Supplementary

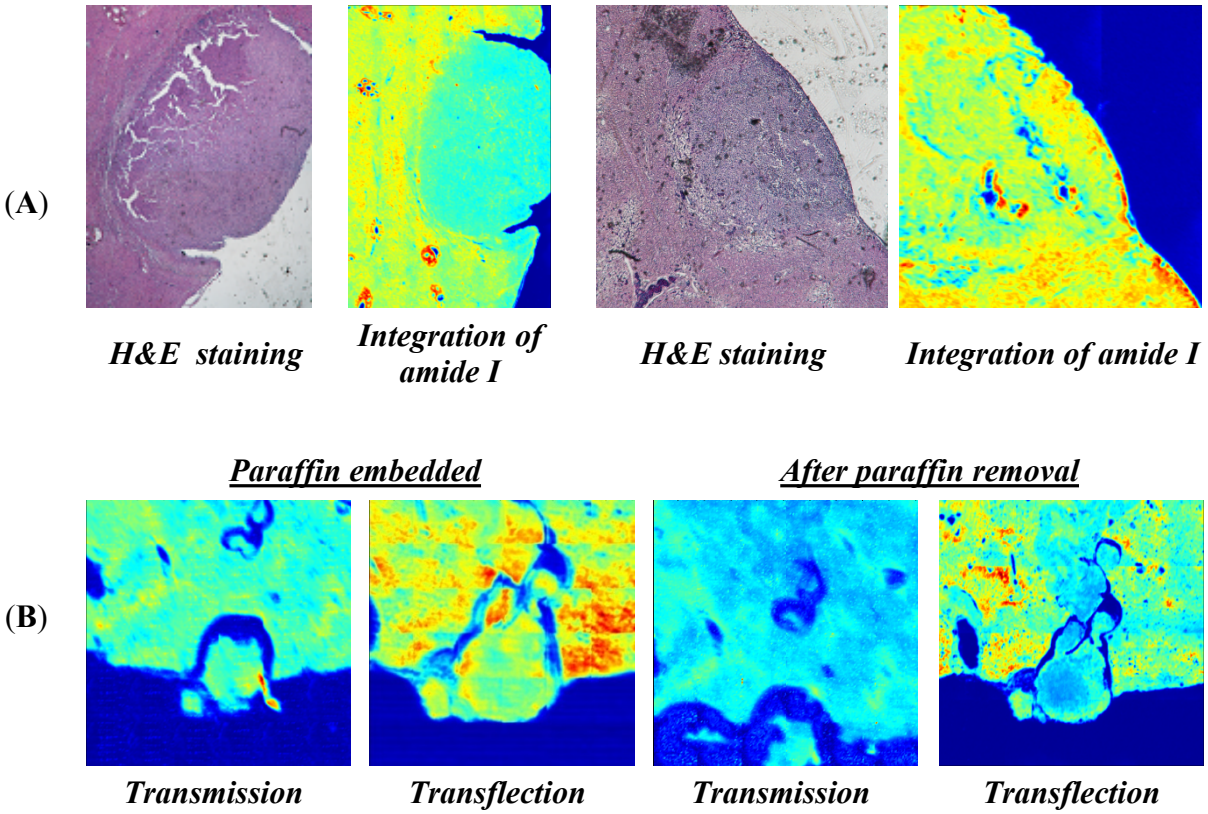
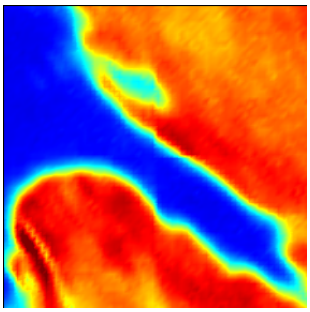


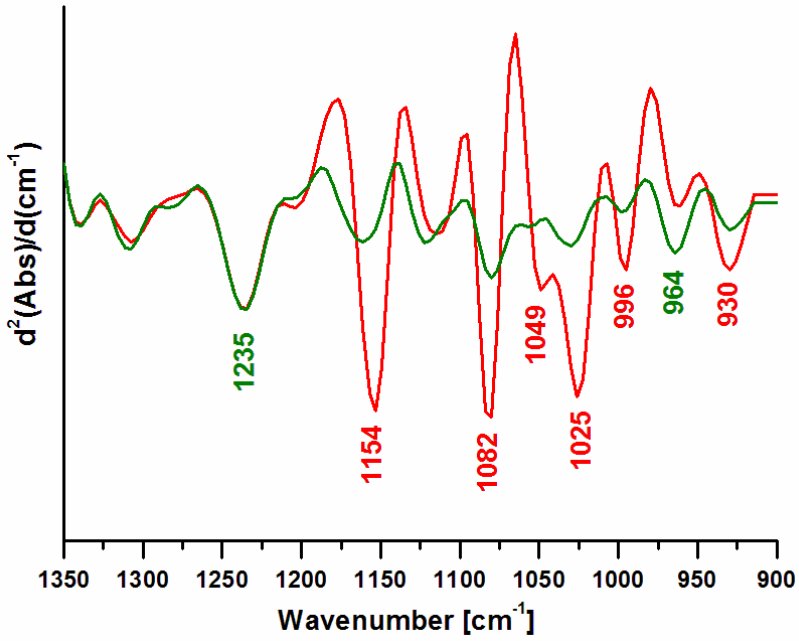
Fig. S1.(**A**) Examples of correlation between measured areas and the results of H&E staining on neighboring sections and (**B**) an example of correlation between areas measured before and after paraffin removal in transmission and transflection.

