Analysis of Proteomics Data using MALDIquant

Sebastian Gibb
Institute for Medical Informatics, Statistics and Epidemiology (IMISE)

University of Leipzig

17. August 2011

Table of Contents

Predicting disease

Proteomics

Proteomics

- Study of the entirety of proteins produced by an organism.
- Foci: identification, structure determination, biomarker, pathways, expression.

Mass Spectrometry

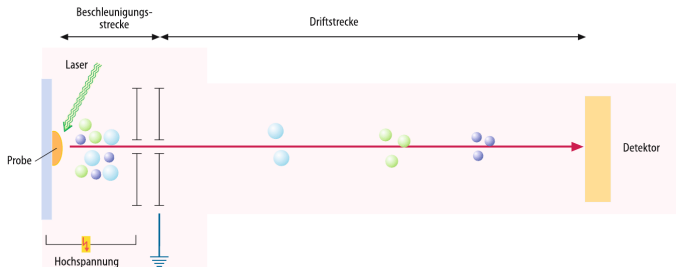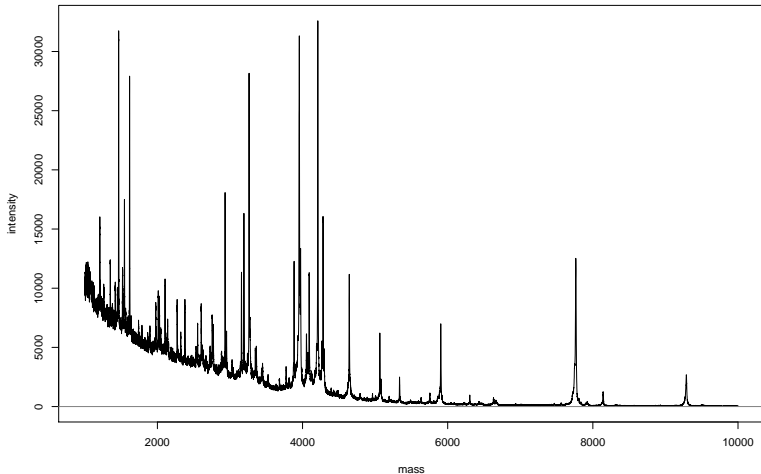| Ion Source: MALDI	Mass Analyzer: TOF	Detector
Matrix-Assisted Laser Desorption/Ionization	Time Of Flight ($t \propto \sqrt{\frac{m}{q}}$)	Quantity Measurement

Abb. 3.14.; S. 67; "Biochemie & Pathobiochemie", Löffler G., 8. Auflage (2007), Springer Medizin Verlag
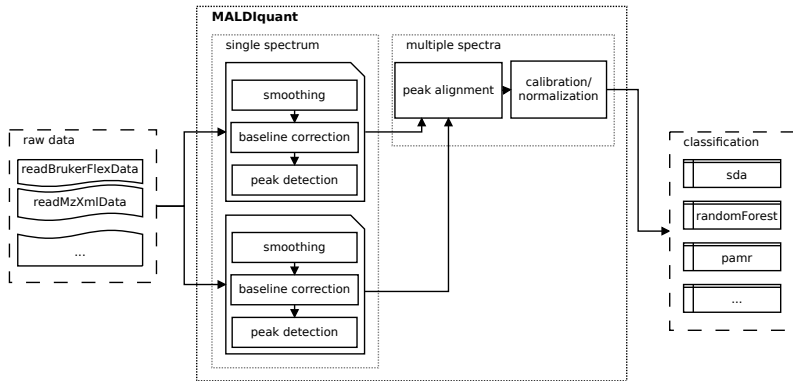
MALDI-TOF Example Spectrum

MALDIquant
Motivation

- Only relatively few open source software solutions available and very few for the R platform.
- No MALDI-TOF package fitting our needs for clinical diagnostics.
- Necessity of handling both technical and biological replicates.
- Unsatisfying quantification of relative intensities (total-ion-current, 0/1)
- Investigation of impact of calibration of spectra on clinical prognosis.
- Modular and easy to customize analysis routines.

