Supporting Information

Title: Metabolomics Based Predictive Biomarkers of Oral Cancer and its Severity in *Human* Patients from North India Using Saliva

Supplementary contents:

Table of content:

-**Figure S7. A.** Two-dimensional score plot of principal component analysis with red colour representing control and green patient from human Saliva. **B.** Biplot of principal component analysis.

-Figure S8. Boxplots show levels of differential metabolites in oral cancer and control individuals.

-Figure S9. Variables importance in projection (VIP).

-Figure S10. ROC curve analysis for the saliva specific metabolites level.

-Figure S11. A. Cross Validation of the multivariate O-PLS-DA model using 10-Fold related to NMR spectroscopy on saliva sample of Oral Cancer patients. **B.** Permutation test of the multivariate PLS-DA model using by separation distance based on sum of squares between and sum of squares within (B/W) ratio for NMR on saliva sample of Oral Cancer patients.

Table S1 Metabolites resulted from saliva sample patients using NMR spectroscopy.

S. No.	Metabolites	Assignment	^{1}H chemical shift (ppm) and multiplicity	^{13}C chemical shift (ppm)	CHEBI No
1	Acetone	CH_3	2.22(s)	-	15347
2	Acetoacetate	CH_3	2.29	-	13705
		CH_2	3.45(s)	-	
3	Adipic acid	CH_2	1.56	-	30832
		CH_2COOH	2.23		
4	Alanine	βCH_3	1.48(d)	19.00	15428
		αCH	3.78(q)	53.41	
5	Aspartic Acid	βCH_2	2.67	39.36	22660
		βCH_2	2.88	39.36	
		αCH	3.88(dd)	54.90	
6	Betaine	n-CH_3	3.22(s)	56.68	91242

		CH	3.89(s)	69.00	
		αCH	3.75	57.00	
7	Butyric acid	CH_2	1.68	-	30772
8	Creatine	CH_3	3.03	39.72	16919
		CH_2	3.92	56.73	
9	Citrate	CH_2	2.48		30769
10	α-Glucose	C4H	3.40	72.48	17634
		C2H	3.54	74.10	
		C3H	3.71	75.58	
		C6H	3.83	63.33	
		C5H	3.85	74.85	
11	β-Glucose	C4H	3.42	72.48	17634
		C2H	3.24	77.09	
		C3H	3. 49	78.24	
		C6H	3.74	63.99	
		C5H	3.48	78.66	
		C1H	4.65	98.78	
		C1H	5.24(d)	94.98	
12	Glutamate	βCH	2.06	29.87	29985
		β′CH	2.13	29.87	
		γCH_2	2.35	36.16	
		αCH	3.78	44.32	
13	Histidine	C4H,ring	7.05	119.49	15971
		C2H,ring	7.77	13.07	
		αCH	3.99	57.50	
		βCH_2		30.013	
14	Ethanol	CH_3	1.17(t)	19.48	16236
		CH	3.64	60.13	
15	Formate	CH	8.44	-	15740
16	Isoleucine	δCH_3	0.92(t)	13.83	17191
		βCH_3	0.99	17.41	
		γCH	1.25	27.18	
		γCH	1.45	27.18	
		βCH	1.96	38.81	
		αCH	3.65	63.40	
17	Leucine	δCH_3	0.92(t)	23.64	15603

		δ′CH_3	0.99	24.93	
		γCH	1.25	27.05	
		βCH_2	1.45	42.65	
		αCH	1.96	56.10	
		CH	3.65	57.21	
18	Phenylalanine	βCH	3.14	39.40	17295
		β′CH	3.27	39.40	
		αCH	3.98	58.98	
		C2H, C6H ring	7.33	132.36	
		C4H, ring	7.36	130.42	
		C3H, C5H, ring	7.42	132.09	
19	Pyruvate	CH_3	2.36	29.55	15361
20	Succinate	CH_3	2.39	36.85	30031
21	Urea	CH2	5.78	-	16199
22	Tryptophan	CH	7.28	-	27897

Table S2. Cross Validation of the multivariate O-PLS-DA model using 5-Fold and 10-Fold related to NMR spectroscopy on saliva sample of Oral Cancer patients.

PLS-DA cross-validation related to polar extract using NMR spectroscopy

Measure using 5-fold	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.96667	0.97778	0.97778	0.97778	0.97778
R2	0.87363	0.91066	0.93114	0.94082	0.95369
Q2	0.83169	0.88065	0.89608	0.8964	0.89953
Measure using 10-fold	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.96667	0.96668	0.96668	0.96778	0.97889
R2	0.87363	0.91066	0.93114	0.94082	0.95369
Q2	0.84608	0.88627	0.90095	0.90229	0.90315

Table S3. Cross Validation of the multivariate O-PLS-DA model using 5-Fold related to NMR spectroscopy on saliva sample of OSCC Stage 1 and OSCC stage 2.

PLS-DA cross-validation related to polar extract using NMR spectroscopy

Measure using 5-fold	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.5	0.4	0.45	0.5	0.55
R2	0.52603	0.66583	0.81193	0.86076	0.92941
Q2	-0.42358	-0.20674	-0.42536	-0.6785	-0.80796

Table S4. ROC curve analysis of potential biomarkers in oral cancer.

Metabolites	AUC	t-TEST
Acetone	0.98123	2.2774×10^{-27}
Tryptophan	0.95358	2.116399×10^{-21}
5-aminopentanoic acid	0.91506	9.1662×10^{-13}
Betaine	0.91506	6.1673×10^{-12}
Aspartic acid	0.90667	8.0693×10^{-14}
Ethanol	0.89778	7.7899×10^{-7}
Acetoacetate	0.89284	1.1811×10^{-12}
Adipic acid	0.8642	5.0659×10^{-11}
Citrate	0.82963	9.6554×10^{-8}
Formate	0.75605	1.5455×10^{-5}

Phenylalanine	0.74519	4.344×10^{-5}
Butyric acid	0.73926	0.0063549
Imidazole	0.72	2.6841×10^{-4}
Creatine	0.66963	0.069028
Glutamic acid	0.66469	5.4578×10^{-4}
Succinic acid	0.63556	0.030161
Alanine	0.6237	0.10795
Histidine	0.60988	0.62005
Propionic acid	0.6	0.95428
Glucose	0.57975	0.061482
Leucine	0.55802	0.94562
Isoleucine	0.54222	0.22714
Pyruvate	0.53481	0.75944
Hypoxanthine	0.52741	0.93639
Galactose	0.50272	0.3261

Table S5. Overview of enrichment metabolites.

