Supplementary Information

NMR-Based Urine Metabolic Profiling of Cri-du-Chat Syndrome

Glenda S. de Oliveira,*[*,a] Danielle Z. S. Furtado,[b] Elenilson G. Alves Filho,[c] Alex Castro,[d] Fernando B. V. M. Leite,[e] Tiago Venâncio[f] and Nilson A. Assunção[b,e]

[a]*Laboratório de Química Biológica, Instituto de Química, Universidade Estadual de Campinas, 13083-862 Campinas-SP, Brazil*

[b]*Escola Paulista de Medicina, Universidade Federal de São Paulo, 04023-062 São Paulo-SP, Brazil*

[c]*Departamento de Engenharia de Alimentos, Universidade Federal do Ceará, 60020-181 Fortaleza-CE, Brazil*

[d]*Laboratório Nacional de Biociências (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), 13083-970 Campinas-SP, Brazil*

[e]*Departamento de Química, Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Universidade Federal de São Paulo, 09913-030 Diadema-SP, Brazil*

[f]*Departamento de Química, Universidade Federal de São Carlos, 13565-905 São Carlos-SP, Brazil*

*e-mail: glendaoliveira.res@gmail.com

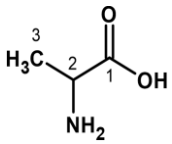
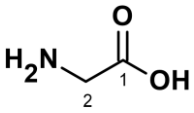
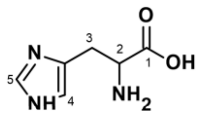
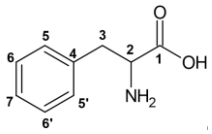
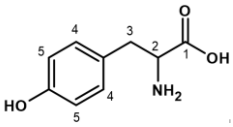
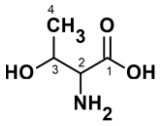
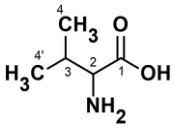
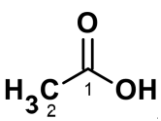
Glenda S. Oliveira: https://orcid.org/0000-0001-9528-9465
Danielle Z.S. Furtado: https://orcid.org/0000-0001-5653-494X
Elenilson G.A. Filho: https://orcid.org/0000-0003-1304-569X
Alex Castro: https://orcid.org/0000-0002-0298-2672
Tiago Venancio: https://orcid.org/0000-0002-5592-3940
Nilson A. Assunção: https://orcid.org/0000-0002-3747-0415

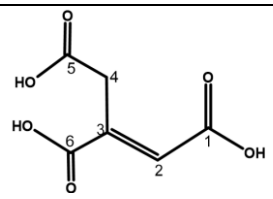
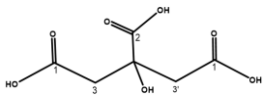
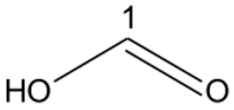
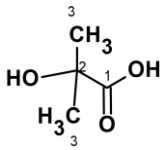
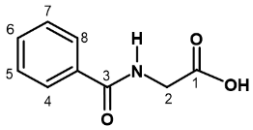
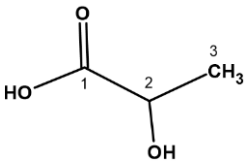
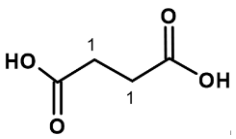
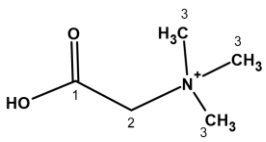
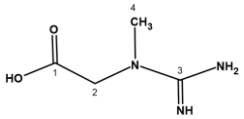
Table S1. Instrumental settings and acquisition parameters for 1D ^{1}H and 2D ^{1}H-^{1}H (COSY) and ^{1}H-^{13}C (HSQC) NMR experiments (600 MHz, D_2O)

