Supplementary material of the article "ASICS: an R package for a whole analysis workflow of 1D ^{1}H NMR spectra"

G. Lefort *et al.*

Contents

S1 Post-quantification figures available in ASICS

The following figures are examples of graphical representations available in **ASICS**. The post-quantification analyses leading to these figures are detailled in the Section 2.4 of the article.

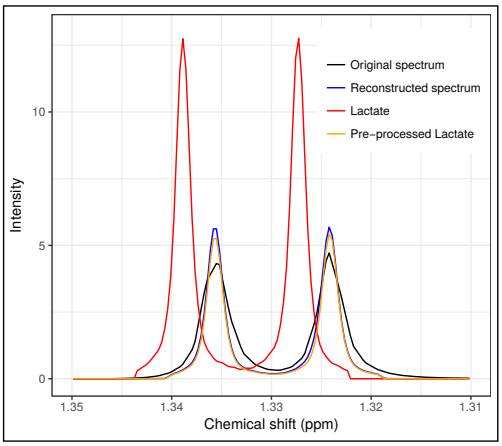

Fig. S1. Zoom in the diagnostic plot of the quantification to visually access the quality of the quantification of one of the lactate peaks and of the reconstructed spectrum, as compared to the original complex mixture spectrum.

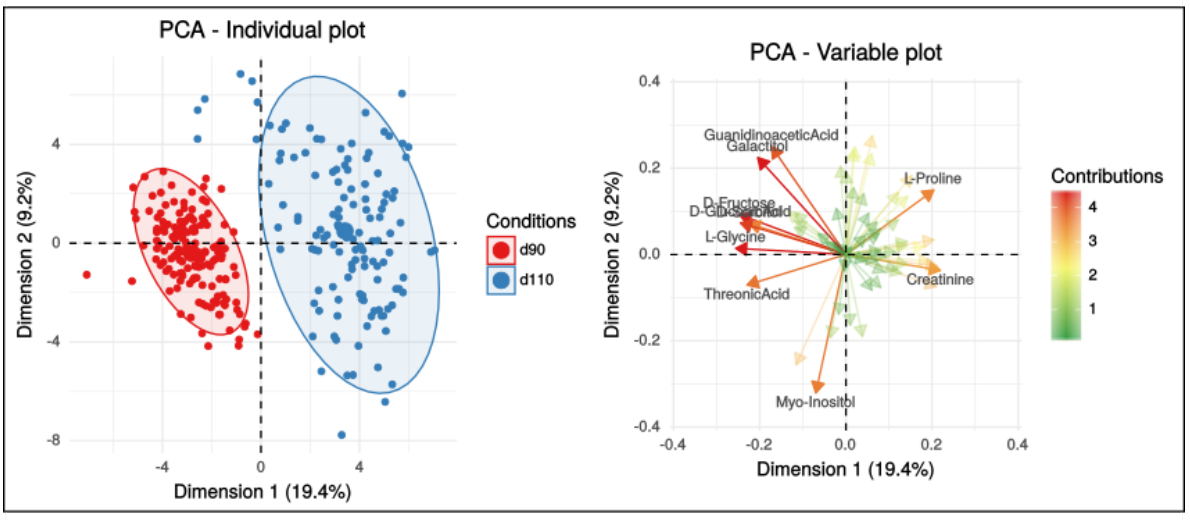

Fig. S2. PCA plots on **ASICS** quantifications for the study on plasmatic metabolome at the end of gestation in piglets. Left: individuals. Right: variables.