MALDIquant
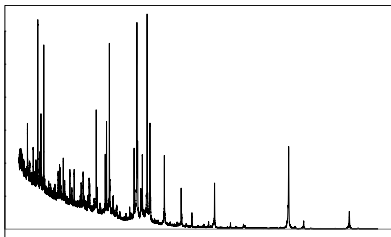
The UNIX philosophy: do one thing and do it well McIlroy (1978)

MALDIquant

The UNIX philosophy: do one thing and do it well McIlroy (1978)

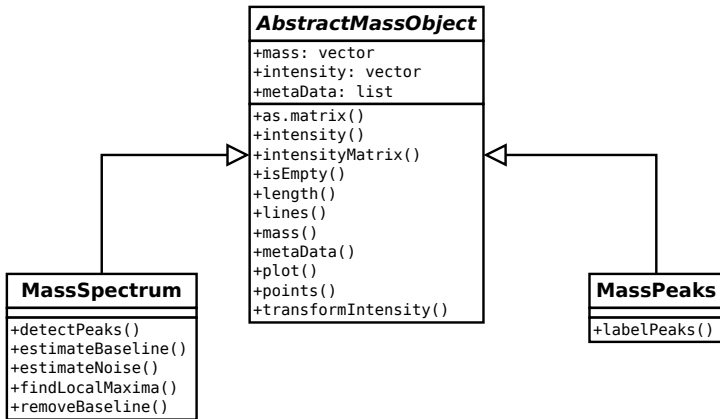

$$\Rightarrow \quad \begin{pmatrix} f_{1,1} & f_{1,2} & f_{1,3} & \ldots & f_{1,n} \\ f_{2,1} & f_{2,2} & f_{2,3} & \ldots & f_{2,n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ f_{m,1} & f_{m,2} & f_{m,3} & \ldots & f_{m,n} \end{pmatrix}$$

MALDIquant
Structure

object-oriented, S4

Example Data

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."*

MALDIquant
hands-on: File Import

```
> library("MALDIquant")
> library("readBrukerFlexData")
> spectra <- mqReadBrukerFlex("/data/fiedler2009/")
> length(spectra)
[1] 480
> spectra[[1]]
S4 class type            : MassSpectrum
Number of m/z values     : 42388
Range of m/z values      : 1000.015 - 9999.734
Range of intensity values: 5 - 101840
File                     : /data/fiedler2009/[...]/Pankreas_HB_L_061019_G10/0_m19/1/1SLin/fid
```
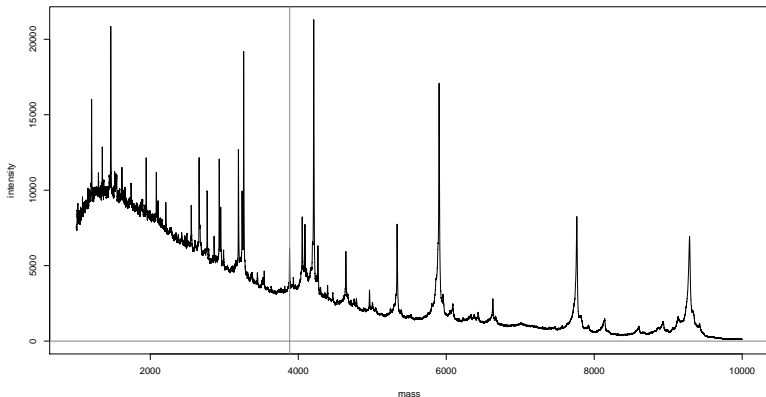
other possibilities:

```
library("MALDIquant")
library("readMzXmlData")
s <- mqReadMzXml("/data/exampleMS/spectrum.mzXML")


library("MALDIquant")
s <- createMassSpectrum(mass=1:5,
                        intensity=runif(5),
                        metaData=list(name="example"))
```
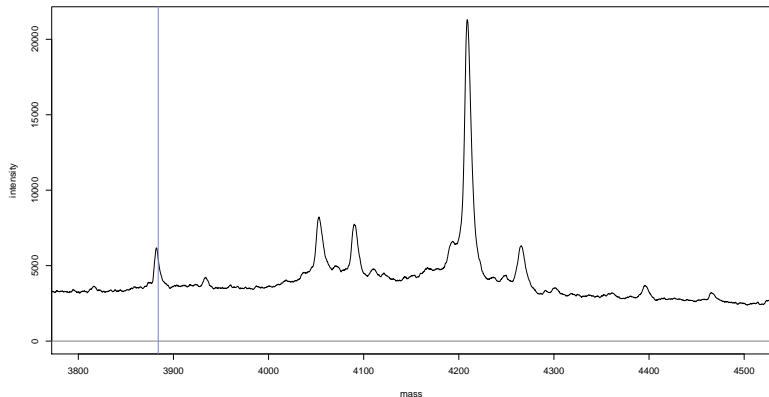
MALDIquant
hands-on: plot

```
> plot(spectra[[1]])
> abline(v=3884, col="blue")
```
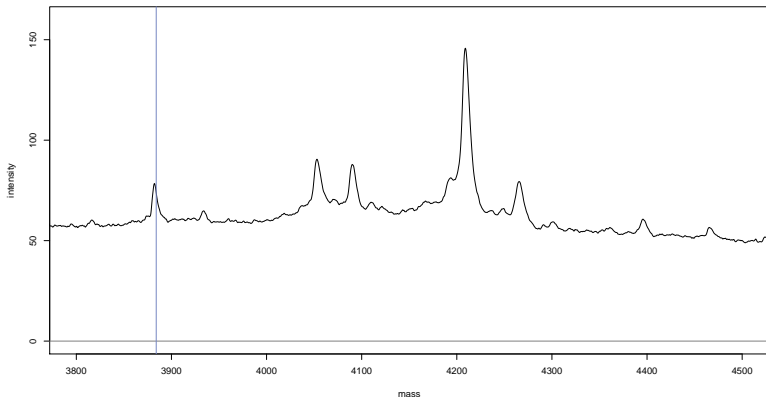
MALDIquant

