Supplementary Information for "Bucket Fuser: statistical signal extraction for 1D ^{1}H NMR metabolomics data"

1 Current GCKD Investigators and Collaborators with the GCKD Study

University of Erlangen: Kai-Uwe Eckardt, Heike Meiselbach, Markus P. Schneider, Mario Schiffer, Hans-Ulrich Prokosch, Barbara Bärthlein, Andreas Beck, André Reis, Arif B. Ekici, Susanne Becker, Dinah Becker-Grosspitsch, Ulrike Alberth-Schmidt, Birgit Hausknecht, Anke Weigel;
University of Freiburg: Gerd Walz, Anna Köttgen, Ulla T. Schultheiß, Fruzsina Kotsis, Simone Meder, Erna Mitsch, Ursula Reinhard;
RWTH Aachen University: Jürgen Floege, Turgay Saritas;
Charité, University Medicine Berlin: Elke Schaeffner, Seema Baid-Agrawal, Kerstin Theisen;
Hannover Medical School: Hermann Haller, Jan Menne;
University of Heidelberg: Martin Zeier, Claudia Sommerer, Johanna Theilinger;
University of Jena: Gunter Wolf, Martin Busch, Rainer Paul;
Ludwig-Maximilians University of München: Thomas Sitter;
University of Würzburg: Christoph Wanner, Vera Krane, Antje Börner-Klein, Britta Bauer;
Medical University of Innsbruck, Division of Genetic Epidemiology: Florian Kronenberg, Julia Raschenberger, Barbara Kollerits, Lukas Forer, Sebastian Schönherr, Hansi Weissensteiner;
University of Regensburg, Institute of Functional Genomics: Peter Oefner, Wolfram Gronwald;
Institute of Medical Biometry, Informatics and Epidemiology, Medical Faculty, University of Bonn: Matthias Schmid, Jennifer Nadal.

2 The Python implementation of the Bucket Fuser algorithm

The Python implementation of the Bucket Fuser algorithm and a comprehensive tutorial are available in supplementary file S1.

3 Supplementary Tables

Table S1: Spearman's correlations to absolutely quantified metabolite concentrations for BF with $\lambda = 1$, BF with $\lambda = 2.5$, BF with $\lambda = 5$, JBA, SRV, and equidistant binnings with bin sizes 0.01 ppm and 0.02 ppm. Here, 3-HB = 3-hydroxybutyrate. Further provided are the spectral positions of the bins with largest absolute Spearman's correlation coefficients and the corresponding metabolite identities.

	BF ($\lambda = 1$)			BF ($\lambda = 2.5$)			BF ($\lambda = 5$)		
	correlation coef.	spectral position	metabolite	correlation coef.	spectral position	metabolite	correlation coef.	spectral position	metabolite
3-HB	0.768	ppm.2.3974-ppm.2.3954	3-HB, carnitine, glutamate	0.720	ppm.4.1694-ppm.4.1664	3-HB, lactate, proline	0.689	ppm.4.1564-ppm.4.1544	3-HB, lactate, proline
Acetate	0.757	ppm.1.9244-ppm.1.9194	acetate	0.983	ppm.1.9274-ppm.1.9244	acetate	0.968	ppm.1.9274-ppm.1.9224	acetate
Acetoacetate	0.670	ppm.2.3814-ppm.2.3794	3-hydroxyisovalerate, pyruvate, succinate	0.664	ppm.3.3794-ppm.2.3774	3-hydroxyisovalerate, pyruvate, succinate	0.610	ppm.3.3794-ppm.2.3744	3-hydroxyisovalerate, pyruvate, succinate
Acetone	0.528	ppm.2.3314-ppm.2.3294	2-hydroxyglutarate, glutamate, proline	0.748	ppm.2.3354-ppm.2.3334	acetone, macromolecules	0.568	ppm.2.2374-ppm.2.2314	acetone, macromolecules
Alanine	0.680	ppm.1.4874-ppm.1.4844	alanine	0.927	ppm.1.4834-ppm.1.4914	alanine	0.947	ppm.1.4814-ppm.1.4784	alanine
Asparagine	0.685	ppm.2.1444-ppm.2.1424	glutamate, glutamine	0.662	ppm.2.1444-ppm.2.1384	glutamate, glutamine	0.635	ppm.2.1514-ppm.2.1284	glutamate, glutamine
Betaine	0.509	ppm.3.2814-ppm.3.2694	glucose, betaine, myo-inositol, trimethylamine-N-oxide	0.637	ppm.3.2784-ppm.3.2704	glucose, betaine, myo-inositol, trimethylamine-N-oxide	0.480	ppm.3.2784-ppm.3.2724	glucose, betaine, myo-inositol, trimethylamine-N-oxide
Carnitine	0.678	ppm.2.1484-ppm.2.1464	glutamine, carnitine	0.419	ppm.2.2014-ppm.2.1094	macromolecule (?)	0.427	ppm.3.0504-ppm.3.0434	creatinine, creatine
Creatine	0.029	ppm.3.9384-ppm.3.9374	creatine	0.007	ppm.3.9404-ppm.3.9374	creatine	0.584	ppm.3.0504-ppm.3.0434	creatinine, creatine
Creatinine	0.772	ppm.3.0494-ppm.3.0474	creatinine	0.893	ppm.4.0624-ppm.4.0594	creatinine	0.881	ppm.4.0634-ppm.4.0584	creatinine
Dimethylamine	0.895	ppm.2.7304-ppm.2.7284	dimethylamine	0.649	ppm.2.7304-ppm.2.7214	dimethylamine	0.649	ppm.2.7304-ppm.2.7214	dimethylamine
Glucose	0.990	ppm.3.4624-ppm.3.4584	glucose	0.990	ppm.3.4724-ppm.3.4671	glucose	0.989	ppm.3.4884-ppm.3.4834	glucose
Glutamine	0.573	ppm.2.4674-ppm.2.4654	glutamine, 3-hydroxyisobutyrate,	0.870	ppm.2.4554-ppm.2.4524	glutamine, 3-hydroxyisobutyrate, carnitine	0.905	ppm.2.4684-ppm.2.4514	glutamine, 3-hydroxyisobutyrate, carnitine
Glycine	0.838	ppm.3.5714-ppm.3.5654	glycine, 3-hydroxyisobutyrate	0.808	ppm.3.5694-ppm.3.5674	glycine, 3-hydroxyisobutyrate	0.741	ppm.3.5704-ppm.3.5604	glycine, 3-hydroxyisobutyrate
Histidine	0.548	ppm.2.1444-ppm.2.1424	glutamate, glutamine	0.764	ppm.7.0894-ppm.7.0504	histidine	0.523	ppm.2.1514-ppm.2.1284	glutamate, glutamine
Isobutyrate	0.845	ppm.1.0724-ppm.1.0704	isobutyrate, 3-hydroxyisobutyrate	0.793	ppm.1.0854-ppm.1.0824	isobutyrate, 3-hydroxyisobutyrate	0.568	ppm.1.0854-ppm.1.0744	isobutyrate, 3-hydroxyisobutyrate
Isoleucine	0.866	ppm.1.0104-ppm.1.0084	isoleucine	0.911	ppm.1.0224-ppm.1.0204	isoleucine	0.808	ppm.1.0554-ppm.1.0524	isoleucine
Lactate	0.988	ppm.4.1294-ppm.4.1274	lactate	0.989	ppm.4.1184-ppm.4.1164	lactate	0.989	ppm.4.1294-ppm.4.1154	lactate
Phenylalanine	0.850	ppm.7.3424-ppm.7.3390	phenylalanine	0.874	ppm.7.3914-ppm.7.3874	phenylalanine	0.888	ppm.7.4314-ppm.7.4294	phenylalanine
Proline	0.754	ppm.2.3434-ppm.2.3384	proline, glutamate	0.937	ppm.1.4524-ppm.1.4504	proline	0.699	ppm.3.3504-ppm.3.3344	proline, glutamate
Pyruvate	0.884	ppm.2.3814-ppm.2.3794	pyruvate, 3-hydroxyisovalerate, succinate	0.968	ppm.2.3794-ppm.2.3774	pyruvate, 3-hydroxyisovalerate, succinate	0.941	ppm.2.3794-ppm.2.3774	pyruvate, proline, glutamate
Threonine	0.494	ppm.2.1444-ppm.2.1424	glutamate, glutamine	0.502	ppm.9.9104-ppm.6.9074	tyrosine	0.490	ppm.6.9114-ppm.6.8934	tyrosine
Trimethylamine N-Oxide	0.282	ppm.3.2814-ppm.3.2694	glucose, betaine, myo-inositol	0.401	ppm.3.2784-ppm.3.2704	trimethylamine-N-oxide, glucose, betaine, myo-inositol	0.403	ppm.3.2784-ppm.3.2704	trimethylamine-N-oxide, glucose, betaine, myo-inositol
Tyrosine	0.931	ppm.6.9004-ppm.6.8854	tyrosine	0.941	ppm.7.1924-ppm.7.1894	tyrosine	0.947	ppm.6.9114-ppm.6.8934	tyrosine
Valine	0.811	ppm.0.9674-ppm.0.9654	leucine, macromolecule	0.947	ppm.1.0544-ppm.1.0624	valine	0.961	ppm.1.0434-ppm.1.0404	valine

