Supplementary Table 1. Random forest classifier analysis of gastric juice 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	52	0	20	0	72.2%	18	0	10	0	64.3%
LG	13	32	15	9	46.4%	2	19	9	1	61.3%
HG	20	8	39	2	56.5%	0	3	24	4	77.4%
GC	11	8	18	33	47.1%	2	3	9	16	53.3%
Overall correct class. Rate					55.7%					64.2%

The random forest classifier was performed using *Cutibacterium acnes* and *Streptococcus oralis*, identified as group-discriminative by MaAsLin2 analysis of gastric juice 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.