

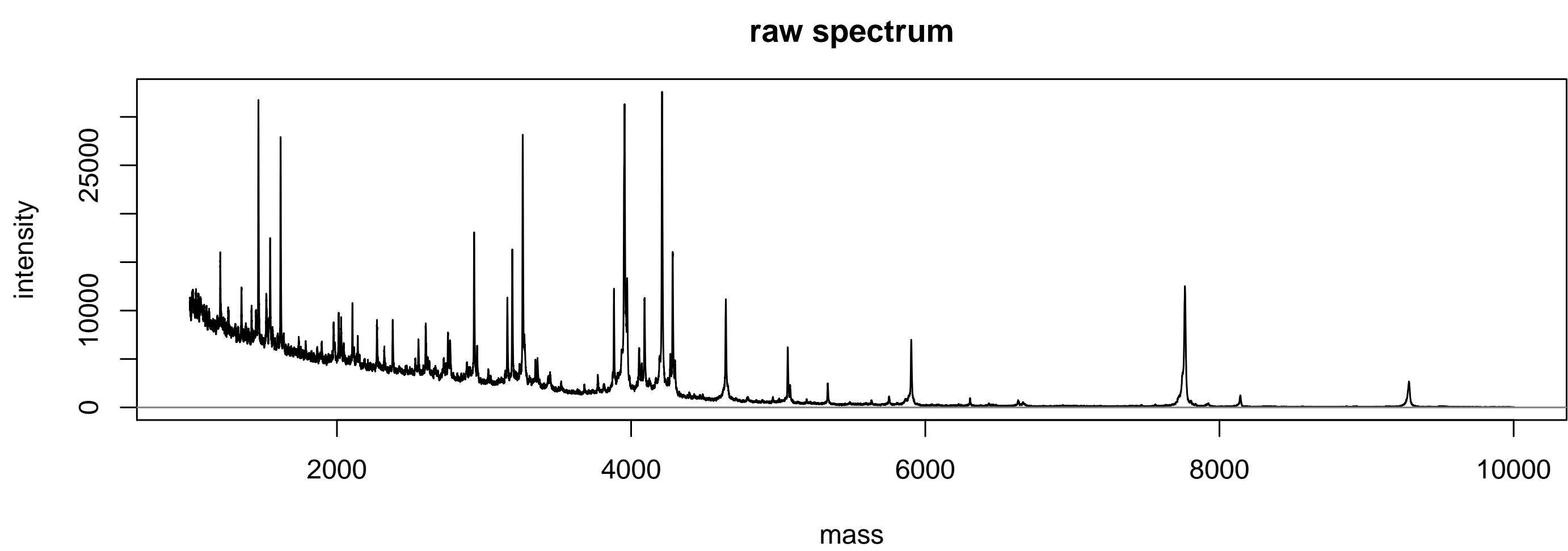
## Overview

MALDI-TOF is a well established technology for mass spectrometric profiling of proteomics data. Here, we introduce the MALDIquant R package that implements an analysis pipeline for quantitative analysis of clinical MALDI-TOF data on the R platform. We provide a brief summary of current and planned capabilities of the MALDIquant software. First, we briefly list our motivation for creating a new analysis pipeline for MALDI-TOF data. Subsequently, we outline the standard preprocessing steps (variance stabilization, baseline correction and peak detection) in the analysis of MALDI-TOF data and show the corresponding R commands using the MALDIquant software. MALDIquant is freely available from the R archive CRAN and is distributed under the GNU General Public License.

## Motivation

- Only relatively few open source software solutions available and very few for the R platform (see also [2]).
- 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.