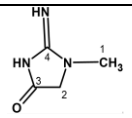
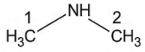
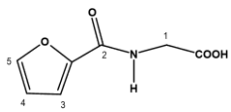
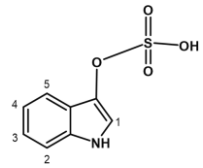
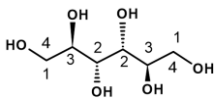
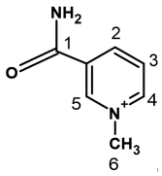
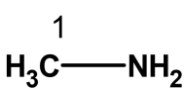
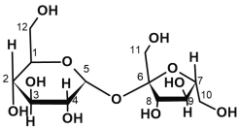
Parameter	^{1}H	^{1}H-^{1}H (COSY)	^{1}H-^{13}C (HSQC)
Pulse sequence (Bruker nomenclature)	Zgesgp	Cosygpprqf	Hsqcetgpprsisp2.3
Number of scans	128	128	256
Acquisition time / s	3.40	F2: 0.21; F1: 0.01	F2: 0.21; F1: 0.003
Time domain data points	64 K	F2: 4096; F1: 256	F2: 2048; F1: 256
Offset frequency / Hz	2820.6	F2: 2,005.97; F1: 2,005.97	F2: 2,005.97; F1: 10,564.50
Spectral width / Hz	9615	F2: 9,014.423, F1: 9,014.423	F2: 9,014.423, F1: 24,904.855
Receiver gain		automatically adjusted	

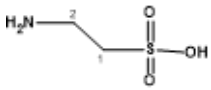
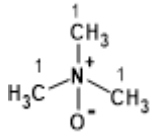
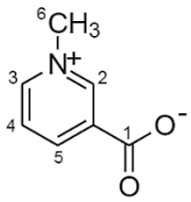
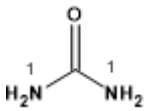
^{1}H: proton nuclear magnetic resonance; COSY: correlation spectroscopy; HSQC: heteronuclear single quantum coherence; F1 and F2: indirect and direct frequency dimensions, respectively; K: thousand data points.

Table S2. ^{1}H and ^{13}C chemical shifts of the metabolites identified in the urine samples from CDCS patients

HMDB Ref. code	Compound	δ ^{1}H (multiplicity with J / Hz)	δ HSQC	δ ^{1}H - Ref.[a] (multiplicity[a] with J / Hz)	δ HSQC - Ref.[a]
			Amino acids		
00161	Alanine 	2-3.94 (o) 3-1.48 (d 7.2)	1.48-19.3	3.90 (q 7.3) 1.52 (d 7.3)	3.77-53.3 1.49-19.0
00123	Glycine 	2-3.56 (s)	3.54 - 44.7	3.54 (s)	3.54-44.3
00177	Histidine 	2-4.00 (dd 7.92; 2.97) 3-3.14 (o) 3'-3.24 (o) 4-7.12 (s) 5 -7.98 (d 1.05)	3.14-30.7 3.24-30.7 7.12-120.2 7.98-139.2	4.00 (dd 7.73; 4.98) 3.29 (dd 16.10; 4.90) 3.20 (dd 15.55; 7.7) 7.13 (s) 8.02 (d 1.13)	4.00-37.2 3.29-30.2 3.20-30.9 7.13-120.0 8.02-138.4
0000159	Phenylalanine 	2-no 5,5'-7.35 (d 6.53) 6,6'-7.41-7.44 (m) 7-7.36 (o)	4.00-57.7 7.35-132.3 7.36-130.6 7.42-132.1	3.98 (dd 7.88; 5.31) 3.19 (m) 7.32 (d 6.98) 7.36 (m) 7.42 (m)	3.98-58.9 3.19-39.1 7.35-132.1 7.36-130.4 7.42-131.8
00158	Tyrosine 	2-3.99 (o) 3-no 4-7.17 (d 8.70) 5-6.87(d 8.79)	3.99-57.8 7.17-133.7 6.87-117.5	3.17-38.17 3.99-58.98 7.17-133.48 6.87 (d 8.68)	3.17-38.2 3.99-59.0 7.17-133.5 6.87-118.9
00167	Threonine 	2-3.59 (d 4.87) 3-4.23 - 4.28 (o) 4-1.33 (d 6.90)	3.59-63.7 4.23-69.2 1.33-23.0	3.57 (d 4.86) 4.24 (m) 1.31 (d 6.58)	3.57-63.5 4.24-68.9 1.31-22.3
00883	Valine 	4-0.99 (d 7.20) 4'-1.04 (d 7.23) 2-no 3-no	no no	0.97 (d 7.01) 1.03 (d 7.05) 2.26 (m) 3.60 (d 4.33)	0.97-19.4 1.03-20.7 2.25-31.9 3.55-63.3
			Organic acids		
00042	Acetic 	2-2.00 (s)	2.00-25. 5	1.90 (s)	1.90-26.1
0000958	(t)-Aconitic	2-6.59 (s) 4-3.45 (s)	6.59-134.2 3.45-46.6	6.58 (s) 3.58 (s)	6.58-134.0 3.44-39.9

