Supporting Information

Diagnosis of prostate cancer by desorption electrospray ionization mass spectrometric imaging of small metabolites and lipids

Shibdas Banerjee[a], Richard N. Zare[a,1], Robert Tibshirani[b], Christian A. Kunder[c], Rosalie Nolley[d], Richard Fan[d], James D. Brooks[d], and Geoffrey A. Sonn[d]

[a]*Department of Chemistry, Stanford University, Stanford, CA 94305*

[b]*Departments of Biomedical Data Sciences, and of Statistics, Stanford University, CA 94305*

[c]*Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305*

[d]*Department of Urology, Stanford University School of Medicine, Stanford, CA 94305*

[1]To whom correspondence should be addressed. E-mail: zare@stanford.edu

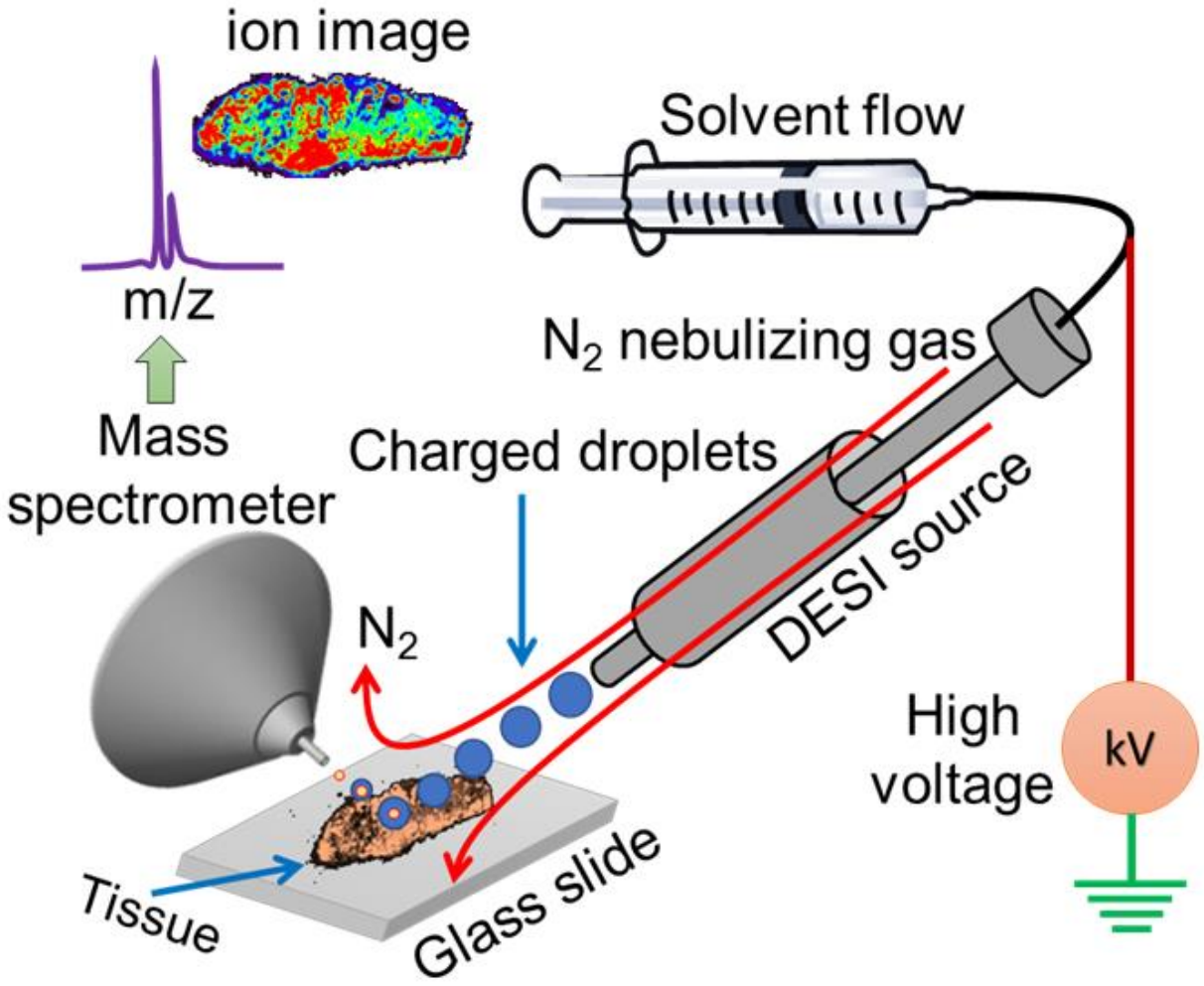

Fig. S1. Schematic diagram of the DESI-MSI set up for intercepting the lipid molecules from the prostate tissue. The tissue section (15 µm thickness) is bombarded with a stream of charged microdroplets generated by electrospraying a 1:1 mixture (v/v) of acetonitrile and dimethylformamide at high voltage (-5 kV) and using nitrogen as the nebulizing gas. These droplets dissolve hundreds of lipids and metabolites present in the tissue by wetting the tissue surface with the droplet-solvent. Then the splashing of this liquid film upon the arrival of subsequent primary droplets results in the formation of secondary microdroplets containing the analyte ions (lipids/metabolites), which are afterwards converted into gaseous ions for mass spectrometric analysis (identification) and characterization. The imaging is performed by scanning the tissue surface in x and y directions through an impinging spray of charged microdroplets and the corresponding analyte ion signals in pixel-to-pixel mass spectra can then be plotted as two-dimensional images allowing a detailed chemical (lipids/metabolites) map to be made of the tissue sample.

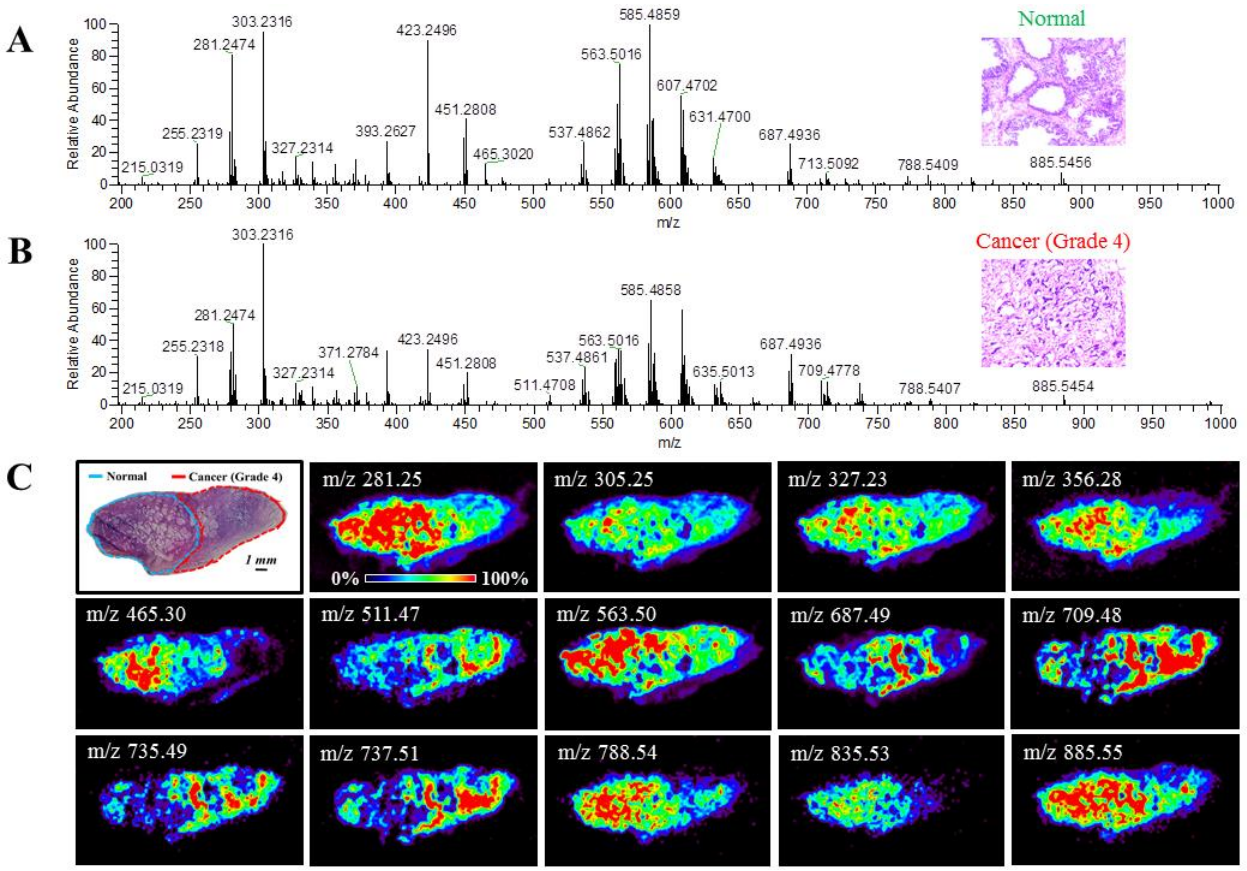

Fig. S2. Negative ion mode DESI-MS data (m/z 200-1000) from a single pixel of prostate tissue (A) which is normal and (B) which is Gleason grade 4 cancer. (C) Spatial distribution of 14 different lipids in a prostate tissue specimen (upper left shows the corresponding H&E staining) that contains both normal (blue outline) and cancer (red outline). In (C) the individual lipid distribution is mapped throughout the tissue whereas in (A) and (B) the abundance of all lipids is displayed in a single pixel. See Table S1 for identification of species with different *m/z* values.