Metabolite Set	Hits	Expect	P value	FDR
Ketone Body Metabolism	3	0.259	0.00172	0.0841
Glucose-Alanine Cycle	3	0.259	0.00172	0.0841
Ammonia Recycling	4	0.619	0.00257	0.0841
Warburg Effect	5	1.14	0.00401	0.0879
Valine, Leucine and Isoleucine Degradation	5	1.18	0.00467	0.0879
Butyrate Metabolism	3	0.379	0.00538	0.0879
Transfer of Acetyl Groups into Mitochondria	3	0.439	0.00823	0.115
Lactose Degradation	2	0.18	0.0125	0.127
Glutamate Metabolism	4	0.958	0.0128	0.127
Phenylalanine and Tyrosine Metabolism	3	0.539	0.0147	0.127

Malate-Aspartate Shuttle	2	0.2	0.0155	0.127
Urea Cycle	3	0.559	0.0162	0.127
Arginine and Proline Metabolism	4	1.04	0.0168	0.127
Citric Acid Cycle	3	0.639	0.0233	0.163
Glycine and Serine Metabolism	4	1.18	0.0258	0.168
Beta-Alanine Metabolism	3	0.679	0.0274	0.168
Alanine Metabolism	2	0.339	0.043	0.248
Glycolysis	2	0.459	0.0745	0.4
Methylhistidine Metabolism	1	0.0798	0.0776	0.4
Cysteine Metabolism	2	0.519	0.0924	0.453

Table S6. Some Information about Different Metabolites.

Metabolites	**p value**	**FC**	**log2(FC)**
Acetone	2.28×10^{-27}	0.090688	2.5262
Tryptophan	2.12×10^{-21}	0.16163	1.7826
Aspartic acid	8.07×10^{-14}	2.9824	1.2325
5-aminopentanoic acid	9.17×10^{-13}	2.6951	1.2526
Acetoacetate	1.18×10^{-12}	0.44268	1.1707
Betaine	6.17×10^{-12}	2.445	1.2279
Adipic acid	5.07×10^{-11}	0.46772	1.0809
Formate	0.0000155	16.488	1.5122
Glutamic acid	0.00054578	0.24188	0.85171
Butyric acid	0.0063549	2.7133	0.89558

Figure S1A. shows a clinical case of Oral submucous Fibrosis (OSMF) of Grade 4 severity. Patient presents with a mouth opening of 3mm. **B**. Intraorally Loss of elasticity and leathery texture of the buccal mucosa, showing "marble -like "appearance of the oral mucosa and blanching of the mucosa is seen due to loss of normal pink color and appearance of white or pale mucosa. **C**. shows a clinical case of Oral squamous cell carcinoma (OSCC) of Stage 2 severity with a ulcer proliferative lesion on right alveolus of mandible (2 x3cm size) according to AJCC classification.

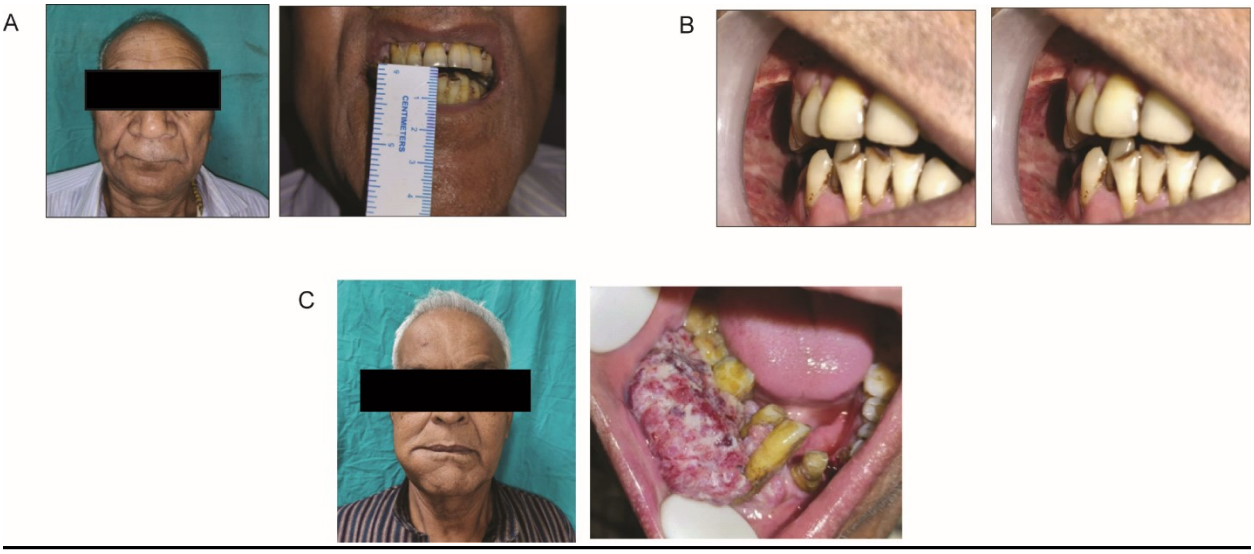
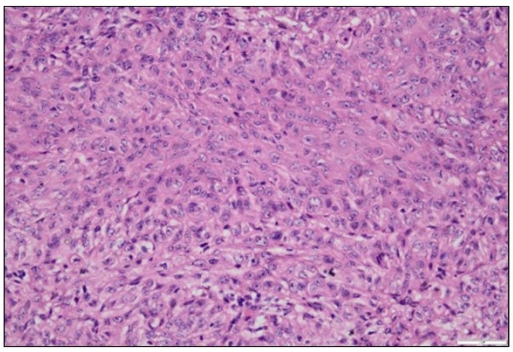
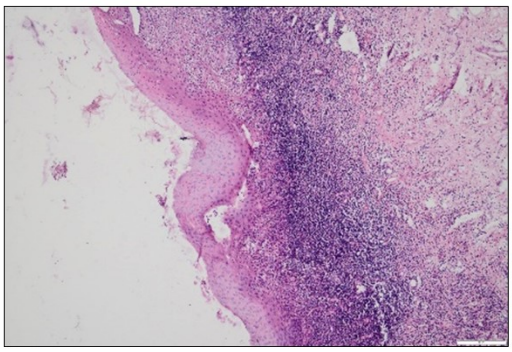


Figure S2. Histopathological slide of OSCC & OSMF



OSCC



OSMF

Figure S3. 1D-cpmgpr 1H nuclear magnetic resonance spectra of saliva samples from oral cancer patient. The spectra are acquired at 800 MHz and T = 310 K.

