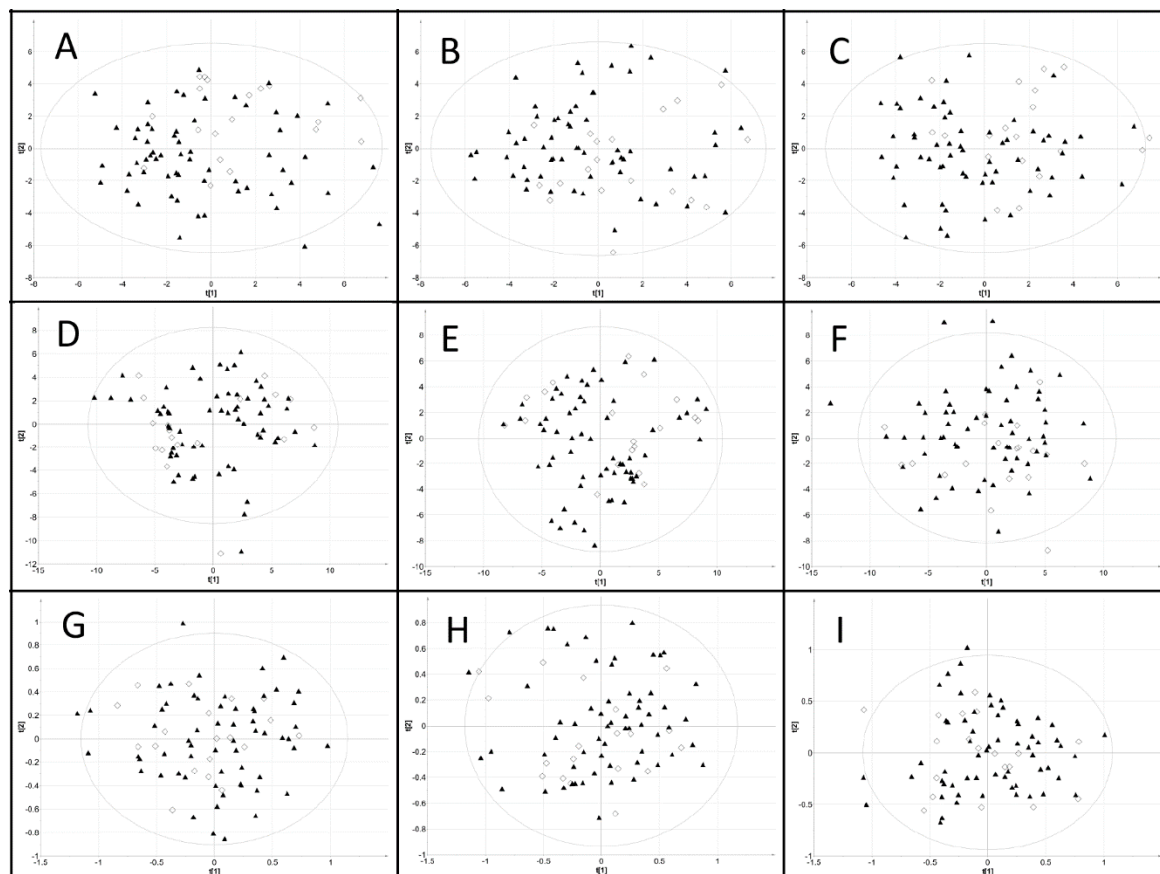# Supplementary Materials: Distinguishing Benign from Malignant Pancreatic and Periampullary Lesions Using Combined Use of 1H-NMR Spectroscopy and Gas Chromatography–Mass Spectrometry

Yarrow J. McConnell, Farshad Farshidfar, Aalim M. Weljie, Karen A. Kopciuk, Elijah Dixon, Chad G. Ball, Francis R. Sutherland, Hans J. Vogel and Oliver F. Bathe