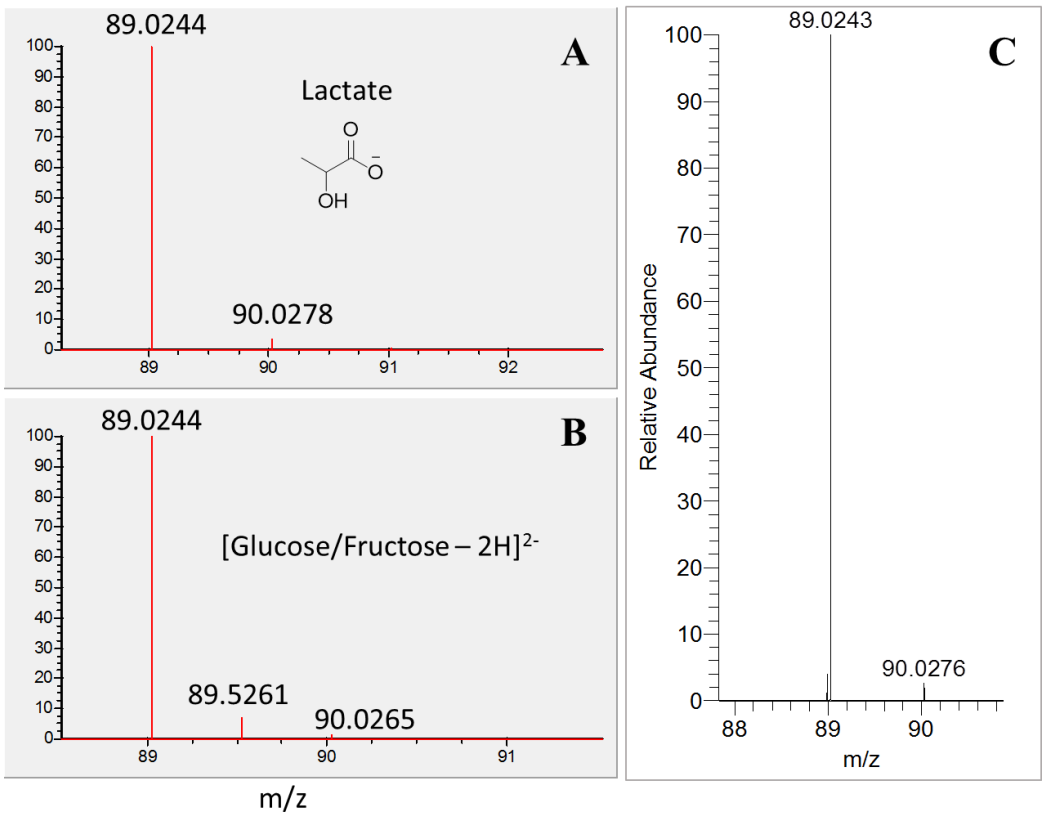

Fig. S3. Left panels show the simulated isotopic distribution in the mass spectra of (A) lactate, and (B) doubly deprotonated glucose/fructose. The right panel (C) shows the isotopic distribution of the species detected at *m/z* 89.0243 (Table S1), unambiguously identifying the species as lactate, not glucose nor fructose.

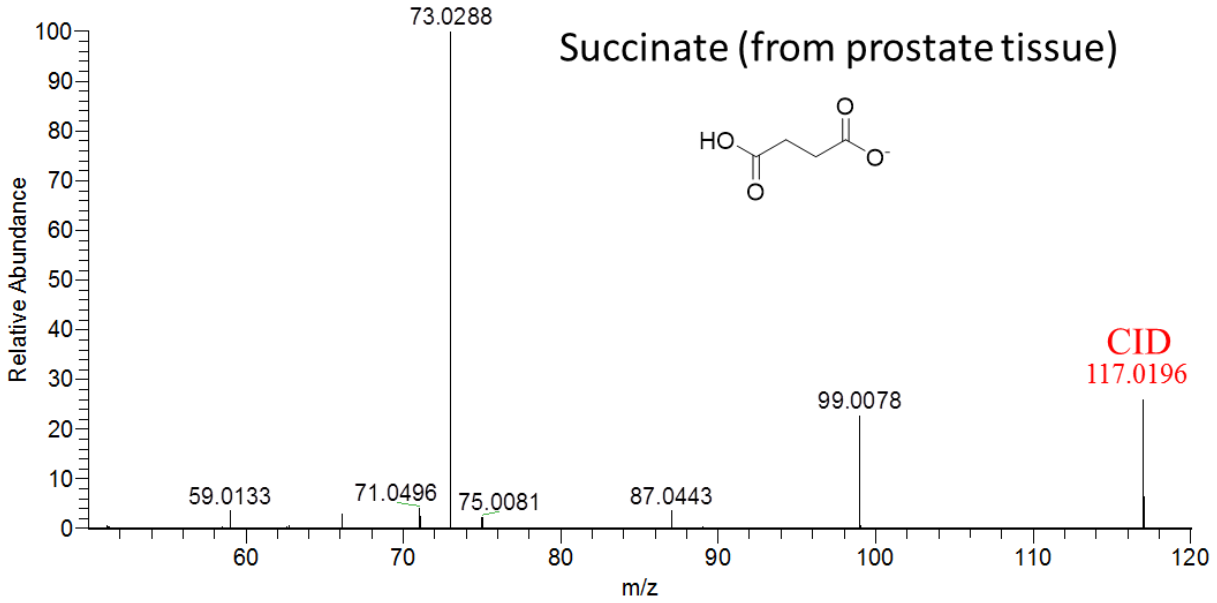

Fig. S4a. Collision induced dissociation (CID) at *m/z* 117.0196 (Table S1) identifies the species as succinate when compared to the CID of the standard given in the MassBank (http://www.massbank.jp/).

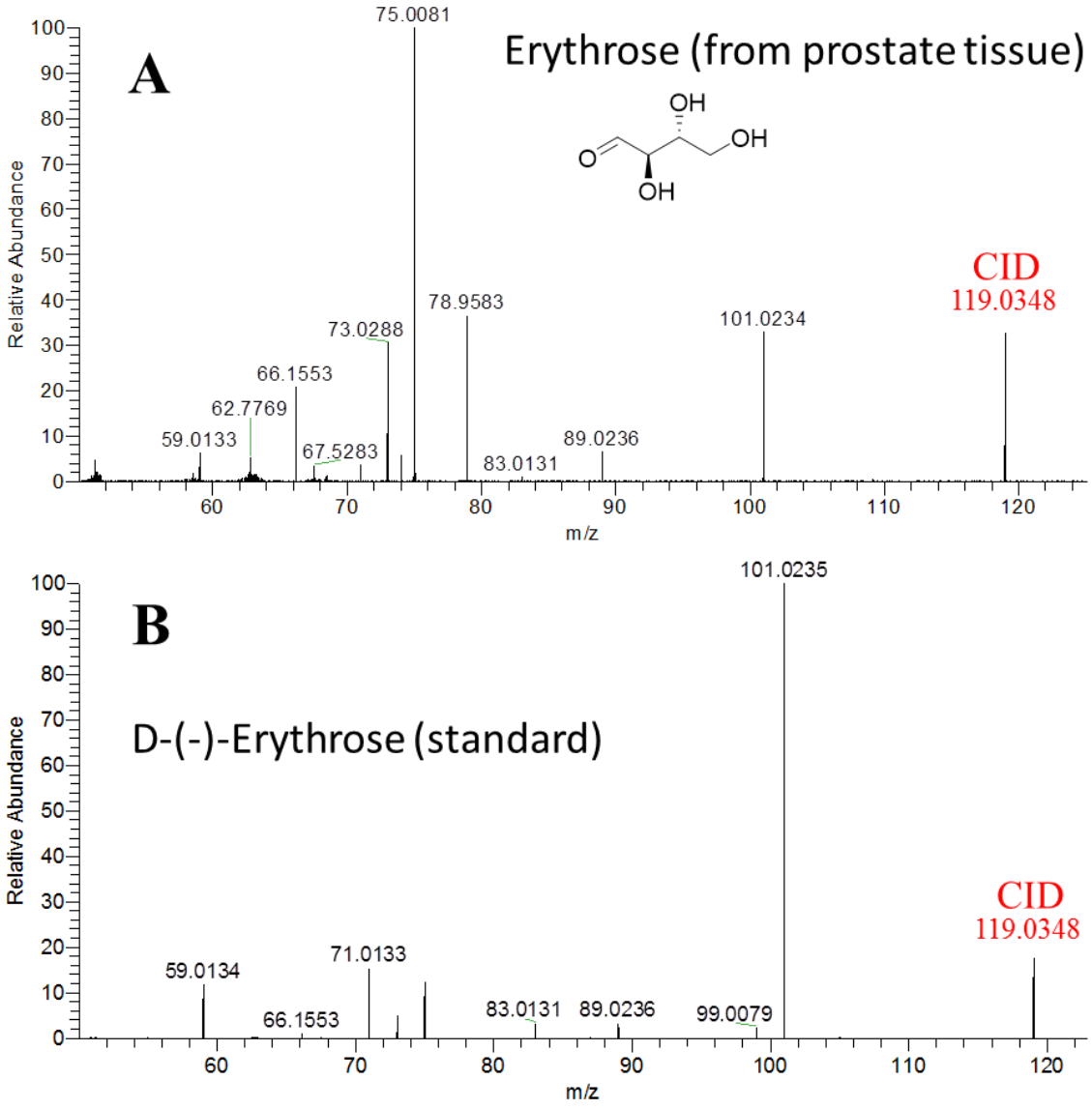

Fig. S4b. Collision induced dissociation (CID) of the species (A) at *m/z* 119.0348 (Table S1) identifies the species mostly as erythrose when compared to the CID of (B) the standard erythrose. A number of ion signal intensities in (A) are different from (B) suggesting the association of other isomeric or isobaric species corresponding to the mass selected ion (m/z 119.0348) in (A).

