Supplementary Table 3. Random forest classifier analysis of serum 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	72	0	0	0	100.0%	28	0	0	0	100.0%
LG	6	45	9	9	65.2%	3	23	1	4	74.2%
HG	7	7	50	5	72.5%	0	0	28	3	90.3%
GC	9	4	4	53	75.7%	4	0	0	26	86.7%
Overall correct class. Rate					78.6%					87.5%

The random forest classifier was performed using *Pseudomonas yamanorum*, *Cutibacterium acnes*, *Ralstonia insidiosa*, and *Pseudomonas antarctica*, identified as group-discriminative by MaAsLin2 analysis of serum 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.

https://doi.org/10.3904/kjim.2024.339