					
00094	Citric 	3. 2.54 (d 16.25) 3'. 2.69 (d 16.12)	48.2 48.2	2.52 (d 16.05) 2.66 (d 16.05)	2.52-48.7 2.66-48.7
00142	Formic 	1-8.45 (s)	8.45-148.9	8.39 (s)	8.39-172.4
0242161	2-Hydroxyisobutyric 	3-1.36 (s)	no	1.34 (s)	1.34-29.3
00714	Hippuric 	2-3.96 (o) 6-7.82 (dd 8.54; 1.40) 4.8-7.54-7.56 (m) 7.5-7.64 (tt 7.57; 1.23)	47.1 130.3 132.1 135.4	3.96 (d 5.84) 7.83-129.93 7.54-131.54 7.64-134.91	3.96-46.7 7.83-129.9 7.54-131.5 7.64-134.9
00190	Lactic 	3-1.33 (d 7.20) 2-4.12 (d 7.20)	1.33-23.1	1.37 (d 7.2) 4.42 (q 7.2)	22.9 71.4
00254	Succinic 	1. 2.40 (s)	33.5	2.38 (s)	2.38-36.8
			Other compounds		
00043	Betaine 	2. no 3-3.27 (s)	3.27-56.5	3.25 (s) 3.88 (s)	3.25-55.8 3.88-68.6
00064	Creatine 	4-3.02 (s) 2-3.92 (s)	3.02-39.2 3.92-57.2	3.02 (s) 3.92 (s)	3.02-39.5 3.92-56.4
00562	Creatinine	1-3.04 (s)	3.04-33.3	3.03 (s)	3.04-33.0

		2-4.06 (s)	4.06-59.7	4.05 (s)	4.06-59.2
	Dimethylamine				
00087		1.2-2.72 (s)	2.72-37.8	2.71 (s)	2.71-37.3
–	Fatty acids	1-no 2-2.26 (m) 3-2.04 (m) 4.5-5.35 (m) 6 - 0.93 (m)	175.8 236.7 30.1 132.5 17.8	– 2.35 (t 7.3) 2.02 (m) 5.35 (m) 0.89 (t 6.8)	173.5 34.0 27.1 129.9 14.1
000439	2-Furoylglycine 	1-no 3-7.18 (d 4.12) 4-6.64 (dd 3.68; 1.61) 5-7.69-7.69 (m)	no no no no	3.94 (d 5.93) 7.18 (d 3.13) 6.64 (dd 3.49; 1.73) 7.68-7.71 (m)	3.94-45.7 7.18-117.9 6.64-114.8 7.68-148.8
00682	Indoxyl sulfate 	2-7.36 (s) 6-7.50 (d 8.33) 7-7.27 (o) 8-7.18 (o) 9-7.70 (d 8.30)	7.36-119.2 7.50-115.4 7.27-125.3 7.19-124.4 7.70-120.6	7.36 (s) 7.50 (d 8.24) 7.28 (m) 7.20 (m) 7.71 (d 7.96)	7.36-118.8 7.50-114.9 7.28-125.1 7.20-122.5 7.71-120.0
00765	Mannitol 	1-3.87 (dd 11.88; 2.76) 2-3.80 (d 8.92) 3-3.74 (o) 4-3.68 (o)	3.87-66.4 3.80-72.5 3.74-74.2 3.68-66.2	3.86 (dd 11.87; 2.80) 3.77 (d 8.72) 3.73 (m) 3.64 (dd 11.76; 6.62)	3.86-66.0 3.77-71.9 3.73-73.5 3.64-66.0
000699	1- Methylnicotinamide 	2 - 8,89 (d 8.59) 3-8.17 (t 7.25) 2-no 4 - 8,97 (d 6,83) 5-9.27 (s) 5-no	no no no no no	8.89 (d 8.20) 8.17 (t 7.10) 8.96 (d 6.14) 9.27 (s) 4.47 (s)	8.89-146.3 8.17-130.8 8.96-145.0 9.27-147.8 4.47-51.3
0000164	Methylamine 	1-2.60 (s)	no	2.59 (s)	2.59-27.6
000258	Sucrose 	7-5.41 (d 3.97 3-4.22 (d 8.69) 4-4.05 (o)	5.41-95.2 4.22-76.5 4.05-77.2	5.40 (d 3.89) 4.21 (d 8.75) 4.04 (t 8.59)	5.40-94.8 4.21-79.2 4.04-76.7