	JBA			SRV			EB (0.01 ppm)			EB (0.02 ppm)		
	correlation coef.	spectral position	metabolite	correlation coef.	spectral position	metabolite	correlation coef.	spectral position	metabolite	correlation coef.	spectral position	metabolite
3-HB	0.768	2.3964	3-HB,	0.600	ppm.2.3994-2.3924	3-HB,	0.547	ppm.2.395	3-HB,	0.498	ppm.2.33	glutamate, proline
Acetate	0.966	1.9254	acetate	0.908	ppm.1.9344-1.9184	acetate	0.946	ppm.1.925	acetate	0.892	ppm.1.93	acetate
Acetoacetate	0.670	2.3804	3-hydroxyisovalerate, pyruvate, succinate	0.611	ppm.2.2884-6.2754	noise	0.614	ppm.6.285	noise	0.603	ppm.6.29	noise
Acetone	0.530	2.4324	glutamate, glutamine, pyruvate, succinate	0.472	ppm.2.4344-2.4294	acetone, macromolecules	0.455	ppm.2.235	acetone, macromolecules	0.350	ppm.2.33	2-hydroxyglutarate, glutamate, proline
Alanine	0.722	1.4859	3-hydroxyisobutyrate, carnitine	0.915	ppm.1.5054-1.4704	alanine	0.926	ppm.1.495	alanine	0.905	ppm.1.49	alanine
Asparagine	0.685	2.4559	glutamine,	0.698	ppm.2.1474-2.1414	glutamate, glutamine	0.683	ppm.2.145	glutamate, glutamine	0.659	ppm.2.15	glutamate, glutamine
Betaine	0.689	3.2724	glucose, betaine, myo-inositol, trimethylamine-N-oxide	0.481	ppm.3.9104-3.9064	glucose, betaine, myo-inositol, trimethylamine-N-oxide	0.630	ppm.3.275	glucose, betaine, myo-inositol, trimethylamine-N-oxide	0.221	ppm.4.07	creatine, myo-inositol
Carnitine	0.692	2.4464	glutamine, carnitine	0.412	ppm.1.8284-1.8164	noise	0.444	ppm.3.045	creatine, creatinine	0.445	ppm.3.05	creatine, creatinine
Creatine	0.009	3.9379	creatine	0.842	ppm.3.0454-3.0384	creatine	0.763	ppm.3.935	creatine	0.626	ppm.3.05	creatine, creatinine
Creatinine	0.443	4.0704	creatinine, myo-inositol	0.749	ppm.4.0654-4.0504	creatinine	0.703	ppm.4.065	creatinine, myo-inositol	0.619	ppm.3.05	creatinine, creatinine
Dimethylamine	0.508	1.7729	noise, macromolecules	0.684	ppm.2.7324-2.7244	dimethylamine	0.587	ppm.1.765	noise, macromolecules	0.592	ppm.1.77	noise, macromolecules
Glucose	0.987	3.4599	glucose	0.989	ppm.4.9243.4774	glucose	0.990	ppm.3.475	glucose	0.988	ppm.3.47	glucose
Glutamine	0.612	2.4559	glutamine, 3-hydroxyisobutyrate, carnitine	0.779	ppm.2.1474-2.1414	glutamine, glutamate	0.897	ppm.2.455	glutamine, 3-hydroxyisobutyrate, carnitine	0.878	ppm.2.45	glutamine, 3-hydroxyisobutyrate, carnitine
Glycine	0.332	3.0579	creatine, creatinine	0.778	ppm.3.5724-3.5634	glycine, 3-hydroxyisobutyrate	0.655	ppm.3.565	glycine, 3-hydroxyisobutyrate	0.543	ppm.3.57	glycine, 3-hydroxyisobutyrate
Histidine	0.453	2.4559	glutamine	0.571	ppm.1.4724-2.1414	glutamate, glutamine	0.558	ppm.2.145	glutamate, glutamine	0.631	ppm.7.07	histidine
Isobutyrate	0.445	1.7244	leucine	0.687	ppm.1.0734-1.0664	isobutyrate, 3-hydroxyisobutyrate	0.618	ppm.1.085	isobutyrate, 3-hydroxyisobutyrate	0.522	ppm.1.05	value
Isoleucine	0.735	0.9724	macromolecules, unknown (?), leucine	0.762	ppm.1.0594-1.0314	value	0.792	ppm.1.045	value	0.790	ppm.1.05	value
Lactate	0.970	4.1204	lactate	0.983	ppm.1.4434-4.1004	lactate	0.990	ppm.1.115	lactate	0.985	ppm.4.11	lactate
Phenylalanine	0.819	7.3424	phenylalanine	0.818	ppm.4.4594-7.4124	phenylalanine	0.810	ppm.7.385	phenylalanine	0.860	ppm.7.39	phenylalanine
Proline	0.676	3.3434	proline, unknown (?)	0.871	ppm.1.4534-1.4384	proline	0.680	ppm.3.345	proline, xylose (?)	0.609	ppm.7.39	proline, xylose (?)
Pyruvate	0.884	2.3804	pyruvate, 3-hydroxyisovalerate, succinate	0.931	ppm.2.3824-2.3734	pyruvate, 3-hydroxyisovalerate, succinate	0.857	ppm.2.375	pyruvate, 3-hydroxyisovalerate, succinate	0.692	ppm.2.375	noise
Threonine	0.120	2.2014	tyrosine	0.472	ppm.9.1744-8.8864	tyrosine	0.450	ppm.895	tyrosine	0.474	ppm.6.89	tyrosine
Trimethylamine N-Oxide	0.449	3.2724	glucose, betaine, myo-inositol	0.279	ppm.3.16143.1554	dimetyl sulfone (?), Co-EDTA2+	0.400	ppm.3.275	trimethylamine-N-oxide, glucose, betaine, myo-inositol	0.231	ppm.8.09	noise
Tyrosine	0.816	7.1944	tyrosine	0.935	ppm.6.91746.8864	tyrosine	0.939	ppm.6.895	tyrosine	0.924	ppm.6.91	tyrosine
Valine	0.725	0.9724	leucine, macromolecule	0.924	ppm.1.0594-1.0314	value	0.952	ppm.1.055	value	0.952	ppm.1.05	value

Table S2: Number of metabolite features extracted from 1D ^{1}H NMR spectra of (a) urinary and (b) plasma AKI data sets for different binning approaches. ^{a}Number of bins after exclusion of 21 bins corresponding to filter residues and free EDTA. ^{b}Number of bins after exclusion of 16 bins corresponding to filter residues and free EDTA. ^{c}Number of bins after exclusion of 29 bins corresponding to filter residues and free EDTA. ^{d}Number of bins after exclusion of 12 bins corresponding to filter residues and free EDTA. ^{e}Number of bins after exclusion of 28 bins corresponding to filter residues and free EDTA. ^{f}Number of bins after exclusion of 12 bins corresponding to filter residues and free EDTA. ^{g}Number of bins after exclusion of 20 bins corresponding to filter residues and free EDTA. ^{h}Number of bins after exclusion of 36 bins corresponding to filter residues and free EDTA. ^{i}Number of bins after exclusion of 28 bins corresponding to filter residues and free EDTA. ^{j}Number of bins after exclusion of 15 bins corresponding to filter residues and free EDTA.

(a) Urinary AKI data set.

	BF ($\lambda = 1$) plateaus	BF ($\lambda = 1$) non-plateaus	BF ($\lambda = 2.5$) plateaus	BF ($\lambda = 2.5$) non-plateaus	BF ($\lambda = 5$) plateaus	BF ($\lambda = 5$) non-plateaus	SRV	JBA	EB (0.01 ppm)	EB (0.02 ppm)
No. regions	180	268	498	383	574	309	585	731	700	350

(b) Plasma AKI data set.

	BF ($\lambda = 1$) plateaus	BF ($\lambda = 1$) non-plateaus	BF ($\lambda = 2.5$) plateaus	BF ($\lambda = 2.5$) non-plateaus	BF ($\lambda = 5$) plateaus	BF ($\lambda = 5$) non-plateaus	SRV	JBA	EB (0.01 ppm)	EB (0.02 ppm)
No. regions	333^a	273^b	503^c	249^d	723^e	224^f	533^g	520^h	712^i	355^j