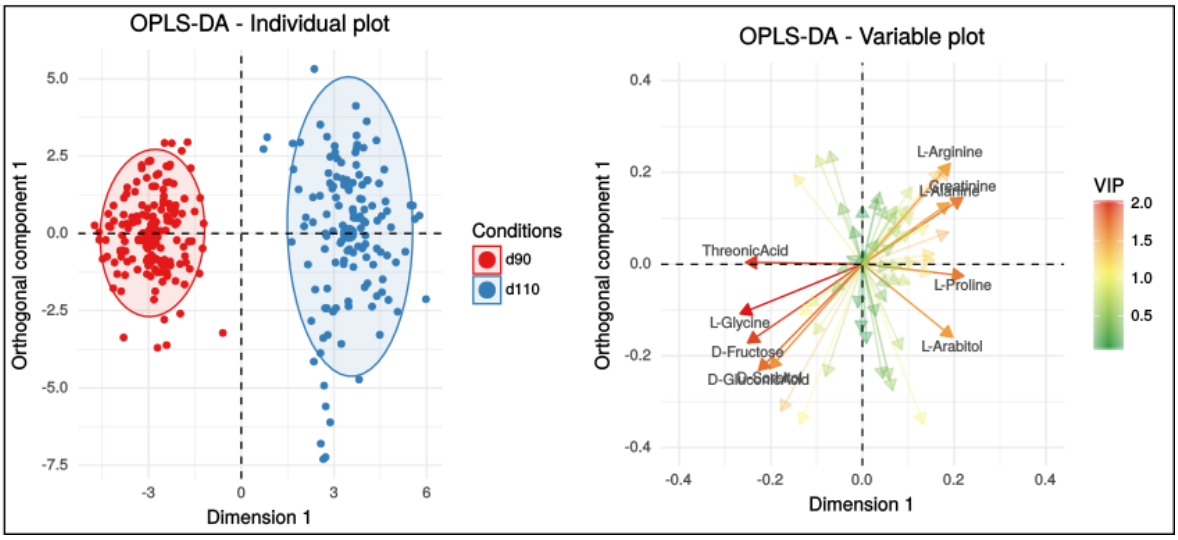

Fig. S3. OPLS-DA plots on **ASICS** quantifications for the study on plasmatic metabolome at the end of gestation in piglets. Left: individuals. Right: variables.

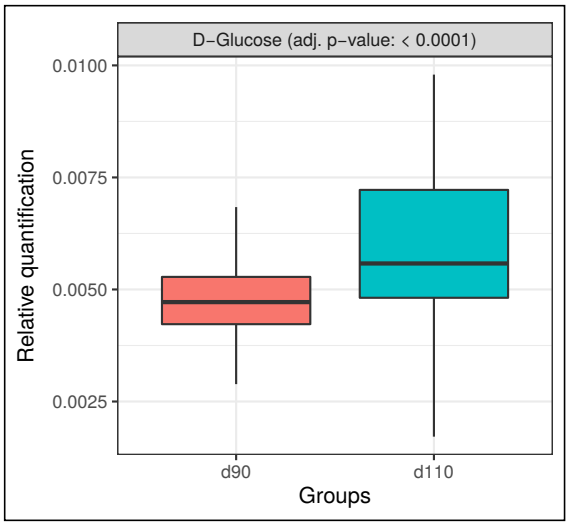

Fig. S4. Boxplots of the estimated glucose quantifications at the two gestational ages for the study on plasmatic metabolome in piglets.

S2 Description of study and data on plasmatic metabolome at the end of gestation in piglets

The PORCINET project (ANR-09-GENM-005) proposed to study the fetal development in late gestation in pigs. The experiment authorization number for the experimental farm GenESI (Genetics, testing and innovative systems experimental unit) is A 17661. The procedures performed in this study and the treatment of animals complied with European Union legislation (Directive 2010/63/EU) and French legislation in the Midi-Pyrénées Region of France (Decree 2001-464). The ethical committee of the Midi-Pyrénées Regional Council approved the experimental design (authorization MP/01/01/01/11). The experimental design was previously described in Voillet *et al.* (2014). In the present article, only Large White fetuses from PORCINET were taken into account. A total of 283 piglets collected at two gestational ages were considered. All sows ($n = 20$) were anesthetized

at 90 days (n = 155 fetuses) or 110 days (n = 128 fetuses) after conception (average gestation term: 114 days). Their fetuses were quickly obtained by caesarean section. Blood (approximately 5 mL) was immediately collected from the umbilical artery via a 21-gauge needle and a 5-mL syringe and placed in heparinized tubes. Plasma was prepared by low-speed centrifugation (2,000×g for 10 min at 4°C) and stored at −20°C until further analysis. Glucose, fructose, and lactate were chosen as indicators of carbohydrate metabolism. Methods and results for blood parameters quantification are described in Table S1.

Table S2. Blood parameter concentration from umbilical artery of purebred fetuses at 90 days and 110 days of gestation

Gestational age (days)	**90**	**110**	**Kruskal-Wallis p-value**	**Method used for analysis**
Arterial glucose, mmol/L	1.76 ± 0.16	2.07 ± 0.52	0.184	Enzymatically (Glucose RTU kit: #61269, Biomérieux, Marcy l'étoile, France) (Gondret *et al.*, 2013)
Arterial fructose, mmol/L	6.05 ± 0.80	2.44 ± 1.36	0.007	Enzymatically (D-Fructose kit: #984302, Thermo Fisher Scientific, Vantaa, Finland) (Gondret *et al.*, 2013)
Arterial lactate, mmol/L	2.24 ± 0.55	3.40 ± 3.55	0.264	Enzymatically (Lactate PAP kit: #61192, Biomérieux, Marcy l'étoile, France) (Gondret *et al.*, 2013)

