

Analyst

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

High-resolution FTIR imaging of colon tissues for elucidation of individual cellular and histopathological features

Jayakrupakar Nallala¹, Gavin Rhys Lloyd², Neil Shepherd³, Nick Stone^{1*}

¹ *Biomedical Physics, School of Physics, University of Exeter, EX4 4QL, UK*

² *Biophotonics Research Unit, Gloucestershire Royal Hospital, Gloucester, GL1 3NN, UK*

³ *Department of Pathology, Gloucestershire Hospitals NHS Foundation Trust, Gloucester,*

GL1 3NN, UK

*e-mail: n.stone@exeter.ac.uk

Abstract:

Novel technologies that could complement current histopathology based cancer diagnostic methods are under examination. In this endeavour mid-infrared spectroscopic imaging is a promising candidate that can provide valuable bio-molecular information from unstained cells and tissues in a rapid and a non-destructive manner. With this imaging technique, the biochemical information obtained from smaller areas of the tissues can be of clinical significance and hence the measured pixel size. Until recently it was difficult to obtain spectral data from pixels below around 5 microns square. High NA objectives have been utilised to reduce the ideal diffraction limit, enabling for the first time elucidation of subcellular features. In this context, the ability of high-resolution imaging, obtained using novel high-magnification optics retro-fitted onto a bench top FTIR imaging system, to characterise histopathological features in colonic tissues has been tested. Formalin fixed paraffin embedded colon tissues from three different pathologies were imaged directly using the conventional and the high-magnification imaging set-ups. To circumvent chemical de-paraffinization protocols, an extended multiplicative signal correction (EMSC) based electronic de-paraffinization was carried out on all the infrared images. Multivariate analysis of the high-magnification infrared imaging data showed a detailed information of the histological features of the colon tissue in comparison to conventional imaging. Furthermore, high-magnification imaging has enabled a label-free characterization of the mucin rich goblet cell features in an unprecedented manner. The current study demonstrates the

applicability of high-magnification FTIR imaging to characterise complex tissues on a smaller scale that could be of clinical significance.

Introduction:

Infrared (IR) spectroscopy is a promising analytical tool that allows the bio-molecular characteristics of cells and tissues to be studied without the need for any external stains (1). The bio-molecular contents in cells and tissues change constantly some of which are indicative of functional abnormalities especially in disorders like cancer. IR spectroscopy can provide information about these variations that can be used to identify, understand and in some cases even predict the possibility of the occurrence of a disease. In this regard, IR spectroscopy has been investigated as a potential cancer diagnostic tool in various tissue types including some highly prevalent cancers such as lung (2), colon (3-7), oral cavity (8), liver (9), prostate (10-12), breast (13-15), lymphnode (16, 17), cervix (18), etc. Using IR spectroscopic imaging approaches, biological material can be imaged to obtain bio-molecular information, where each pixel of the image constitutes a spectrum with a specific bio-molecular signature (19).

Obtaining unlabelled, native bio-molecular information from cellular and sub-cellular regions of cells and tissues may change the understanding of disease and add precise information for diagnosis. As an example, obtaining information about the morphological and molecular features at the cellular level could provide early signs of cancer before it is manifested at tissue or organ level. Therefore when working with tissues the measured pixel size is an important factor to consider especially in diseases such as cancer where features on a smaller scale (e.g. cellular scale) are implicated in the development and progression of the disease.

IR spectroscopic imaging capabilities have evolved in terms of measurement time for a given sample size and the sample details that can be resolved. In recent years, the availability of focal plane array (FPA) detectors with the bench top imaging systems have provided the possibility of measuring large tissue areas in shorter time frames (19-23). The combination of these detectors and the collection optics enabled collection of spectra at a pixel size of as little as $5.5 \times 5.5 \mu\text{m}^2$, which can provide information from gross histological features such as epithelial glands, cellular clusters etc., in tissues such as colon; or glandular regions in tissues such as the breast. One of the limitations in obtaining smaller pixel sizes using such benchtop instruments is the low signal-to-noise which is effectively

dependent on the type of optics and the use of conventional light sources such as Global[®] (24). Alternatively, high intensity light sources like Synchrotron or quantum cascade lasers (QCL) can enable the measurement of smaller pixels on the sample by raising the signal to noise of the spectra acquired from each pixel (20, 25). Very recently, changes to the optical system of a bench top instrument with a conventional Global[®] source has enabled to obtain a five-fold magnification at the sample plane thereby improving the effective pixel size from 5.5x5.5 μm^2 to 1.1x1.1 μm^2 . Although this is below the diffraction limit (d) for the vast majority of the mid-IR spectrum, the effect of oversampling, combined with multi-wavelength image analysis on the molecular images obtained is explored here. Considering the Abbe criterion $d = \lambda / 2n\sin\theta$ {where $n\sin\theta$ = numerical aperture (NA) of the objective}, d is of the order of the wavelength (λ) for this system, which is between 2.5 and 11 microns. For a 0.62 NA objective, d ranges from 2 to 8.9 microns respectively.

Here with aim to determine the applicability of this novel system for characterization of biological tissues on a smaller scale, FTIR spectroscopic imaging was carried out in conjunction with multivariate statistical analysis. For this, formalin fixed paraffin embedded (FFPE) colon tissues from three pathology groups (non-tumoral, tubular adenoma and adenocarcinoma) were measured using the high-magnification set-up and compared with the conventional set-up. The capability of the high-magnification imaging set-up was explored with a focus on the identification of goblet cells; one of the important cell types in colon tissues, which are implicated in some of the important functionalities of the colon (26, 27). Together with this, characterization of the normal epithelium and the connective tissue regions from the benign tissues (non-tumoral and tubular adenoma) and tumor epithelium from the cancerous tissue was carried out. In parallel, the signal-to-noise levels of the instrument in conventional and high-magnification mode were calculated using a polymer film in order to establish the limitations and optimum parameters.