4 Supplementary Figures
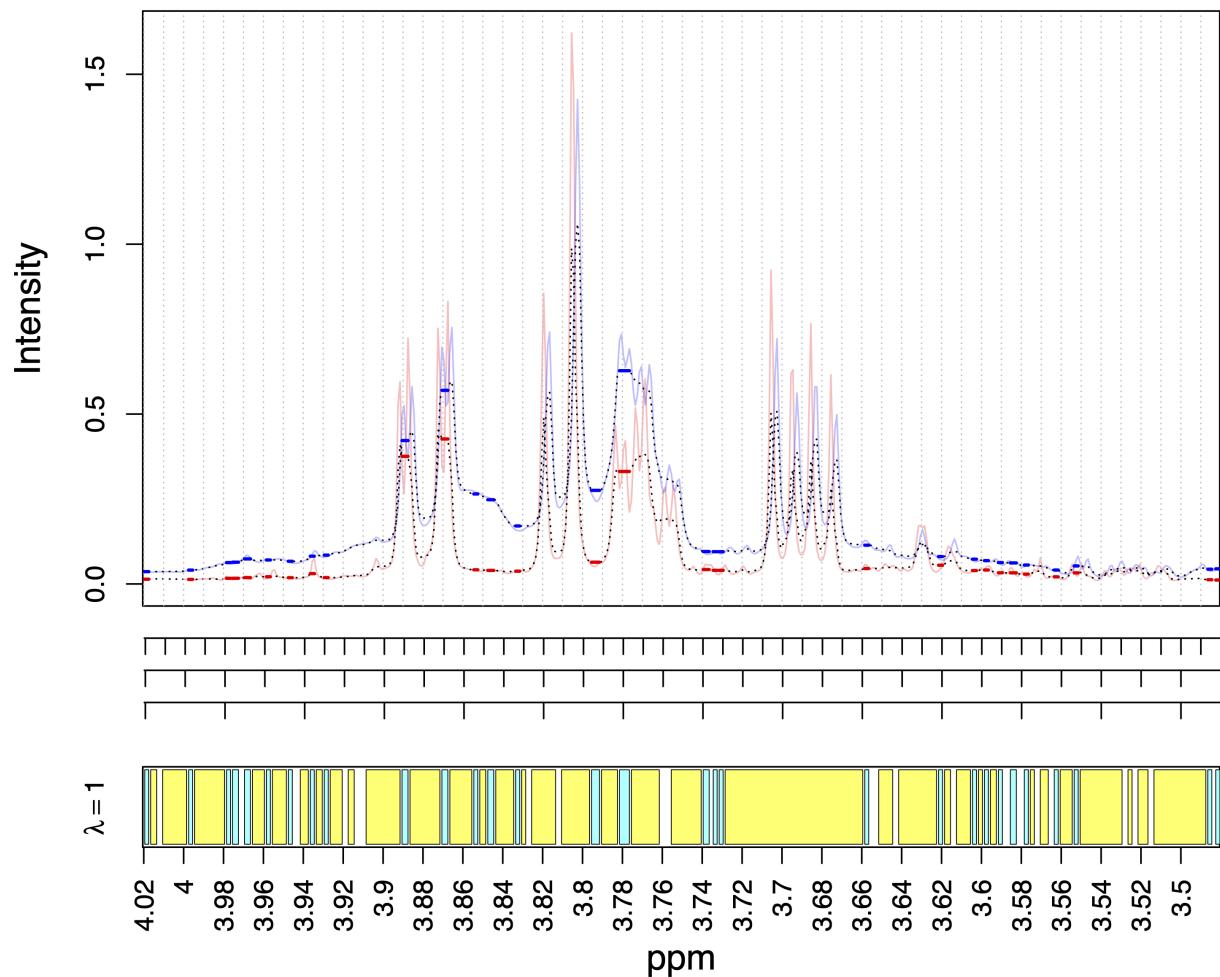
Figure S1: Exemplary NMR spectral regions ranging from 4 ppm to 3.5 ppm together with their corresponding BF fits for $\lambda = 1$. The Figure can be understood analogous to Figure 1a and b in the main article.
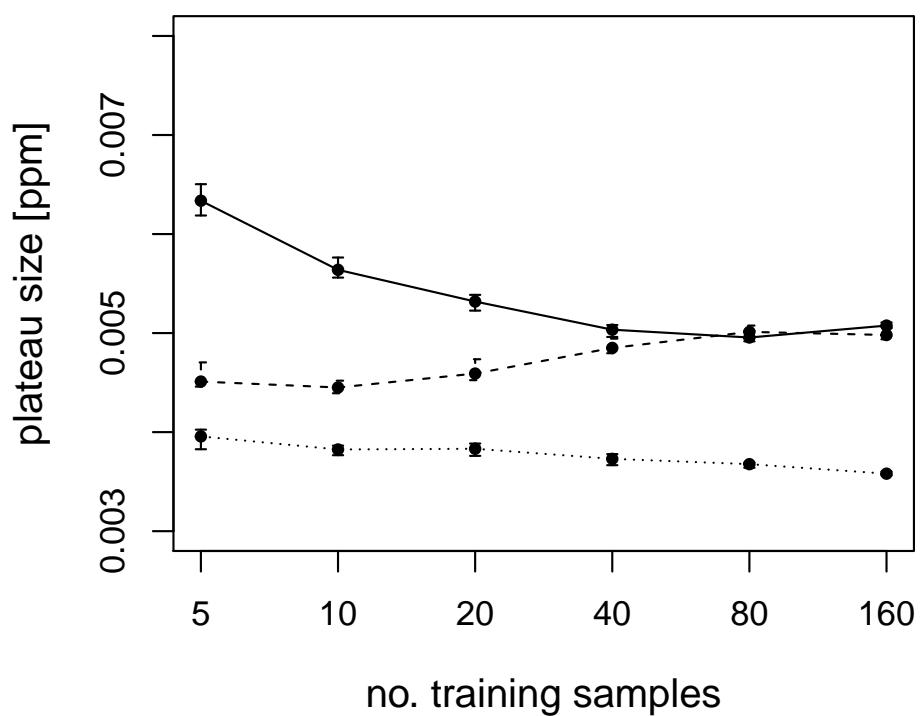
Figure S2: Average plateau widths versus the number of training samples for $\lambda = 1$ (dotted line), $\lambda = 2.5$ (dashed line), $\lambda = 5$ (solid line) in the GCKD data set. The error bars indicate the 25% to 75% percentiles and the dots the corresponding median values across 50 simulation runs.
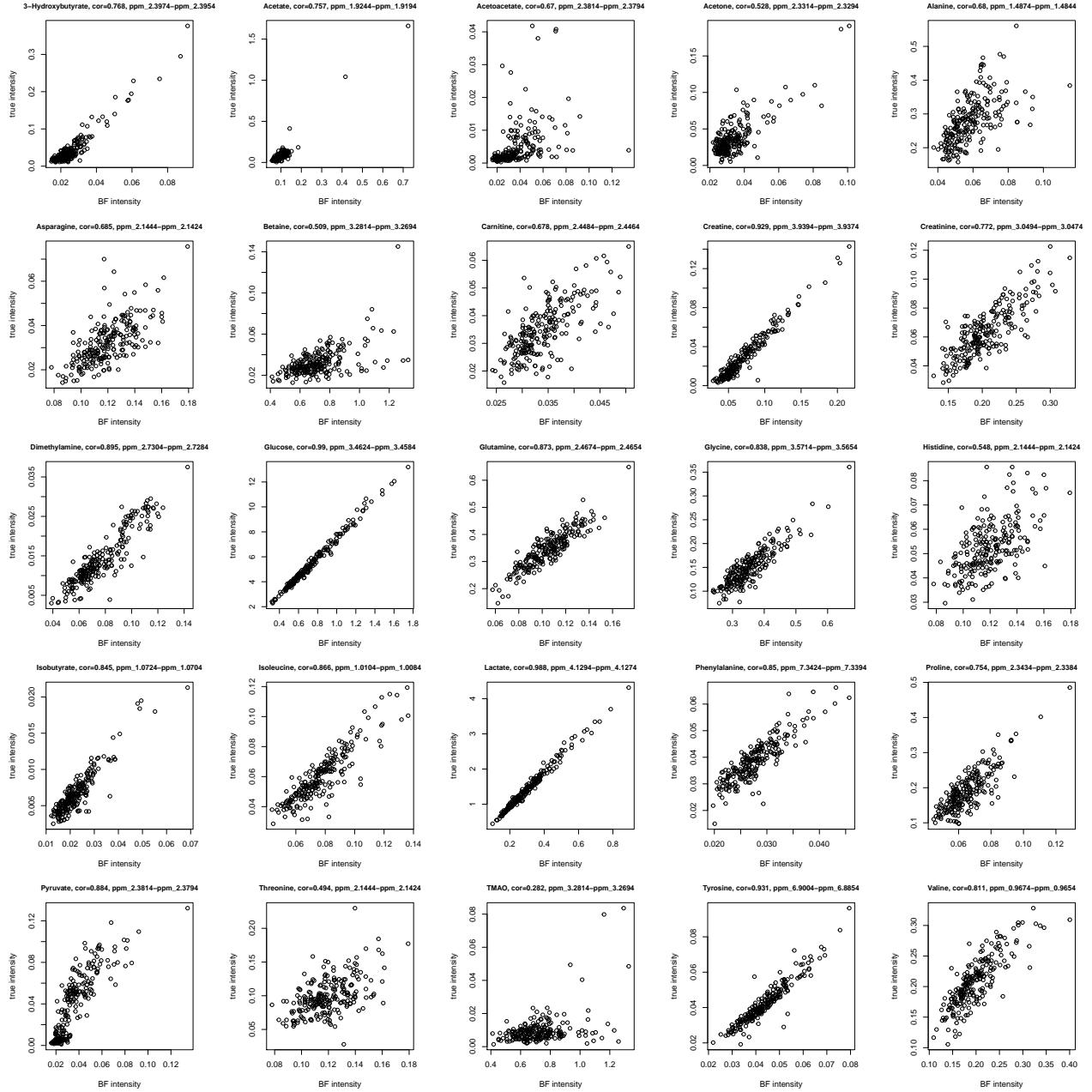
Figure S3: Comparison of the integrals of spectral features constructed by the Bucket Fuser (BF) using $\lambda = 1$ with absolutely quantified metabolite concentrations for 25 metabolites. The x-axis gives the spectral intensities returned by the Bucket Fuser which correlate best with the absolute concentrations shown on the y-axis. The headings indicate the investigated metabolites, followed by Spearman's correlation, and the selected spectral region from the BF binning in ppm. Abbr.: TMAO, trimethylamine-N-oxide.
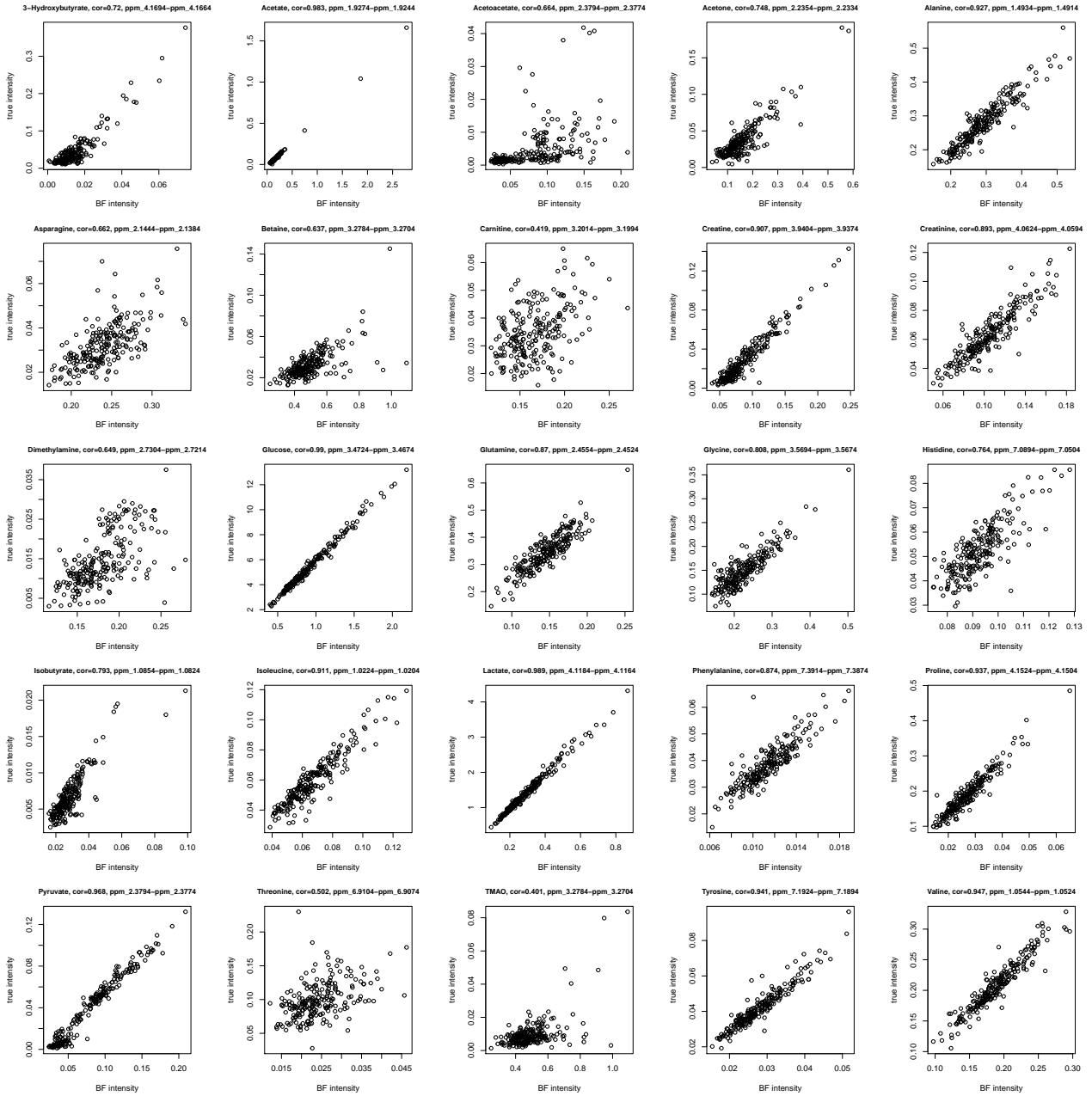
Figure S4: Comparison of the integrals of spectral features constructed by the Bucket Fuser (BF) using $\lambda = 2.5$ with absolutely quantified metabolite concentrations for 25 metabolites. The x-axis gives the spectral intensities returned by the Bucket Fuser which correlate best with the absolute concentrations shown on the y-axis. The headings indicate the investigated metabolites, followed by Spearman's correlation, and the selected spectral region from the BF binning in ppm. Abbr.: TMAO, trimethylamine-N-oxide.
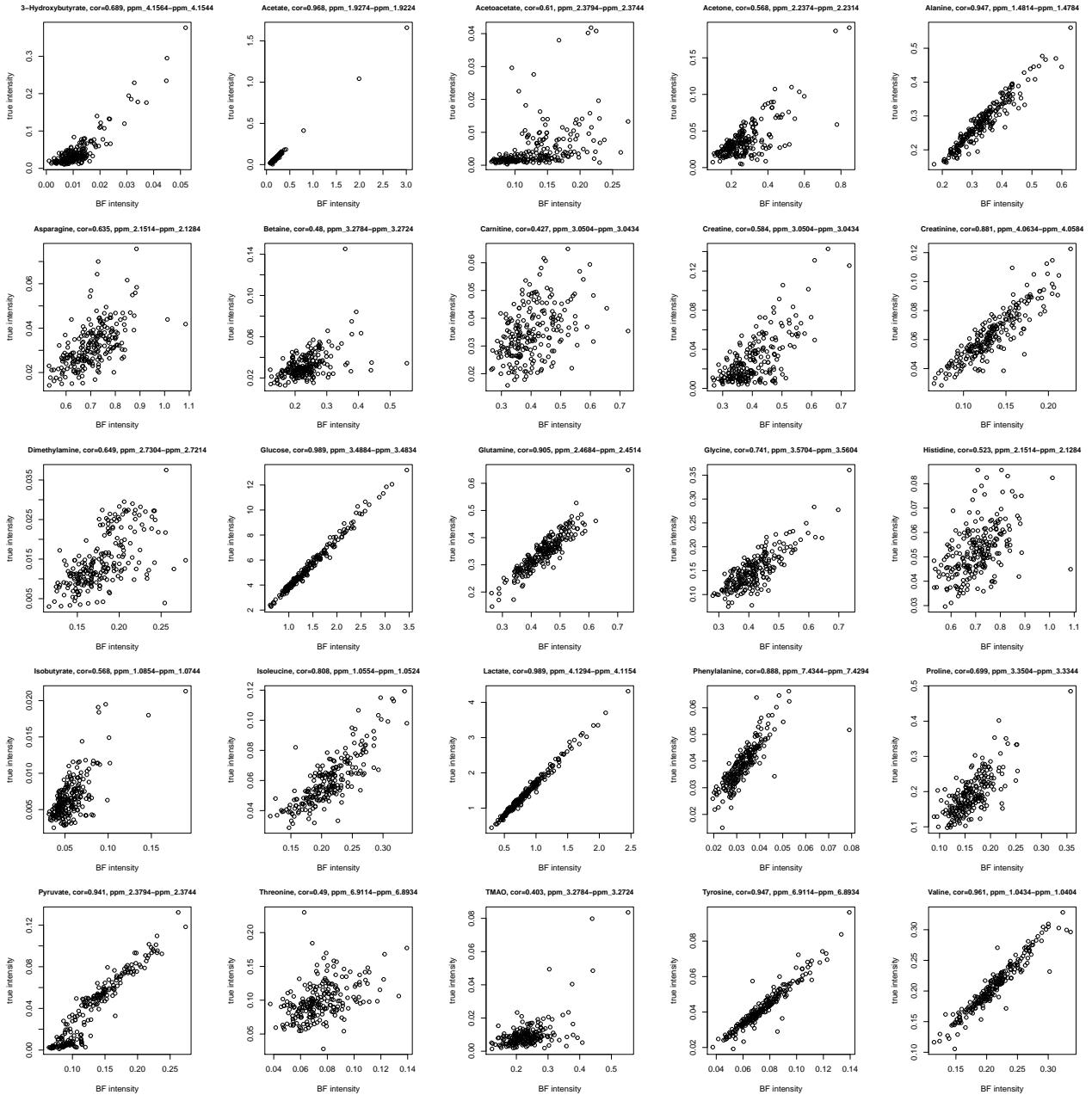
Figure S5: Comparison of the integrals of spectral features constructed by the Bucket Fuser (BF) using $\lambda = 5$ with absolutely quantified metabolite concentrations for 25 metabolites. The x-axis gives the spectral intensities returned by the Bucket Fuser which correlate best with the absolute concentrations shown on the y-axis. The headings indicate the investigated metabolites, followed by Spearman's correlation, and the selected spectral region from the BF binning in ppm. Abbr.: TMAO, trimethylamine-N-oxide.
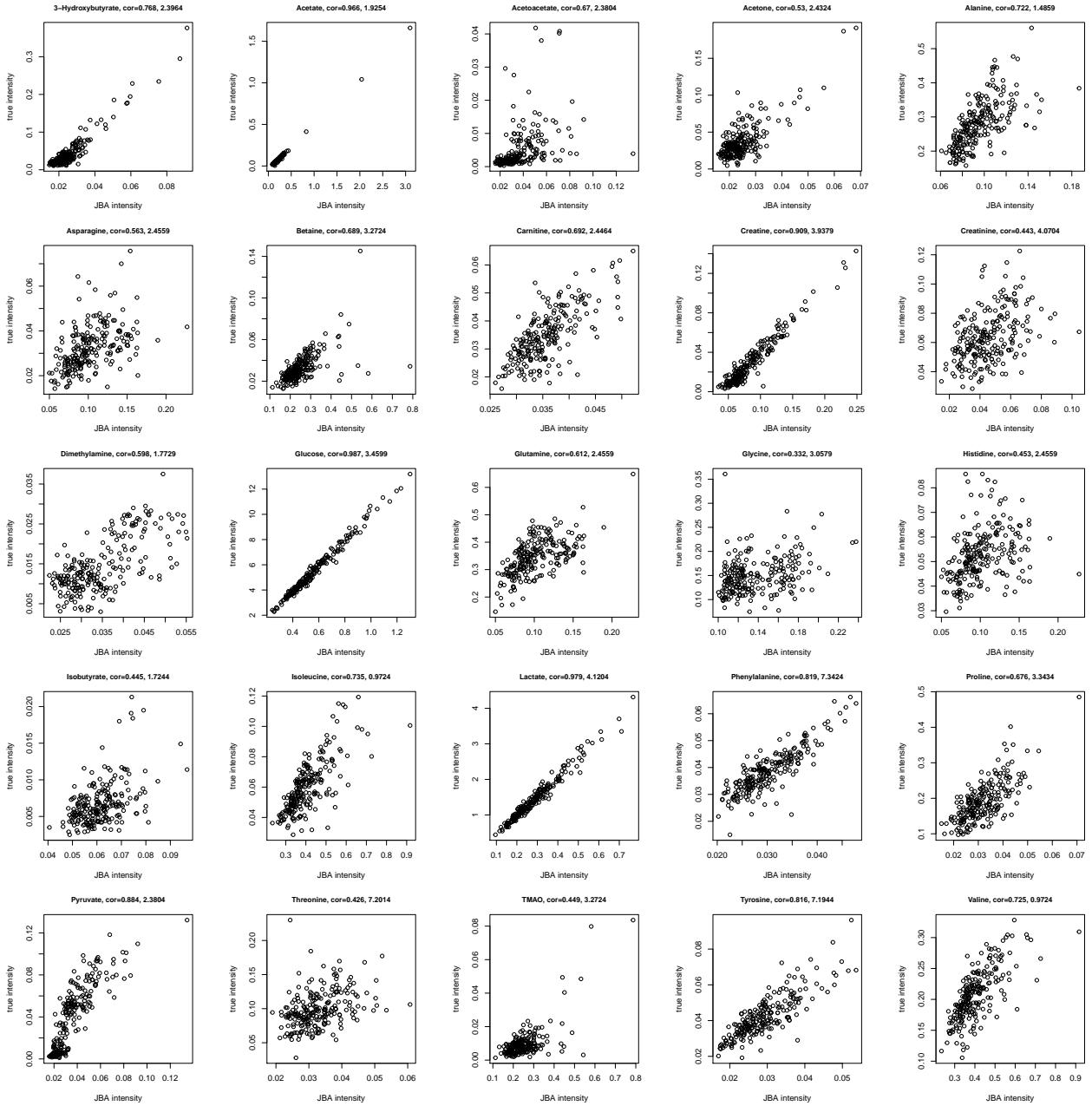
Figure S6: Comparison of the integrals of spectral features constructed by JBA with absolutely quantified metabolite concentrations for 25 metabolites. The x-axis gives the spectral intensities returned by JBA which correlate best with the absolute concentrations shown on the y-axis. The headings indicate the investigated metabolites, followed by Spearman's correlation, and the selected spectral region in ppm. Abbr.: TMAO, trimethylamine-N-oxide.