Transmission



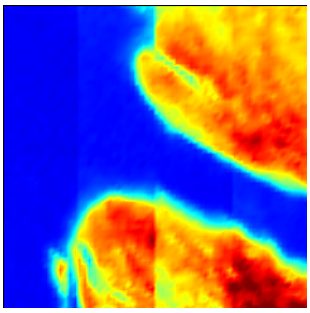
Integration of amide I (1620 – 1680 cm^{-1})

KMC result: distribution of classes



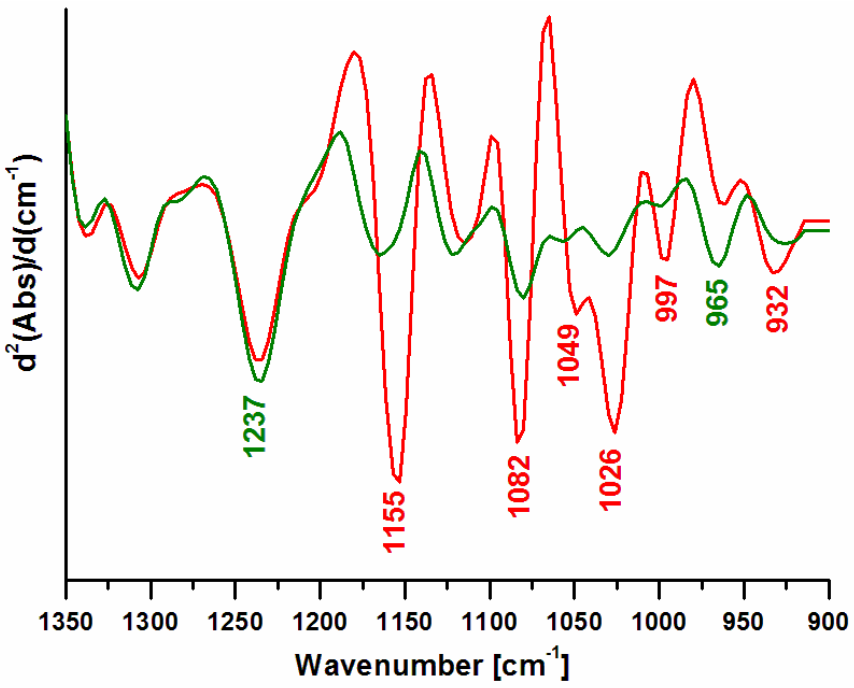
KMC result: average spectra of the normal and cancerous classes.

Transflection



Integration of amide I (1620 – 1680 cm^{-1})

KMC result: distribution of classes



KMC result: average spectra of the normal and cancerous classes.

Fig. S2. Examples of correlation between measured areas from the samples prepared on CaF_2 (transmission) and Kelvley (transflection) from the same individual: integration of amide I bands along with cluster analysis results (2^{nd} derivatives) in the range 1350 – 900 cm^{-1}.

