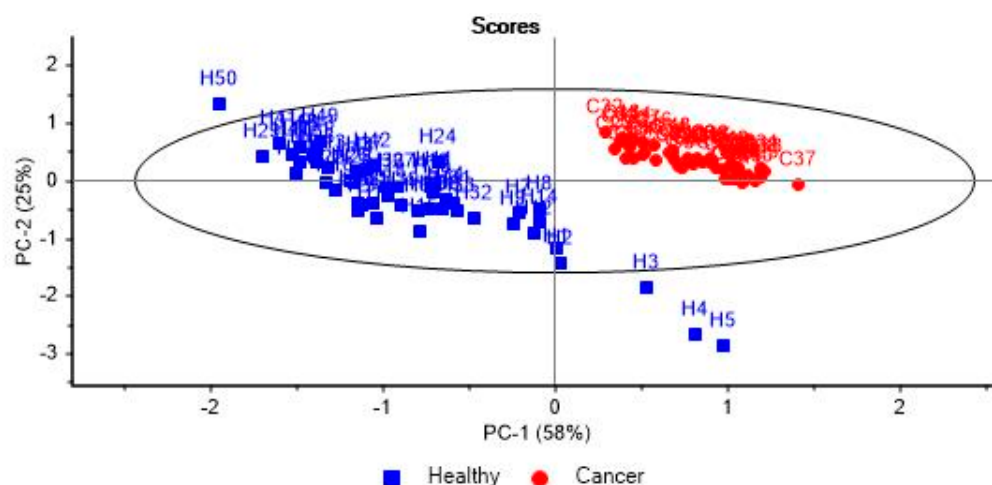
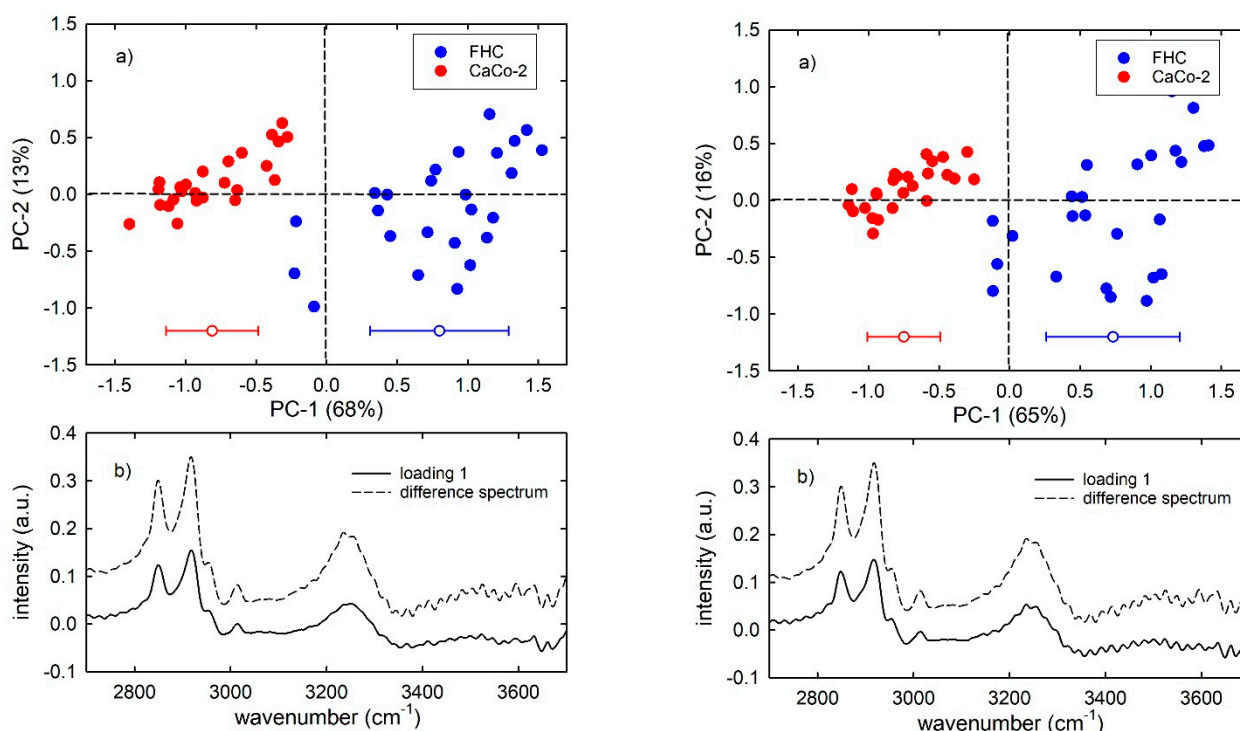


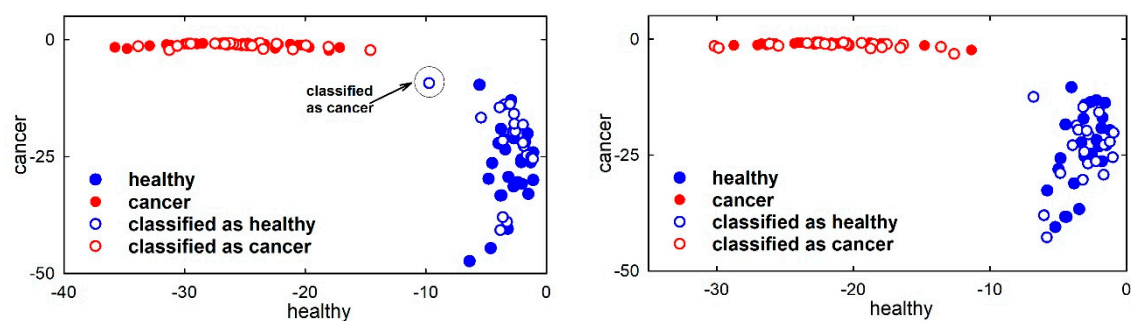
## Classification of healthy and cancer colon cells grown on glass coverslip by means of Fourier transform infrared spectroscopy and multivariate methods.