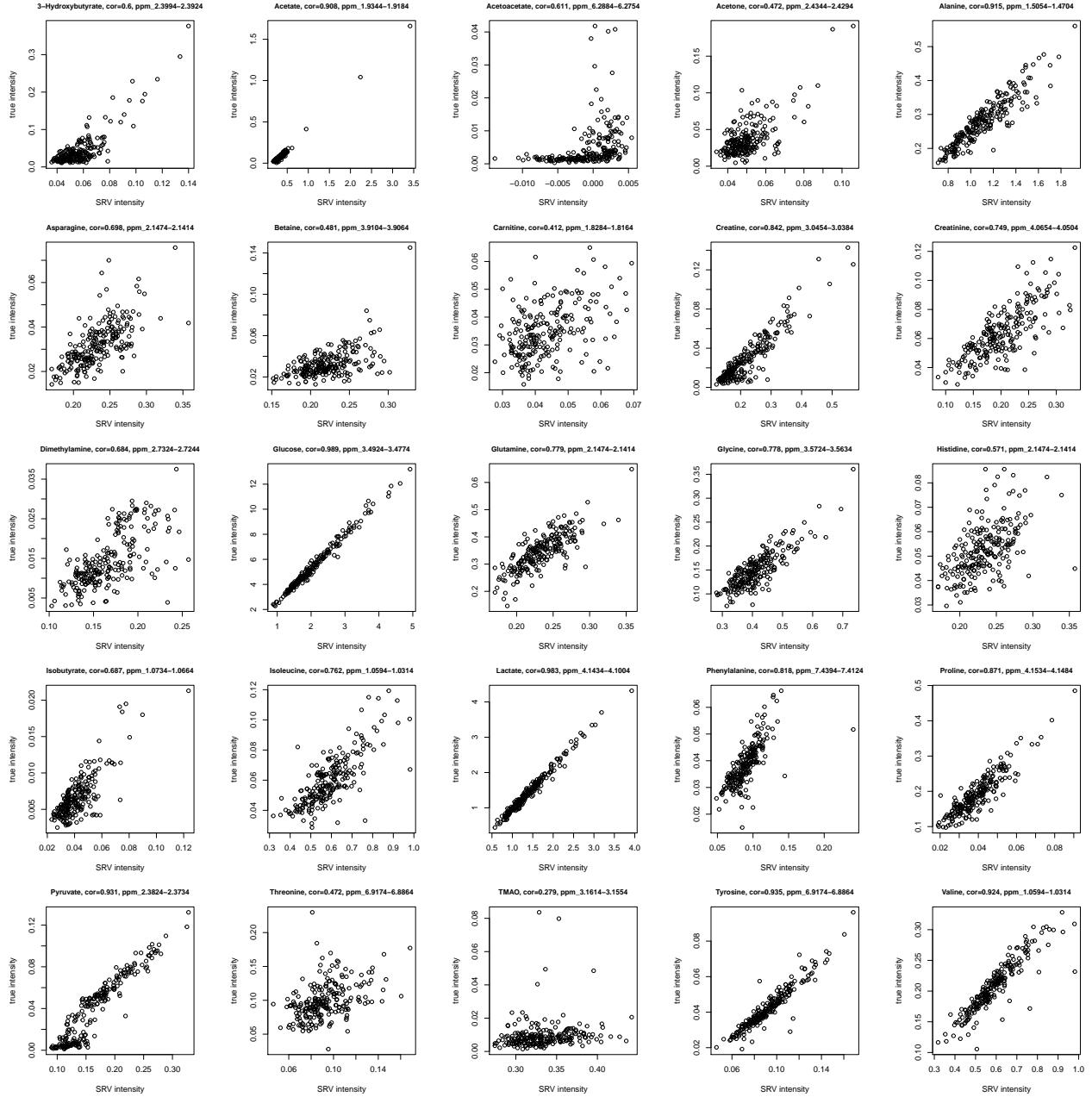
Figure S7: Comparison of the integrals of spectral features constructed by SRV with absolutely quantified metabolite concentrations for 25 metabolites. The x-axis gives the spectral intensities returned by the Bucket Fuser which correlate best with the absolute concentrations shown on the y-axis. The headings indicate the investigated metabolites, followed by Spearman's correlation, and the selected spectral region from the BF binning in ppm. Abbr.: TMAO, trimethylamine-N-oxide.
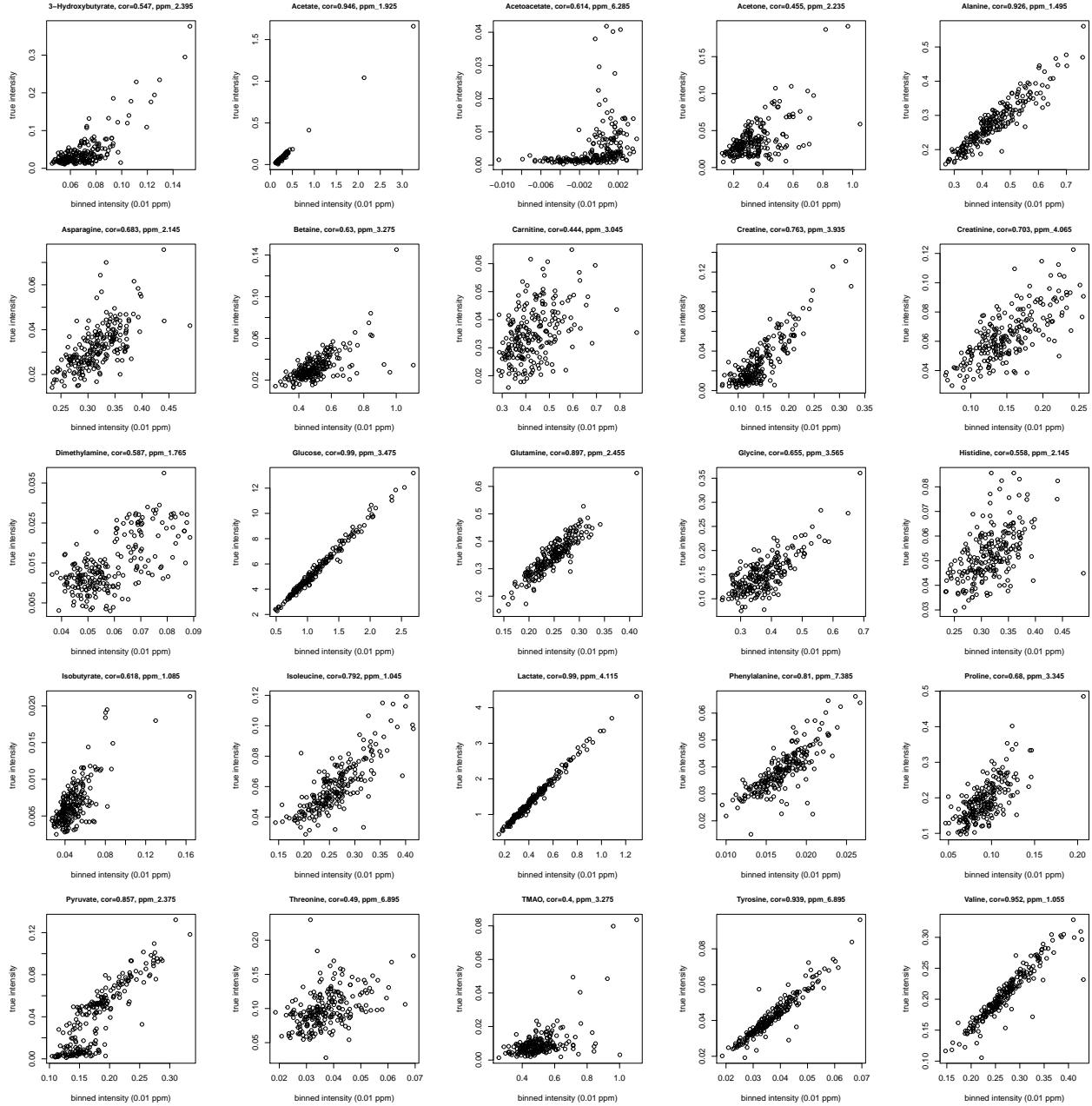
Figure S8: Comparison of the integrals of spectral features from an equidistant binning (bin size = 0.01 ppm) with absolutely quantified metabolite concentrations for 25 metabolites. The x-axis gives the spectral intensities returned by the binning which correlate best with the absolute concentrations shown on the y-axis. The headings indicate the investigated metabolites, followed by Spearman's correlation, and the selected spectral region in ppm. Abbr.: TMAO, trimethylamine-N-oxide.
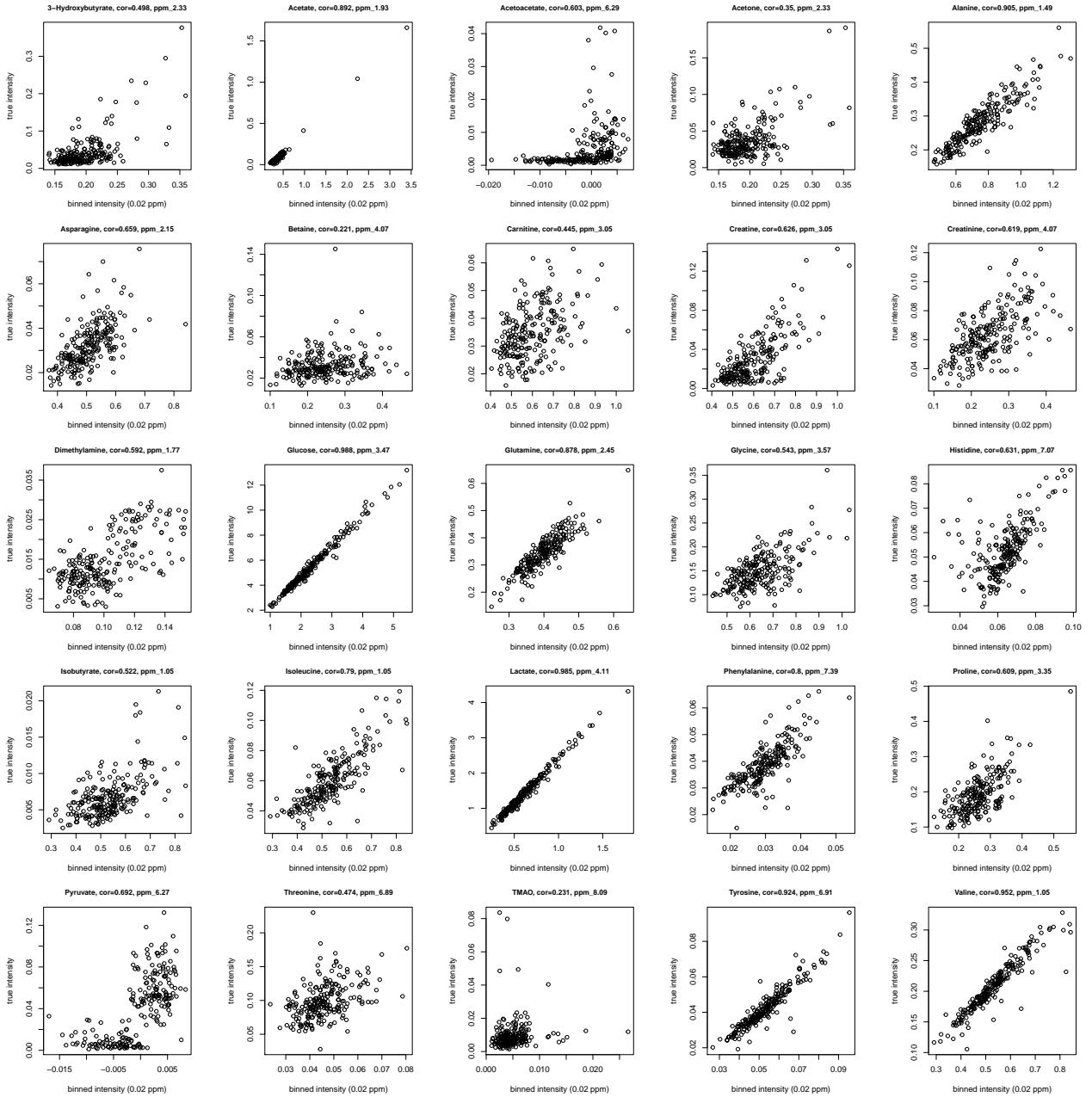
Figure S9: Comparison of the integrals of spectral features from an equidistant binning (bin size = 0.02 ppm) with absolutely quantified metabolite concentrations for 25 metabolites. The x-axis gives the spectral intensities returned by the binning which correlate best with the absolute concentrations shown on the y-axis. The headings indicate the investigated metabolites, followed by Spearman's correlation, and the selected spectral region in ppm. Abbr.: TMAO, trimethylamine-N-oxide.
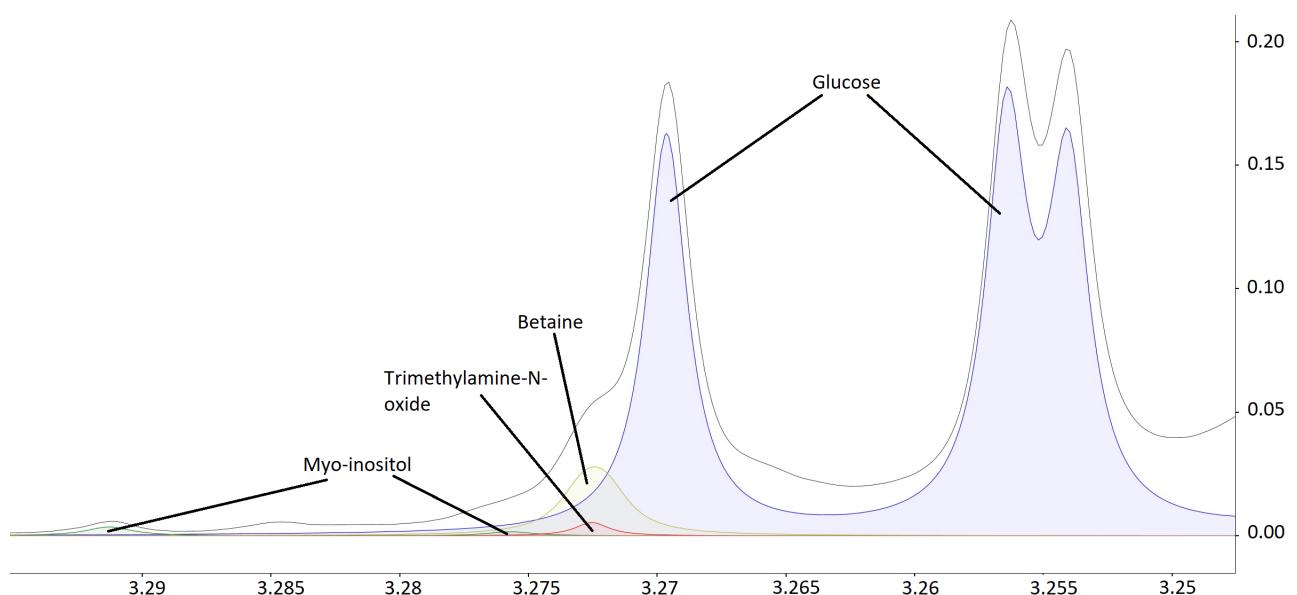
Figure S10: Exemplary NMR spectral region from 3.29 to 3.25 ppm, comprising metabolite signals of D-glucose, betaine, trimethylamine-N-oxide, and myo-inositol. One exemplary GCKD NMR spectrum is shown in black, reference spectra of the pure compounds D-glucose, betaine, trimethylamine-N-oxide, and myo-inositol manually fitted to the GCKD NMR spectrum, are shown in blue, light green, red, and dark green, respectively.
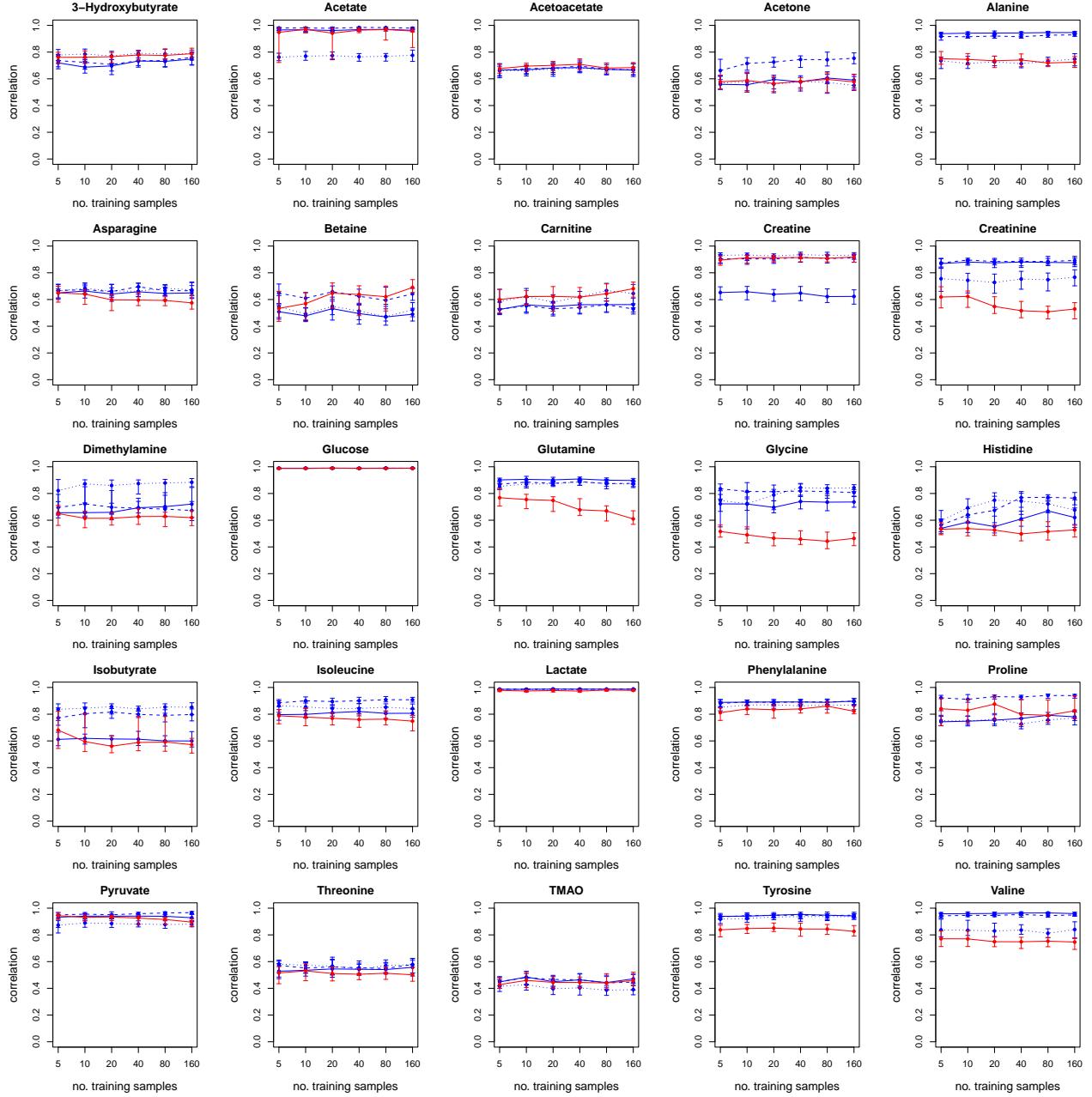
Figure S11: Spearman's correlations between absolutely quantified metabolite concentrations and integrals of corresponding spectral features from different binning approaches in dependence of the number of training samples n for 25 absolutely quantified metabolites in the GCKD data set. The points give the median Spearman's correlation obtained on test data and the whiskers the corresponding 25% and 75% quartiles. The dotted, dashed, and solid blue lines correspond to BF with $\lambda = 1$, $\lambda = 2.5$, and $\lambda = 5$, respectively, the solid red line to JBA. The metabolite identity is given in the header of the plots. Abbr.: TMAO, trimethylamine-N-oxide.