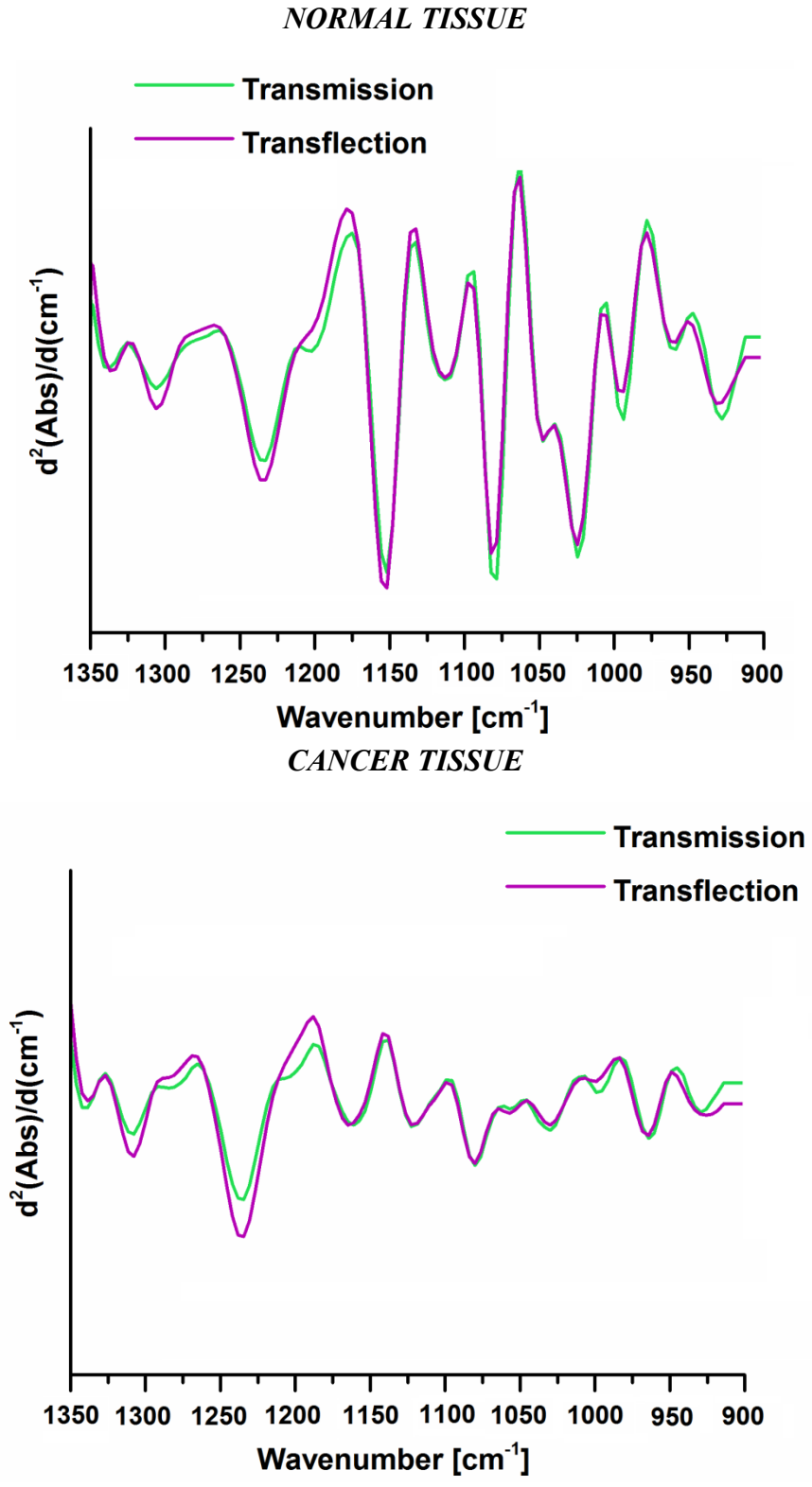
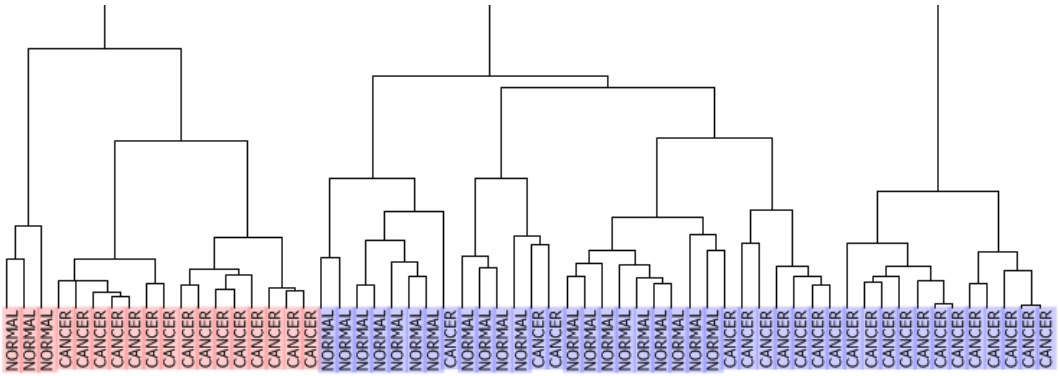


Fig.3S. 2nd derivatives of spectra corresponding to normal and cancer classes (Fig. 2S) measured in transmission and transflection modes. Despite the differences in the measurement modes, there are only marginal differences normal and tumor affected areas.

Transmission set of normal and cancerous spectra from paraffin embedded sections

(1)


NORMAL NORMAL NORMAL CANCER CANCER CANCER CANCER CANCER CANCER CANCER CANCER CANCER CANCER CANCER CANCER CANCER NORMAL NORMAL NORMAL NORMAL NORMAL CANCER NORMAL NORMAL NORMAL CANCER CANCER NORMAL NORMAL NORMAL NORMAL NORMAL NORMAL NORMAL CANCER

(2)