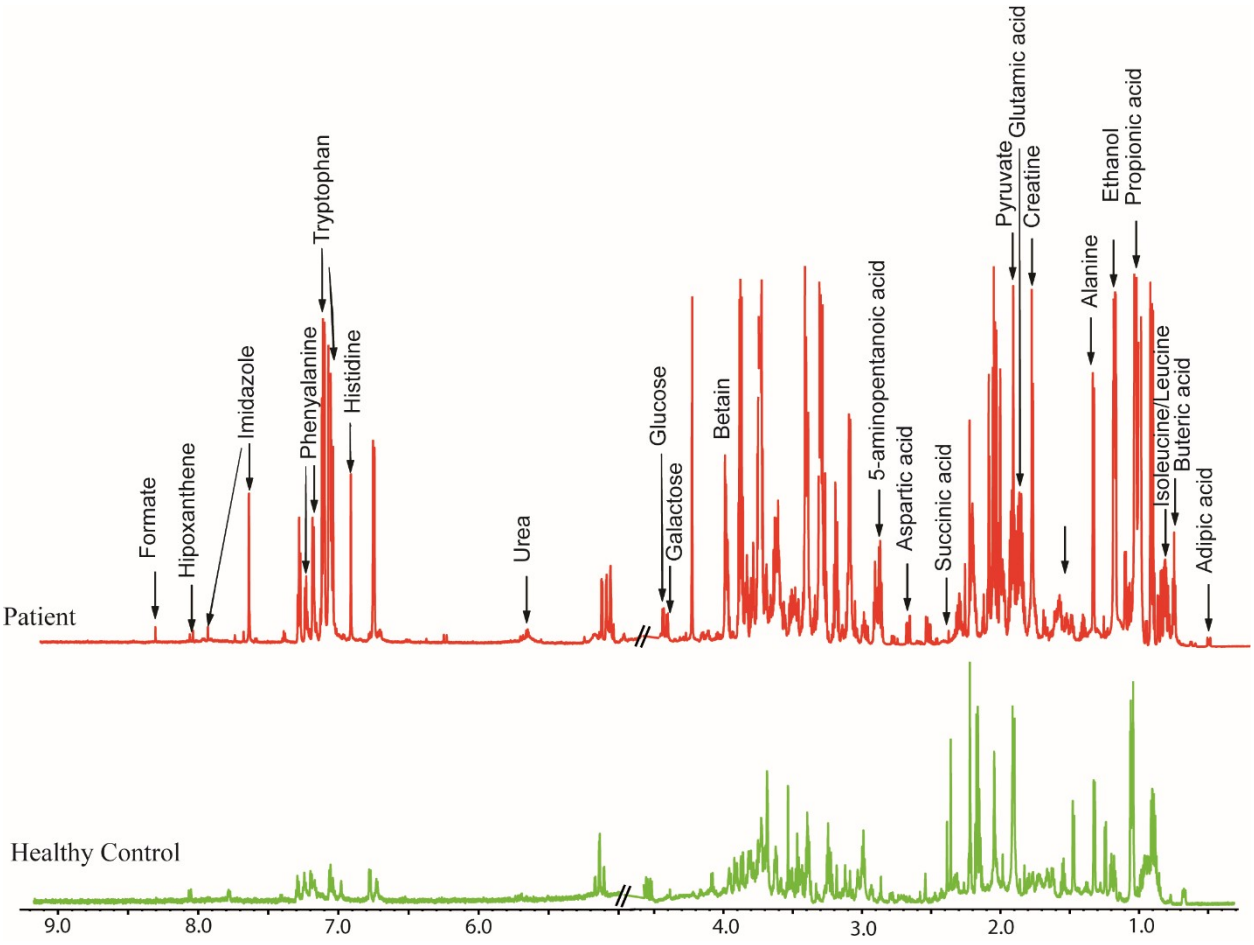


Figure S4. 2D HSQC nuclear magnetic resonance spectra of saliva samples from oral cancer patient. The spectra are acquired at 800 MHz and T = 310 K.