00251	Taurine 	1-3.43 (t 6.48) 2-3.26 (o)	3.43-40.6 3.26-55.4	3.42 (t 6.62) 3.25 (t 6.62)	3.42-38.3 3.25-50.4
00925	Trimethyaminel *N*-oxide 	1-3.27 (s)	3.27-62.7	3.25 (s)	3.25-62.2
00875	Trigonelline 	2-9.13 (s) 3-8.84 (d 7.53) 5-8.83 (d 5.58) 4-8.08 (t 7.96) 6-4.44 (s)	9.13-148.9 8.84-148.0 8.83-148.5 8.08-131.1 4.44-51.3	9.10 (s) 8.84 (d 7.42) 8.83 (d 5.70) 8.07 (t 7.80) 4.43 (s)	9.10-148.4 8.84-147.7 8.83-147.5 8.07-130.4 4.43-51.0
00294	Urea 	1-5.76 (ls)	no	5.78 (ls)	no

[a]Ref., reference values from HMDB. HMDB: Human Metabolome Database; *δ*: chemical shift (ppm); HSQC: heteronuclear single quantum coherence; *J*: coupling constant (Hz); s: singlet; d: doublet; dd: doublet of doublets; m: multiplet; o: overlapped signal. d (doublet); dd (doublet of doublet); ls (large singlet); m (multiplet); no (not observed); o (signals overlap); q (quartet); s (singlet) and t (triplet).