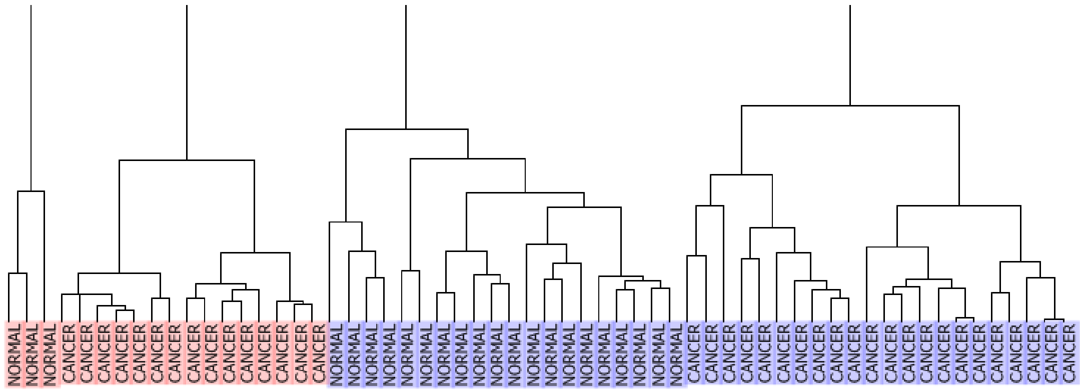
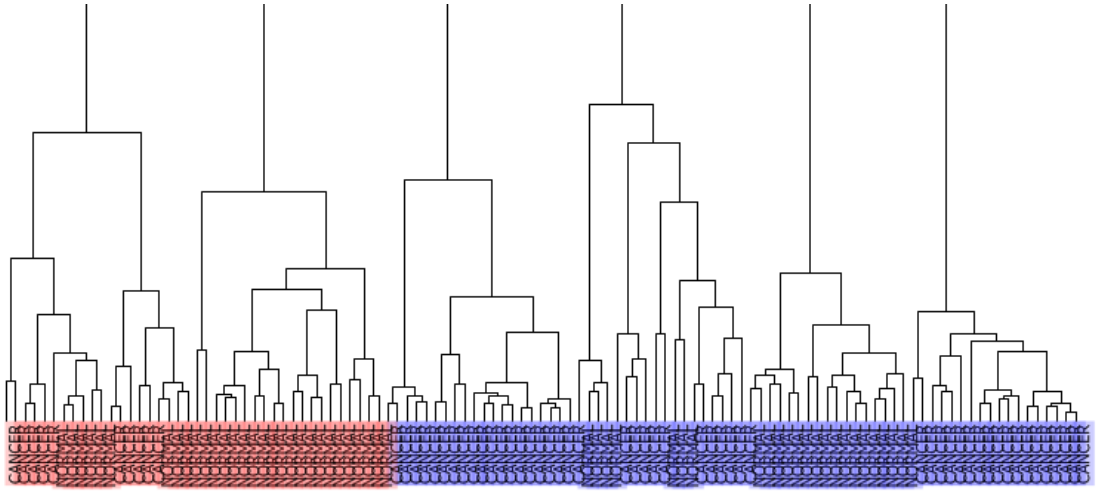


Fig.4S. Results of unsupervised hierarchical cluster analysis (UHCA), performed on the set of normal and cancerous spectra measured in transmission mode for paraffin embedded sections, in the ranges: (**1**) 1680 – 1530 cm^{-1} and (**2**) 1350 – 900 cm^{-1}& 1680 – 1530 cm^{-1}. The separation is less accurate than with the use of only the range 1350 – 900 cm^{-1} (Fig. 2).

Transflection set of normal and cancerous spectra from paraffin embedded sections.

(1)



(2)

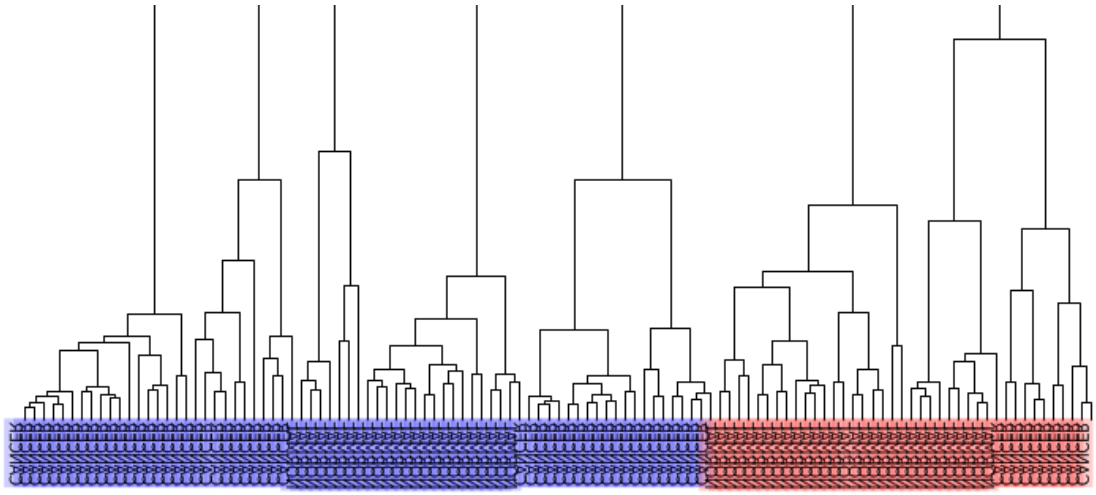
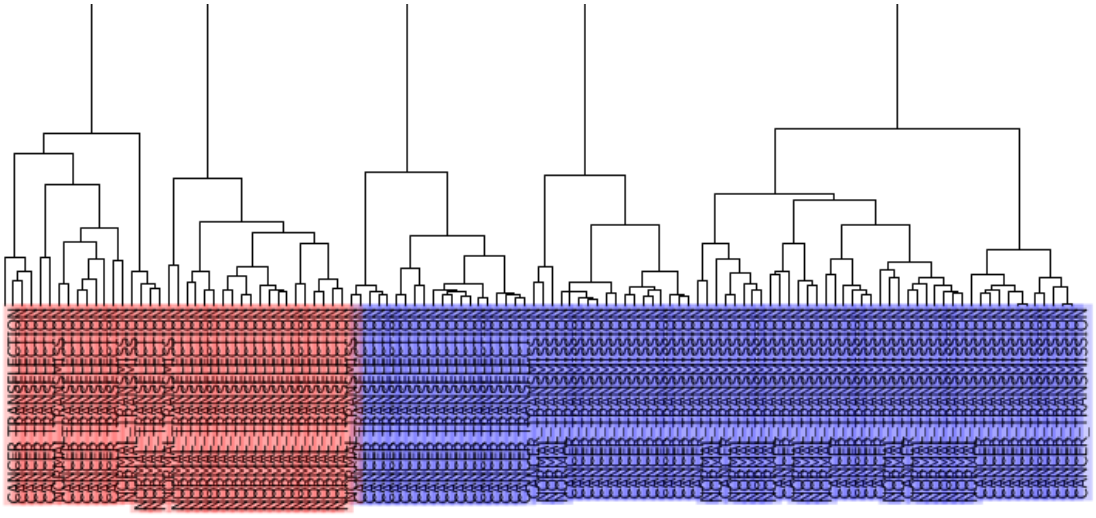


