

Supplemental Information to:

FTIR spectroscopy of biofluids revisited: An automated approach to spectral biomarker identification

Julian Ollesch, Steffen L. Drees, H. Michael Heise, Thomas Behrens, Thomas Bruening, and Klaus Gerwert

The baseline correction algorithm that was used on each of the five spectral regions (Fig. 6) is available for download at <https://code.google.com/p/airpls/> as on April 25th, 2013.¹⁻³ The parameterization of the airPLS.m function was

```
[~,baselines(i)]=airPLS([dataset],a(i),b(i),c(i),d(i));
```

with the parameters

```
a=[50e11 10e10 10e4 10e9 10e1];
```

```
b=[2 2 2 2 10];
```

```
c=[0.05 0.05 0.05 0.05 0.5];
```

```
d=[0.05 0.05 0.05 0.05 0.5];
```

The regions were ordered by increasing wavenumber, e.g., a(i) parameterizes region (v) etc. (see Fig. 6).

1. P. H. C. Eilers, *Anal. Chem.*, 2003, **75**, 3631–3636.
2. H. F. M. Boelens, R. J. Dijkstra, P. H. C. Eilers, F. Fitzpatrick, and J. A. Westerhuis, *J. Chromatogr. A*, 2004, **1057**, 21–30.
3. F. Gan, G. Ruan, and J. Mo, *Chemom. Intell. Lab. Syst.*, 2006, **82**, 59–65.

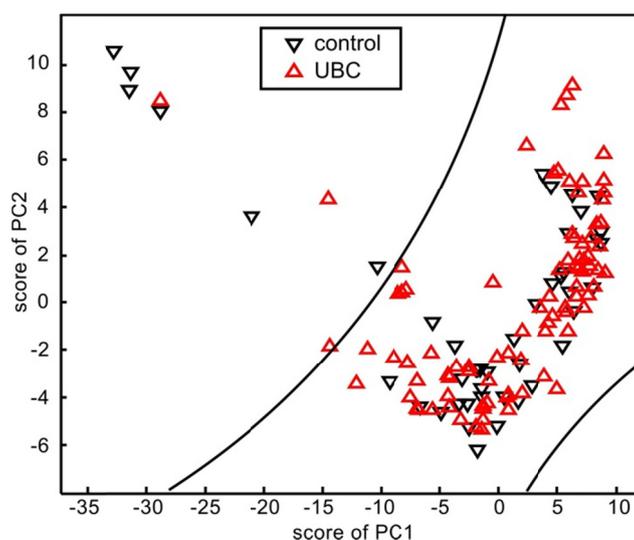


Figure S1: For illustration, the LDA classifier separating controls versus UBC based on the scores of the first and second principal components (PC) of spectra comprised of the full spectral feature set is displayed with a misclassification error rate of 32 %.

In the following, our full spectral dataset from 135 probands as used for feature selection is presented. Apart from the serum dataset, some wavenumber variables were found classification relevant within the absorbance and first derivative spectra, but most did not meet the selection frequency threshold of ≥ 7 of 12 feature selection cycles. For the spectra of EDTA- and citrate-plasma (Fig. S2 and S3), relevant variables were exclusively identified in the second derivative spectra.

