Electronic Supplementary Material (ESI) for Analyst.

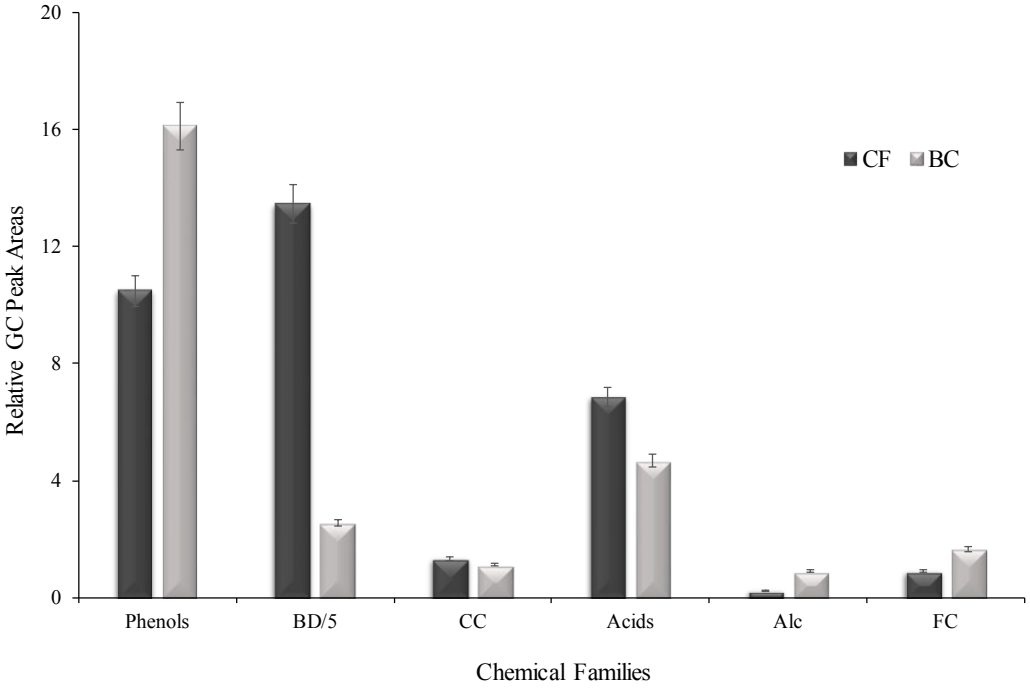

Fig. S1 – Major chemical families identified in BC and CF tissue samples. Legend: BD-benzene derivates; CC-carbonyl compounds; Alc-alcohols; FC-furanic compounds.

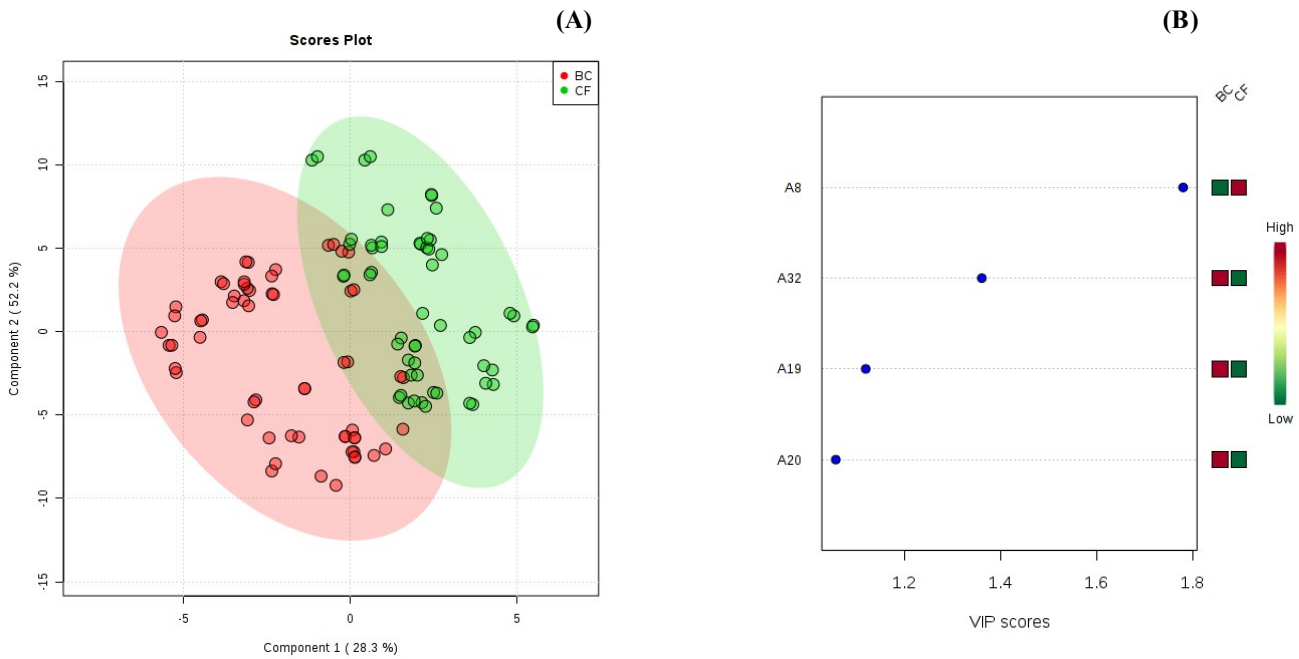

Fig. S2 – (A) Loading score plots of partial least square discriminant analysis (PLS-DA) and (B) variable importance in projection (VIP) scores of tissue samples from BC and CF subjects. For identification please see Table 2.

Table S1 – Performance of training and cross validation model results obtained for the PLS-DA analysis.

		Training set			Cross validation		
Algorithm	Scaling	AUC area	Sensitivity	Accuracy	AUC area	Sensitivity	Accuracy
PLS-DA	Log10	0.923	0.867	0.920	0.901	0.867	0.917

Table S2 – Summary of logistic regression model for each metabolite.

Metabolites	Estimate	SD error	z value	Pr(>|z|)	Odds
Intercept	-0.077	0.273	-0.28		
limonene	0.966	0.180	5.36	<0.001	2.63
acetic acid	-0.752	0.259	-2.909	0.004	0.47
furfural	-0.076	0.226	-0.334	0.738	0.93
decanoic acid	-0.133	0.187	-0.712	0.477	0.88

Table S3 –Sample class labels prediction and probabilities scores for 20 samples.

Name	Probability Scores	Predicted Class
O1_1	0.982	BC
O1_2	0.981	BC
O29_1	0.887	BC
O29_2	0.885	BC
O10_1	0.829	CF
O10_2	0.840	CF
C15_1	0.842	CF
C15_2	0.831	CF
C6_1	0.822	CF
C6_2	0.817	CF
C8_1	0.787	CF
C8_2	0.782	CF
C2_1	0.729	CF
C2_2	0.714	CF
O5_1	0.606	CF
O5_2	0.637	CF
O18_1	0.624	CF
O18_2	0.518	CF
C25_1	0.566	CF
C25_2	0.595	CF