14
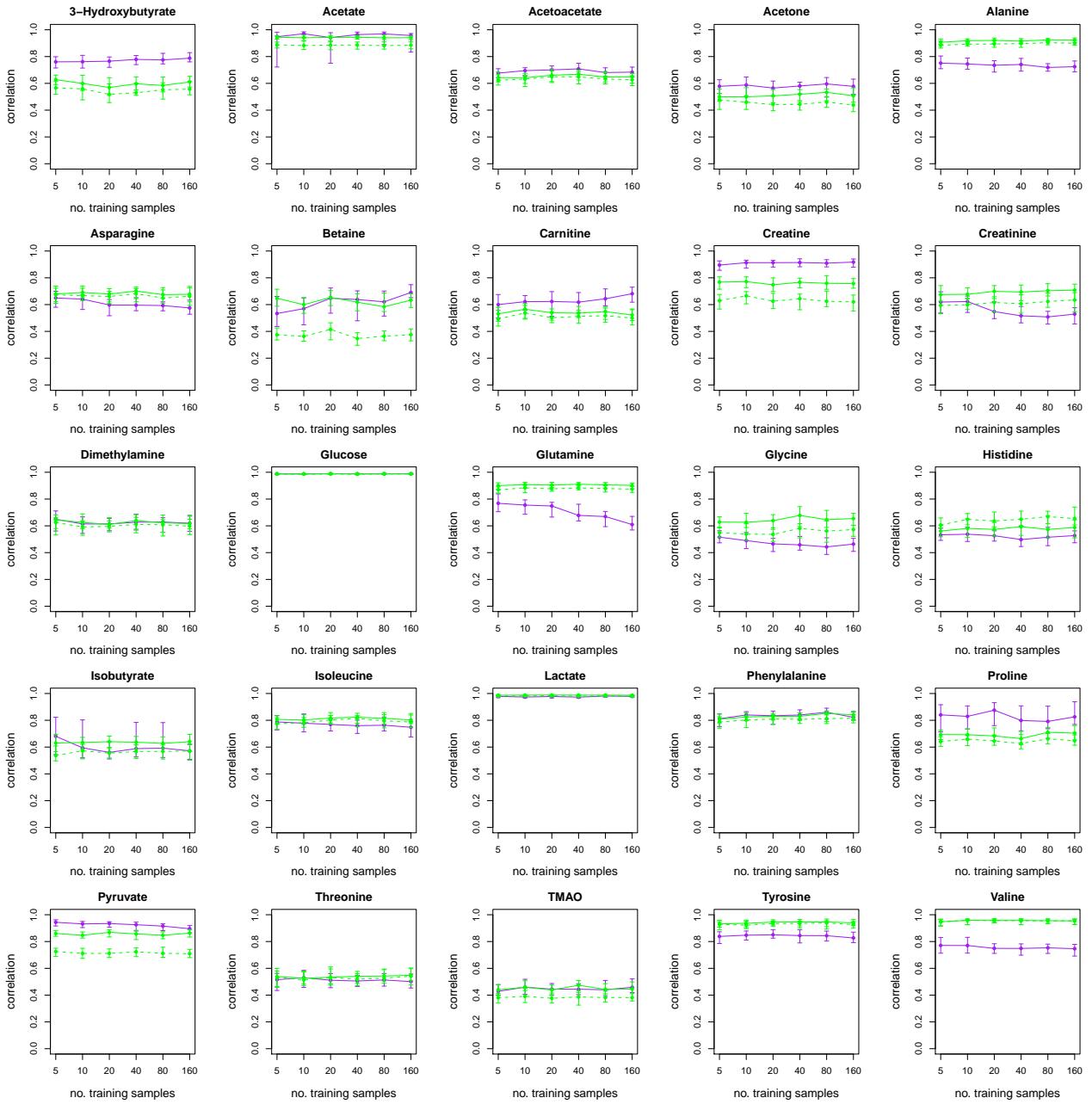
Figure S12: Spearman's correlations between absolutely quantified metabolite concentrations and integrals of corresponding spectral features from different binning approaches in dependence of the number of training samples n for 25 absolutely quantified metabolites in the GCKD data set. The points give the median Spearman's correlation obtained on test data and the whiskers the corresponding 25% and 75% quartiles. The solid purple lines correspond to SRV and the green solid and dashed lines to an equidistant binning with bin size 0.01 ppm and 0.02 ppm, respectively. The metabolite identity is given in the header of the plots. Abbr.: TMAO, trimethylamine-N-oxide.
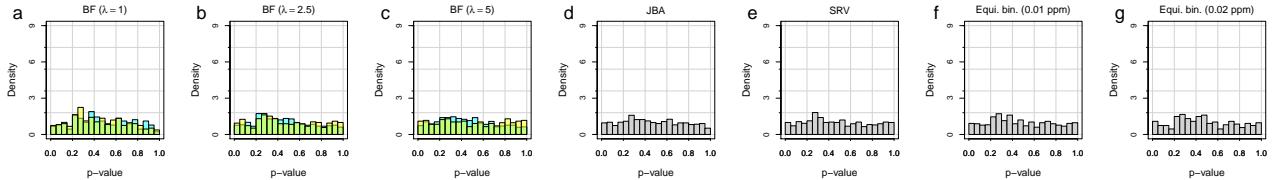
Figure S13: *P*-value distributions for the comparison AKI versus non-AKI after cardiac surgery based on permuted urine data. Figure (a) to (g) correspond to the different binning approaches, BF with $\lambda = 1$, BF with $\lambda = 2.5$, BF with $\lambda = 5$, JBA, SRV, equidistant binning with bin size 0.01 ppm, and equidistant binning with bin size 0.02 ppm, respectively. For the BF method, the same color coding as, e.g., in Figure 1 was applied. Please note that light green bars correspond to the overlap of cyan and yellow bars.
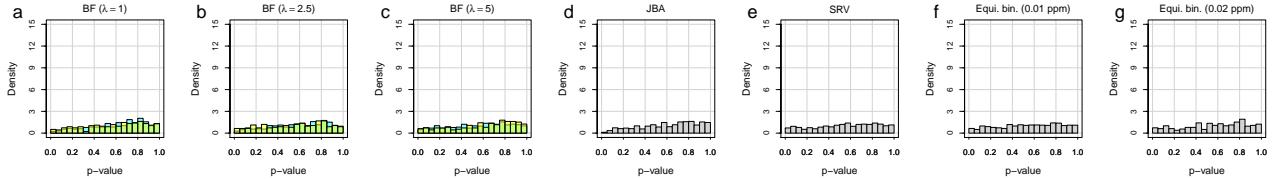