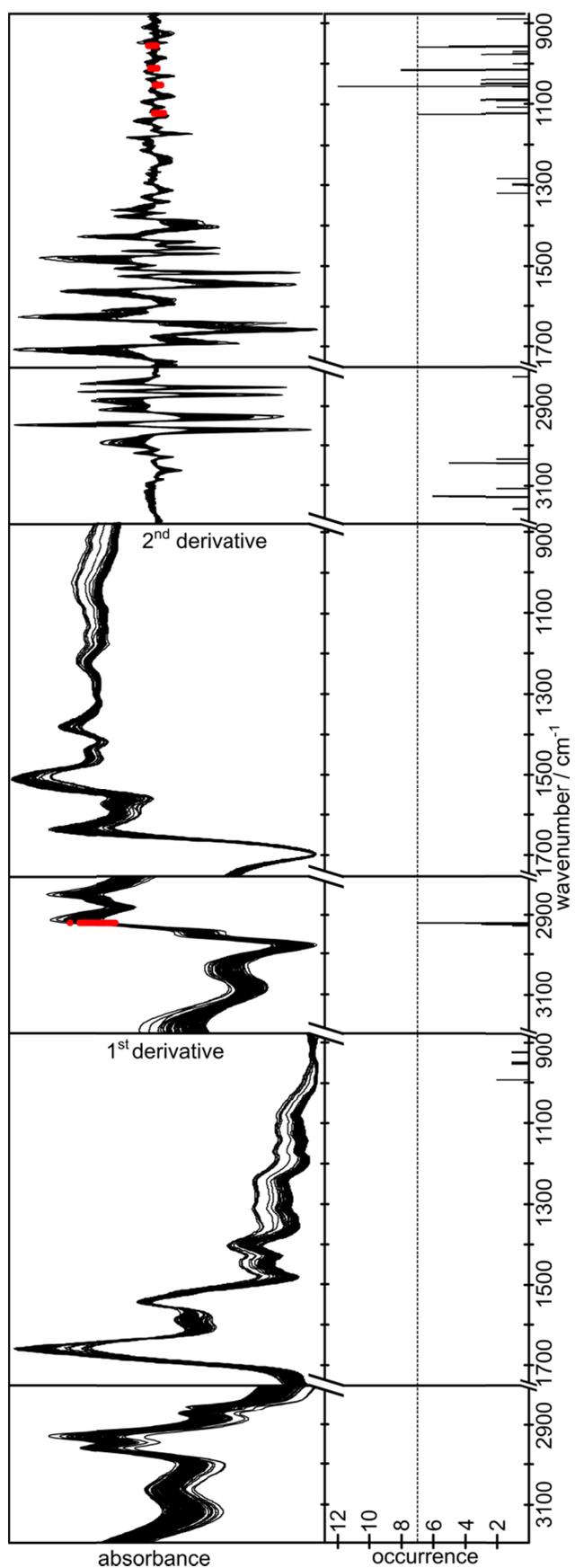


Figure S2: Serum absorbance, 1st derivative, 2nd derivative and selection frequency of features important for classification (labeled in red). Wavenumber segments were scaled individually for optimum display.

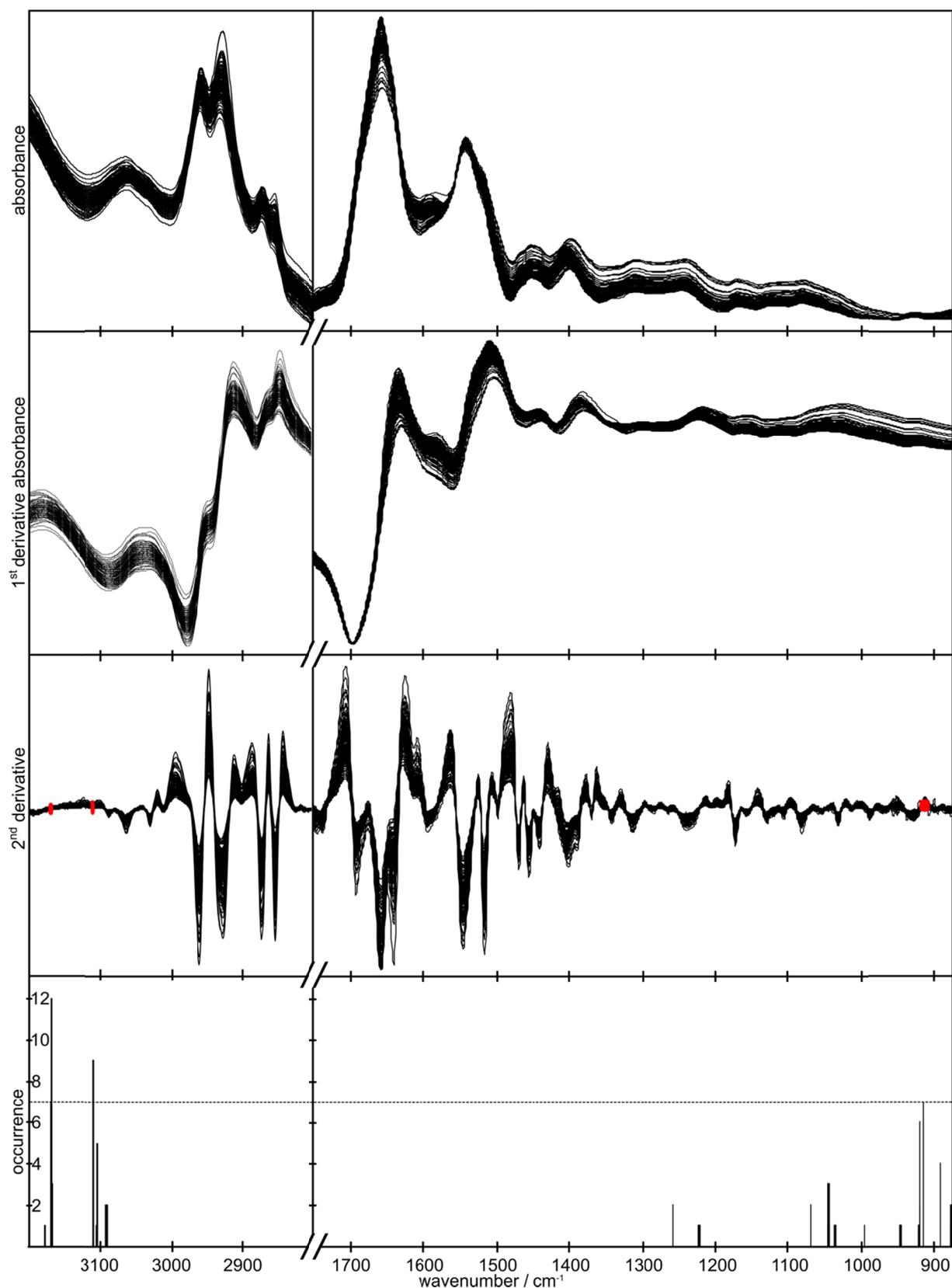


Figure S3: EDTA-plasma absorbance, 1st derivative, 2nd derivative and selection frequency of features important for classification. No features were identified within absorbance or 1st derivative spectra. Wavenumber range segments were scaled individually for optimum display.