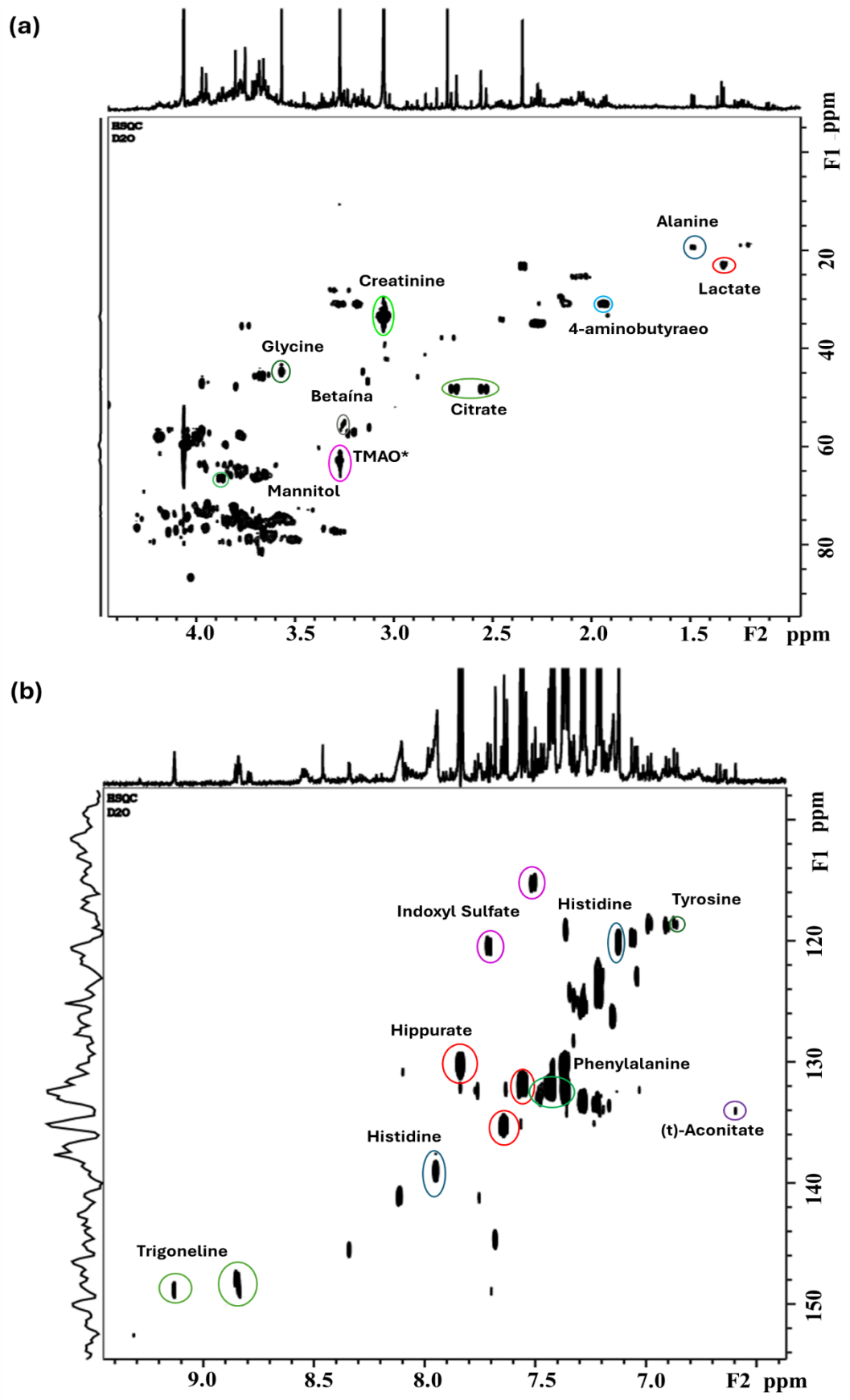


Figure S1. Presentative ^{1}H-^{13}C HSQC NMR spectra (600 MHz, D_2O) of CDCS patient urine sample. Showing the (a) aliphatic region (δ 0.90-4.50 ppm) and (b) aromatic region (δ 6.20-9.40 ppm). *Trimethylamine *N*-oxide (TMAO).

Table S3. Statistical results from the OPLS-DA analysis. VIP scores greater than 1 are presented for the children, adolescents, and adults groups

Metabolite	V1	V2
	Children	
4-Hydroxyphenylacetate	1.55	0.10
Tryptophan	1.53	0.13
Histidine	1.46	0.60
Methylguanidine	1.31	0.71
Hypoxanthine	1.30	0.50
Succinate	1.26	0.48
Phenylalanine	1.26	0.016
Dimethylamine	1.23	1.27
Methylamine	1.21	1.16
	Adolescents	
Phenylalanine	2.31	0.56
Acetate	1.87	0.49
Sucrose	1.55	0.38
Trigonelline	1.52	0.90
Tyrosine	1.51	0.88
Taurine	1.36	0.059
Histidine	1.32	0.99
	Adults	
Histidine	1.99	0.88
Taurine	1.99	0.23
Pyruvate	1.88	0.51
Glycine	1.78	0.86
Tyrosine	1.71	0.79
Hippurate	1.43	0.79
Alanine	1.42	1.05
Tryptophan	1.39	0.82
Methylmalonate	1.30	0.51

OPLS-DA: orthogonal partial least squares discriminant analysis; VIP: variable importance in projection; V1 and V2: predictive and orthogonal components of the OPLS-DA model, respectively.

Table S4. OPLS-DA model performance for children, adolescents, and adults cohorts. R^2X represents the proportion of variance explained by each component, R^2Y represents the proportion of class variance explained by the model, and Q^2 indicates predictive ability estimated by cross-validation. Predictive (p1) and orthogonal (o1, o2) components are shown, together with cumulative model values

Component	R^2X	R^2Y	Q^2
Children			
p1 (predictive)	—	0.624	0.502
o1 (orthogonal)	—	—	–0.116
Model (cumulative)	—	0.624	0.502
Adolescents			
p1 (predictive)	—	0.430	0.291
o1 (orthogonal)	—	—	—
Model (cumulative)	—	0.430	0.291
Adults			
p1 (predictive)	0.147	0.764	0.489
o1 (orthogonal)	0.287	0.157	0.177
o2 (orthogonal)	0.093	0.061	0.083

OPLS-DA: orthogonal partial least squares discriminant analysis; R^2X and R^2Y: explained variance of X and Y matrices, respectively; Q^2: predictive ability estimated by cross-validation; p1: predictive component; o1/o2: orthogonal components.

OPLS-DA was applied separately to the children, adolescent, and adult cohorts to explore group-related metabolic differences. The children and adolescent models showed moderate discrimination with acceptable (children) or modest (adolescents) predictive ability, consistent with exploratory metabolomic analyses. The adult model exhibited strong apparent separation. However, given the intrinsically small number of adult CDCS patients, reflecting the reduced life expectancy associated with the CDCS, this cohort is necessarily limited in size. As a result, these findings should be interpreted with caution, since high R^2Y values in small datasets may reflect model instability and an increased risk of overfitting rather than strong, generalizable predictive performance.

Given the small number of adult CDCS individuals, the OPLS-DA model was evaluated using permutation testing to assess whether the observed separation could arise by chance. The permutation test (n = 1,000) yielded $R^2Y = 0.982$ ($p = 0.013$) and $Q^2 = 0.748$ ($p = 0.002$), indicating that both the model fit and predictive performance were significantly better than those obtained from randomly permuted class labels. While these results confirm that the adult group separation is not random, the limited sample size implies that the model may be sensitive to individual samples, reinforcing that should be interpreted as exploratory rather than strongly generalized.