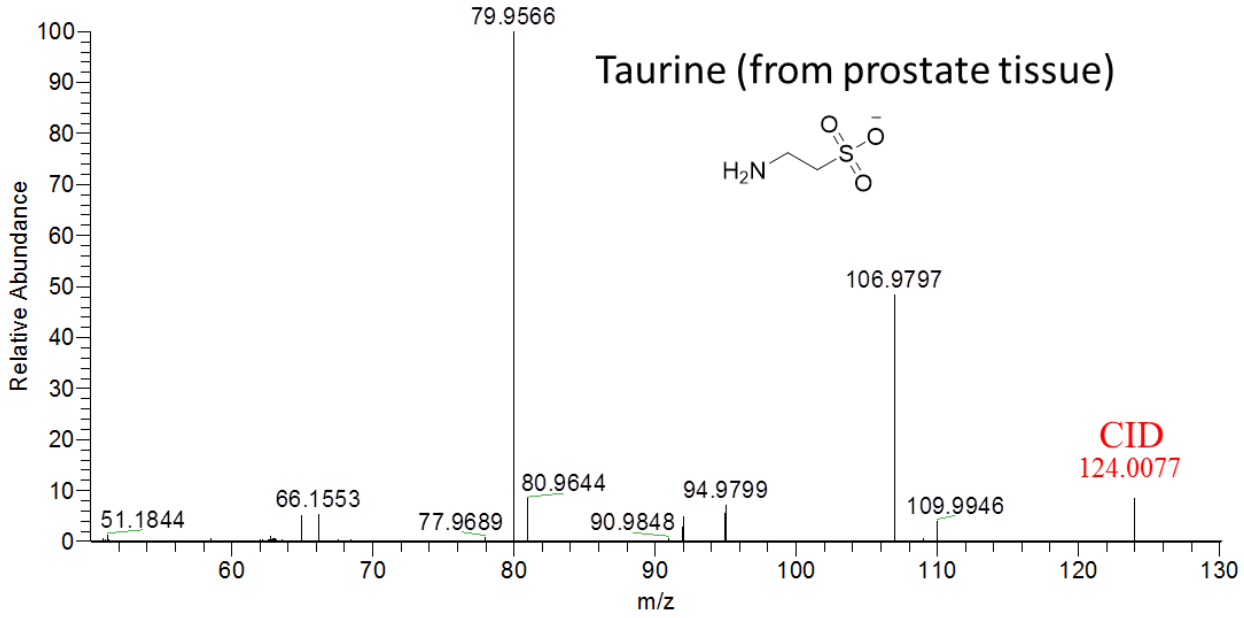

Fig. S4c. Collision induced dissociation (CID) at *m/z* 124.0077 (Table S1) identifies the species as taurine when compared to the CID of the standard given in the MassBank (http://www.massbank.jp/).

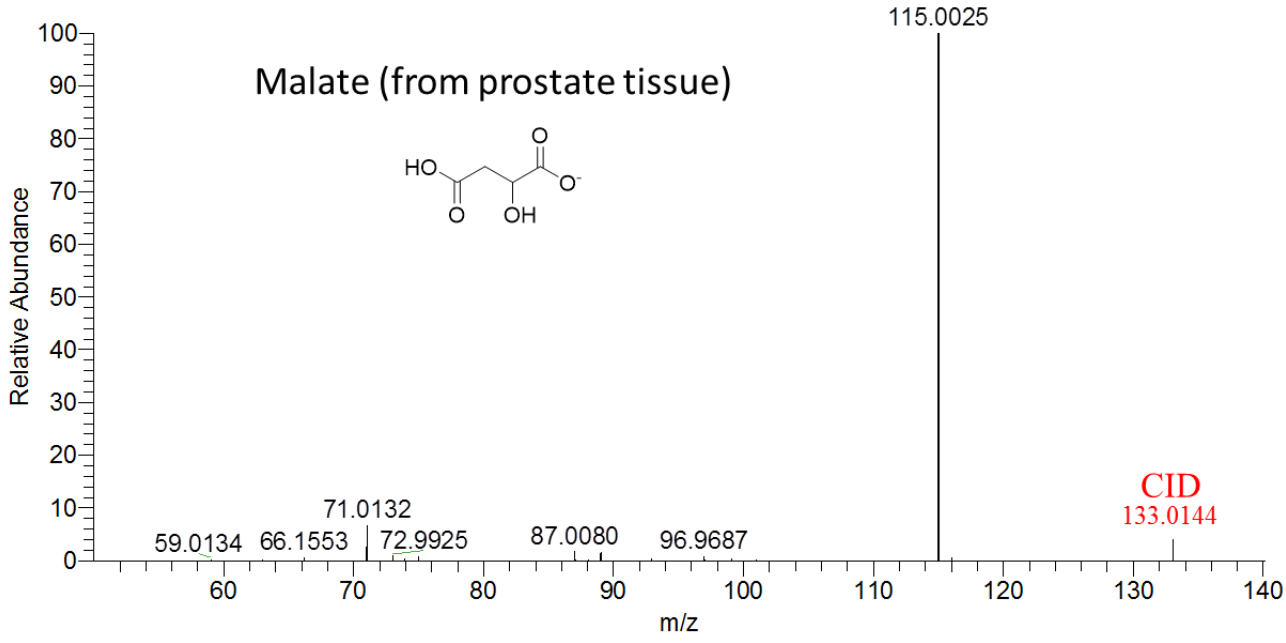

Fig. S4d. Collision induced dissociation (CID) at *m/z* 133.0144 (Table S1) identifies the species as malate when compared to the CID of the standard given in the MassBank (http://www.massbank.jp/).

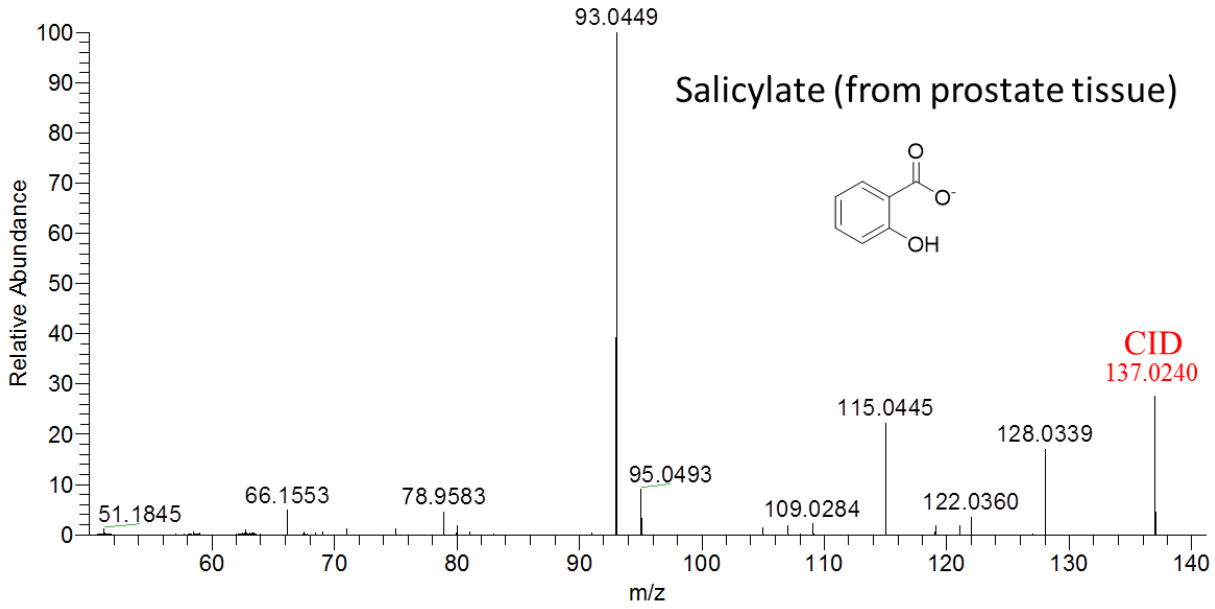

Fig. S4e. Collision induced dissociation (CID) at *m/z* 137.0240 (Table S1) identifies the species as salicylate when compared to the CID of the standard given in the MassBank (http://www.massbank.jp/).

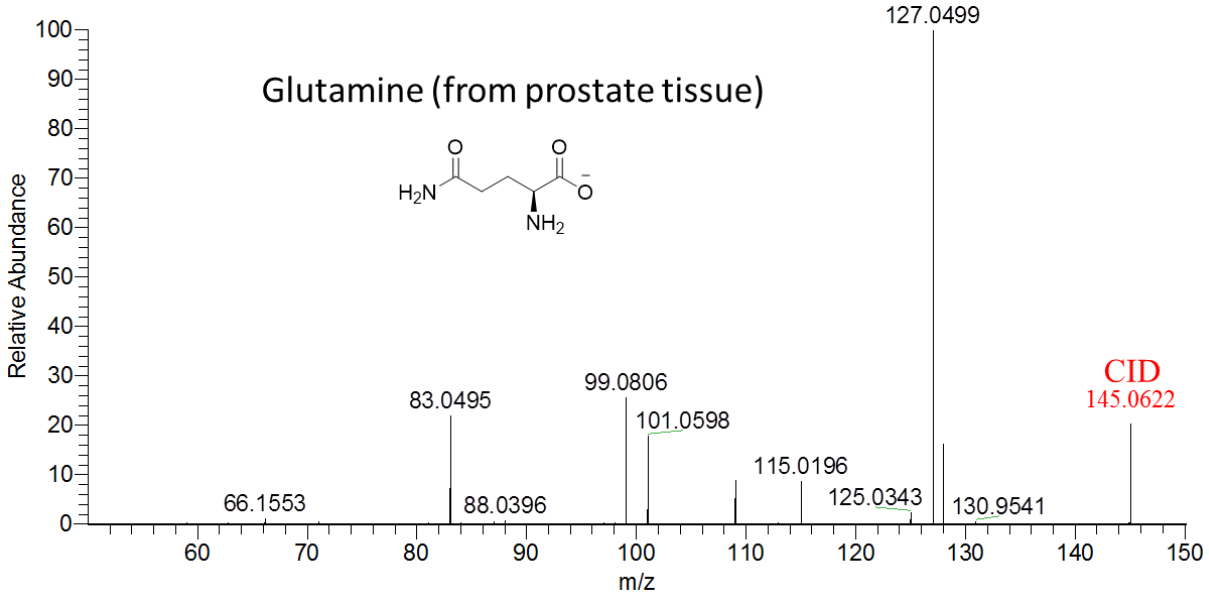

Fig. S4f. Collision induced dissociation (CID) at *m/z* 145.0622 (Table S1) identifies the species as salicylate when compared to the CID of the standard given in the MassBank (http://www.massbank.jp/).

