

**Supplementary Table 2. Random forest classifier analysis of saliva samples by groups**

Act. group	Pred. group (train set)				Correctly classified	Pred. group (test set)				Correctly classified	
	Control	LG	HG	GC		Control	LG	HG	GC		
Control	68	0	0	0	100.0%	32	0	0	0	100.0%	
LG	2	57	0	13	79.2%	1	20	3	4	71.4%	
HG	0	2	59	8	85.5%	3	1	20	7	64.5%	
GC	2	1	1	67	94.4%	1	3	0	25	86.2%	
Overall correct class. Rate					89.6%						80.8%

The random forest classifier was performed using *Cutibacterium acnes*, *Ralstonia insidiosa*, *Streptococcus oralis*, and *Pseudomonas antarctica*, identified as group-discriminative by MaAsLin2 analysis of saliva samples. To address sample imbalance, SMOTE was applied, generating a balanced dataset of 100 samples.

Pred., predicted group; Act., actual group; LG, low-grade dysplasia; HG, high-grade dysplasia; GC, gastric cancer.