hands-on: plot m/z 3884

```
> plot(spectra[[1]], xlim=c(3800, 4500))
> abline(v=3884, col="blue")
```

MALDIquant
hands-on: Variance Stabilization and Smoothing

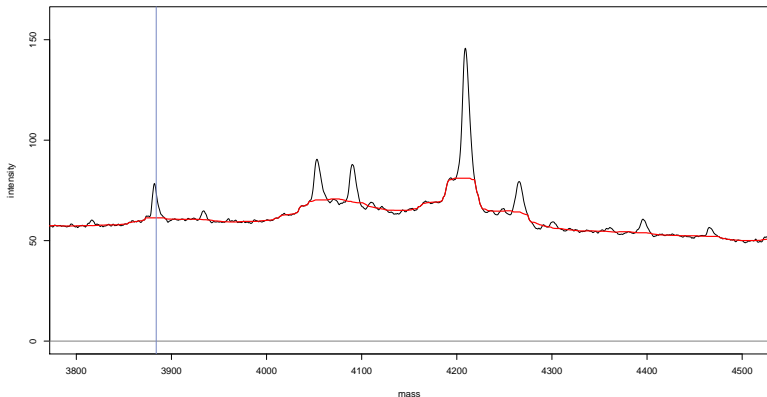```
> spectra <- lapply(spectra, transformIntensity, fun=sqrt)
> movAvg <- function(y) {return(filter(y, rep(1, 5)/5, sides=2));}
> spectra <- lapply(spectra, transformIntensity, fun=movAvg)
```
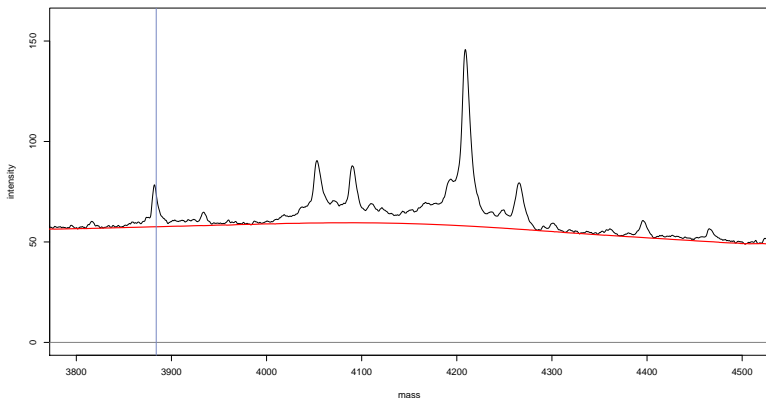
## MALDIquant
### hands-on: Baseline Correction – don't harm the data