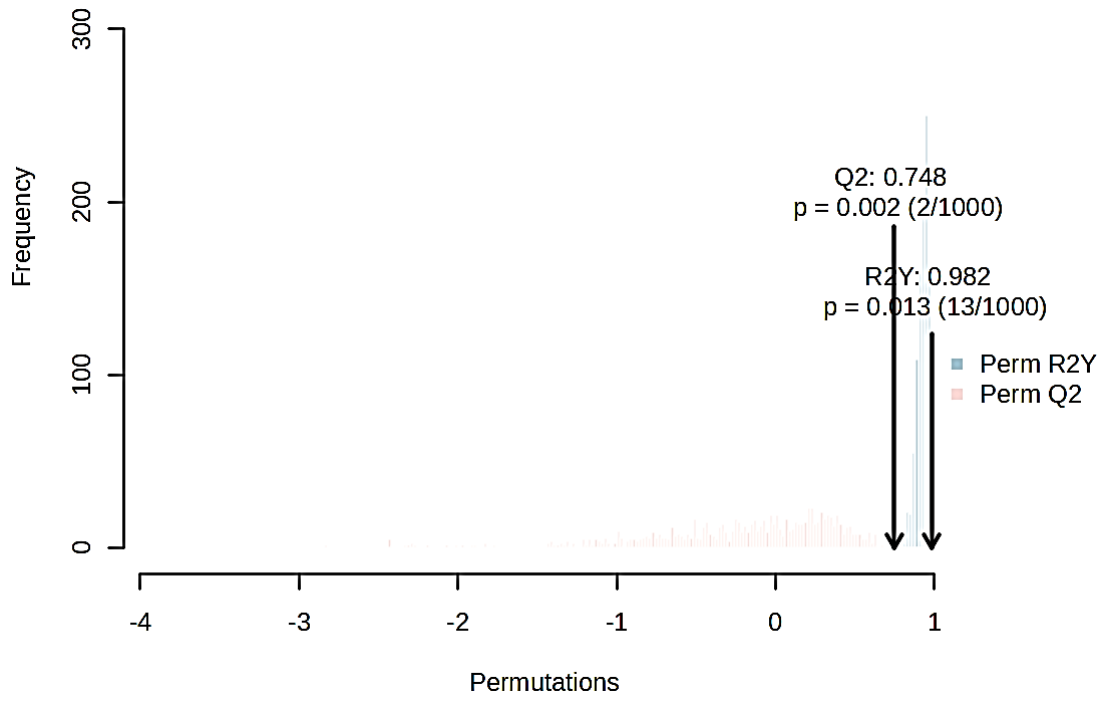


Figure S2. Permutation test for the adult OPLS-DA model. The observed model values (R^2Y = 0.982 and Q^2 = 0.748), lie outside the permutation distributions, demonstrating that the model fit and predictive performance are significantly better than expected by chance.

Table S5. Permutation-validated OPLS-DA performance (adults)

Metric	Observed value	Permutations (n)	Empirical *p*-value
		Adults	
R^2Y	0.982	1,000	0.013 (13/1000)
Q^2	0.748	1,000	0.002 (2/1000)

OPLS-DA: orthogonal partial least squares discriminant analysis; R^2Y: explained variance of the Y matrix; Q^2: predictive ability estimated by cross-validation; n: number of permutations.

Table S6. Statistical data from fold-change values of significant metabolites in children, adolescents and adults

Metabolite	FC	Log2(FC)	*p*-value
	Children		
Tryptophan	3.98	1.99	0.0012
4-Hydroxyphenylacetate	3.12	1.64	0.012
Methylamine	3.00	1.58	0.012
Hypoxanthine	2.75	1.46	0.028
Histidine	2.31	1.21	0.012
Dimethylamine	2.29	1.20	0.012
Phenylalanine	2.25	1.17	0.028
Valine	1.98	0.98	0.033
Methylguanidine	1.95	0.96	0.029
	Adolescents		
Phenylalanine	2.86	1.51	0.0011
	Adults		
Taurine	3.69	1.88	0.030
Histidine	2.68	1.42	0.030

FC: fold change (ratio of metabolite levels between CDCS patients and control group); Log_2(FC): base-2 logarithm of fold change; *p*-value: probability value obtained from statistical analysis. Only statistically significant metabolites are shown.