Figure S1. Principal Component Analysis (PCA) Scores Scatter plots for three training and validation set allocations of 1H-NMR, GC-MS and combined datasets. All variables are included in the analyses. A, B, and C: three case allocations for 1H-NMR data; D, E and F: three case allocations for GC-MS data; G, H and I: three case allocations for combined 1H-NMR and GC-MS data. Sample sizes were 80 per allocation. Number of variables in each PCA (k) was 83 for 1H-NMR spectroscopy; 146 for GC-MS; and 183 for combined. A= number of components.
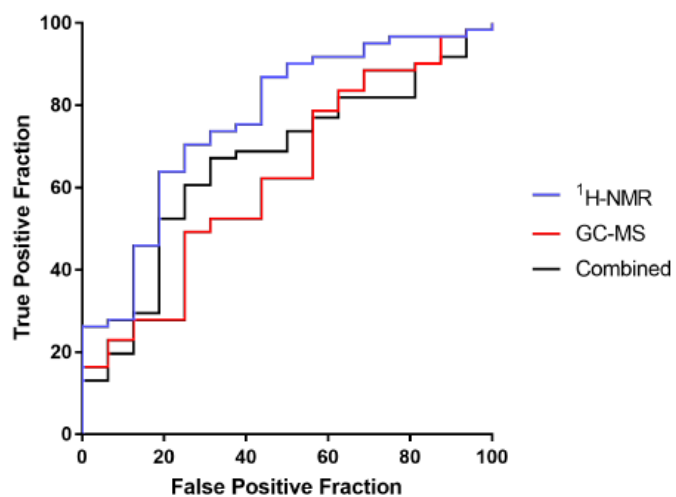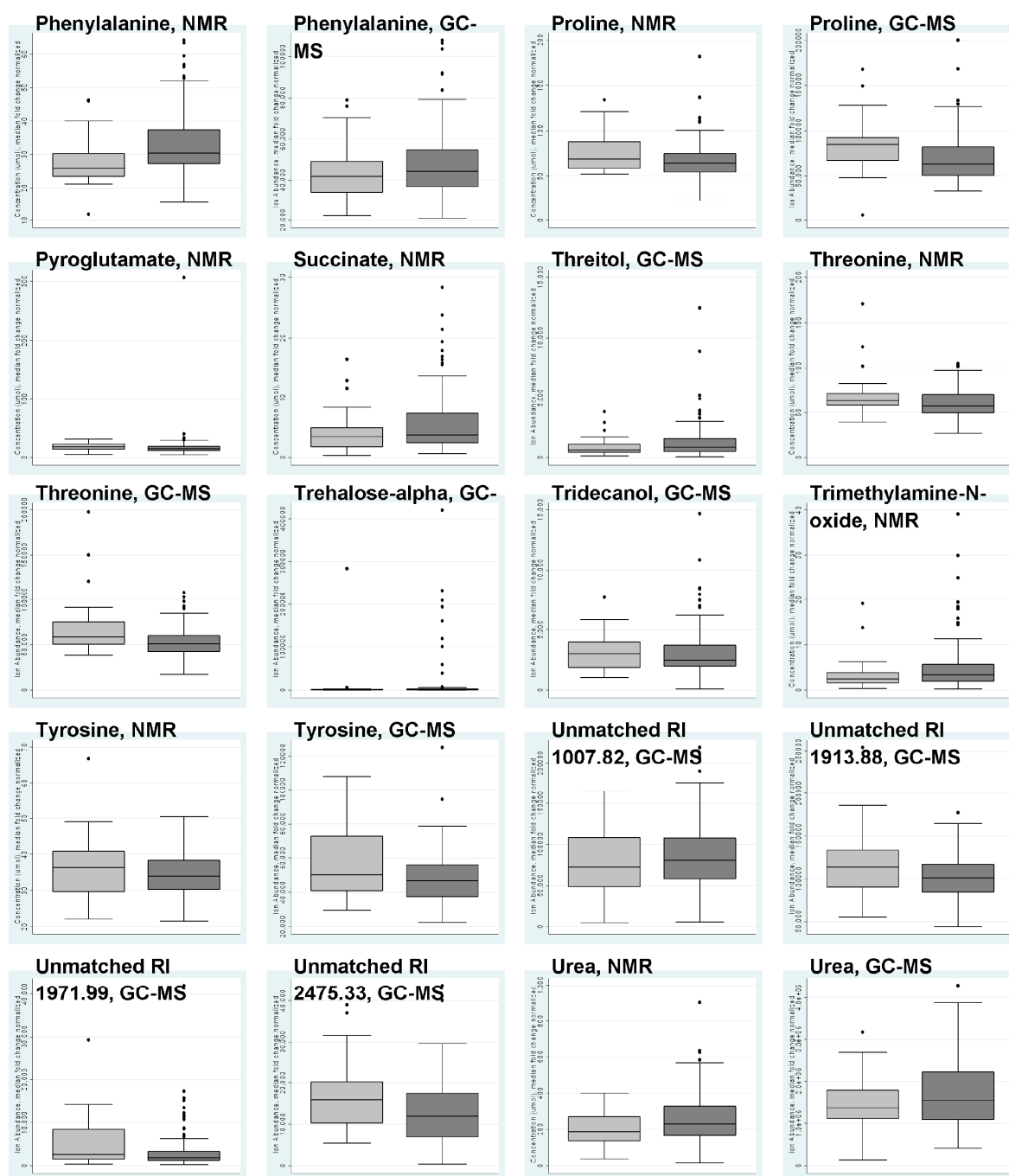
Figure S2. Receiver Operating Characteristic (ROC) curve for detection of benign and malignant periampullary lesions. ^{1}H-NMR: blue, GC-MS: red, Combined model: black.
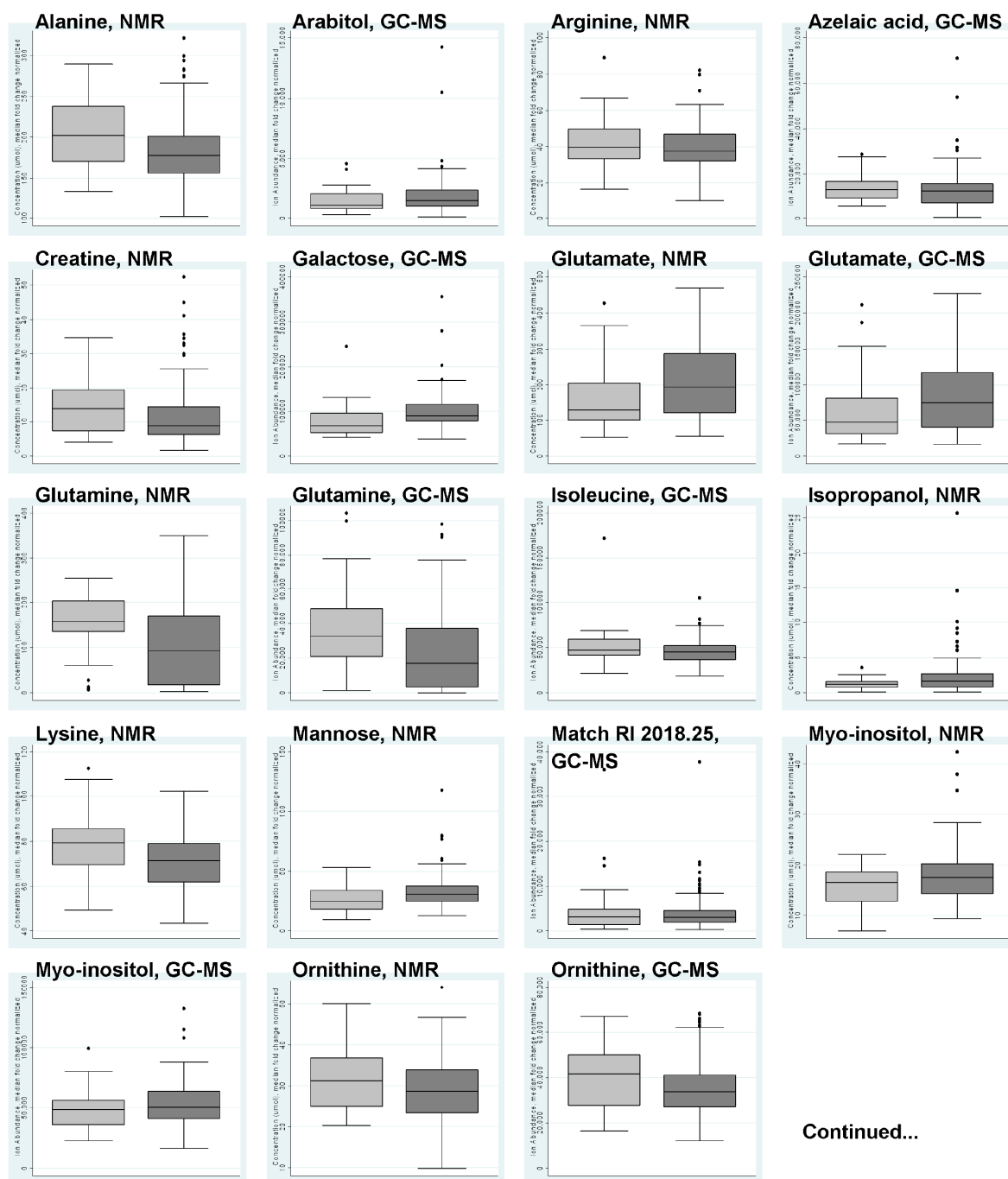
Figure S3. *Cont.*