For proton nuclear magnetic resonance (^{1}H NMR) spectroscopy analysis, sample preparation was performed as follows: D2O (500 μL) was added to plasma (200 μL) and mixed, the sample was then centrifuged for 10 min at 3,000 × g at room temperature, and the supernatant (600 μL) was transferred to 5-mm nuclear magnetic resonance (NMR) tubes for ^{1}H NMR analysis. All ^{1}H NMR spectra were acquired on a Bruker Avance DRX-600 spectrometer (Bruker SA, Wissembourg, France) operating at 600.13 MHz for ^{1}H resonance frequency and equipped with a pulsed-field gradients z system, an inverse ^{1}H−^{13}C−^{15}N cryoprobe attached to a cryoplatform (the preamplifier cooling unit), and a temperature control unit maintaining the sample temperature at 300 ± 0.1°K. The ^{1}H NMR spectra of plasma samples were acquired at 300K using the Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse sequence with presaturation with a total spin-echo delay ($2n\pi$) of 240 ms to attenuate broad signals from proteins and lipoproteins, which otherwise display a wide signal and hide the narrower signals of low molecular weight metabolites. The ^{1}H signal was acquired by accumulating 128 transients over a spectral width of 20 ppm (note: chemical shift units kept ppm), collecting 32,000 data points. The interpulse delay of the CPMG sequence was set at 0.4 ms with $n = 300$ as defined in the following sequence: [90-(τ-180-τ)n acquisition]. A 2-s relaxation delay was applied. The Fourier transformation was calculated on 64,000 points. All ^{1}H NMR spectra were phased and baseline corrected. The ^{1}H chemical shifts were calibrated on the resonance of lactate at 1.33 ppm. Then plasma spectra were data-reduced before statistical analysis using AMIX software (Analysis of Mixtures version 3.8; Bruker Analytische Messtechnik; Rheinstetten, Germany). The spectral region δ 0.5 to 10.0 ppm was segmented into consecutive non overlapping regions of 0.01 ppm (buckets) and normalized according to the total signal intensity in every spectrum. The region around δ 4.8 ppm corresponding to water resonance (5.1–4.5 ppm) was excluded from the pattern recognition analysis to eliminate artifacts of residual water.

S3 Quantification methods comparison

Table S3. An overview ofopen source NMR data processing solutions. Name

Name	**Software**	**Pre-processing**	**Alignment**	**Identification**	**Quantification**	**Data analysis**	**Parallel environment**	**Computational time**
ASICS	R	Yes	Yes	Yes	Yes	Yes	Yes	$\sim$ 1'30 min
Autofit (Weljie *et al.*, 2006)	Chenomx	Yes	No	Yes	Yes	No	No	$<$ 1min
batman (Hao *et al.*, 2012)	R	No	No	Yes	Yes	No	Yes	$\sim$ 2 days
Bayesil (Ravanbakhsh *et al.*, 2015)	Web	Yes	No	Yes	Yes	No	No	$\sim$ 10 min
rDolphin (Cañueto *et al.*, 2018)	R	No	No	Yes	Yes	No	No	$\sim$ 1'30 min

S4 Supplementary results

S4.1 Differences between gestational ages (day 90 and day 110) of fetuses

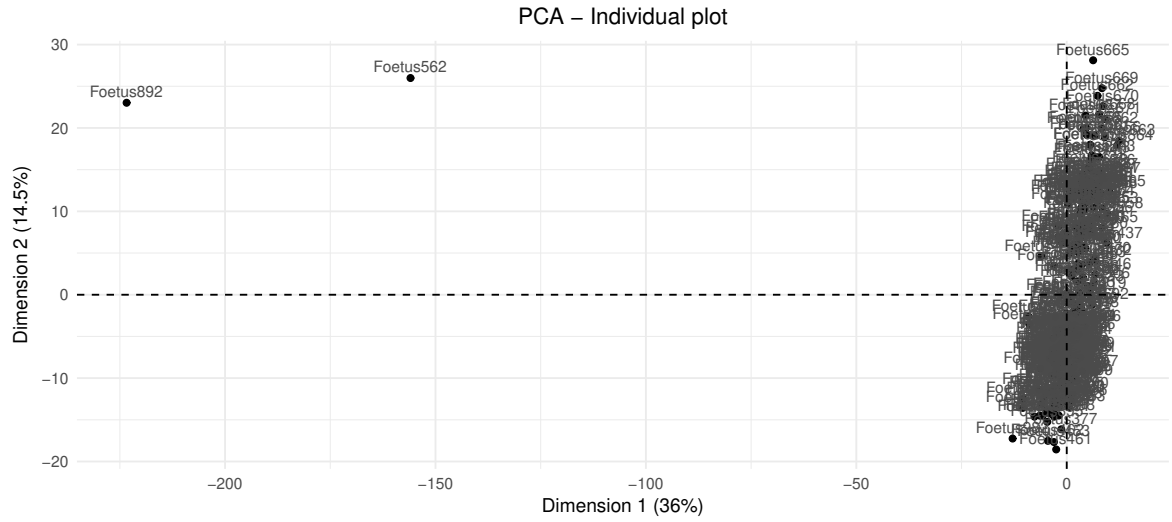

Fig. S5. PCA on buckets (axes 1 and 2, projection of individuals). Two outliers are identified that were removed from the analysis.