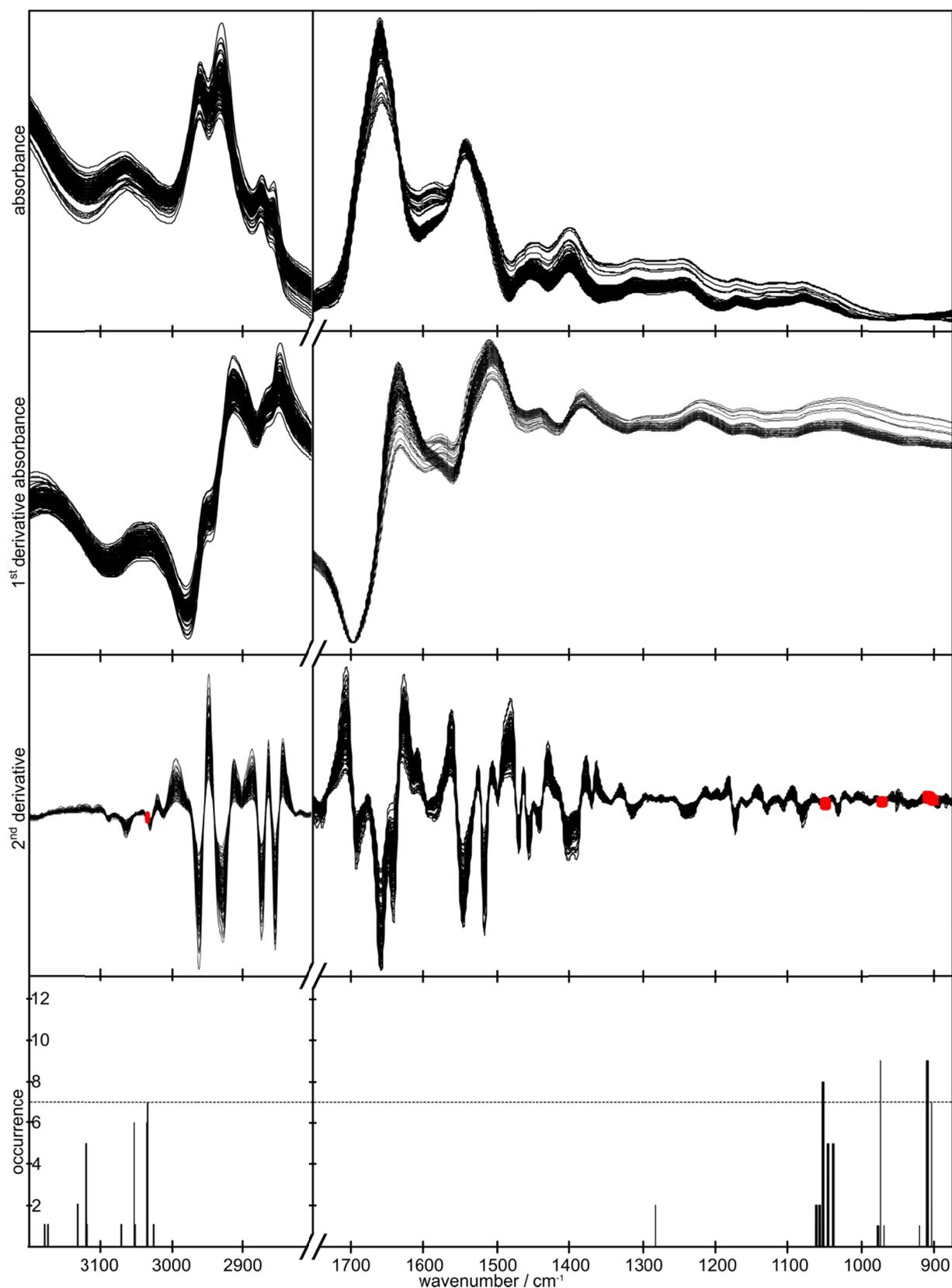


Figure S4: Sodium citrate stabilized plasma absorbance, 1st derivative, 2nd derivative and selection frequency of features important for classification. No features were identified within absorbance or 1st derivative spectra. Wavenumber range segments were scaled individually for optimum display.