Figure S3. Whisker plots for consistently contributing metabolites from NMR and GC-MS. Whisker box limits are 1st and 3rd quartiles, whisker bar limits are upper and lower adjacent values. Middle line is median. Y-axis displays normalized concentration/ion abundance data: NMR = μmol, GC-MS = ion abundance. Benign disease: light gray, malignant disease: dark gray.
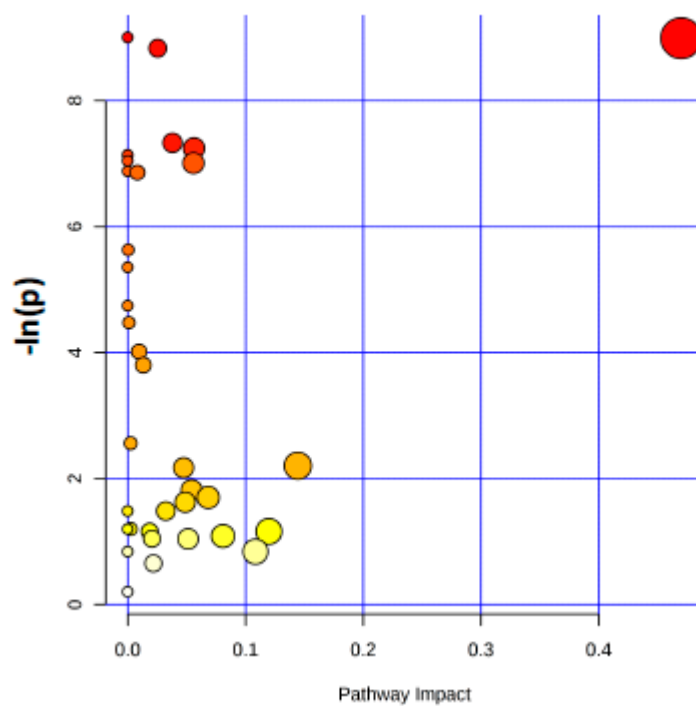
Figure S4. Metaboanalyst 2.0 pathway analysis bubble plot. The plot depicts each pathway's significance (expressed as "–ln(p-value)" on Y-axis) versus its degree of involvement in the biological process (as identified by "impact" on X-axis).