Fig.5S. Results of unsupervised hierarchical cluster analysis (UHCA), performed on the set of normal and cancerous spectra measured in transflection mode for paraffin embedded sections, in the ranges: (**1**) 1680 – 1530 cm^{-1} and (**2**) 1350 – 900 cm^{-1}& 1680 – 1530 cm^{-1}. The separation is less accurate than with the use of only the range 1350 – 900 cm^{-1} (Fig. 2).

Transmission & transflection set of normal and cancerous spectra from paraffin embedded sections.

(1)



(2)

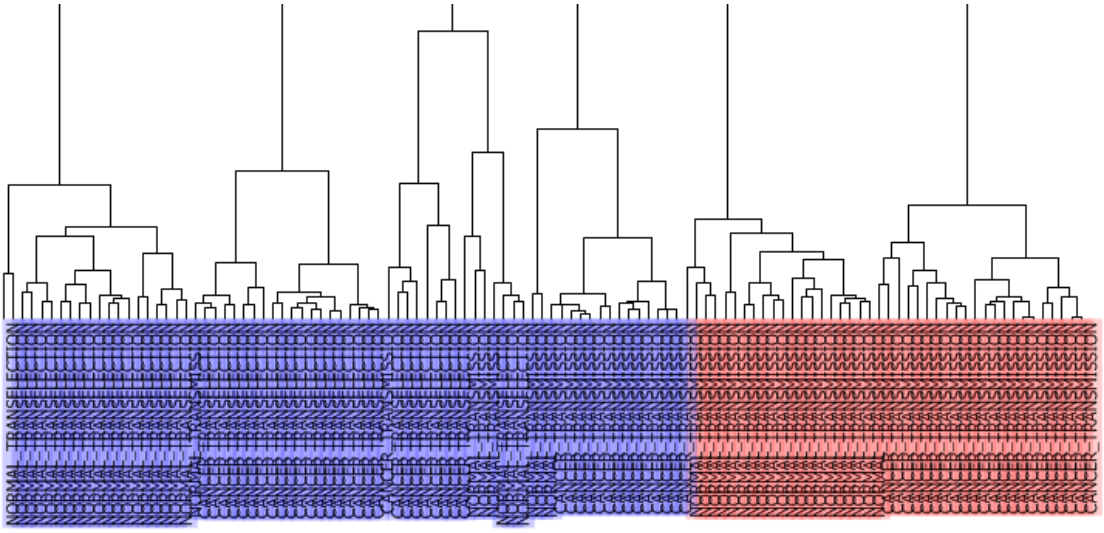
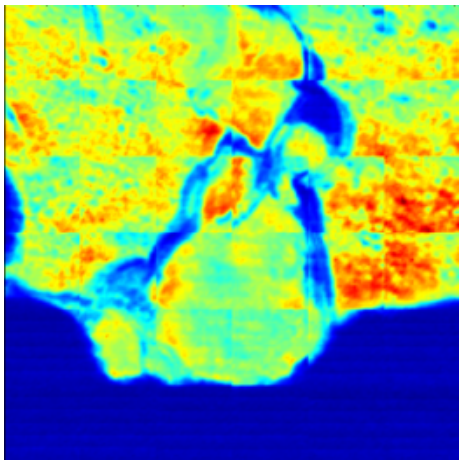
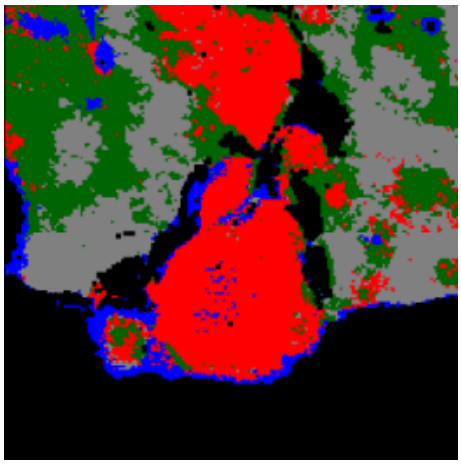


Fig.6S. Results of unsupervised hierarchical cluster analysis (UHCA), performed on the set of normal and cancerous spectra measured in transmission and transflection mode for paraffin embedded sections, in the ranges: (**1**) 1680 – 1530 cm^{-1} and (**2**) 1350 – 900 cm^{-1}& 1680 – 1530 cm^{-1}. The separation is less accurate than with the use of only the range 1350 – 900 cm^{-1} (Fig. 2).