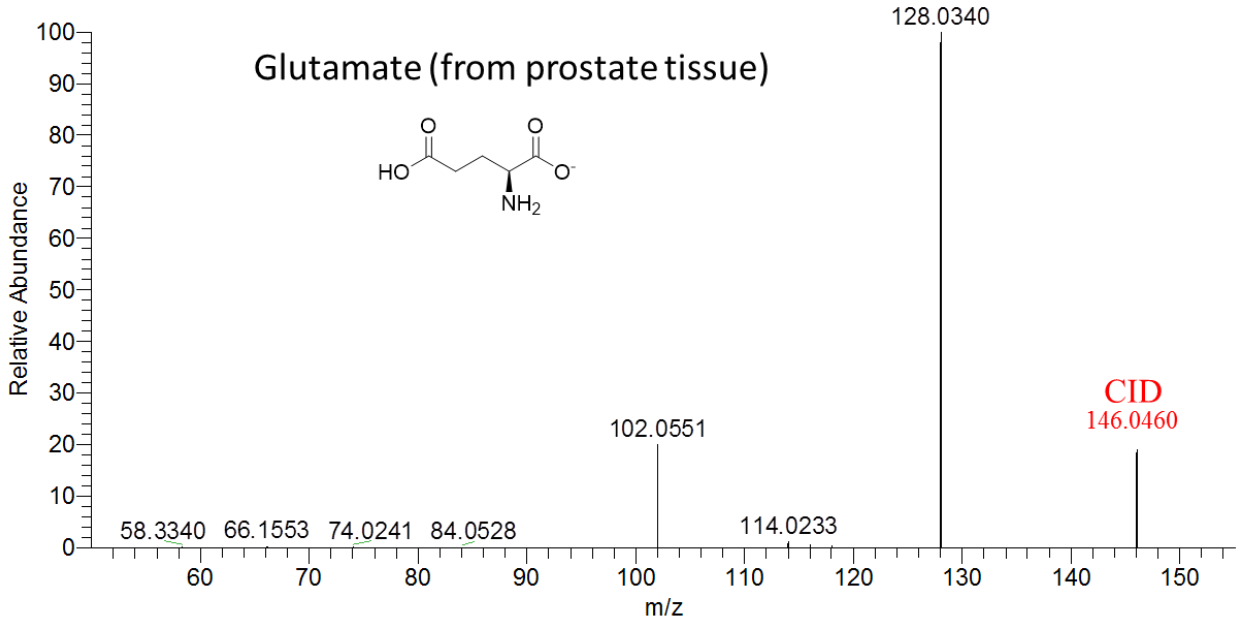

Fig. S4g. Collision induced dissociation (CID) at *m/z* 146.0460 (Table S1) identifies the species as glutamate when compared to the CID of the standard given in the MassBank (http://www.massbank.jp/).

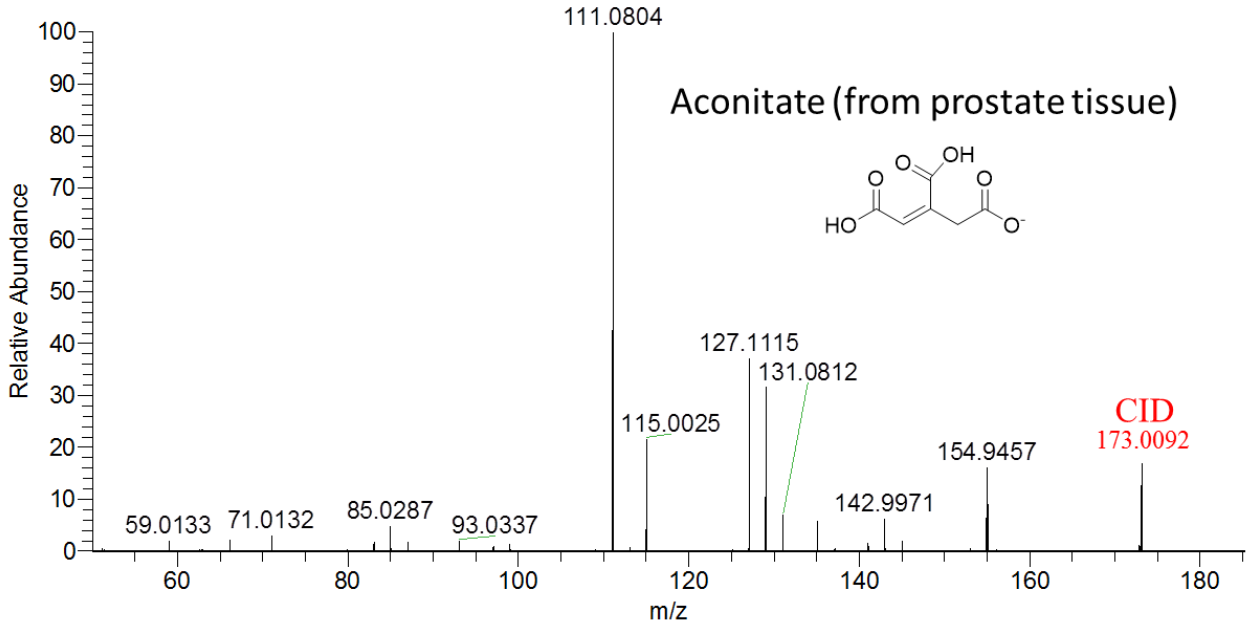

Fig. S4h. Collision induced dissociation (CID) at *m/z* 173.0092 (Table S1) identifies the species as cis/trans-aconitate when compared to the CID of the standard given in the MassBank (http://www.massbank.jp/).

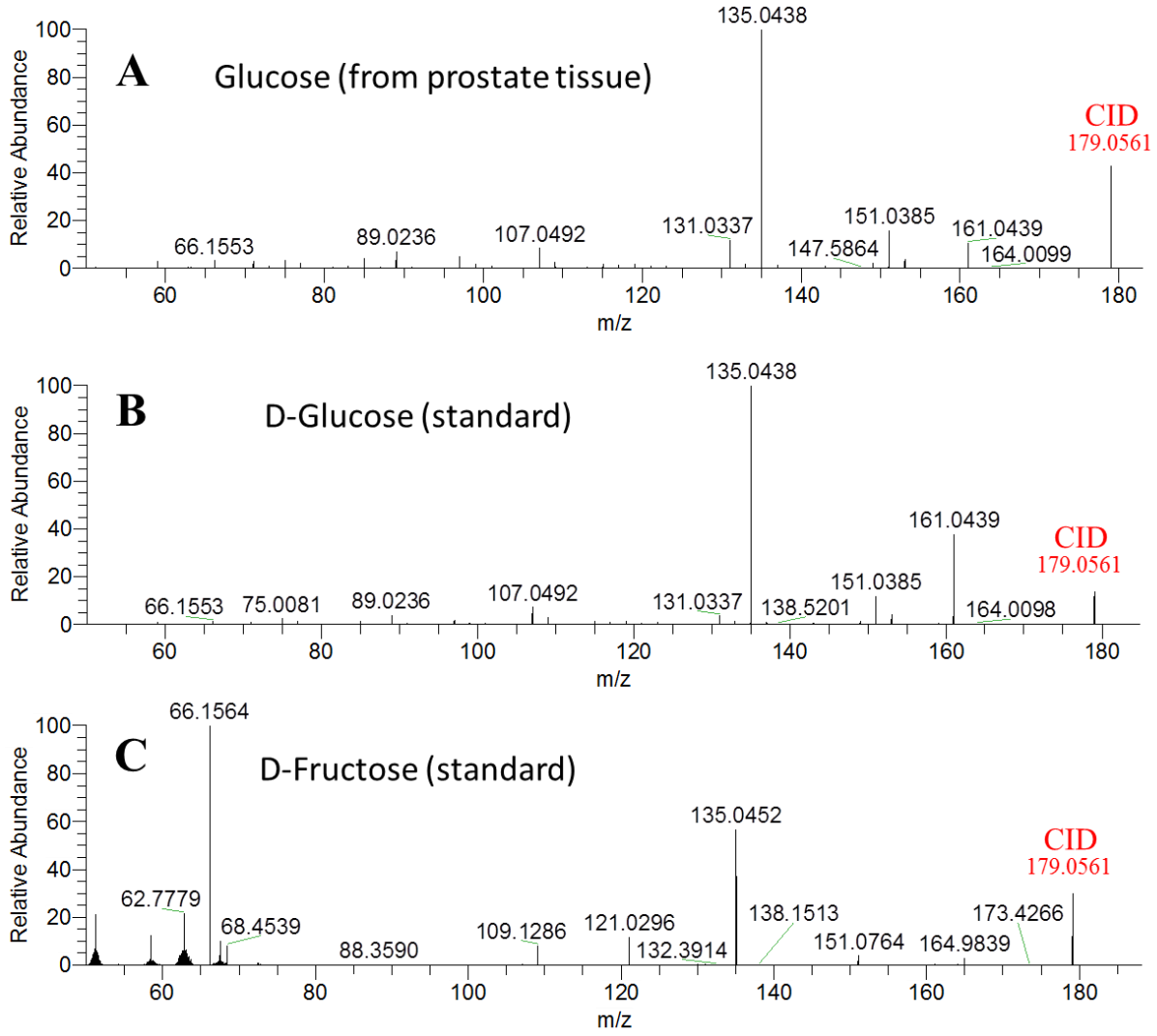

Fig. S4i. Collision induced dissociation (CID) at *m/z* 179.0561 (Table S1) identifies the species mostly as glucose when compared to the standard CID data of isomeric (B) D-glucose, and (C) D-Fructose.

