## Electronic Supplementary Material (ESI) for Analytical Methods. This journal is © The Royal Society of Chemistry 2014 **Electronic Supplementary Information**

## Classification of cervical cytology for HPV infection using biospectroscopy and variable selection techniques

Kássio M.G. Lima<sup>1,2</sup>, Ketan Gajjar<sup>1,3</sup>, George Valasoulis<sup>4</sup>, Maria Nasioutziki<sup>5</sup>, Maria Kyrgiou<sup>6</sup>, Petros Karakitsos<sup>7</sup>, Evangelos Paraskevaidis<sup>4</sup>, Pierre L. Martin Hirsch<sup>1,3</sup>, Francis L. Martin<sup>1\*</sup> <sup>1</sup>Centre for Biophotonics, LEC, Lancaster University, Lancaster LA14YQ, UK <sup>2</sup>Institute of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Natal 59072-970, RN-Brazil

<sup>3</sup>Department of Obstetrics and Gynaecology, Central Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK

<sup>4</sup>Department of Obstetrics and Gynaecology, University Hospital of Ioannina, Ioannina, 45500, Greece

<sup>5</sup>Molecular & Diagnostic Cytopathology Laboratory, Second Department of Obstetrics and *Gynecology, Aristotle University of Thessaloniki, Hippokration Hospital, Thessaloniki, Greece* <sup>6</sup>Division of Surgery and Cancer, Faculty of Medicine, Imperial College, UK <sup>7</sup>Department of Cytopathology, "Attikon" Hospital, University of Athens, Greece

## Number of pages = 6Number of Figures = 5



**Figure S1 -** Mean spectrum (green) per category with respective standard deviation bands (blue) obtained for lrHPV (**A**); and, hrHPV (**B**) cervical cytology specimens. To demonstrate that  $H_2O$  was significantly excluded from these samples, (**C**) shows regions (Amide I and 3400 cm<sup>-1</sup>, likely to be influenced if there was aqueous contamination. Amide I lefthand shoulder shows no splitting and there is no rounding to the peak at 3400 cm<sup>-1</sup>, suggesting that the samples are  $H_2O$ -free.



**Figure S2** - Mean spectrum (green) per category with respective standard deviation bands (blue) obtained for  $\leq 29$  y (A); and,  $\geq 30$  y (B) cervical cytology specimens. To demonstrate that H<sub>2</sub>O was significantly excluded from these samples, (C) shows regions (Amide I and 3400 cm<sup>-1</sup>, likely to be influenced if there was aqueous contamination. Amide I lefthand shoulder shows no splitting and there is no rounding to the peak at 3400 cm<sup>-1</sup>, suggesting that the samples are H<sub>2</sub>O-free.



**Figure S3 -** Mean spectrum (green) per category with respective standard deviation bands (blue) obtained for  $\leq 29$  y (A); and,  $\geq 30$  y (B) normal cervical cytology (NCC) specimens. To demonstrate that H<sub>2</sub>O was significantly excluded from these samples, (C) shows regions (Amide I and 3400 cm<sup>-1</sup>, likely to be influenced if there was aqueous contamination. Amide I lefthand shoulder shows no splitting and there is no rounding to the peak at 3400 cm<sup>-1</sup>, suggesting that the samples are H<sub>2</sub>O-free.



**Figure S4** - Mean spectrum (green) per category with respective standard deviation bands (blue) obtained for  $\leq 29$  y (A); and,  $\geq 30$  y (B) low-grade cervical cytology (LG) specimens. To demonstrate that H<sub>2</sub>O was significantly excluded from these samples, (C) shows regions (Amide I and 3400 cm<sup>-1</sup>, likely to be influenced if there was aqueous contamination. Amide I lefthand shoulder shows no splitting and there is no rounding to the peak at 3400 cm<sup>-1</sup>, suggesting that the samples are H<sub>2</sub>O-free.



**Figure S5** - Mean spectrum (green) per category with respective standard deviation bands (blue) obtained for HPV16/18 (A); HPV31/35 (B); and, HPV others (C) based on HPV types in cervical cytology specimens. To demonstrate that  $H_2O$  was significantly excluded from these samples, (D) shows regions (Amide I and 3400 cm<sup>-1</sup>, likely to be influenced if there was aqueous contamination. Amide I lefthand shoulder shows no splitting and there is no rounding to the peak at 3400 cm<sup>-1</sup>, suggesting that the samples are  $H_2O$ -free.