Figure S14: *P*-value distributions for the comparison AKI versus non-AKI after cardiac surgery based on permuted plasma data. Figure (a) to (g) correspond to the different binning approaches, BF with $\lambda = 1$, BF with $\lambda = 2.5$, BF with $\lambda = 5$, JBA, SRV, equidistant binning with bin size 0.01 ppm, and equidistant binning with bin size 0.02 ppm, respectively. For the BF method, the same color coding as, e.g., in Figure 1 was applied. Please note that light green bars correspond to the overlap of cyan and yellow bars.
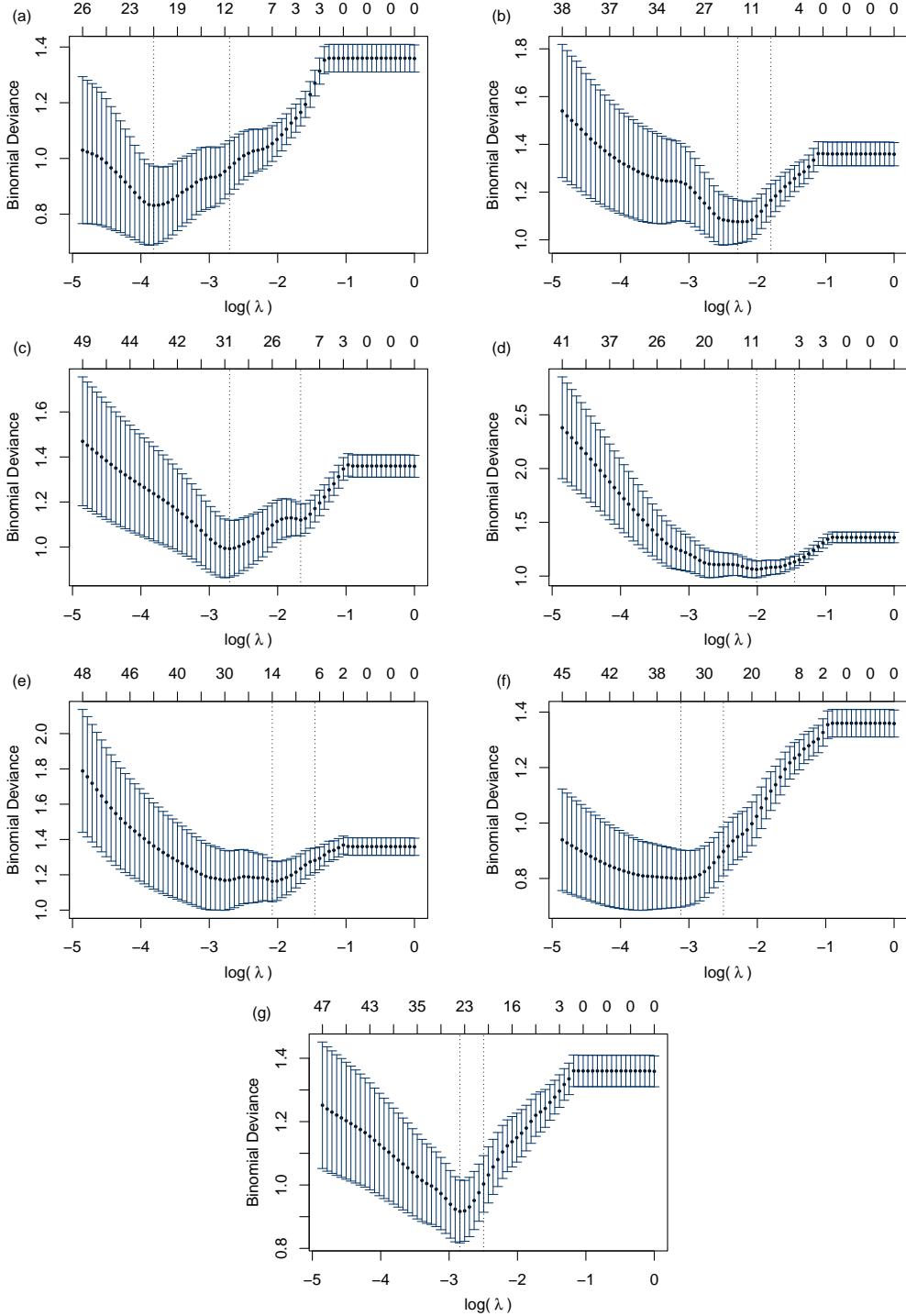
Figure S15: Hyper-parameter calibration of binomial zero-sum regression for the plasma AKI data set. Figures correspond to BF ($\lambda = 1$) (a), BF ($\lambda = 2.5$) (b), BF ($\lambda = 5$) (c), JBA (d), SRV (e), EB (0.01 ppm) (f), and EB (0.02 ppm) (g). The x-axis gives $\log(\lambda)$ corresponding to the size of the l_1 regularization term of zero-sum regression (not to be confused with the BF regularization) and the y-axis gives the binomial deviance in a leave-one-out cross validation. The values at the top of the figures give the number of selected NMR features.
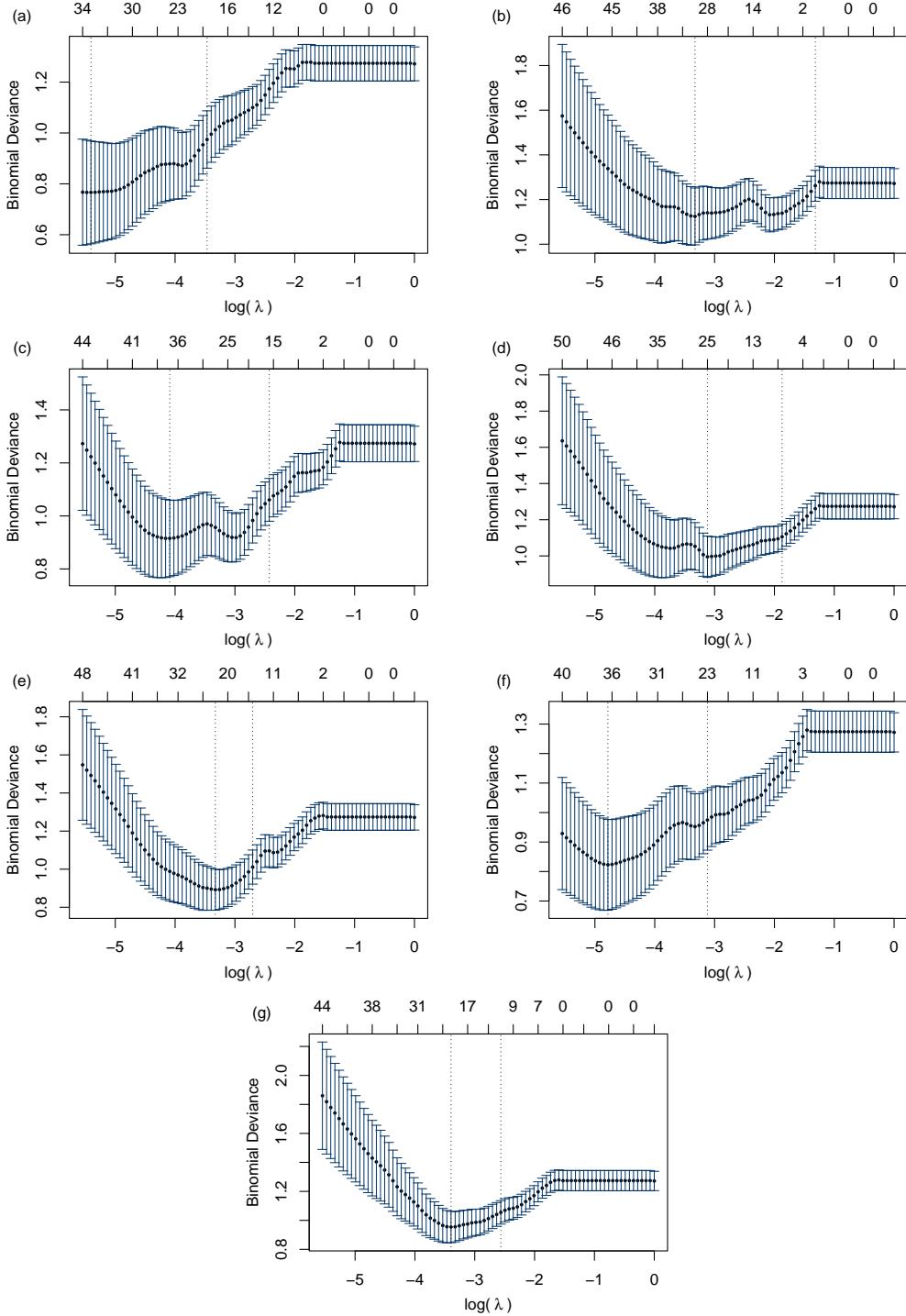
Figure S16: Hyper-parameter calibration of binomial zero-sum regression for the urinary AKI data set. Figures correspond to BF ($\lambda = 1$) (a), BF ($\lambda = 2.5$) (b), BF ($\lambda = 5$) (c), JBA (d), SRV (e), EB (0.01 ppm) (f), and EB (0.02 ppm) (g). The x-axis gives $\log(\lambda)$ corresponding to the size of the l_1 regularization term of zero-sum regression (not to be confused with the BF regularization) and the y-axis gives the binomial deviance in a leave-one-out cross validation. The values at the top of the figures give the number of selected NMR features.