Table S1. Performance measurements in triplicate analysis of ^{1}H-NMR dataset, 50/50 training/test dataset (n = 80/77).

| | OPLS-DA: Filtered | | | | | | OPLS-DA: Minimal[c] | | | | | | | | |
| | Training | | | | Test | | Training | | | | | | | Test | |
Trial	X[a]	R2Y	Q2	p[b]	Core CV[c]	AUROC	SE	X[a]	R2Y	Q2	p[b]	AUROC(i)	Cancer	Non-cancer	AUROC	SE
A	26	0.334	0.174	6.4×10^{-4}	0.487	0.71	0.07	13	0.251	0.122	6.6×10^{-3}		Isopropanol Mannose Phenylalanine Succinate TrimethylamineNoxide TwoHydroxyisobutyrat	Alanine Betaine Glutamine Histidine Lysine Proline Threonine	0.79	0.06
B	28	0.378	0.259	9.8×10^{-6}	0.484	0.73	0.07	14	0.308	0.224	6.1×10^{-5}		Mannose MyoInositol Phenylalanine Urea	Creatine Ethanol Glutamine Histidine Lysine Ornithine Proline Pyroglutamate Threonine Xanthine	0.71	0.05
C	24	0.282	0.140	3.3×10^{-3}	0.491	0.80	0.05	12	0.230	0.128	5.2×10^{-3}	0.756	*Choline Isopropanol Mannose Succinate TrimethylamineN-oxide*	Alanine Creatine Inosine Lysine Ornithine Pyroglutamate Threonine	0.75	0.06
Average	*26*	*0.331*	*0.191*	*1.3×10^{-3}*		*0.75*	*0.06*	*13*	*0.263*	*0.158*	*3.8×10^{-3}*				*0.75*	*0.06*