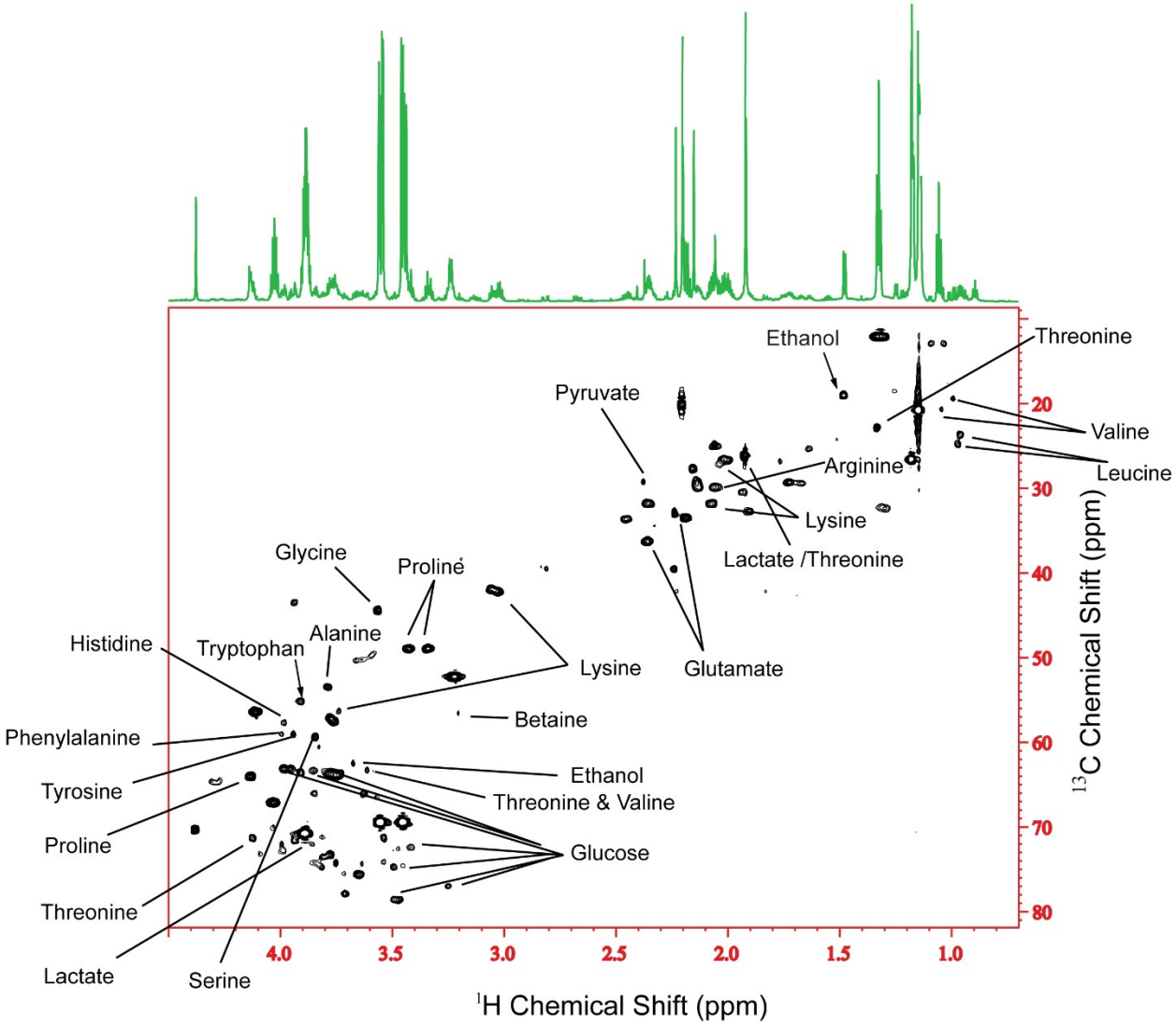


Figure S5. 2D COSY nuclear magnetic resonance spectra of saliva samples from oral cancer patient. The spectra are acquired at 800 MHz and T = 310 K.