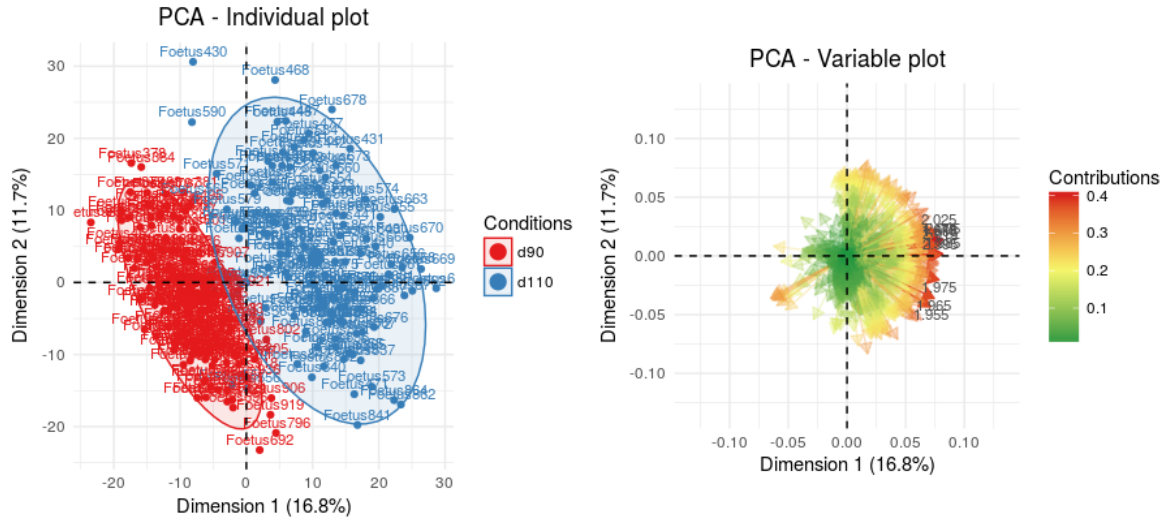

Fig. S6. PCA on buckets (axes 1 and 2, after the two outliers have been removed). Left: individuals. Right: variables.

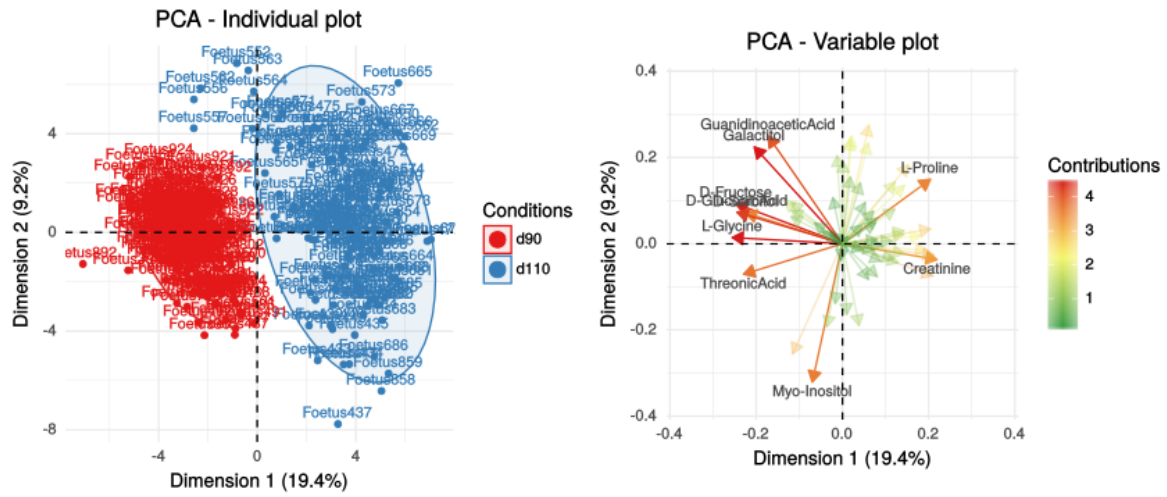

Fig. S7. PCA on **ASICS** quantification (axes 1 and 2, after the two outliers have been removed). Left: individuals. Right: variables.

Table S4. Metabolites selected by OPLS-DA as relevant to discriminate ages of gestation for both approaches. [a] buckets for metabolites extracted with the bucket approach and identified by an expert. [b] VIP for metabolites extracted with the **ASICS** approach.