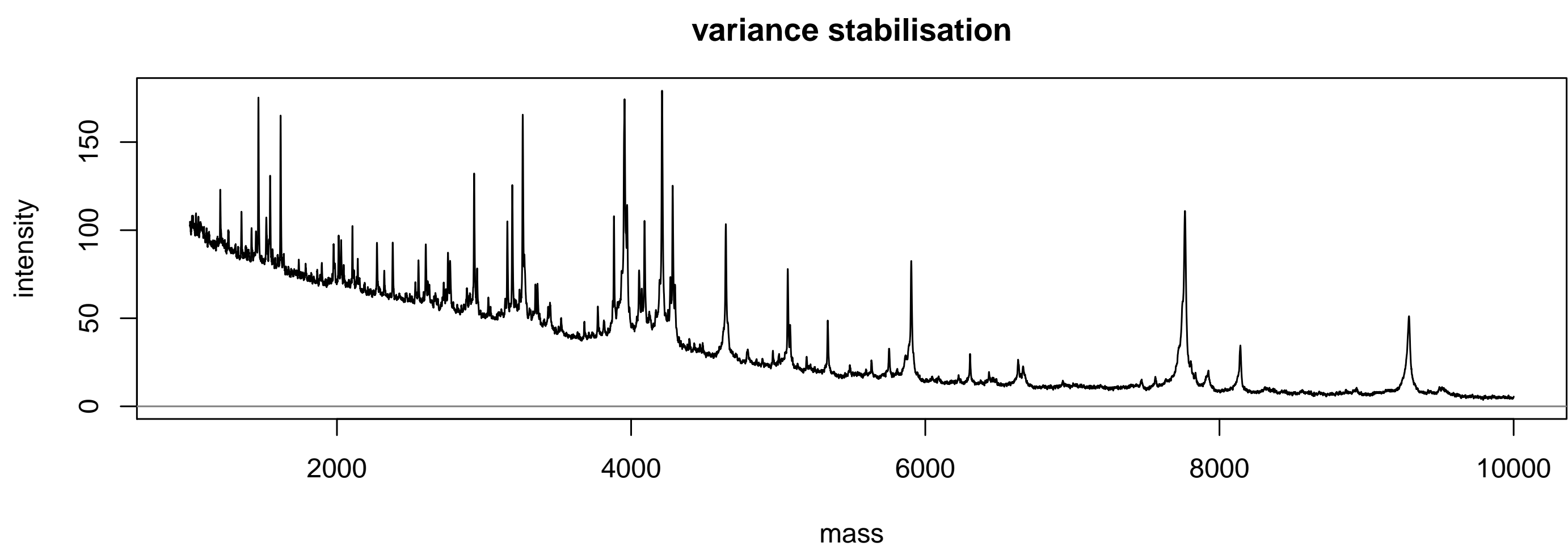
## Import of Raw Data



```
R code:
library("MALDIquant")
library("readBrukerFlexData")
s <- mqReadBrukerFlex("/data/exampleMS/fid")
plot(s)
```