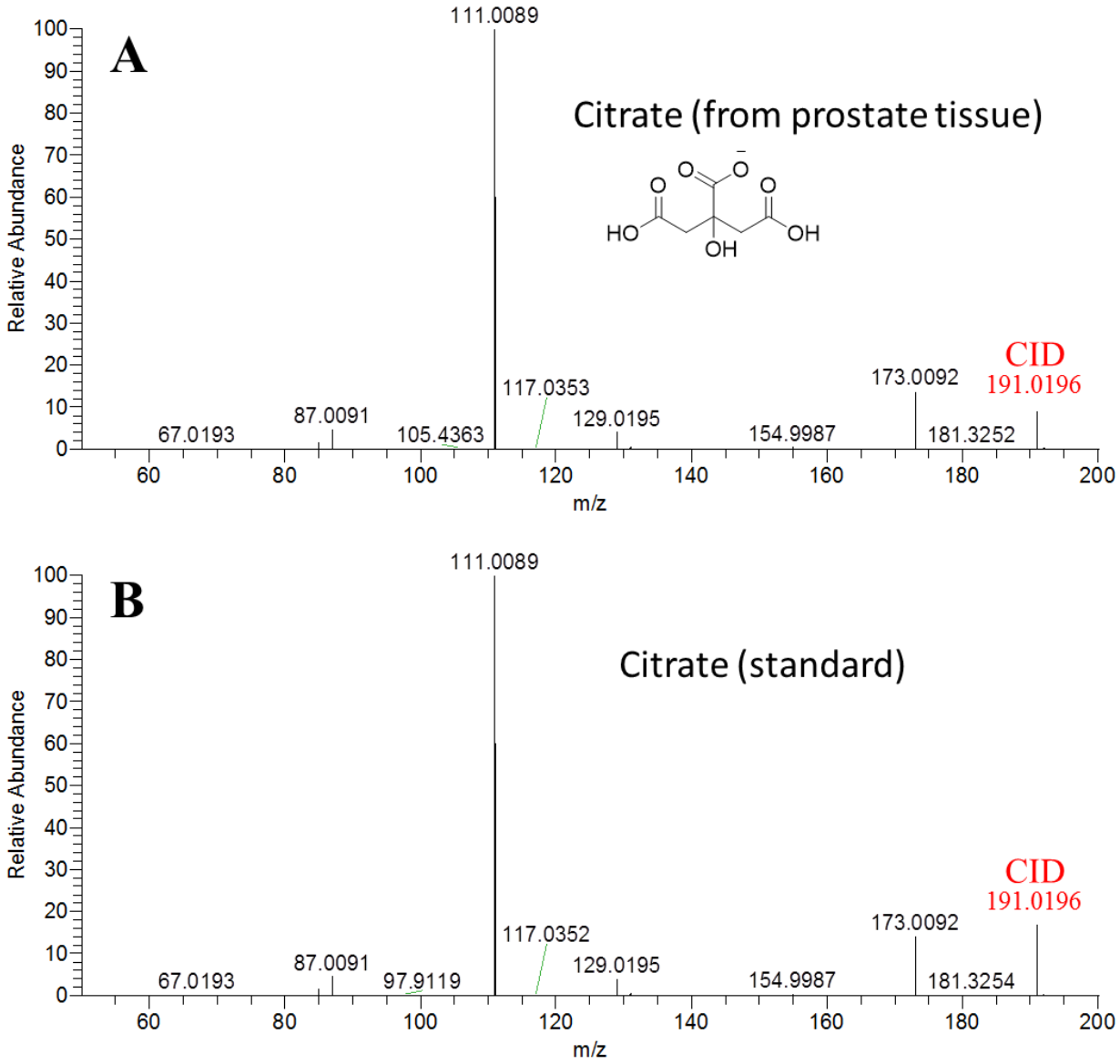

Fig. S4j. Collision induced dissociation (CID) of the species (A) at *m/z* 191.0196 (Table S1) identifies the species mostly as citrate when compared to the CID of (B) standard citrate. However, the ion signal at m/z 191.0196 in A might also be contributed by a minor amount (negligible) of isomeric isocitrate as analyzed by consulting the lipid database (www.lipidmaps.org/).

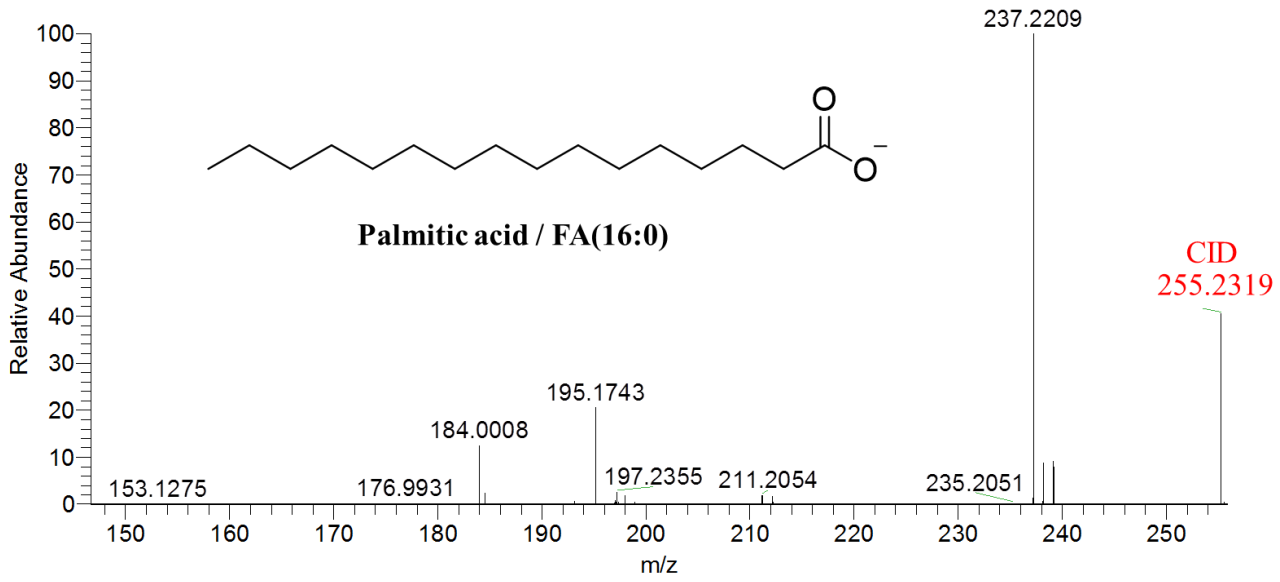

Fig. S4k. Collision induced dissociation (CID) of the species at *m/z* 255.2319 (Table S1) identifies the species as palmitic acid when compared to the CID of the standard in the lipid database (www.lipidmaps.org/)

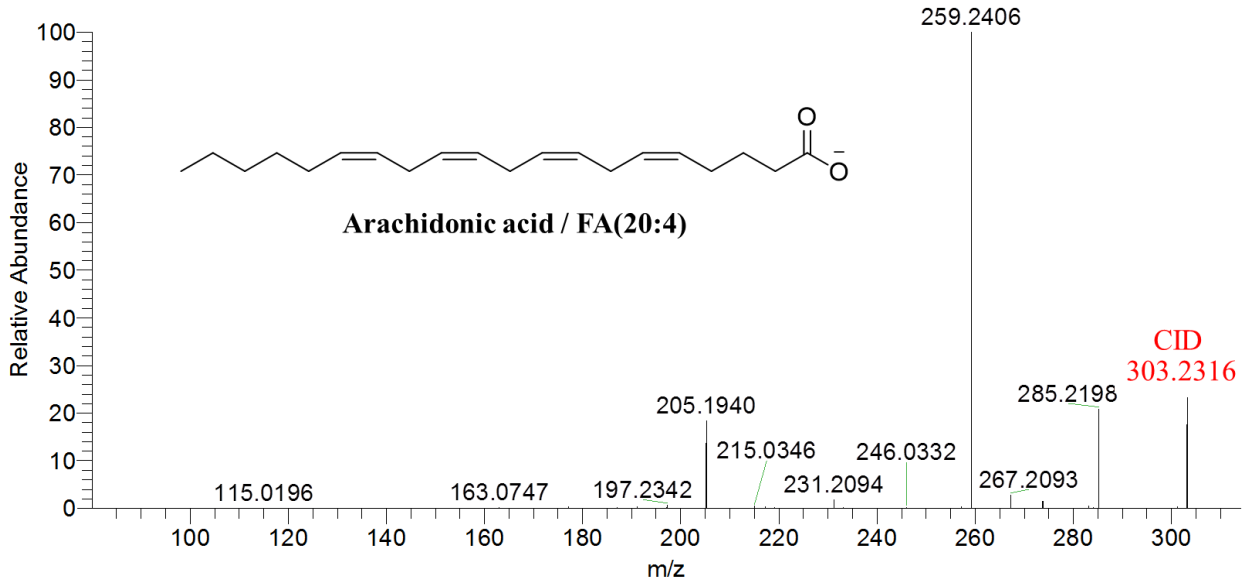

Fig. S4l. Collision induced dissociation (CID) of the species at *m/z* 303.2316 (Table S1) identifies the species as arachidonic acid when compared to the CID of the standard in the lipid database (www.lipidmaps.org/)

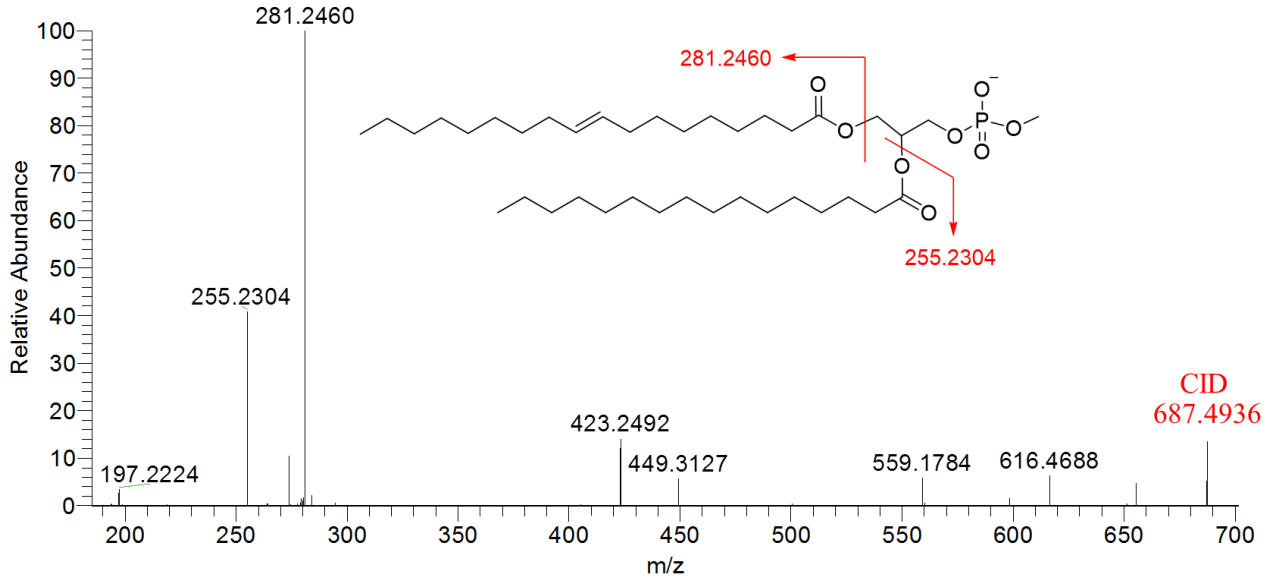

Fig. S4m. Collision induced dissociation (CID) at *m/z* 687.4936 (Table S1) identifies the species as the methyl ester of PA(18:1/16:0). Inset shows the fragmentation profile of the proposed species. We could not find any lipid corresponding to this m/z value among more than 40,000 biologically relevant lipids in the lipid database (http://www.lipidmaps.org/).