```
> bl <- estimateBaseline(spectra[[1]], method="Median")
> plot(spectra[[1]], xlim=c(3800, 4500));
> lines(bl, col="red");
```

MALDIquant

hands-on: Baseline Correction – SNIP

```
> bl <- estimateBaseline(spectra[[1]], method="SNIP")
> plot(spectra[[1]], xlim=c(3800, 4500)); lines(bl, col="red");
```


C. G. Ryan, E. Clayton, W. L. Griffin, S. H. Sie, and D. R. Cousens. SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. *Nucl. Instrument. Meth. B*, 34:396–402, 1988

MALDIquant

hands-on: Baseline Correction – SNIP

```
> spectra <- lapply(spectra, removeBaseline)
> lines(spectra[[1]], col="blue")
```
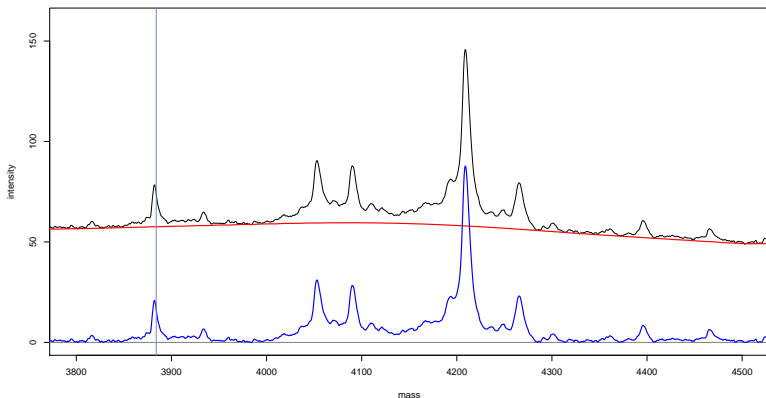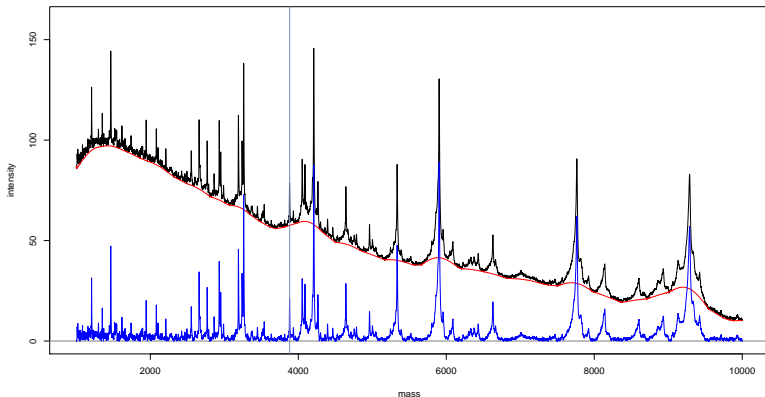

C. G. Ryan, E. Clayton, W. L. Griffin, S. H. Sie, and D. R. Cousens. SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. *Nucl. Instrument. Meth. B*, 34:396–402, 1988

MALDIquant
hands-on: Baseline Correction − SNIP

MALDIquant
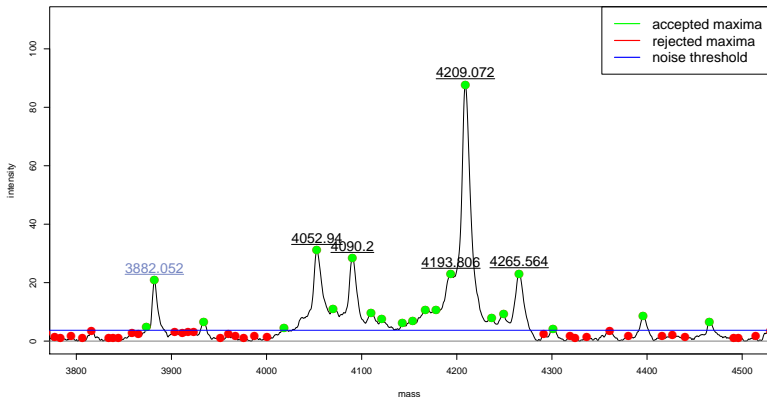### hands-on: Peak Detection