Paraffin embedded section

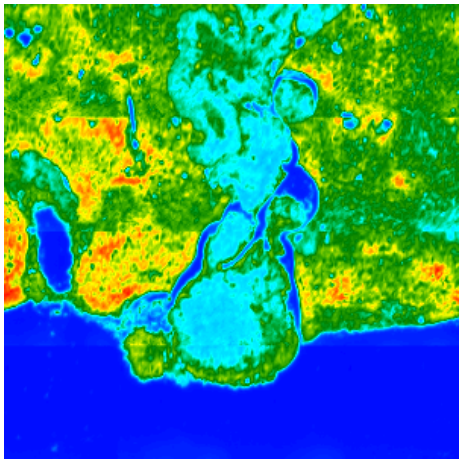


Integration of amide I (1620 – 1680 cm^{-1})

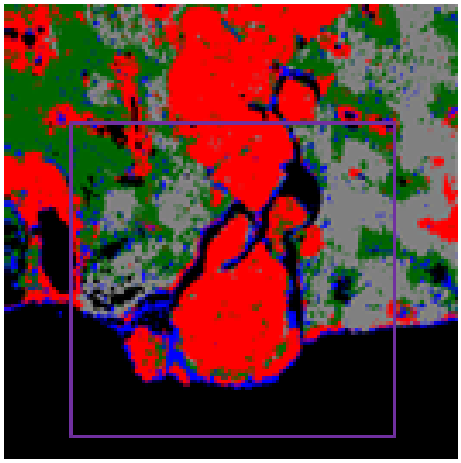


KMC result: distribution of classes

Deparaffined section



Integration of amide I (1620 – 1680 cm^{-1})



KMC result: distribution of classes

Fig.7S. Direct comparison of the KMC results (distribution of classes) for areas of the same sample measured before and after removing paraffin. Since the area measured after paraffin removal is bigger, the corresponding part has been marked with purple square. Distribution of the classes is highly similar.

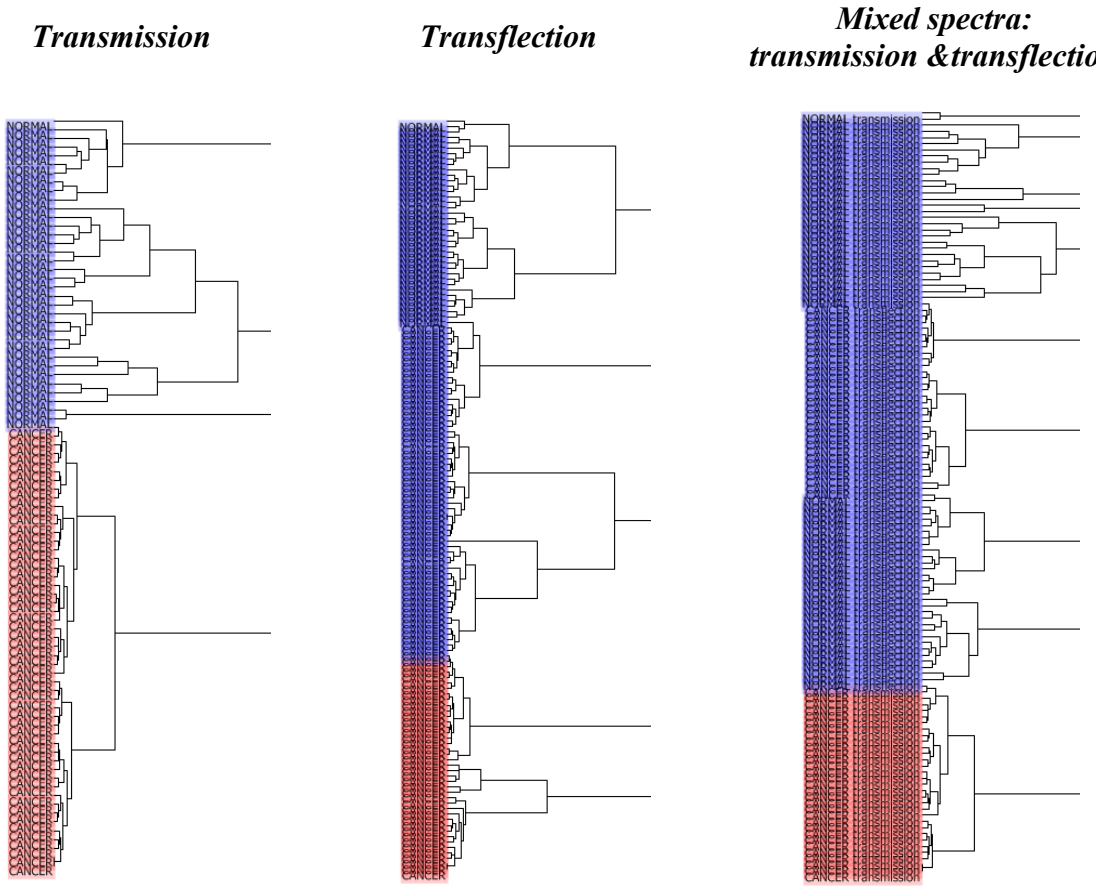
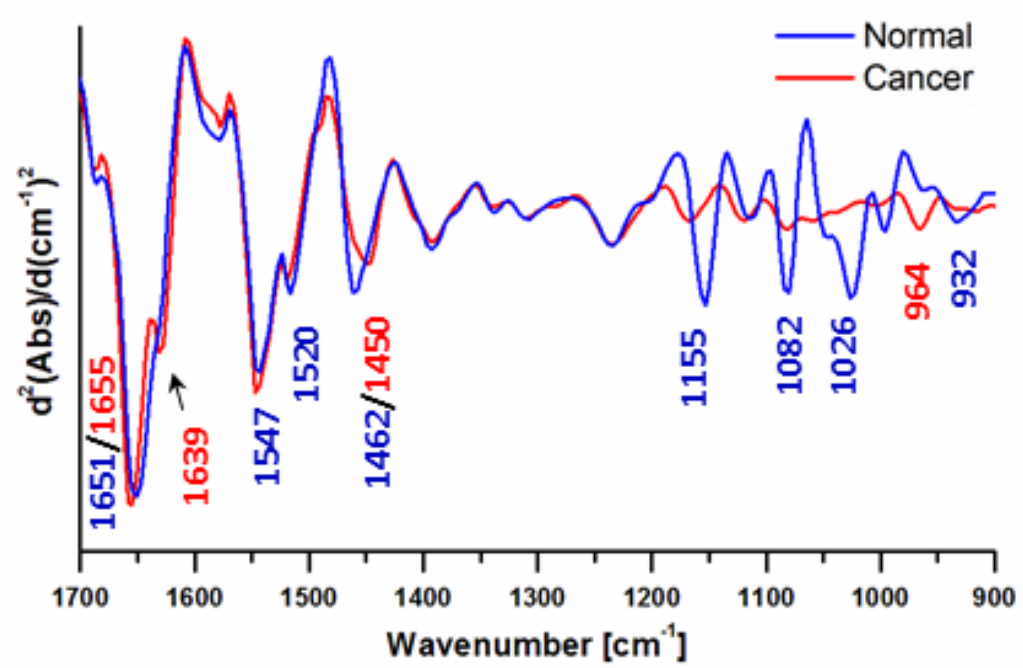