5 Relationship between the choice of logarithmic base and the regularization parameter λ

Data values on a logarithmic scale with base b can be translated to base b' by

$$\log_b(y_{ij}) = \frac{\log_{b'}(y_{ij})}{\log_{b'}(b)} \ .$$

With $Y_{ij} = \log_b(y_{ij})$, $Y'_{ij} = \log_{b'}(y_{ij})$ and $c = \log_{b'}(b)$, we obtain

$$Y_{ij} = \frac{1}{\log_{b'}(b)} Y'_{ij} = \frac{1}{c} Y'_{ij} \ .$$

Therefore, Eq. (1) from the main manuscript becomes

$$
\begin{aligned}
\hat{B} &= \arg\min_B \left\{ ||Y - B||_F^2 + \lambda\sqrt{n} \sum_{j=1}^{p-1} ||B_{\cdot j} - B_{\cdot j+1}||_2 \right\} \\
&= \arg\min_B \left\{ ||\frac{1}{c}Y' - B||_F^2 + \lambda\sqrt{n} \sum_{j=1}^{p-1} ||B_{\cdot j} - B_{\cdot j+1}||_2 \right\} \\
&= \arg\min_B \left\{ ||Y' - cB||_F^2 + c^2\lambda\sqrt{n} \sum_{j=1}^{p-1} ||B_{\cdot j} - B_{\cdot j+1}||_2 \right\} \\
&= \arg\min_B \left\{ ||Y' - cB||_F^2 + c\lambda\sqrt{n} \sum_{j=1}^{p-1} ||cB_{\cdot j} - cB_{\cdot j+1}||_2 \right\}
\end{aligned}
$$