```
> spectra[[1]]
S4 class type            : MassSpectrum
Number of m/z values     : 42388
Range of m/z values      : 1000.015 - 9999.734
Range of intensity values: 0 - 709.207
File                     : /data/fiedler2009/[...]/Pankreas_HB_L_061019_G10/0_m19/1/1SLin/fid


> peaks <- lapply(spectra, detectPeaks)
> peaks[[1]]
S4 class type            : MassPeaks
Number of m/z values     : 198
Range of m/z values      : 1011.059 - 9423.422
Range of intensity values: 19.273 - 709.207
File                     : /data/fiedler2009/[...]/Pankreas_HB_L_061019_G10/0_m19/1/1SLin/fid
```

MALDIquant
hands-on: Peak Detection

```
> plot(spectra[[1]], xlim=c(3800, 4500)); points(peaks[[1]], col="green")
> top5 <- intensity(p) %in% sort(intensity(p)[mass(p)>3800 & mass(p)<4500], decreasing=TRUE)[1:5]
> labelPeaks(peaks[[1]], index=top5); labelPeaks(peak[[1]], mass=3884, col="blue")
```

MALDIquant
Single Spectrum Workflow

```
> library("MALDIquant")
> library("readBrukerFlexData")
> spectra <- mqReadBrukerFlex("/data/fiedler2009/")

> movAvg <- function(y) {return(filter(y, rep(1, 5)/5, sides=2));}

> spectra <- lapply(spectra, transformIntensity, fun=sqrt)
> spectra <- lapply(spectra, transformIntensity, fun=movAvg)

> spectra <- lapply(spectra, removeBaseline)
> peaks <- lapply(spectra, detectPeaks)
```
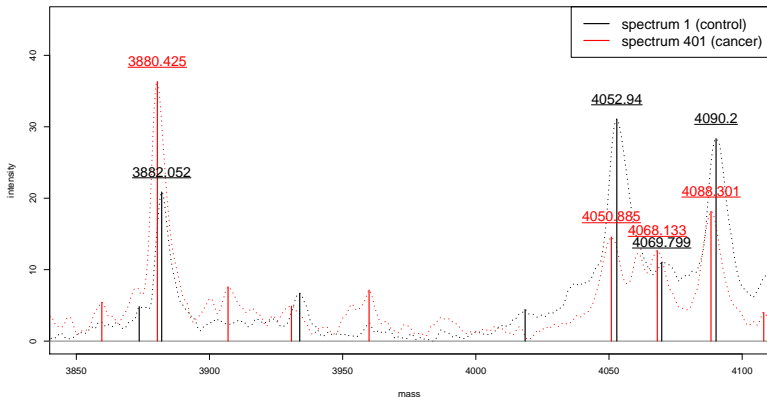
MALDIquant
Comparison of Multiple Spectra

MALDIquant
Conclusion

MALDIquant is free software (GPLv3).

Currently available:

- Easy to use and full established single spectrum workflow.
- Easy to customize.
- Easy data exchange.

Shortly available:

- Peak alignment.
- Calibration/Normalization routines.

MALDIquant
Thanks

Alexander B. Leichtle, helpful discussions
(Institute for Clinical Chemistry, Bern University Hospital)

Korbinian Strimmer, supervision
(IMISE, University of Leipzig)

Thanks for your attention!

Download of MALDIquant software:
http://strimmerlab.org/software/maldiquant/