Metabolite	**Buckets**[a]	**VIP**[b]	**Change at 110 days**
2-Oxoisovalerate	[1.12, 1.14]		↗
3-Methyl-2-oxovaleric acid	[1.09, 1.11]		↗
Betaine	[3.26, 3.27]		↗
Citrate	[2.51, 2.53], [2.54, 2.57], [2.67, 2.72]		↗
Creatine	[3.04, 3.05]		↗
Creatinine	[3.04, 3.05]	1.71	↗
D-Fructose	[3.55, 3.56], [3.57, 3.58], [3.69, 3.73], [3.78, 3.82], [3.89, 3.90], [3.98, 4.05]	1.96	↘
D-Gluconic acid		1.77	↘
D-Glucose	[3.23, 3.26], [3.38, 3.51], [3.74, 3.78], [5.22, 5.25]	1.04	↗
D-Glucose-6-Phosphate		1.07	↘
D-Sorbitol		1.58	↘
Galactitol		1.39	↘
Glycerophosphocholine	[4.28, 4.35]		↗
Guanidinoacetic acid		1.07	↘
Isovaleric acid		1.13	↗
L-Alanine	[1.46, 1.49]	1.48	↗
L-Arabitol		1.54	↗
L-Arginine		1.49	↗
L-Glutamic acid	[2.00, 2.09], [2.33, 2.38]		↘
L-Glycine		2.08	↘
L-Isoleucine	[0.92, 0.95], [1.00, 1.02]		↗
L-Leucine	[0.95, 0.98]	1.48	↗
L-Lysine	[1.68, 1.76], [2.99, 3.03]		↗
L-Phenylalanine	[7.31, 7.38], [7.40, 7.42]		↗
L-Proline	[1.96, 2.00], [3.32, 3.36]	1.72	↗
L-Threonine	[4.26, 4.28]	1.47	↗
L-Tyrosine	[6.88, 6.90], [7.18, 7.20]		↗
L-Valine	[0.98, 1.00], [1.03, 1.05], [2.20, 2.22], [2.25, 2.30]	1.32	↗
Lactate	[1.32, 1.35], [4.09, 4.14]	1.23	↗
Lipids	[0.88, 0.89], [0.90, 0.92]		↗
N-Acetylglycine		1.23	↗
S-Acetamidomethylcysteine		1.17	↗
Threonic acid		1.99	↘
Xylitol		1.17	↘
Unidentified buckets	72 buckets	-	-

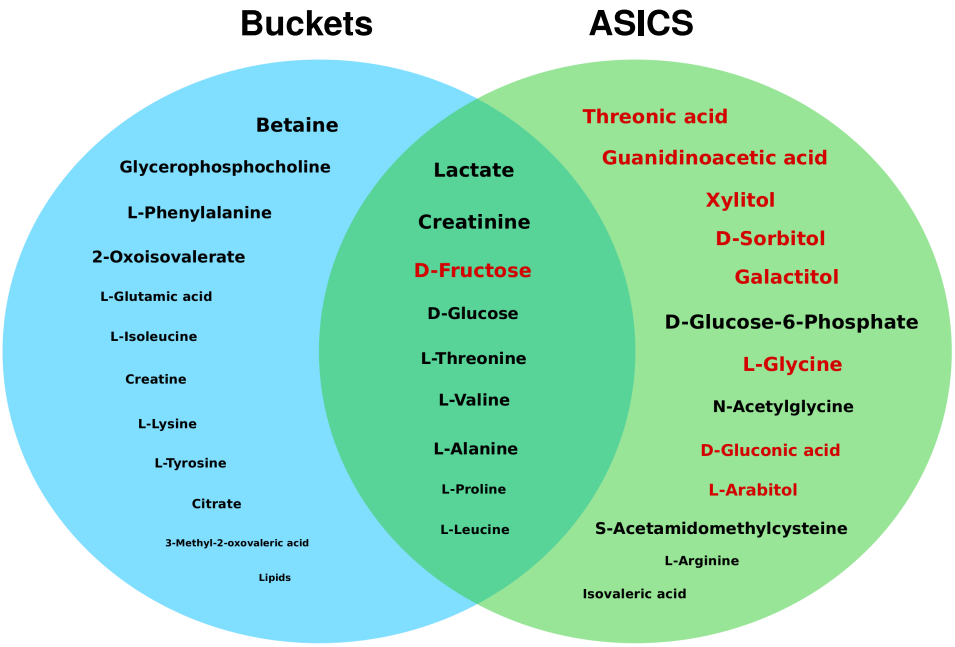

Fig. S8. Venn diagram comparing selected metabolites from analyses made on buckets (left) and on **ASICS** quantifications (right). Font size corresponds to average intensity of the associated buckets. A name is written in red if all peaks for this metabolite fall in the 3.5–4.2 ppm region (a region with a high density of peaks).

