

**Supplementary Table 4. Random forest classifier analysis of urine 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	43	11	9	62.3%	2	20	2	7	64.5%	
HG	3	5	56	4	82.4%	4	1	25	2	78.1%	
GC	9	5	5	52	73.2%	3	1	3	22	75.9%	
Overall correct class. Rate					79.6%						79.2%

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