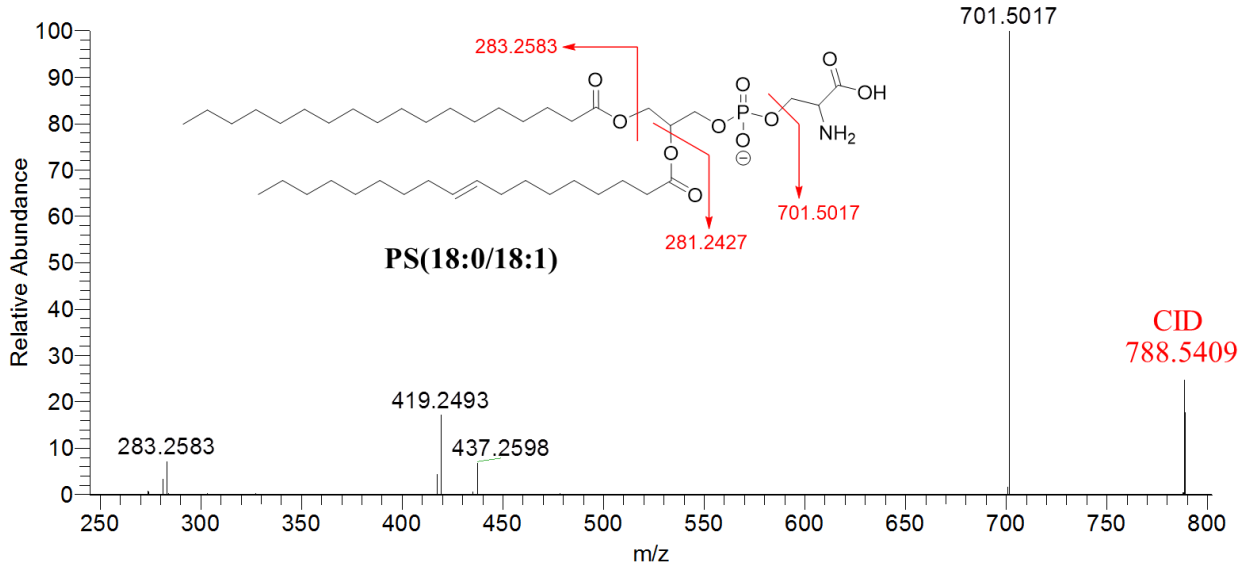

Fig. S4n. Collision induced dissociation (CID) at *m/z* 788.5409 (Table S1) identifies the species as PS(18:0/18:1). Inset shows the fragmentation profile of the proposed species, which also matches with the standard in the lipid database (www.lipidmaps.org/)

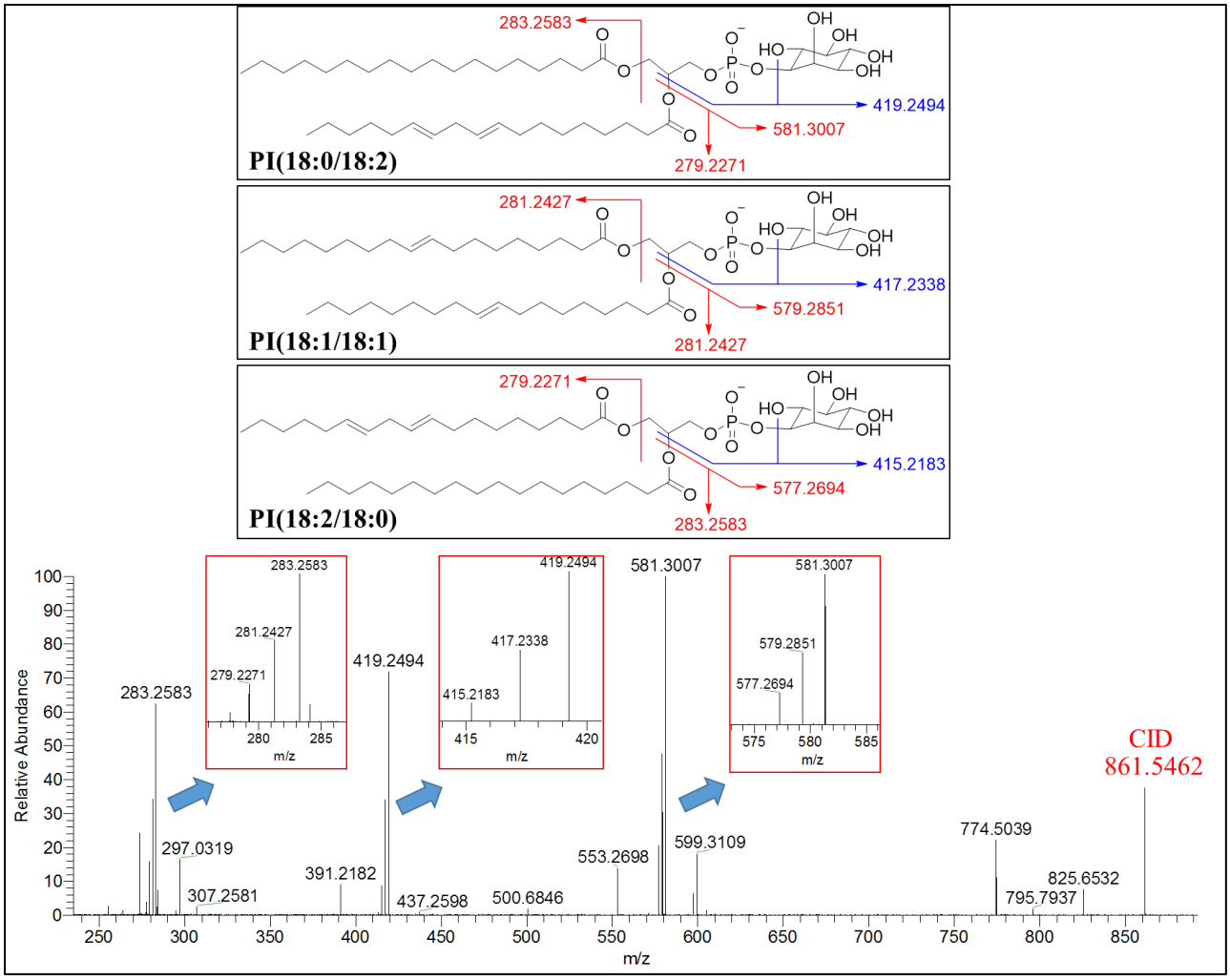

Fig. S4o. Collision induced dissociation (CID) at *m/z* 861.5462 (Table S1) identifies the species as the mixture of PI(18:0/18:2), PI(18:1/18:1), and PI(18:2/18:0) (approx. 4:2:1 ratio). Insets show the fragmentation profiles of proposed species, which also match with standards in the lipid database (www.lipidmaps.org/)

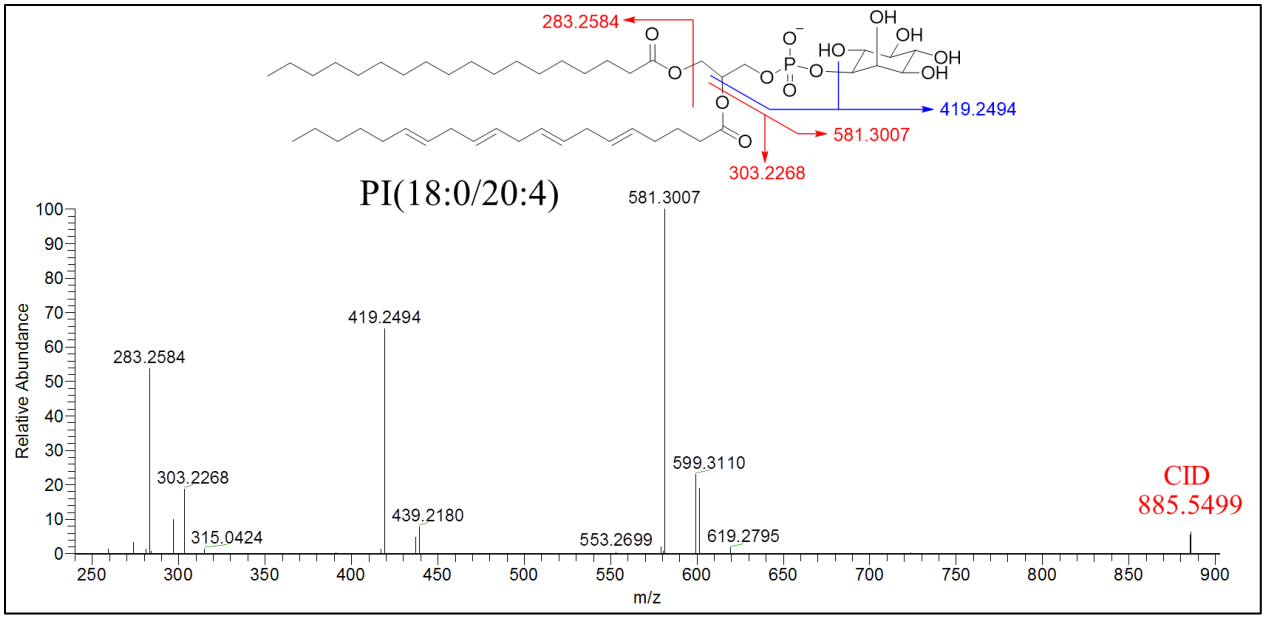

Fig. S4p. Collision induced dissociation (CID) at *m/z* 885.5499 (Table S1) identifies the species as PI(18:0/20:4). Inset shows the fragmentation profile of the proposed species, which also matches with the standard in the lipid database (www.lipidmaps.org/)