## Variance Stabilization and Smoothing

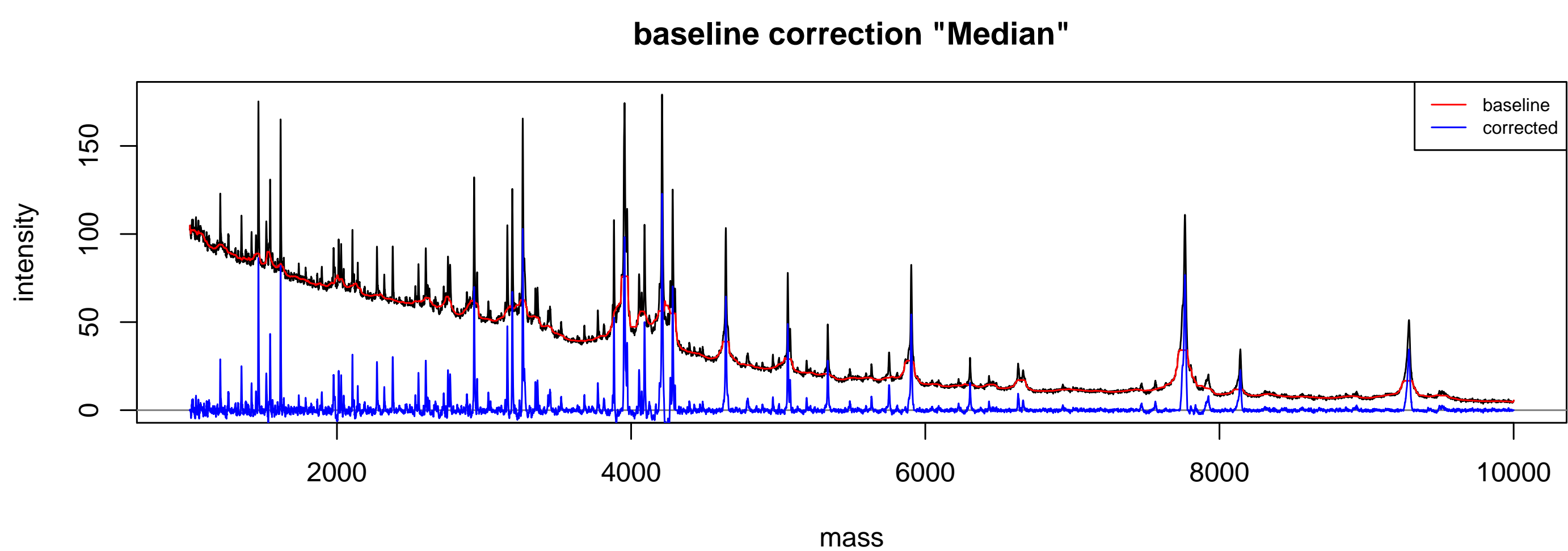
To stabilize variance in the intensity values of MALDI-TOF data sets a number of different transformations are established. Here we use a square root transformation, followed by a moving average filter to smooth the transformed intensities.



```
R code:
s1 <- transformIntensity(s, fun=sqrt)
movAvg <- function(x) {return(filter(y, rep(1, 5)/5, sides=2));}
s1 <- transformIntensity(s1, fun=movAvg)
plot(s1)
```