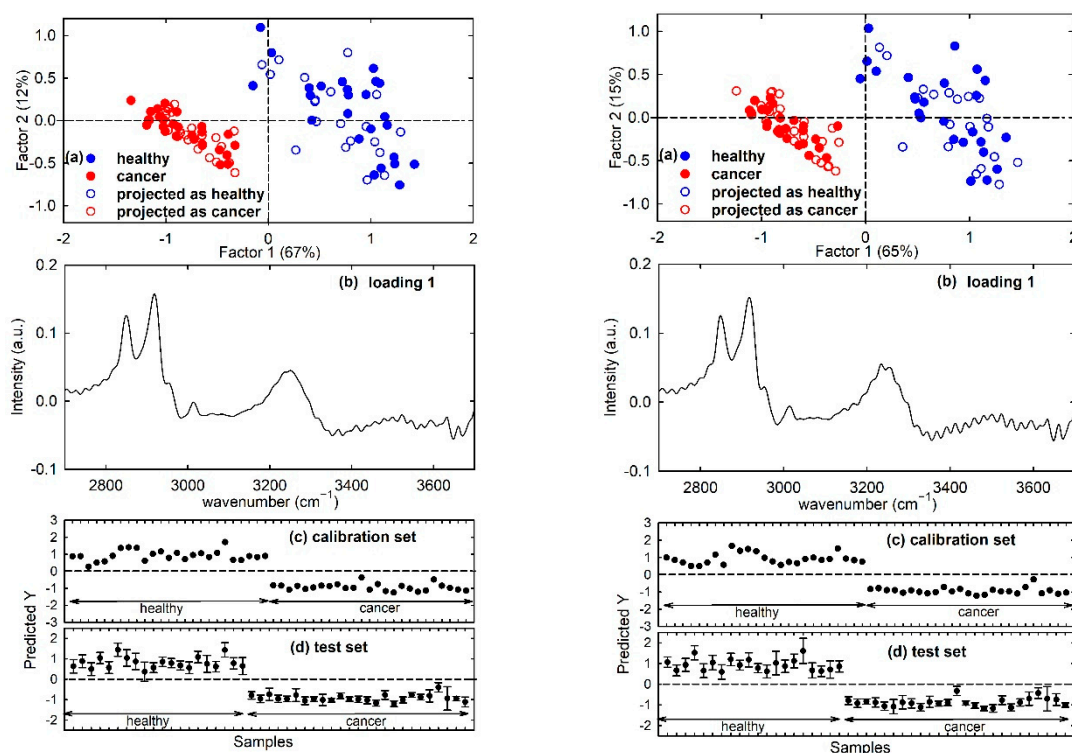
**Figure S1.** PC2 vs. PC1 score plot for the full dataset of 100 FTIR spectra: 50 spectra measured from healthy FHC cells (blue squares) and 50 spectra from cancer CaCo-2 cells (red dots). The continuous line corresponds to the Hotelling's  $T^2$  ellipse. The spectra outside of the ellipse are considered as outlier spectra and they were removed from the subsequent analysed dataset.



**Figure S2.** PC2 vs. PC1 score plot (a) for the healthy FHC cells (blue dots) and the cancerous CaCo-2 cells (red dots) FTIR spectra from two calibration sets of FTIR spectra randomly chosen from the original dataset. Both calibration sets (one on the left side and the other on the right side) include spectra which are different from those of the calibration set of the main article. The mean and standard deviation of the PC1 score values are shown on the bottom side of (a): results of the t-test point out that the two distributions of score values are significantly different ( $p < 0.001$ ). Difference of mean spectrum values and loading 1 spectrum are reported in (b).



**Figure S3.** Discrimination plot obtained from the results of Linear Discriminant Analysis model developed using 3 principal components from two sets of calibration and test samples of FTIR spectra randomly chosen from the original dataset. Both calibration sets (one on the left side and the other on the right side) include spectra which are different from those of the calibration set of the main article. The scatter plot shows the discrimination of healthy FHC cell spectra (filled blue circles) and cancerous Caco-2 cell spectra (filled red circles). The projections of the test samples on the PCA-LDA model are shown as hollow circles. One misclassified sample of the test set is enclosed inside a circle and labelled in the test set on the right side.



**Figure S4.** Scores plot (Factor 2 vs. Factor 1) of the developed PLS-DA model, showing the calibration (filled circles) and projected test set (hollow circles) samples for the healthy FHC cell (blue circles) and cancer CaCo-2 cell (red circles) spectra in (a) from two sets of calibration and test samples of FTIR spectra randomly chosen from the original dataset. The root mean square error for cross-validation (RMSECV) with three factors is 0.27 for both calibration sets. Loading 1 plot for the PLS model in (b). Predicted Y values with respect to the reference values (1: healthy FHC cells; -1: cancer CaCo-2 cells) after PLS regression for the spectra of the calibration set and predicted Y values, with corresponding deviations, from the PLS-DA model for the spectra of the test set are reported in (c) and (d), respectively, with black dots.