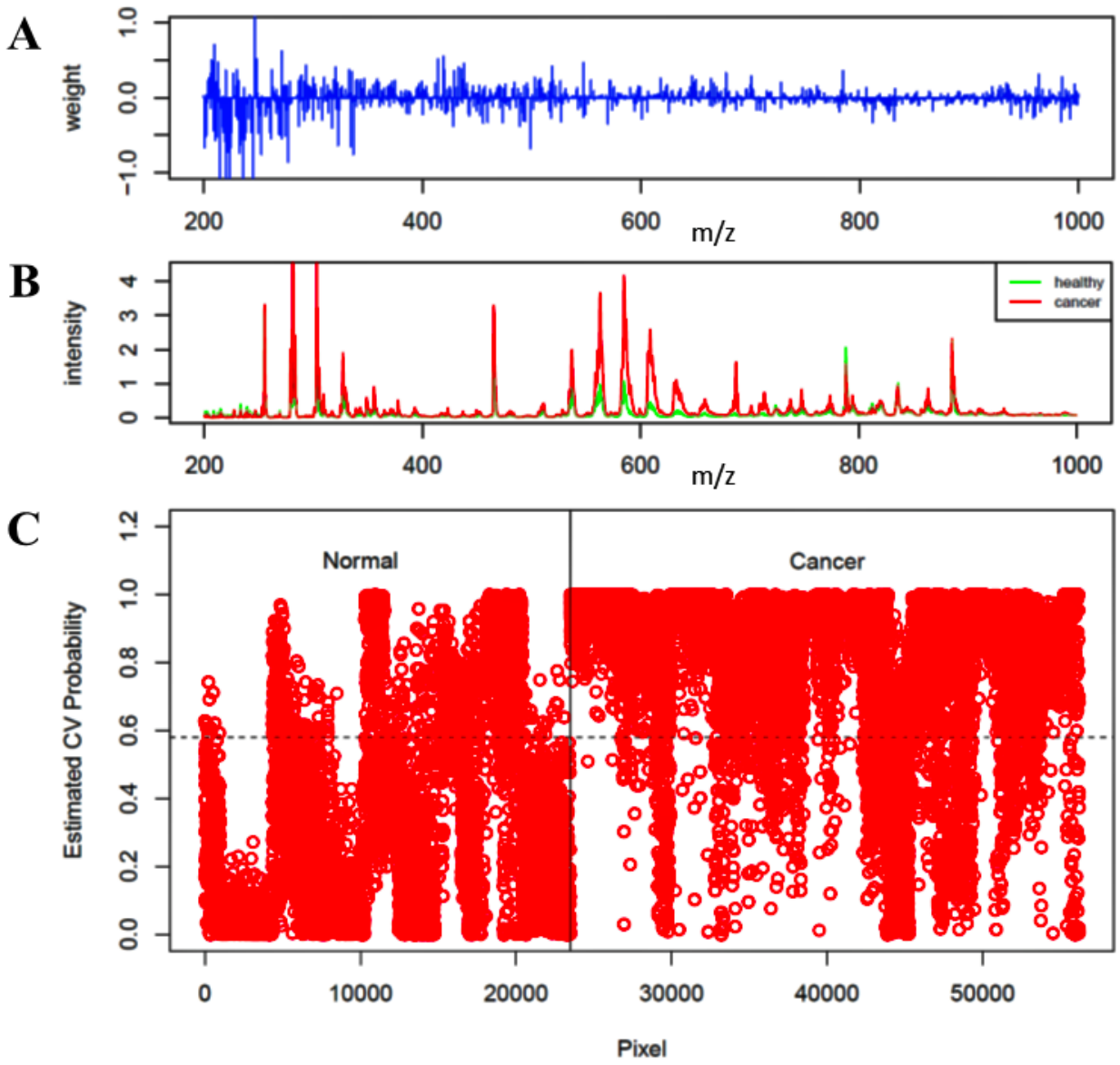

Fig. S5. The Lasso method yields a model with parsimonious sets of mass spectral features for distinguishing cancer and normal tissues from 45 specimens including 17 pure normal and 28 pure cancer tissues (total 56,196 pixels). (A) A mathematical weight for statistically informative features embedded in ion abundances (m/z 200-1000) was calculated by the Lasso to characterize a certain class (cancer *vs* normal). The ion with a positive weight indicates higher chance of cancer and the ion with negative weight indicates lower chance of cancer. (B) The average mass spectra calculated from all pixels in the training data set are shown in red for cancer and green for normal. (C) The cross-validated cancer probability was estimated based on lipid abundances (*m/z* 200-1000) for all pixels in the training set samples and represented in red circles. Each panel (left for normal and right for cancer) indicates the true class of each pixel displayed in arbitrary order.

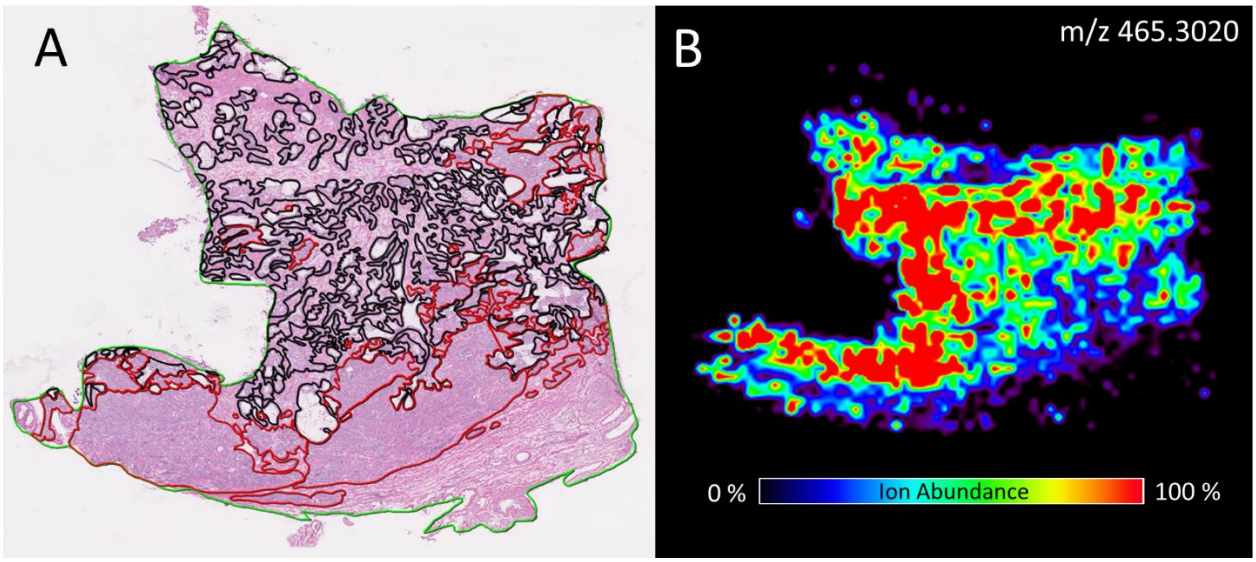

Fig. S6. Comparison of the (A) optical image (H&E), and (B) cholesterol sulfate ion (*m/z* 465.3020) image (DESI-MSI) of a typical prostate biopsy specimen. In the H&E (A) cancer areas have been demarcated by red, benign areas by black, and stroma areas by green.

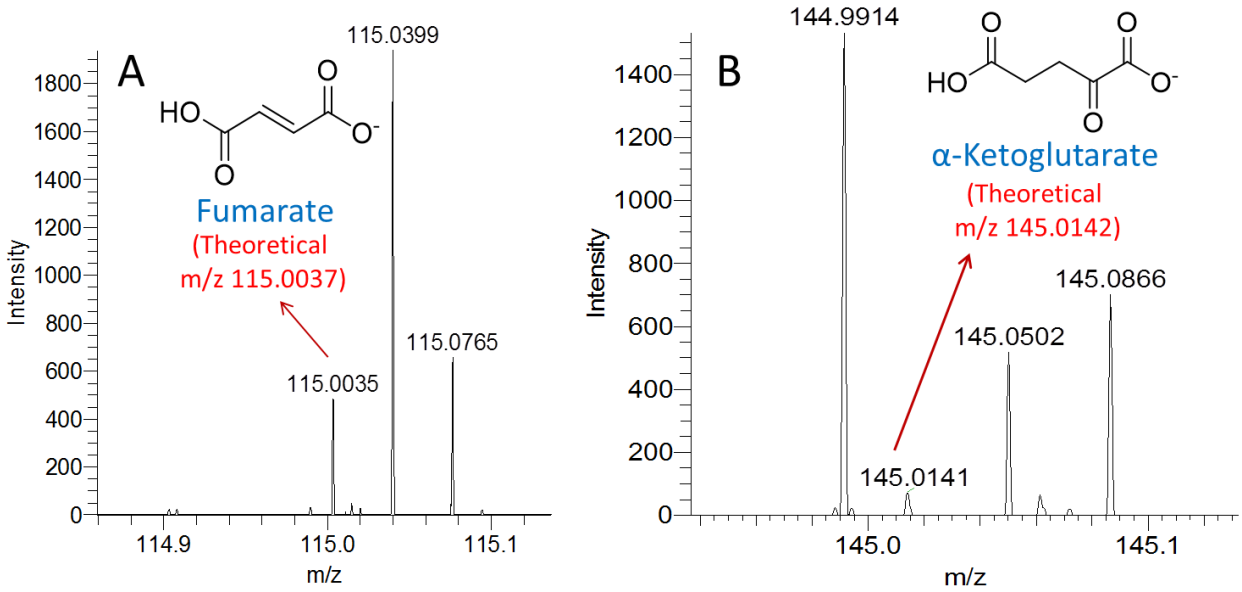

Fig. S7. Fumarate and α-ketoglutarate are intermediates of Krebs cycle (Fig. 2). By zooming the DESI mass spectrum (Fig. 1A), we detected the ion signals of (A) fumarate (*m/z* 115.0035; error ~1.7 ppm) and (B) α-ketoglutarate (*m/z* 145.0141; error ~0.7 ppm). The detected signal intensities of these species were too low to construct ion images and therefore they are absent in Fig. 1B.