Fig.8S. Unsupervised hierarchical cluster analysis (UHCA) in the range 1700 – 900 cm^{-1} showing the distinction between normal and cancerous spectra in three data sets: transmission (including normal and cancerous spectra measured in transmission), transflection (including normal and cancerous spectra measured in transflection) and mixed spectra (including normal and cancerous spectra measured in transmission and transflection) after paraffin removal. The separation is less accurate than with the use of only the range 1350 – 900 cm^{-1} (Fig. 6) for transflection and mixed spectra data sets. It remains identical for transmission.

TRANSMISSION



TRANSFLECTION

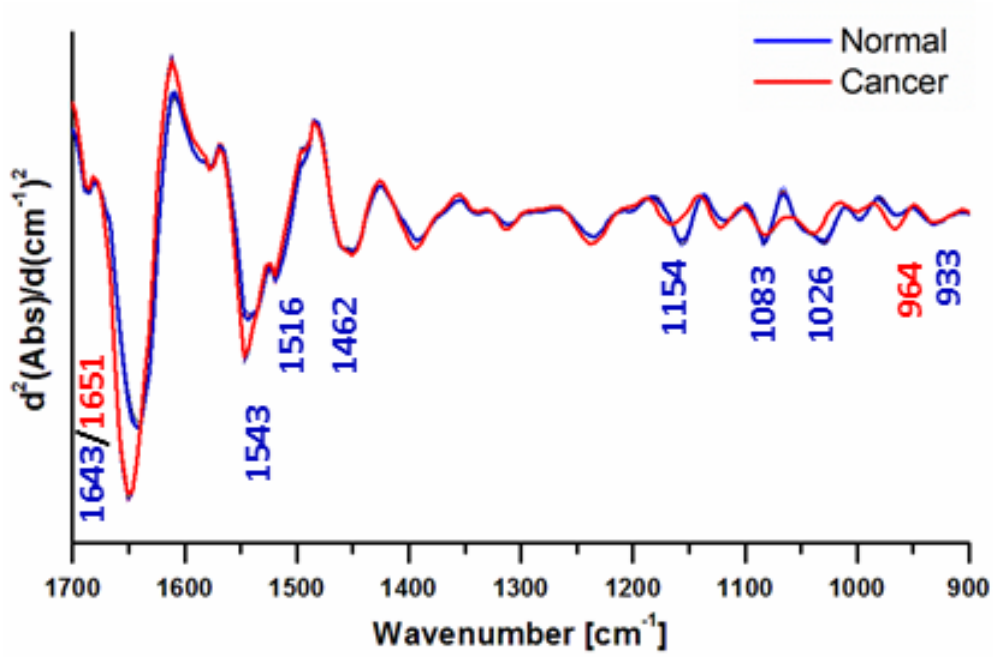
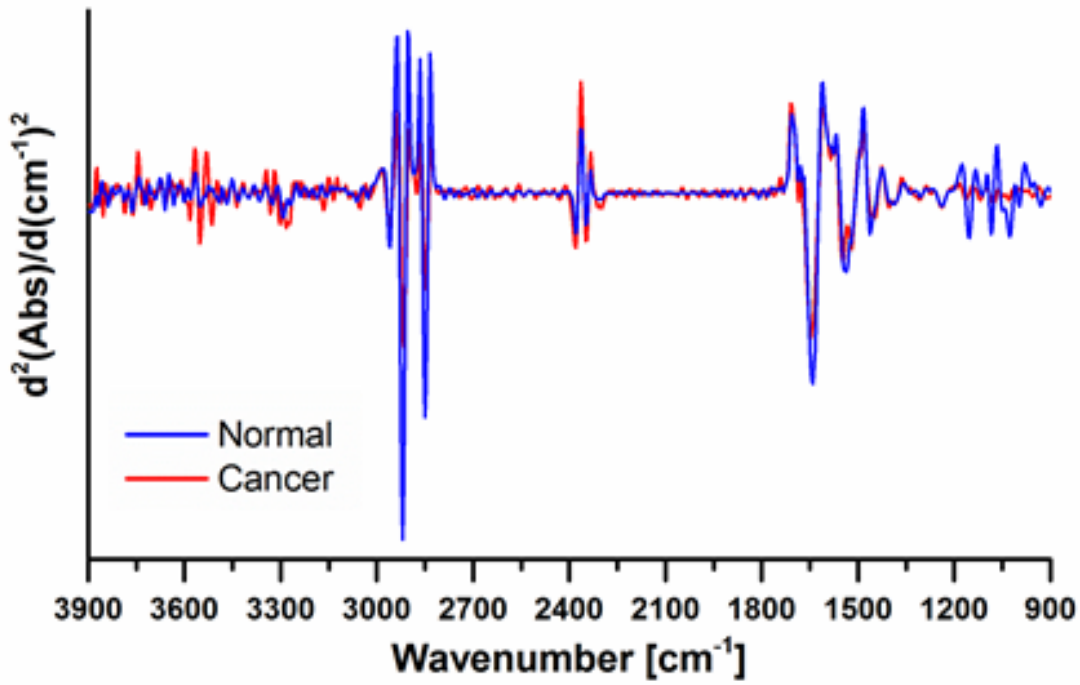