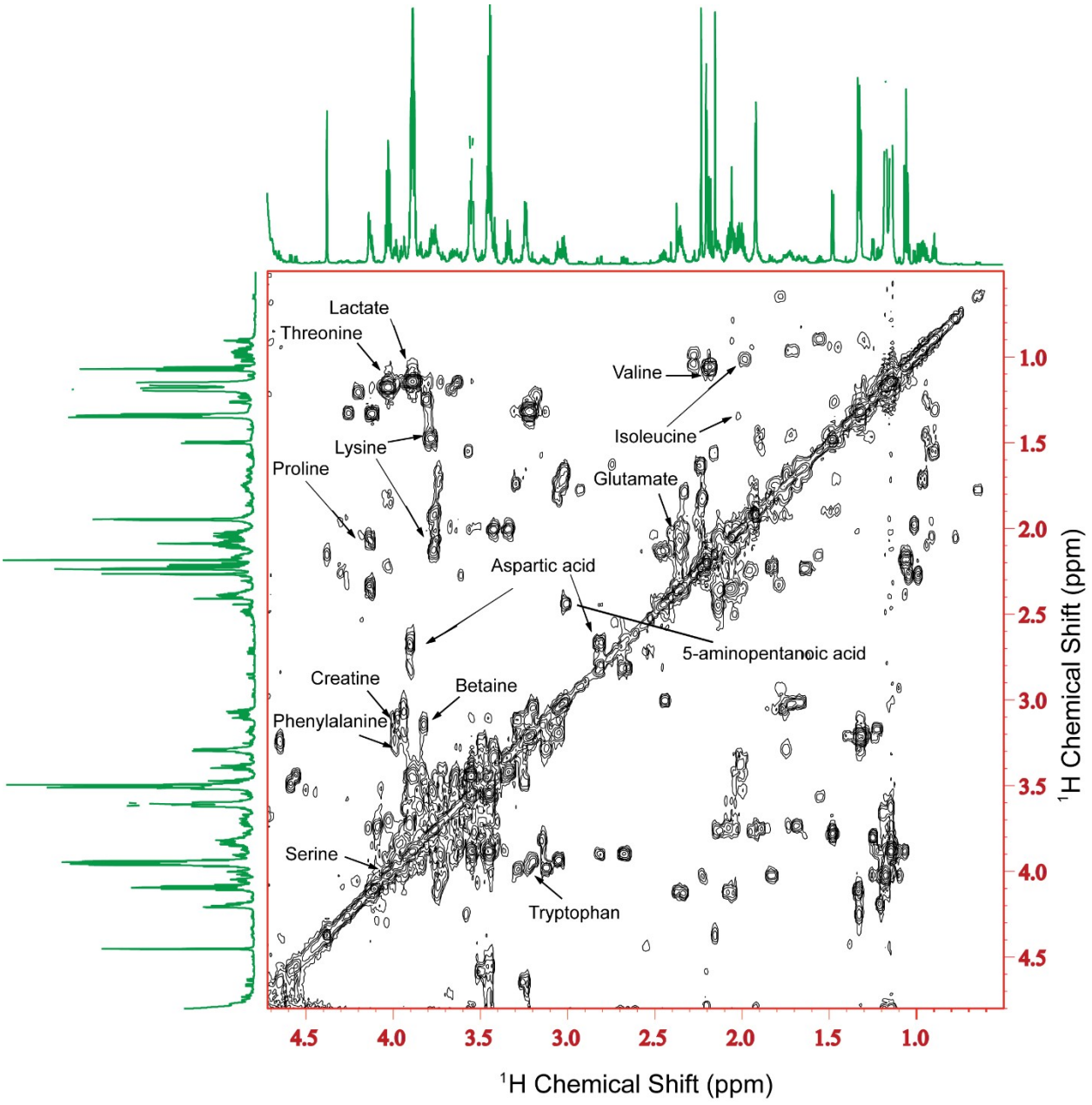
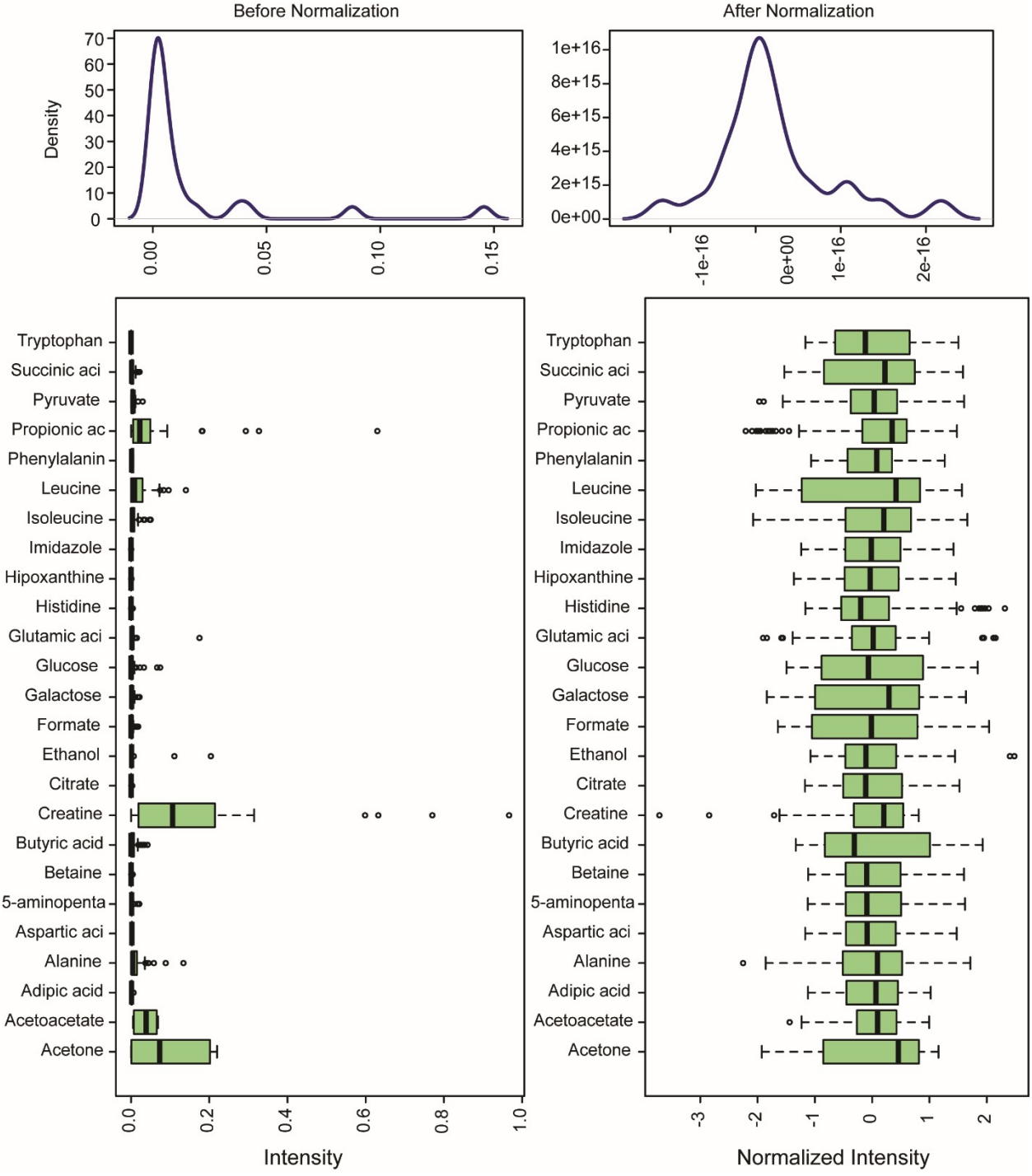
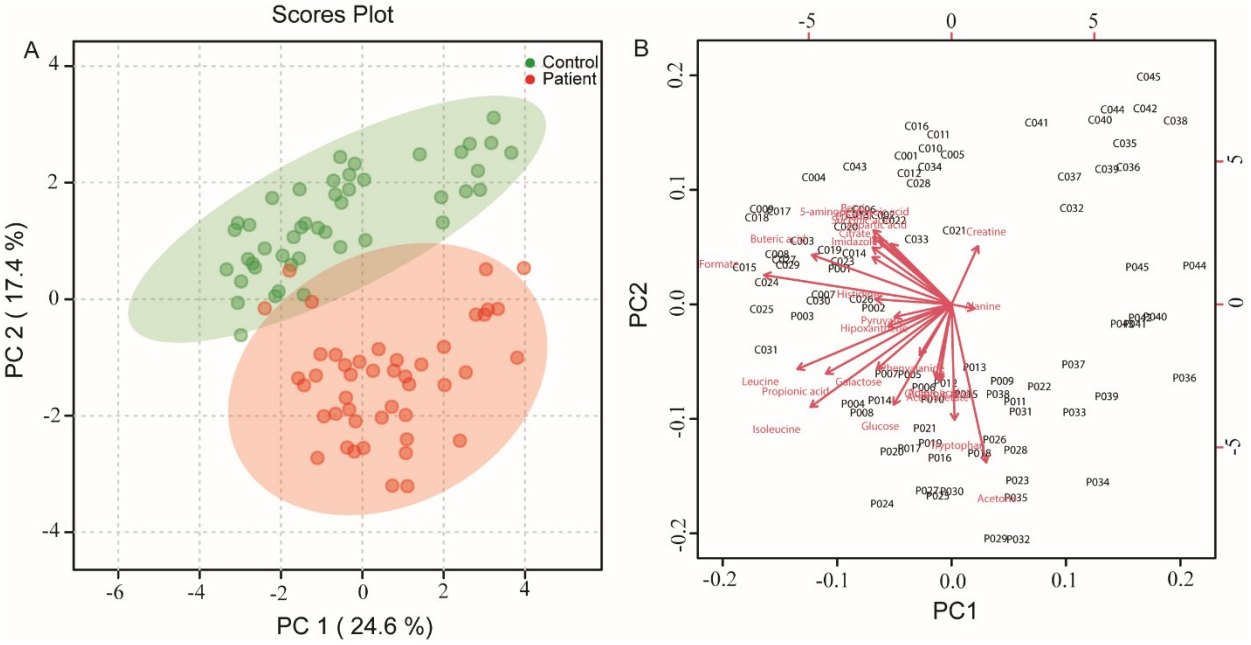


Figure S6. Data normalization by Pareto scaling and log transformation.