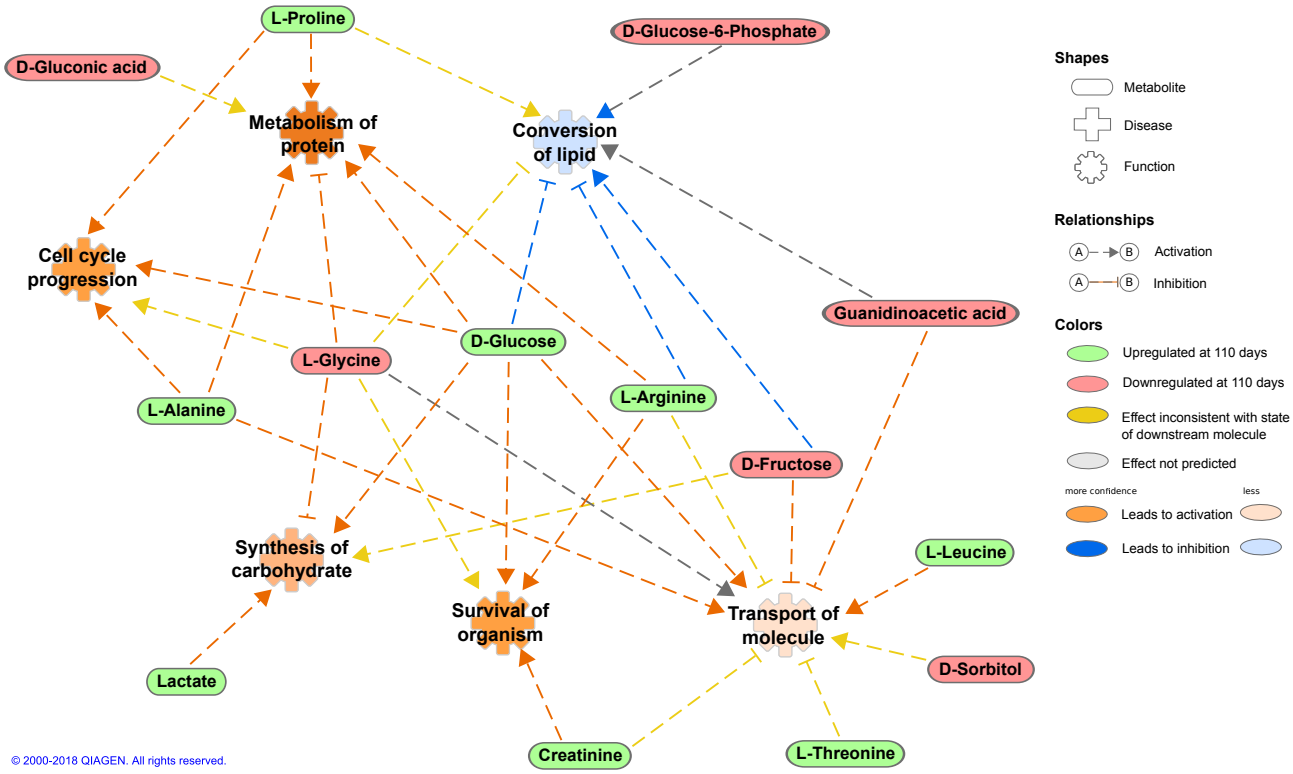

Fig. S9. Metabolomic pathway based on the metabolites identified by **ASICS** as obtained with Ingenuity Pathway Analysis© (IPA©, Ingenuity Systems; QIAGEN, Inc., Valencia, CA, USA, `https://analysis.ingenuity.com/pa`). IPA contains a large bibliographic database (Ingenuity Pathways Knowledge Base©). 13 out of 22 of the identified metabolites are present in the network, among which 6 (guanidinoacetic adic, sorbitol, glucose-6-phosphate, glycine, gluconic acid, and arginine) were identified only by **ASICS**.

S4.2 Differences between groups (T2DM)

Table S5. Metabolites selected by OPLS-DA as relevant to discriminate T2DM patients for both approaches. [a] buckets for metabolites extracted with the bucket approach and identified by an expert. [b] VIP for metabolites extracted with the **ASICS** approach.