Fig.9S. 2nd derivatives of spectra corresponding to normal and cancer areas measured in transmission and transflection modes in the range 1700 – 900 cm^{-1}. Clearly visible difference of band intensities from the range 1400 – 900 cm^{-1} relative the amide I band. As the differential spectral changes are located in this range, the higher relative intensity of these bands (relative to the amide I) may explain their greater influence on the result of the analysis in the range of 1700 – 900 cm^{-1} for transmission (Fig. 7S).

3900 – 900 cm^{-1}



3700 – 3300 cm^{-1}

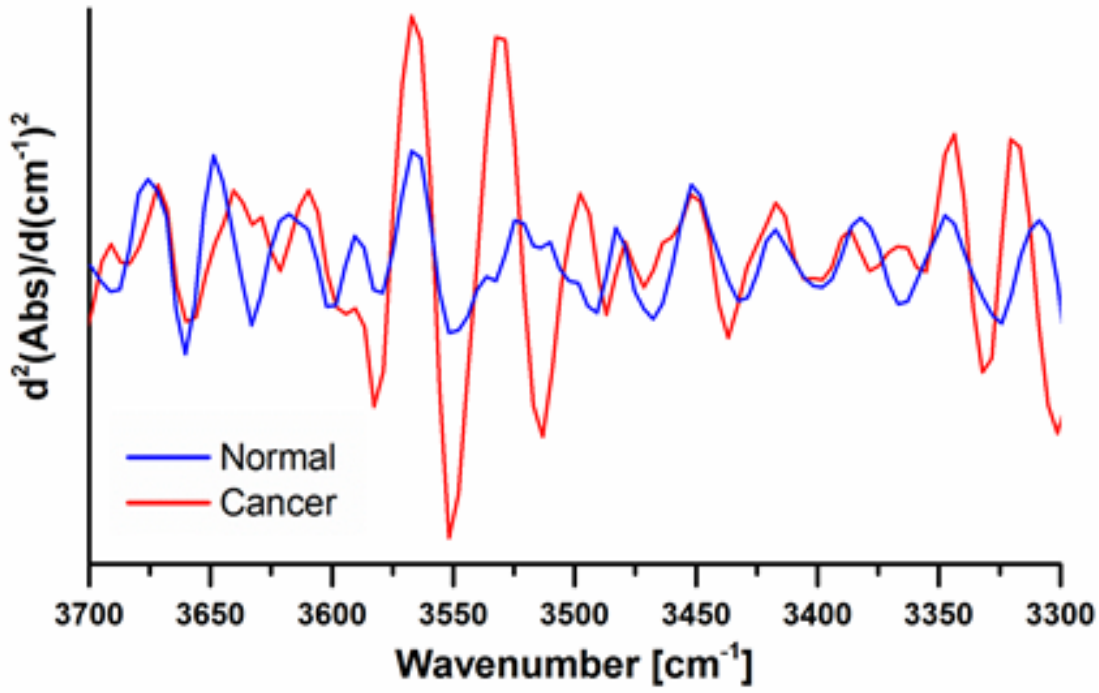


Fig.10S. 2nd derivatives of spectra corresponding to normal and cancer areas measured in transmission in the range 3900 – 900 cm^{-1} with highlighted area between 3700 – 3300 cm^{-1} (showing the O – H stretching mode).