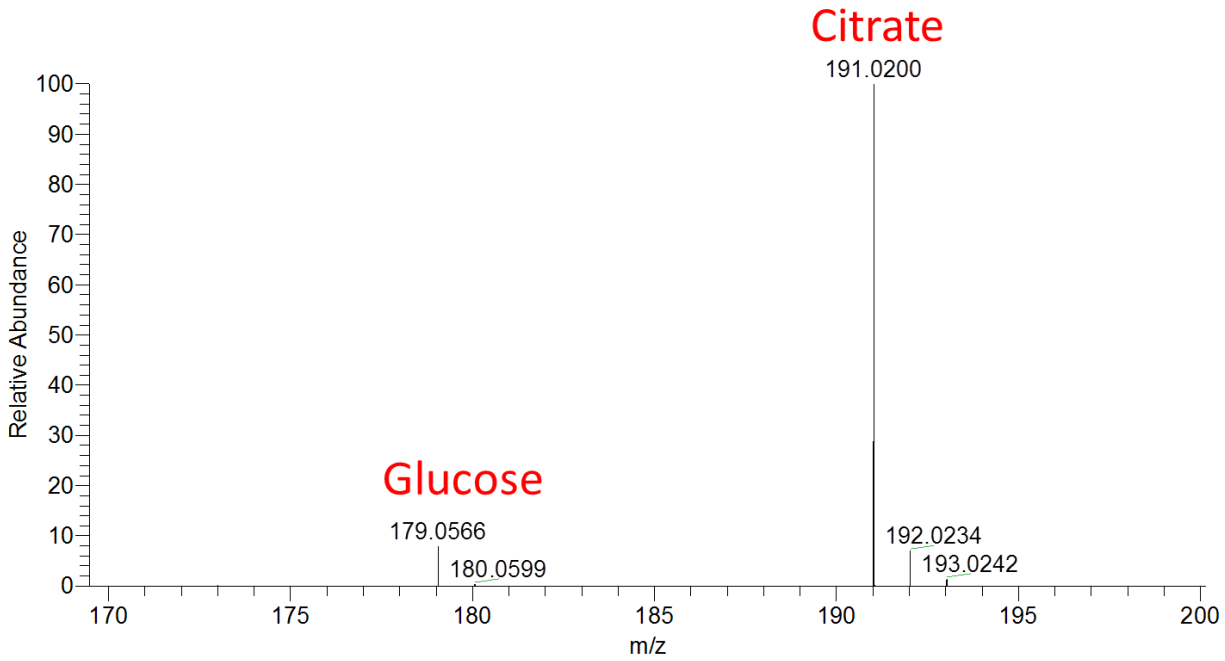

Fig. S8. Mass spectrum obtained by electrospraying a mixture of glucose and citrate in the concentration ratio 20:1 (glucose:citrate). Higher ion signal intensity from citrate indicates much better ionization efficiency of citrate than that of glucose.

Table S1. Identification of ions appearing in the mass spectral data (see Figs. S1 and S2) and considered in the statistical analysis.

Sl. No.	*Measured m/z**	*Proposed formula*	*Theoretical m/z†*	*Δm/z*	*Error (ppm)*	*Attribution#*
1	87.0089	$C_3H_3O_3$	87.0088	-0.0001	-1.15	Pyruvic acid
2	89.0243	$C_3H_5O_3$	89.0244	+0.0001	+1.12	Lactic acid
3	117.0196	$C_4H_5O_4$	117.0193	-0.0003	-2.56	Succinic acid
4	119.0348	$C_4H_7O_4$	119.0349	+0.0001	+0.84	Erythrose
5	124.0077	$C_2H_6NO_3S$	124.0074	-0.0003	-2.42	Taurine
6	125.0014	$C_2H_6O_4P$	125.0009	-0.0005	-4.00	Dimethylphosphate (DMP)
7	133.0144	$C_4H_5O_5$	133.0142	-0.0002	-1.50	Malic acid
8	137.0240	$C_7H_5O_3$	137.0244	+0.0004	+2.92	Salicylic acid
9	145.0622	$C_5H_9N_2O_3$	145.0619	-0.0003	-2.07	Glutamine
10	146.0460	$C_5H_8NO_4$	146.0459	-0.0001	-0.68	Glutamic acid
11	167.0219	$C_5H_3N_4O_3$	167.0211	-0.0008	-4.79	Uric acid
12	173.0092	$C_6H_5O_6$	173.0092	0	0	Aconitic acid
13	179.0561	$C_6H_{11}O_6$	179.0561	0	0	Glucose/Fructose
14	181.0710	$C_6H_{13}O_6$	181.0717	+0.0007	+3.86	Sorbitol
15	191.0196	$C_6H_7O_7$	191.0197	+0.0001	+0.52	Citric acid/Isocitric acid
16	215.0319	$C_6H_{12}ClO_6$	215.0328	-0.0009	-4.18	[Glucose/Fructose]Cl^-
17	255.2319	$C_{16}H_{31}O_2$	255.2329	-0.0010	-3.92	Palmitic acid/FA(16:0)
18	281.2474	$C_{18}H_{33}O_2$	281.2486	-0.0012	-4.27	Oleic acid/FA(18:1)
19	303.2316	$C_{20}H_{31}O_2$	303.2329	-0.0013	-4.29	Arachidonic acid/FA(20:4)
20	305.2473	$C_{20}H_{33}O_2$	305.2486	-0.0013	-4.26	FA(20:3)
21	327.2314	$C_{22}H_{31}O_2$	327.2329	-0.0015	-4.58	Docosahexaenoic acid/FA(22:6)
22	356.2789	$C_{20}H_{38}NO_4$	356.2806	-0.0017	-4.77	N-palmitoyl threonine
23	393.2627	$C_{23}H_{37}O_5$	393.2647	-0.0020	-5.08	Norcholic acid/Noravicholic acid/Norhyocholic acid
24	423.2496	$C_{20}H_{40}O_7P$	423.2517	-0.0021	-4.96	PA(17:0/0:0)
25	451.2808	$C_{22}H_{44}O_7P$	451.283	-0.0022	-4.87	PA(19:0/0:0)
26	465.3020	$C_{27}H_{45}O_4S$	465.3044	-0.0024	-5.16	Cholesterol sulfate
27	511.4708	$C_{32}H_{63}O_4$	511.4732	-0.0024	-4.69	Palmitic acid dimer
28	537.4862	$C_{34}H_{65}O_4$	537.4888	-0.0026	-4.84	[Oleic acid + palmitic acid]
29	563.5016	$C_{36}H_{67}O_4$	563.5044	-0.0028	-4.97	Oleic acid dimer
30	585.4858	$C_{38}H_{65}O_4$	585.4888	-0.0030	-5.12	[Oleic acid + arachidonic acid]
31	607.4702	$C_{38}H_{64}NaO_4$	607.4732	-0.0030	-4.94	Arachidonic acid dimer
32	631.4700	$C_{35}H_{68}O_7P$	631.4708	-0.0008	-1.27	PA(16:0/16:1)
33	687.4936	$C_{38}H_{72}O_8P$	687.497	-0.0034	-4.94	Methyl ester of PA(16:0/18:1)
34	709.4778	$C_{40}H_{70}O_8P$	709.4814	-0.0036	-5.07	PA(20:4/17:0)
35	713.5092	$C_{40}H_{74}O_8P$	713.5127	-0.0035	-4.90	PA(18:2/19:0), PA(18:1/19:1) PA(18:0/19:2)
36	735.4922	$C_{42}H_{72}O_8P$	735.497	-0.0048	-6.53	PA(20:4/19:1)
37	737.5093	$C_{42}H_{74}O_8P$	737.5127	-0.0034	-4.61	PA(20:4/19:0)
38	773.5303	$C_{42}H_{78}O_{10}P$	773.5338	-0.0035	-4.52	PG(18:1/18:1), PG(20:2/16:0)

39	788.5409	$C_{42}H_{79}NO_{10}P$	788.5447	-0.0038	-4.82	PS(18:0/18:1)
40	819.5144	$C_{46}H_{76}O_{10}P$	819.5182	-0.0038	-4.64	PG(18:1/22:6), PG(22:5/18:2)
41	835.5299	$C_{43}H_{80}O_{13}P$	835.5342	-0.0043	-5.15	PI (18:1/16:0)
42	861.5462	$C_{45}H_{82}O_{13}P$	861.5499	-0.0037	-4.29	PI(18:0/18:2), PI(18:1/18:1), PI(18:2/18:0)
43	885.5456	$C_{47}H_{82}O_{13}P$	885.5499	-0.0043	-4.86	PI(18:0/20:4)

*****High mass accuracy/mass resolution measurements were obtained using an Orbitrap mass spectrometer.

[†]Theoretical *m/z* was based on the monoisotopic mass of the respective deprotonated species.

[#]Tentative assignments are based on high mass accuracy, isotopic distribution, and tandem mass spectrometry (using the collision induced dissociation technique). FA, fatty acid; PA, phosphatidic acid; PG, glycerophosphoglycerols; PS, glycerophosphoserines; PI, glycerophosphoinositols. (M:N/X:Y) denotes the number of carbons (M and X) and the number of double bonds (N and Y) in each of the two fatty-acid chains. As the position and stereochemistry of the double bond in FA complicates the structural elucidation, they are often tentatively assigned in FA and glycerophospholipids (GP).

Table S2. List of *m/z* values of some unidentified small metabolites* considered in the statistical analysis.

Sl. No.	*m/z values*	*Species labels*
44	126.9979	Unknown-1
45	127.0165	Unknown-2
46	127.0398	Unknown-3
47	162.9819	Unknown-4
48	164.056	Unknown-5
49	171.1021	Unknown-6
50	183.0041	Unknown-7
51	184.9836	Unknown-8
52	185.0215	Unknown-9
53	187.0413	Unknown-10
54	188.0169	Unknown-11

*These species consistently appeared with significant abundances in the mass spectral data of all patients. Though we could not identify these species, they appeared to be statistically significant.

Table S3. List of statistical weights* corresponding to the ratio of two small metabolites ion-signals.

Sl. No.	Species 1	Species 2	Statistical weights of (Species 1/Species 2)
1	Pyruvate	Erythrose	-0.978280293483076
2	Lactate	Taurine	0.521863156492476
3	Lactate	Unknown-3	-0.67300411456976
4	Lactate	Malate	-1.04098548082797
5	Lactate	Uric acid	1.59551220725463
6	Succinate	Erythrose	-0.0389180216322248
7	Succinate	Taurine	1.71198325162579
8	Succinate	Salicylate	-0.0886298574011376
9	Succinate	Unknown-7	-0.19308241990539
10	Succinate	Citrate	0.671640971272858
11	Erythrose	Unknown-1	-1.19733921375036
12	DMP	Salicylate	-1.09971833395516
13	Unknown-1	Unknown-2	1.72900160556438
14	Unknown-2	Malate	-0.0110749248597451
15	Unknown-2	Salicylate	-0.84923796548107
16	Glutamine	Glutamic acid	-1.06176762761876
17	Unknown-4	Uric acid	-0.23007413181499
18	Unknown-5	Unknown-11	0.4906097034672
19	Unknown-5	Citrate	0.180594125454462
20	Uric acid	Citrate	0.229473178923721
21	Unknown-6	Citrate	0.92501991536386
22	Aconitate	Glucose	-0.357907305086132
23	Aconitate	Unknown-8	-0.44477242765184
24	Glucose	Sorbitol	1.73879635472162
25	Glucose	Unknown-9	-0.676143475153528
26	Glucose	Unknown-11	1.15893497995136
27	Glucose	Citrate	0.0707798130713905
28	Sorbitol	Citrate	0.75232437548441

*****Probability of cancer $\propto$ (Statistical weight) log ($I_{species\text{-}1}/I_{species\text{-}2}$)