(S1)

Thus, an estimate of B' which corresponds to the input data Y' can be obtained via

$$
\hat{B}' = \arg\min_B \left\{ ||Y' - B'||_F^2 + \lambda'\sqrt{n} \sum_{j=1}^{p-1} ||B'_{\cdot j} - B'_{\cdot j+1}||_2 \right\}
$$

(S2)

with $\lambda' = c\lambda = \log_{b'}(b)\lambda$. As outlined in the main manuscript, regularization values of $\lambda = \{1, 2.5, 5\}$ performed well given the specific input data (of given resolution and quality) and that the data were $\log_2$ transformed. If data are provided using a different logarithmic base, a re-adjustment of regularization parameters might be necessary. In our case this suggests $\lambda' = \log_{10}(2)\{1, 2.5, 5\} = \{0.301, 0.753, 1.505\}$ for $\log_{10}$-transformed data.

6 Convergence analyses of the BF algorithm

We confirmed the convergence of the BF algorithm for different step sizes ρ (Suppl. Fig. S17) and evaluated respective residuals between model fits (Suppl. Fig. S18) on the GCKD plasma data set. Details are summarized in the figure captions.
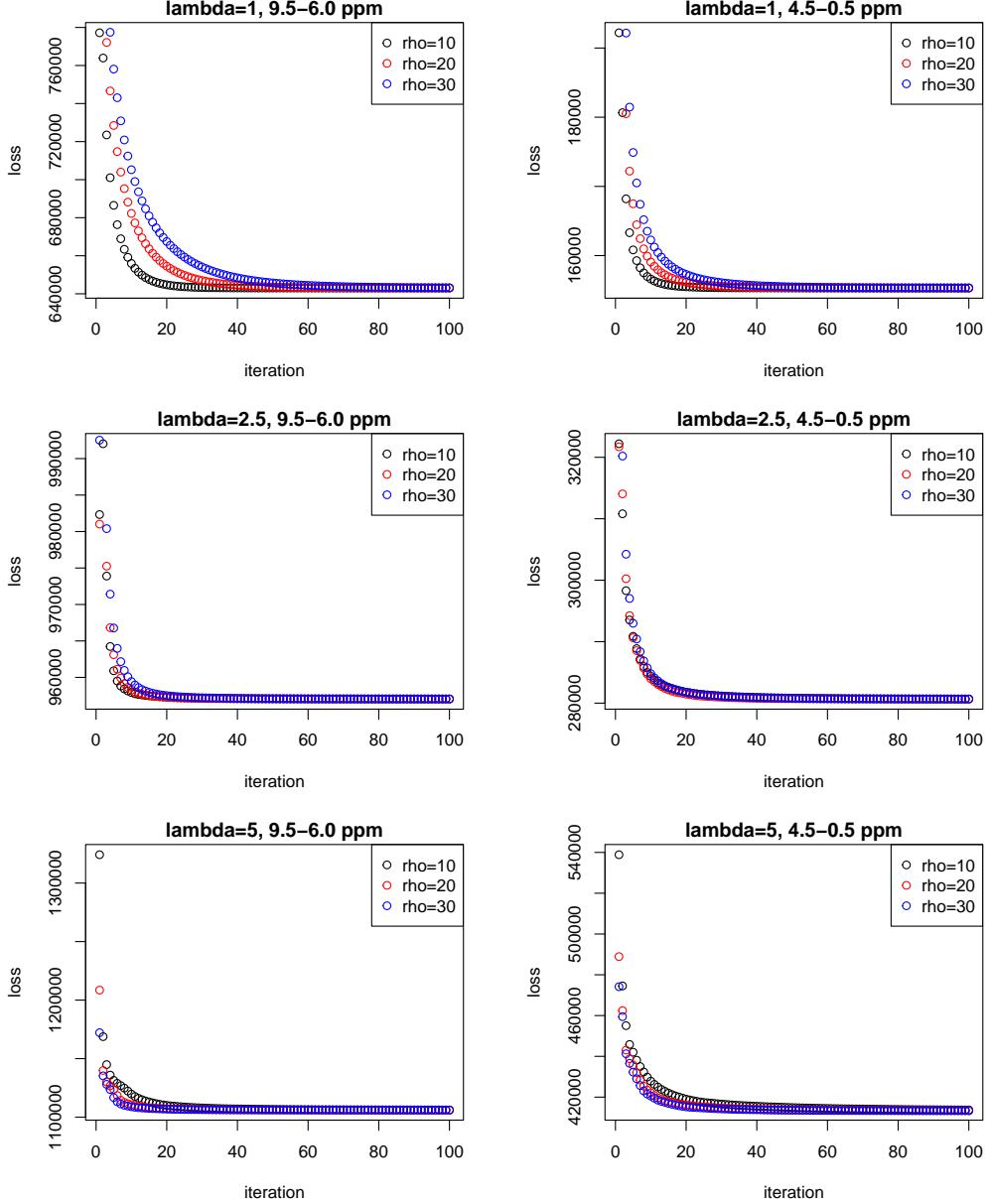

Figure S17: Losses (y-axis) versus iteration steps (x-axis). The left column corresponds to the Bucket Fuser optimization using data in the spectral region from 9.5 to 6.0 ppm and the right column to the region from 4.5 to 0.5 ppm. The rows correspond to different regularization parameters λ; $\lambda = 1$ (first row), $\lambda = 2.5$ (second row), $\lambda = 5$ (third row). The colors correspond to different step sizes ρ; $\rho = 10$ (black circles), $\rho = 20$ (red circles), and $\rho = 30$ (blue circles).
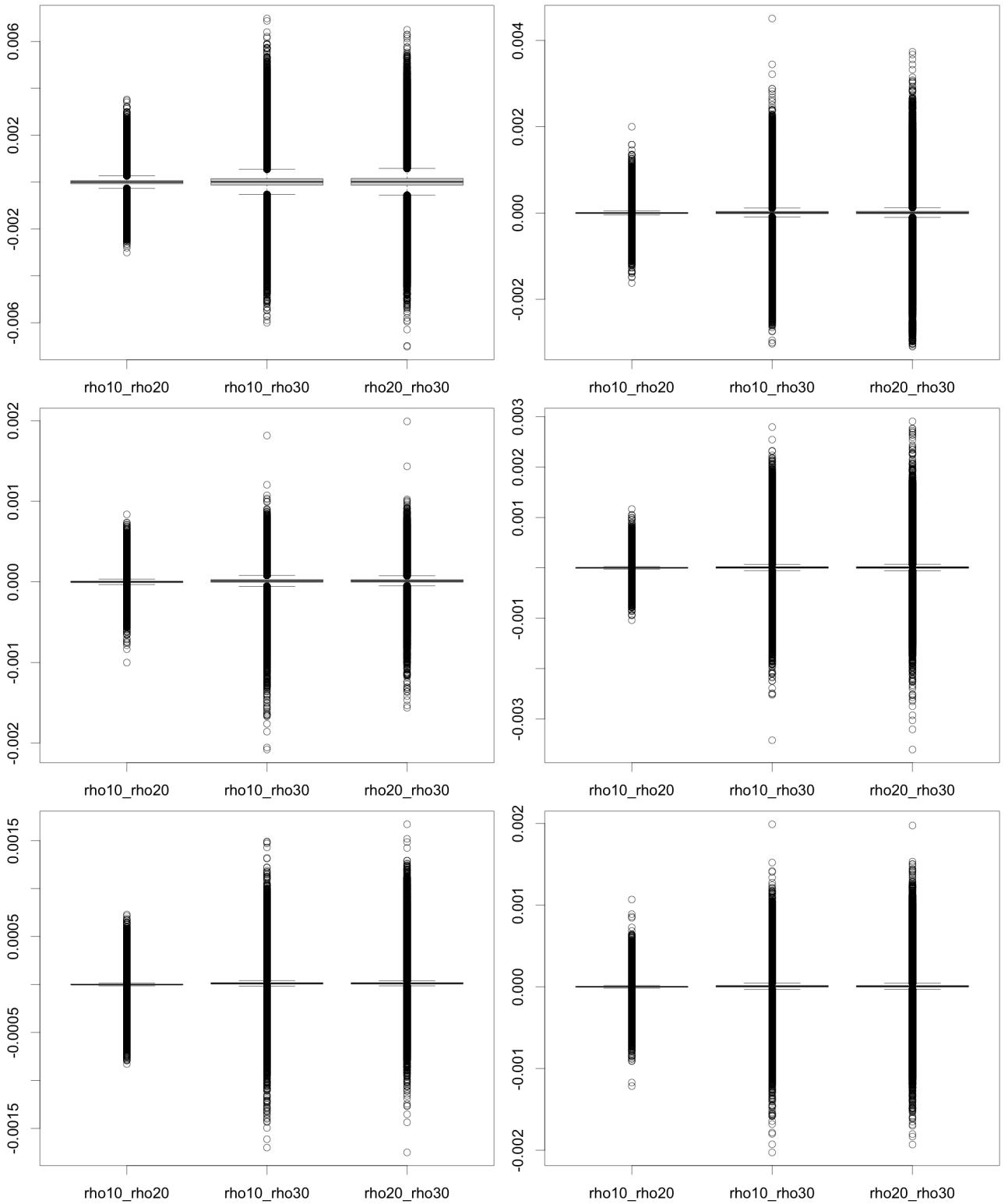
Figure S18: Residuals between different estimates of B using step sizes of $\rho \in \{10, 20, 30\}$. The sub-figures correspond to Fig. S17; input data from 9.5 to 6.0 ppm (left column), input data from 4.5 to 0.5 ppm (right column), $\lambda = 1$ (first row), $\lambda = 2.5$ (second row), $\lambda = 5$ (third row). The boxplots represent from left to right the residuals $B(\rho = 10) - B(\rho = 20)$, $B(\rho = 10) - B(\rho = 30)$, and $B(\rho = 20) - B(\rho = 30)$ in each of the subplots.