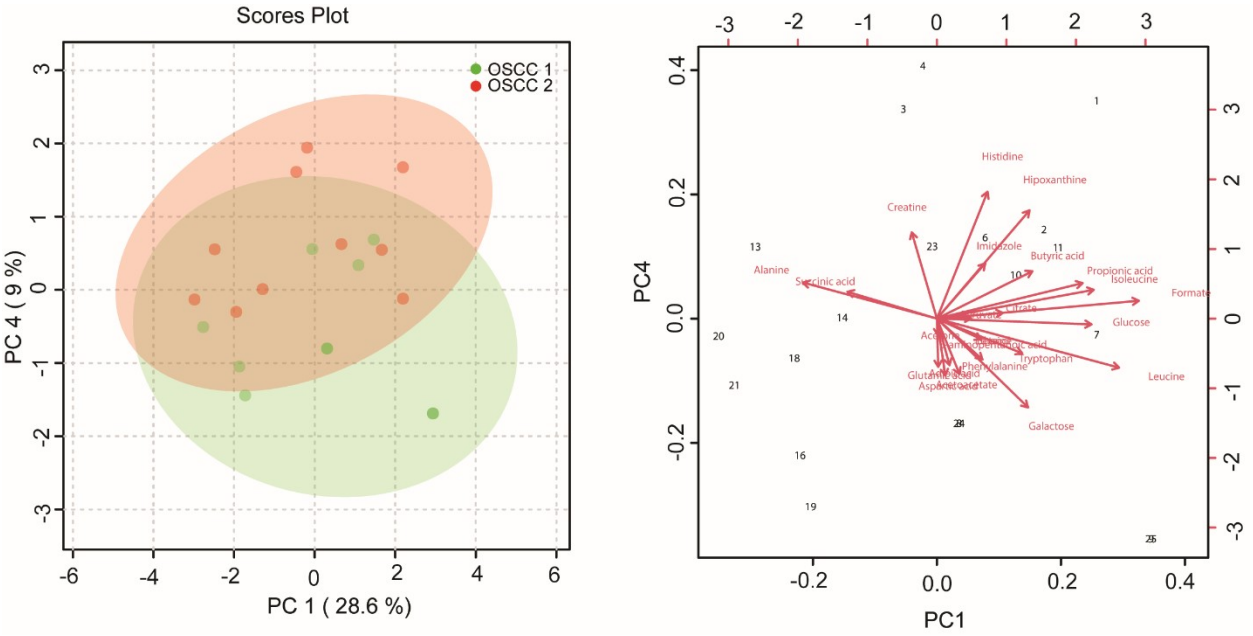
Data normalization before statistical analysis data was Pareto scaled and log transformed to minimize both the induced and uninduced discrepancy within the data to obtain Gaussian distribution for better interpretation of results.



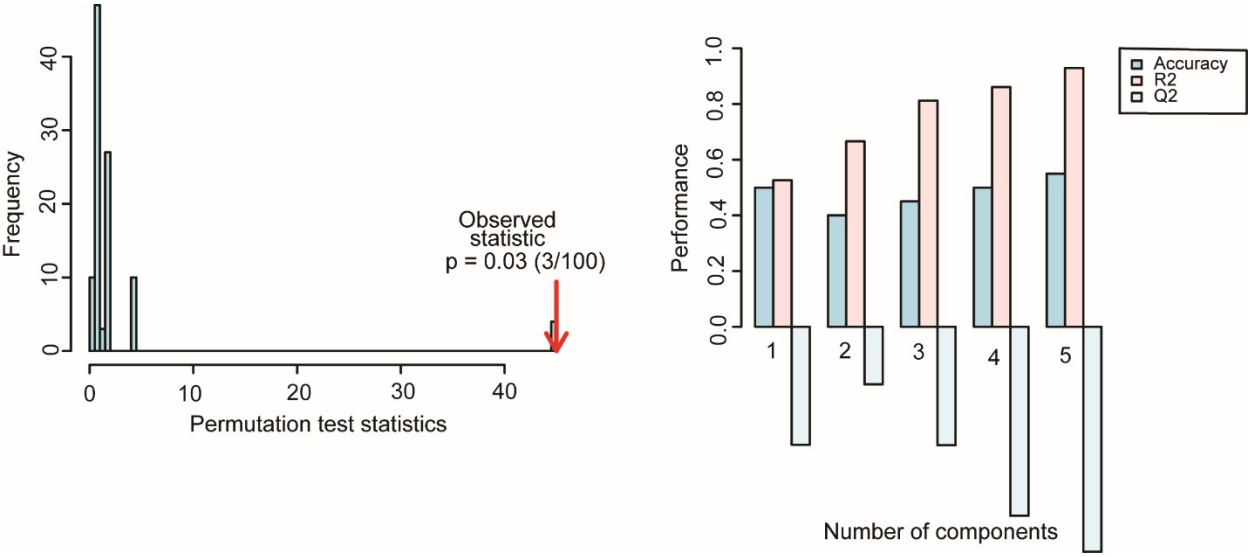
- **Figure S7. A.** Two-dimensional score plot of principal component analysis with red colour representing Patient and green Control from human Saliva with [PERMANOVA] F-value: 44.674; R-squared: 0.33672; p-value (based on 999 permutations): 0.001 **B.** Biplot of principal component analysis.