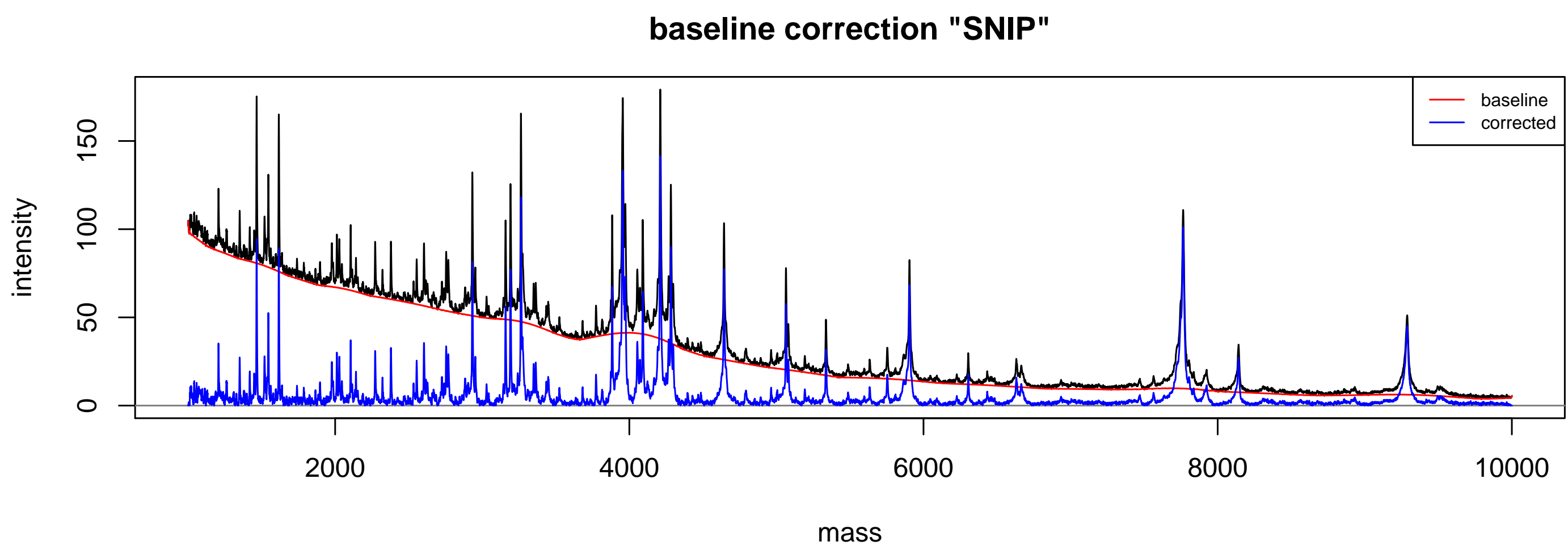
## Baseline Correction

MALDIquant supports commonly used baseline correction algorithms (e.g. moving median or convex hull).



```
R code:
b <- estimateBaseline(s1, "Median");
s2 <- removeBaseline(s1, b);
plot(s1); lines(b, col="red"); lines(s2, col="blue")
```

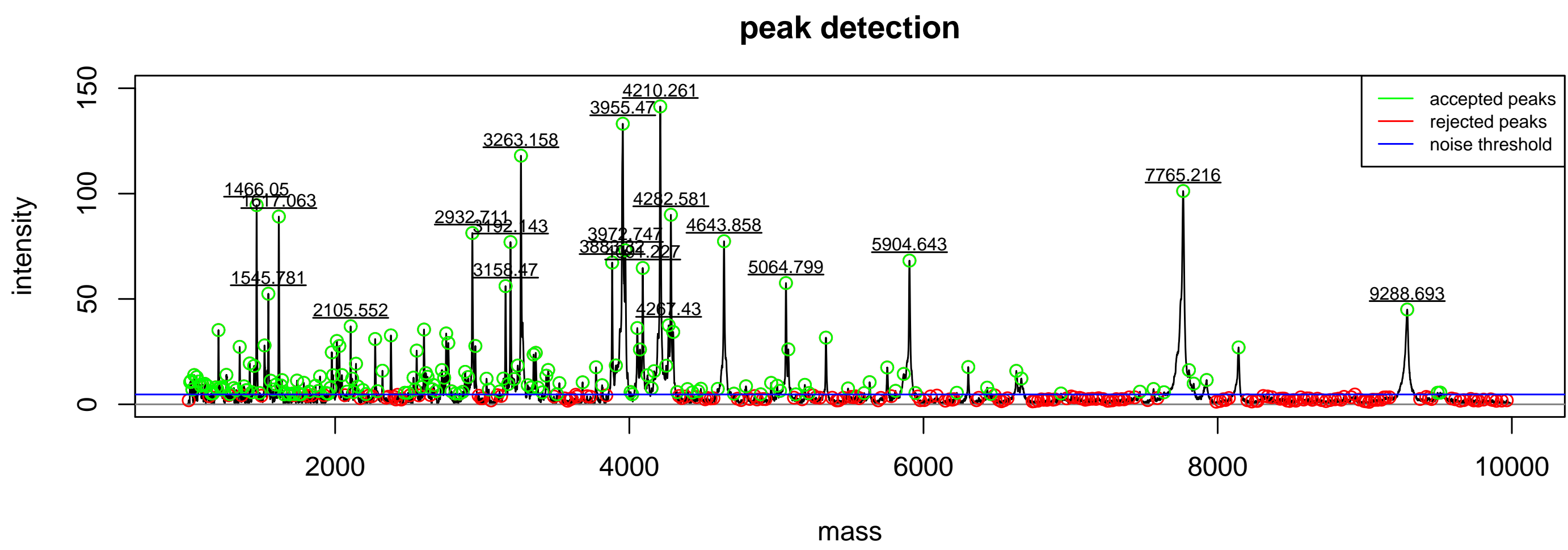
When quantifying peak intensities it is desirable to employ a baseline correction that produces only positive intensity values. The default algorithm for removing the baseline in MALDIquant is SNIP [3].