Metabolite	**Buckets**[a]	**VIP**[b]	**Change in T2DM**	**Change in T2DM in Salek *et al.* (2007)**	**Change in T2DM in Yousri *et al.* (2015)**
2PY	[6.66, 6.69], [8.32, 8.34]		↗	↗	
2-Deoxycytidine		1.67	↘		
2-Oxoisovalerate	[1.13, 1.14]		↘	↗	
3-Hydroxybutyrate	[1.20, 1.22], [4.14, 4.17]		↗	↗	↗
Acetoacetate	[2.26, 2.27]	1.59	↗	↗	↗
Acetone		1.69	↗		
Allantoin	[5.37, 5.40]		↗	↘	
Betaine	[3.25, 3.26], [3.90, 3.91]	2.13	↗		
Butyrate	[0.88, 0.89], [1.50, 1.55], [2.12, 2.13], [2.15, 2.18]		↗	↗	
Creatinine	[3.04, 3.05], [4.05, 4.07]	1.53	↘	↘	↘
D-Fucose		1.29	↘		
D-Glucose	[3.38, 3.41], [3.45, 3.55], [3.83, 3.92], [5.22, 5.25]	1.41	↗		↗
D-Glucose-6-Phosphate		1.44	↗		
D-Mannose		1.72	↘		↗
Dihydrothymine		1.23	↗		
Dimethylglycine	[2.92, 2.94]		↗	↗	
Fumaric acid	[6.52, 6.54]		↘	↘	
GABA		1.26	↗		
Glycerol		2.13	↗	↗	
Guanidinoacetic acid	[3.78,3.79]	2.41	↘		
Hippuric acid	[3.94, 3.98], [7.54, 7.58], [7.62, 7.66], [7.82, 7.85]	1.49	↘	↘	
Indoxylsulfate	[7.20, 7.24], [7.26, 7.30], [7.69, 7.70]	1.64	↗	↗	
L-Alanine	[3.76, 3.77], [3.79, 3.80]	1.38	↗	↗	↗
L-Arabitol		2.41	↘		↘
L-Isoleucine	[0.94, 0.95], [1.00, 1.01], [1.25,1.27]		↗	↘	↗
L-Lysine	[1.37, 1.44], [1.70, 1.71], [1.74,1.75], [1.93, 1.96], [3.00, 3.02]		↗		
Lactate	[4.11, 4.14]		↗	↗	↗
Malic acid	[2.33, 2.35], [2.37, 2.39], [2.65, 2.66], [4.31, 4.32]		↘	↘	↗
Methanol		1.69	↘		
Phenylacetylglycine	[3.67, 3.70], [3.72, 3.75], [7.42, 7.47]		↗	↘	
Phosphocholine		1.37	↗		
Syringic acid		1.54	↗		
TMAO	[3.25, 3.26]		↗	↗	
Trigonelline	[4.40, 4.45], [8.05, 8.06], [8.08, 8.09], [8.82, 8.86], [9.12, 9.14]	1.35	↘	↘	
Uracil		1.86	↘		
Unidentified buckets	114 buckets	-	-	-	-

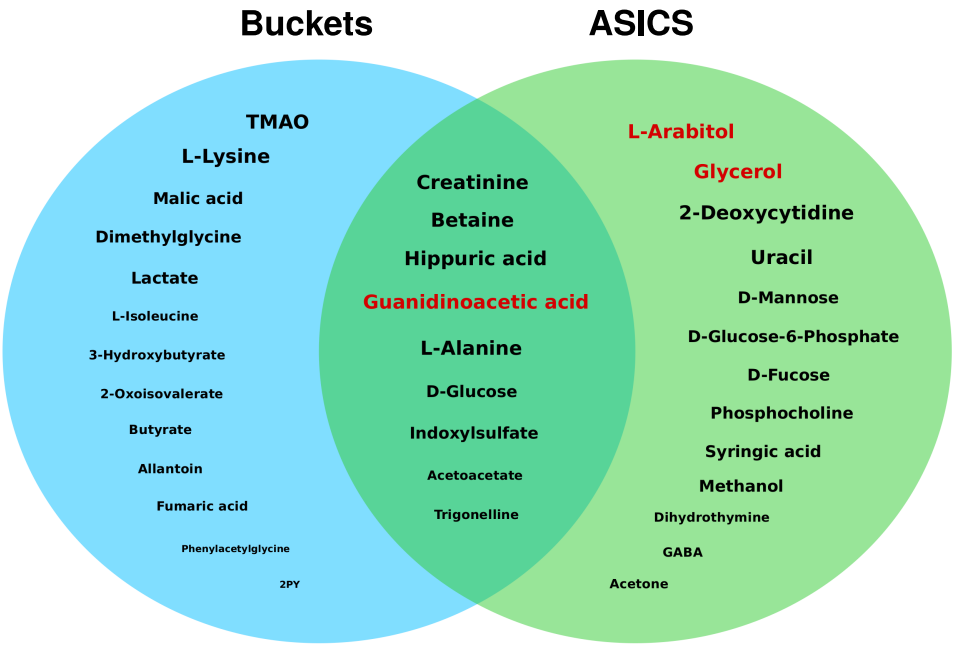

Fig. S10. Venn diagram comparing selected metabolites from analyses made on buckets (left) and on **ASICS** quantifications (right). Font size corresponds to average intensity of the associated buckets. A name is written in red if all peaks for this metabolite fall in the 3.5–4.2 ppm region (a region with a high density of peaks).