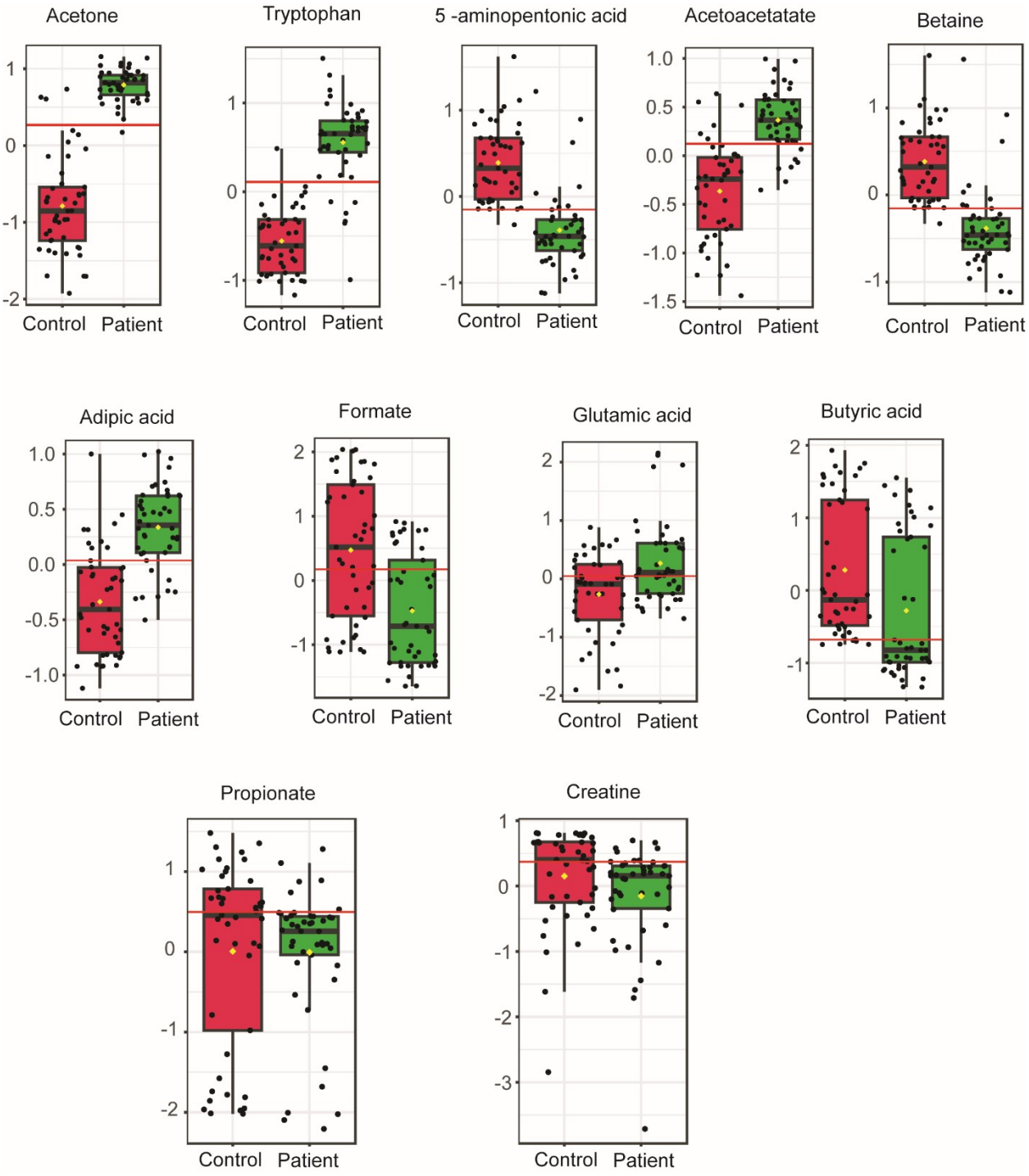
- **Figure S8. A.** Two-dimensional score plot of principal component analysis with red colour representing OSCC Stage 2 and green OSCC Stage 1 from human Saliva with [PERMANOVA] F-value: 0.23925; R-squared: 0.013117; p-value (based on 999 permutations): 0.799 **B.** Biplot of principal component analysis.



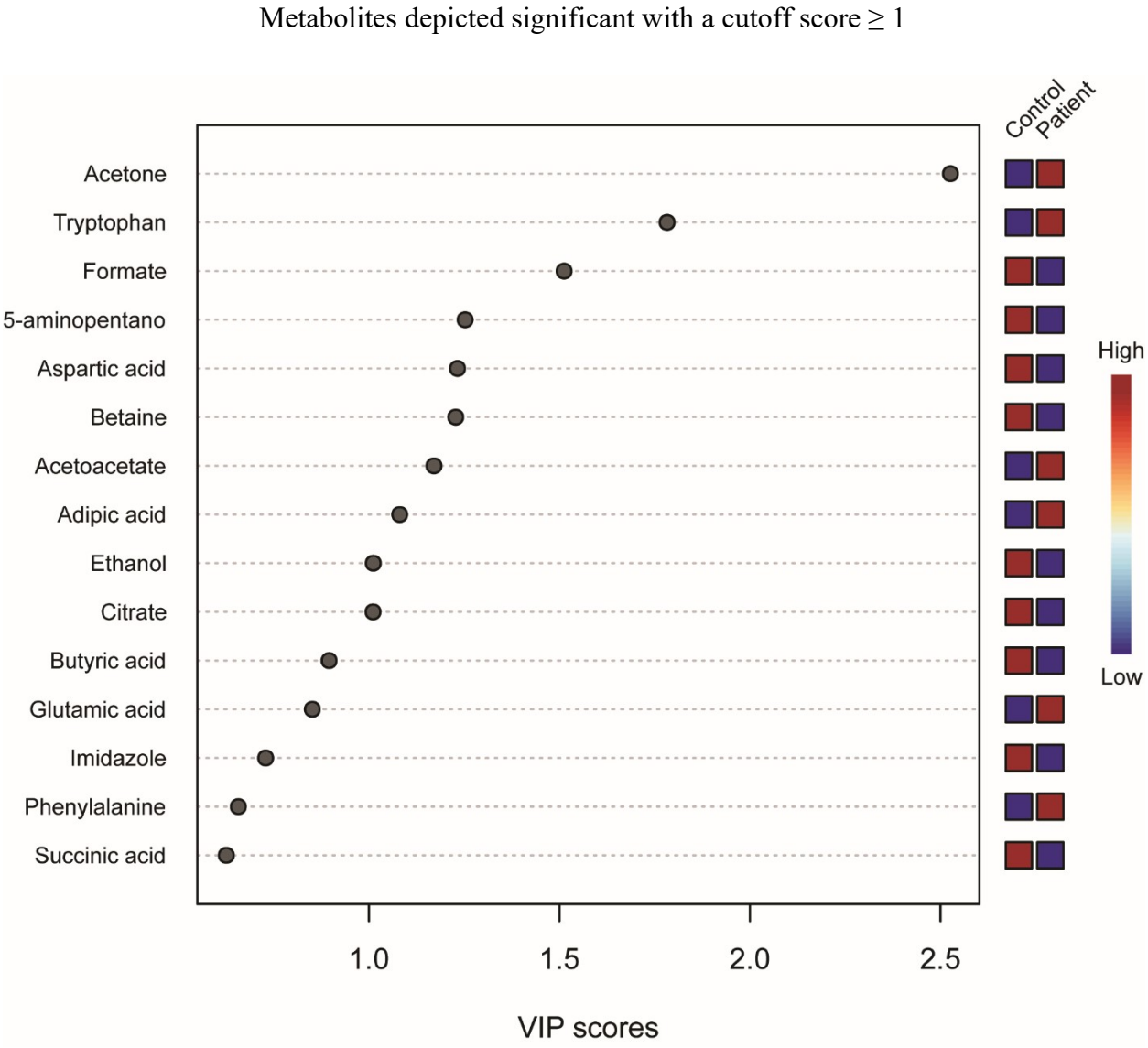
-Figure S9. A. Cross Validation of the multivariate O-PLS-DA model using 5-Fold related to NMR spectroscopy on saliva sample of OSCC Stage 1 & OSCC Stage 2. **B.** Permutation test of the multivariate PLS-DA model using by separation distance based on sum of squares between and sum of squares within (B/W) ratio for NMR on saliva sample OSCC Stage 1 & OSCC Stage 2.