[a] Number of metabolites/features in the filtered (*t*-test *p*-value <0.3) training set

[b] *p*-value for CV-ANOVA testing

[c] Minimal list of metabolites determined by multistep iterative process: selection of metabolites contributing most to class difference by multivariate O2PLS analysis controlling for technical and clinical covariates; then iterative OPLS-DA modeling with sequential exclusion of metabolites based on their contribution to test model parameters- to maximize AUROC.

Table S2. Performance measurements in triplicate analysis of GC-MS dataset, 50/50 training/test dataset (n = 80/77).

| | OPLS-DA: Filtered | | | | | | | OPLS-DA: Minimal[c] | | | | | | | | |
| | Training | | | | | Test | | Training | | | | | | | Test | |
Trial	X[a]	R2Y	Q2	p[b]	Core CV[c]	AUROC	SE	X[a]	R2Y	Q2	p[b]	AUROC(i)	Cancer	Non-cancer	AUROC	SE
A	33	0.440	0.239	4.8×10^{-5}	0.736	0.69	0.07	17	0.336	0.165	1.3×10^{-3}	0.770	Arabitol5TMS Galactose1MEOX MatchRI170922 Threitol4TMS Trehalosealpha UnmatchedRI11763	Isoleucine2TMS MatchRI129451 Ornithine4TMS Proline2TMS Sorbitol6TMS Threonine3TMS Tridecan1oln1TMS UnmatchedRI16313 UnmatchedRI19719 UnmatchedRI24753 UnmatchedRI28215	0.69	0.07
B	43	0.378	0.258	1.3×10^{-5}	0.734	0.52	0.08	21	0.327	0.238	2.9×10^{-5}	0.832	Glutamicacid3TMS Inositolmyo6TMS MatchRI201825 Phenylalanine2TMS UnmatchedRI10078 UnmatchedRI11250 UnmatchedRI11396 UnmatchedRI18425 Urea2TMS	AzelaicAcid2TMS GlutamineDL3TMS Glycerol3TMS Octadecadienoic Ornithine4TMS Tridecan1oln1TMS UnmatchedRI11646 UnmatchedRI13319 UnmatchedRI19138 UnmatchedRI19719 UnmatchedRI22199 UnmatchedRI24753	0.54	0.07
C	33	0.348	0.199	2.6×10^{-4}	0.784	0.70	0.07	17	0.273	0.162	1.2×10^{-3}	*0.755*	*Butanoicacid2am* *Galactose1MEOX* *Inositolmyo6TMS* *Ribitol5TMS* *Threitol4TMS* *UnmatchedRI10078* *Urea2TMS* *UnmatchedRI22127*	AzelaicAcid2TMS GlutamineDL3TMS Isoleucine2TMS MatchRI201825 Ornithine4TMS Tyrosine3TMS UnmatchedRI19138 UnmatchedRI19736 UnmatchedRI24753	0.67	0.08
Average	*36*	*0.389*	*0.232*	*1.1×10^{-4}*	*0.751*	*0.64*	*0.07*	*18*	*0.312*	*0.188*	*8.4×10^{-4}*	*0.786*			*0.67*	*0.07*

[a] Number of metabolites/features in the filtered (*t*-test *p*-value <0.3) training set

[b] *p*-value for CV-ANOVA testing

[c] Minimal list of metabolites determined by multistep iterative process: selection of metabolites contributing most to class difference by multivariate O2PLS analysis controlling for technical and clinical covariates; then iterative OPLS-DA modeling removing metabolites based on their contribution to predictive modeling in test set (YpredPS and T2PS)