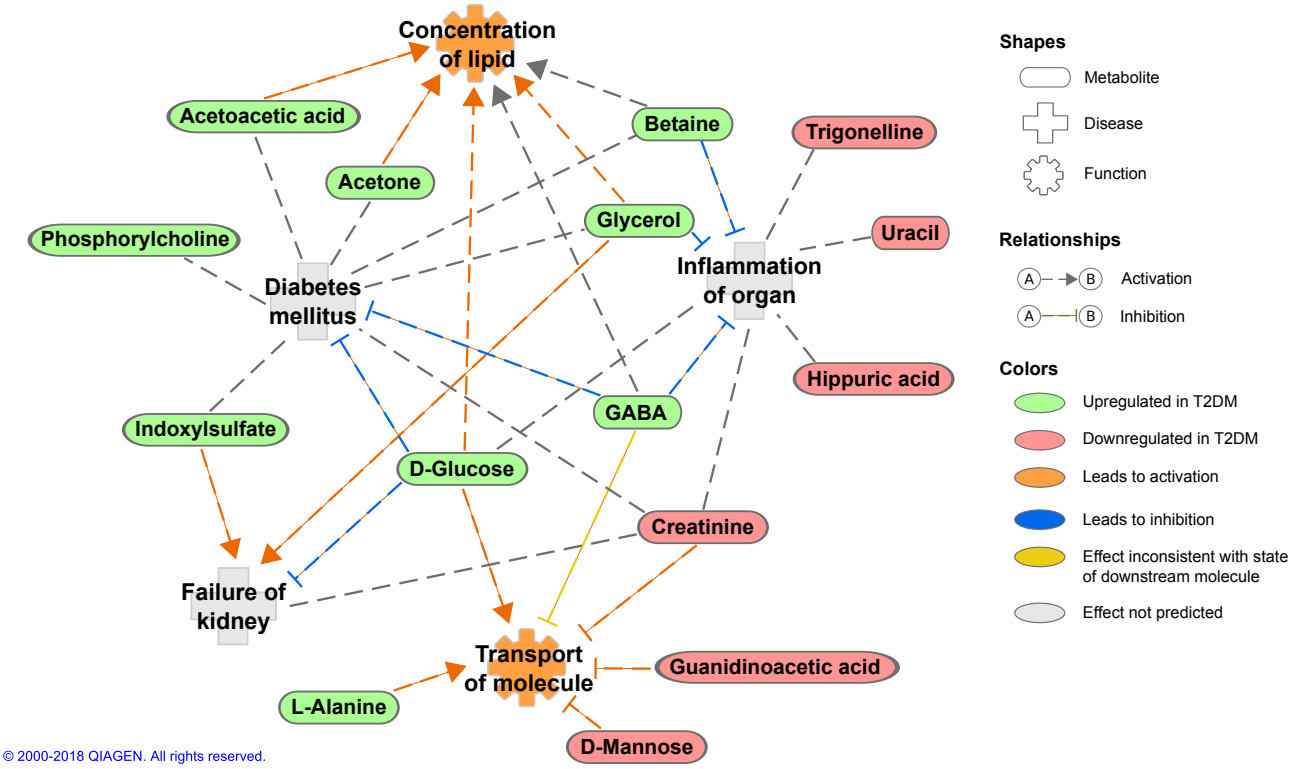

Fig. S11. Metabolomic pathway based on the metabolites identified by **ASICS** as obtained with Ingenuity Pathway Analysis© (IPA©, Ingenuity Systems; QIAGEN, Inc., Valencia, CA, USA, `https://analysis.ingenuity.com/pa`). IPA contains a large bibliographic database (Ingenuity Pathways Knowledge Base©). 15 out of 22 of the identified metabolites are present in the network, among which 6 (acetone, uracil, GABA, glycerol, phosphorylcholine, and mannose) were identified only by **ASICS**.

References

Cañueto, D. *et al.* (2018). rDolphin: a GUI R package for proficient automatic profiling of 1D 1H-NMR spectra of study datasets. *Metabolomics*, **14**(3), 24.

Gondret, F. *et al.* (2013). Spontaneous intra-uterine growth restriction modulates the endocrine status and the developmental expression of genes in porcine fetal and neonatal adipose tissue. *General and Comparative Endocrinology*, **194**, 208–216.

Hao, J. *et al.* (2012). BATMAN – an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics*, **28**(15), 2088–2090.

Ravanbakhsh, S. *et al.* (2015). Accurate, fully-automated NMR spectral profiling for metabolomics. *PLOS ONE*, **10**(5), e0124219.

Salek, R. *et al.* (2007). A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. *Physiological Genomics*, **29**(2), 99–108.

Voillet, V. *et al.* (2014). Muscle transcriptomic investigation of late fetal development identifies candidate genes for piglet maturity. *BMC Genomics*, **15**, 797.

Weljie, A. *et al.* (2006). Targeted profiling: quantitative analysis of 1 H NMR metabolomics data. *Analytical Chemistry*, **78**(13), 4430–4442.

Yousri, N. A. *et al.* (2015). A systems view of type 2 diabetes-associated metabolic perturbations in saliva, blood and urine at different timescales of glycaemic control. *Diabetologia*, **58**(8), 1855–1867.