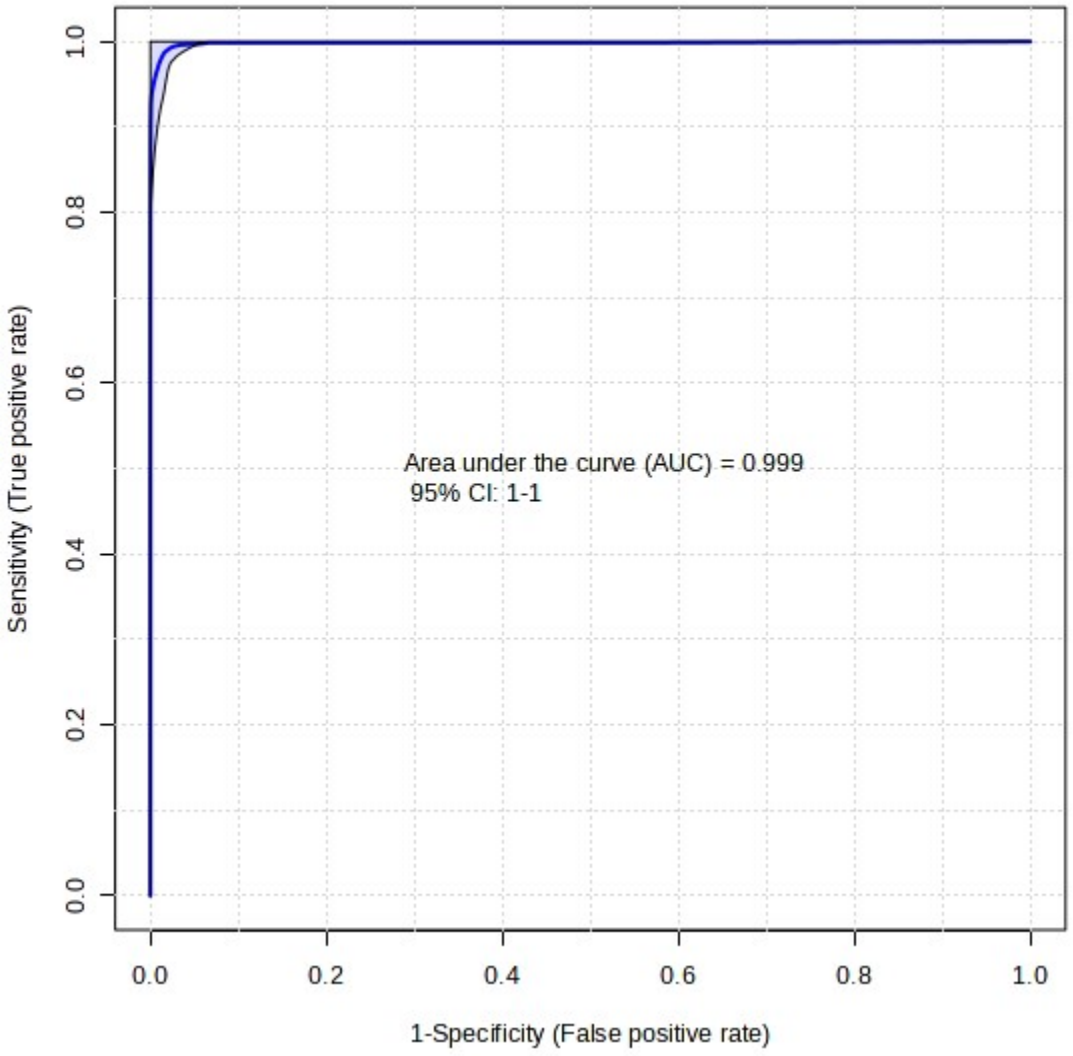
-Figure S10. Boxplots showing levels of differential metabolites in oral cancer and control individuals.



-Figure S11. Variables importance in projection (VIP).



-Figure S12. ROC curve analysis for the saliva specific metabolites level.



-Figure S13. A. Cross Validation of the multivariate O-PLS-DA model using 10-Fold related to NMR spectroscopy on saliva sample of Oral Cancer patients. **B.** Permutation test of the multivariate PLS-DA model using by separation distance based on sum of squares between and sum of squares within (B/W) ratio for NMR on saliva sample of Oral Cancer patients.