Table S3. Performance measurements in triplicate analysis of Combined dataset, 50/50 training/test dataset (*n* = 80/77).

| Trial | OPLS-DA: Filtered | | | | | | | OPLS-DA: Minimal[c] | | | | | | | | |
| | Training | | | | | Test | | Training | | | | | | | | Test | |
	X[a]	R2Y	Q2	*p*[b]	Core CV[c]	AUROC	SE	X[a]	R2Y	Q2	*p*[b]	AUROC(i)	Cancer	Non-cancer	AUROC	SE
A	50	0.451	0.238	4.1×10^{-5}	-	0.67	0.08	24	0.341	0.177	7.8×10^{-4}	0.802	Arabitol5TMS-G Isopropanol-N Mannose-N MatchRI170922-G MyoInositol-N Phenylalanine-C Ribitol5TMS-G Succinate-N Threitol4TMS-G Trehalosealpha-G TwoHydroxyisobutyrat Urea-C	Histidine-C Lysine-N MatchRI129451-G Methionine-C Ornithine-C Proline-C Sorbitol6TMS-G Threonine-C UnmatchedRI16313-G UnmatchedRI22199-G UnmatchedRI24753-G UnmatchedRI28215-G	0.67	0.07
B	62	0.402	0.239	3.7×10^{-7}	-	0.63	0.08	26	0.367	0.231	5.2×10^{-5}	0.839	Glucuronicacid1M-G Mannose-N MyoInositol-N Phenylalanine-C UnmatchedRI10078-G UnmatchedRI10325-G UnmatchedRI11396-G UnmatchedRI18377-G UnmatchedRI18425-G Urea-C	Arginine-N AzelaicAcid2TMS-G Creatine-N Glutamine-C Glycerol-C Histidine-C Lysine-N Methanol-N Ornithine-C Proline-C UnmatchedRI11646-G UnmatchedRI13319-G UnmatchedRI19736-G UnmatchedRI22199-G UnmatchedRI24753-G Xanthine-N	0.66	0.07

[a] Number of metabolites/features in the filtered (*t*-test *p*-value <0.3) training set

[b] *p*-value for CV-ANOVA testing

[c] Minimal list of metabolites determined by multistep iterative process: selection of metabolites contributing most to class difference by multivariate O2PLS analysis controlling for technical and clinical covariates; then iterative OPLS-DA modeling removing metabolites based on their contribution to predictive modeling in test set (YpredPS and T2PS)

Table S3. *Cont.*

| | OPLS-DA: Filtered | | | | | | | OPLS-DA: Minimal[c] | | | | | | | | |
| | Training | | | | Test | | Training | | | | | | | | Test | |
Trial	X[a]	R2Y	Q2	*p*[b]	Core CV[c]	AUROC	SE	X[a]	R2Y	Q2	*p*[b]	AUROC(i)	Cancer	Non-cancer	AUROC	SE
C	47	0.369	0.219	9.3×10^{-5}	-	0.73	0.06	19	0.311	0.204	1.7×10^{-4}	0.795	Choline-N Erythritol4TMS-G Galactose1MEOX-G Isopropanol-N Mannose-N MyoInositol-N Phenylalanine-C Ribitol5TMS-G Threitol4TMS-G TrimethylamineNoxide UnmatchedRI22127-G	Arginine-N Creatine-N Inosine-N Lysine-N MatchRI201825-G Threonine-C UnmatchedRI19138-G UnmatchedRI19736-G	0.68	0.07
Average	*53*	*0.407*	*0.232*	*4.5×10^{-5}*		*0.68*	*0.07*	*23*	*0.34*	*0.204*	*3.3×10^{-4}*	*0.812*			*0.67*	*0.07*

[a] Number of metabolites/features in the filtered (*t*-test *p*-value <0.3) training set

[b] *p*-value for CV-ANOVA testing

[c] Minimal list of metabolites determined by multistep iterative process: selection of metabolites contributing most to class difference by multivariate O2PLS analysis controlling for technical and clinical covariates; then iterative OPLS-DA modeling removing metabolites based on their contribution to predictive modeling in test set (YpredPS and T2PS)