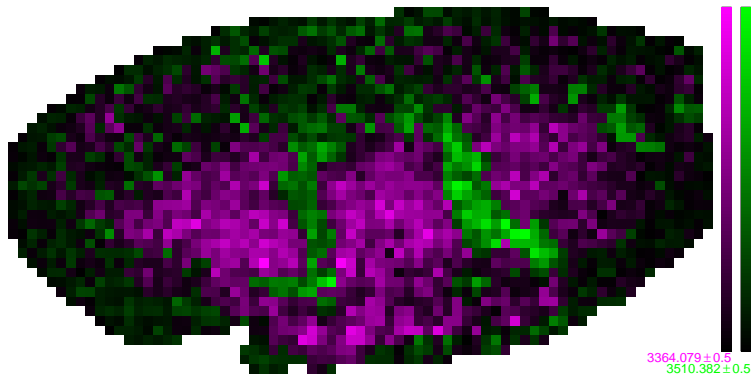
(Epidemiology and Biostatistics, School of Public Health, Imperial College London)

Mass Spectrometry Imaging



This dataset was kindly provided by Dr. Adrien Nyakas (adrien.nyakas@dcu.unibe.ch; <http://dx.doi.org/10.6084/m9.figshare.735961>).

Mass Spectrometry Imaging



This dataset was kindly provided by Dr. Adrien Nyakas (adrien.nyakas@dcb.unibe.ch; <http://dx.doi.org/10.6084/m9.figshare.735961>).