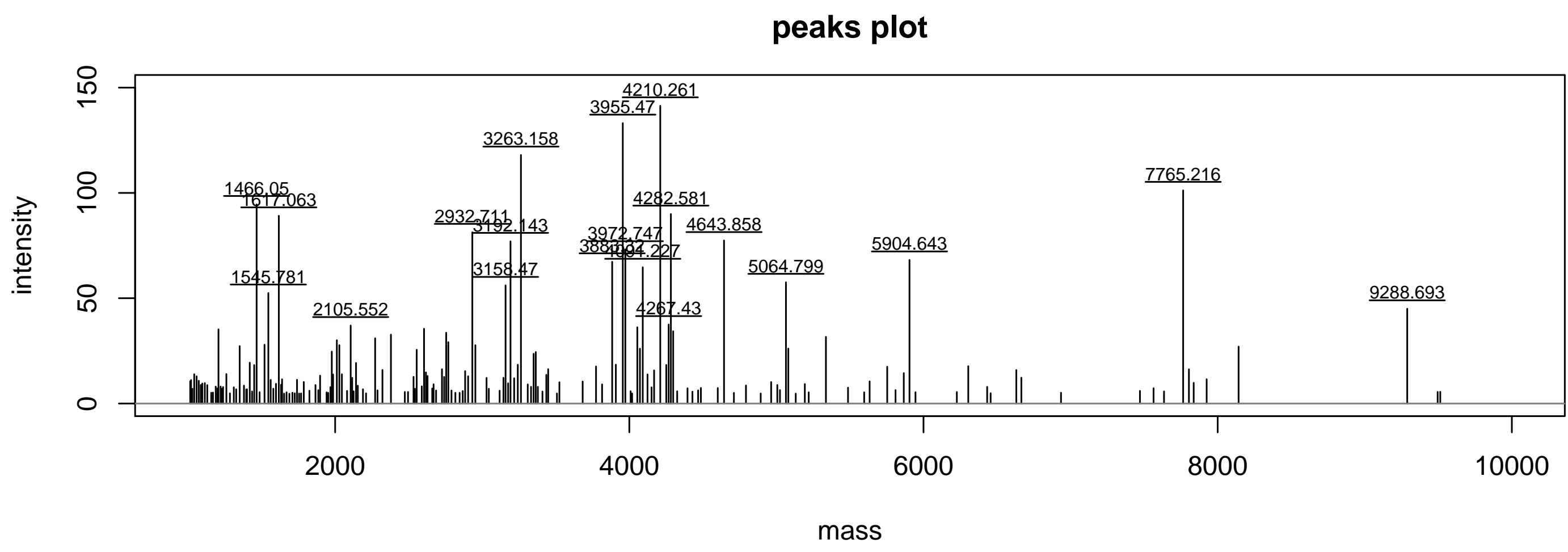
```
R code:
b <- estimateBaseline(s1, "SNIP")
s2 <- removeBaseline(s1, b)
plot(s1); lines(b, col="red"); lines(s2, col="blue")
```

## Peak Picking

For identifying peaks MALDIquant searches for local maxima above a noise threshold. In MALDIquant the default noise threshold is estimated by the median absolute deviation (MAD) of all intensity values multiplied by a user defined signal-to-noise-ratio.



```
R code:
p <- detectPeaks(s2)
plot(s2); points(p)
top20 <- intensity(p) %in% sort(intensity(p), decreasing=TRUE)[1:20]
labelPeaks(p, index=top20)
```



```
R code:
plot(p)
labelPeaks(p, index=top20)
```

## Upcoming Features

MALDIquant is under active development [1]. Features currently being implemented include:

- Handling of technical replicates.
- Peak alignment and and calibration of peak intensities.
- Classification for clinical diagnostics.

The impact of the choice of calibration on classification is investigated in [4].

## Auxiliary R Packages

- readBrukerFlexData – import Bruker \*flex files into MALDIquant
- readMzXmlData – import mzXML files into MALDIquant

## References

- 1 MALDIquant – <http://strimmerlab.org/software/malDIquant/>
- 2 Open Source Tools for Mass Spectrometry Analysis – <http://strimmerlab.org/notes/spectrometry.html>
- 3 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, vol. **34**, pp. 396–402, 1988.
- 4 S. Gibb and K. Strimmer. 2011. *Calibration of MALDI-TOF mass spectrometric data in clinical diagnostics* (in preparation)