Materials and Methods:

The FTIR imaging system consisted of an Agilent 620 FTIR microscope coupled with an Agilent 670 FTIR spectrometer with a Global[®] light source, and a liquid-nitrogen cooled 128x128 FPA detector. IR light was transmitted through a 15x objective (N. A. = 0.62). The same set-up was used for conventional as well as high-resolution imaging; the latter was achieved by an already in-built high-

magnification optics. In the conventional mode, there is $\sim 2x$ de-magnification after the 15x objective, giving a total system magnification of 7.3x, and provides $5.5 \times 5.5 \mu\text{m}^2$ pixel size with a $700 \times 700 \mu\text{m}^2$ field of view (FOV). In the high-mag mode, there is $\sim 2.4x$ extra magnification on top of the 15x, to give a final overall magnification of 36x, with $1.1 \times 1.1 \mu\text{m}^2$ pixel size and a $140 \times 140 \mu\text{m}^2$ FOV. Therefore, in going from $\sim 36x$ to $\sim 7.3x$, a five-fold magnification is obtained enhancing a “pixel size limited” situation, into a “diffraction limited” situation. All measurements were carried out in the mid-IR spectral range of $1000\text{--}3800 \text{ cm}^{-1}$ at a spectral resolution of 4 cm^{-1} . Firstly, IR images of a polystyrene film (envelope window) standard were measured in transmission mode to calculate the signal-to-noise ratio (SNR) of the instrument using both the normal and the high-magnification set-ups. The polystyrene film was measured using 4, 8, 16 and 32 scans respectively using 64 scans for the air background.

Secondly, five FFPE colon tissues from 3 different pathology groups (3 non-tumoral, 1 tubular adenoma and 1 moderately differentiated adenocarcinoma) were selected for this study. Ethical approval was obtained from the local research ethics committee for collection of additional biopsies during routine clinical investigations by colonoscopy. From these tissues blocks, $7 \mu\text{m}$ thick sections were cut using a microtome and transferred onto calcium fluoride (CaF_2) substrates. An adjacent $3 \mu\text{m}$ section was also obtained from each of these tissues, transferred on to glass slide and stained with haematoxylin and eosin (HE) to be used as a morphological reference. IR images were acquired directly from the unstained, FFPE tissue sections in transmission mode without any prior de-paraffinization, in both normal- and high-magnification set-ups. Initially, images were obtained using the conventional resolution. From these images, selected regions of interest were then imaged using the high-magnification set-up. A single FPA tile using the conventional magnification scans an area of $704 \times 704 \mu\text{m}^2$ at a pixel resolution of $5.5 \times 5.5 \mu\text{m}^2$ while the high-magnification scans an area of $121 \times 121 \mu\text{m}^2$ at a pixel resolution of $1.1 \times 1.1 \mu\text{m}^2$. Prior to the tissue measurements in each set-up, clear regions of the CaF_2 substrates were measured as a background respectively at 256 scans which were then ratioed out. In total 10 images (3 non-tumoral, 1 tubular adenoma and 1 tumoral from each set-up) from three distinct pathology groups were acquired.

The FTIR tissue images were pre-processed and analysed in Matlab 2013 (Mathworks USA) using in-house scripts in the fingerprint region of $1000\text{--}1800 \text{ cm}^{-1}$. A modified extended multiplicative signal

correction (EMSC) was used to 'electronically de-paraffinize' the tissues where the influence of paraffin contribution was neutralized. Working on paraffinized tissues have an advantage of reducing the tissue processing steps by avoiding chemical treatments and at the same time provides an index matching medium to reduce the potential effects of scattering artefacts. The use of EMSC to electronically de-paraffinize the tissues is now well-established (28-33) which along with 'electronic de-paraffinization' was used for baseline correction (polynomial order 4) and normalisation of the IR spectra. Once the images are pre-processed using EMSC, outlier spectra (paraffin spectra and other noisy spectra) are depicted as white pixels in the images. The pre-processed spectra of each image (excluding outliers) were independently subjected to *k*-means cluster analysis (using Euclidean distance) to separate different histological/histo-pathological features based on the spectral features. Spectral distances between clusters were calculated (Hierarchical clustering using Ward's linkage algorithm) and represented in the form of a dendrogram in order to visualize the heterogeneity between different histological groups.

The signal-to-noise calculations of both the conventional and the high-magnification set-ups were carried out for the polystyrene film measurements. For this, the maximum peak height minus the minimum peak height, above the baseline, in the selected range of 1480 - 1500 was used for the signal while the maximum peak height minus the minimum peak height in the signal-free range of 1556 - 1568 was used for the noise. The median SNR over all the pixels in the image with error bars of +/- 1 median absolute deviation from the median (MAD) were calculated. Medians were used to reduce the influence of outlier values.

Results and discussion:

In order to ascertain the signal-to-noise response of the instrument, IR signal from a polystyrene film was used as a standard. An IR spectral profile of the polystyrene film is shown in SI 1. The SNR between the conventional and the high-magnification systems at different numbers of scans were compared as shown in figure 1. The plot shows that there is overall a two times difference between the signal-to-noise levels of the two modalities. The plots also show an increasing signal trend with increasing number of scans consistently for both conventional and high-magnification set-ups, except

for 32 scans in conventional magnification which does not show any improved SNR compared to 16 scans.

Once this trend was ascertained on the polystyrene film, tissue measurements using the conventional set-up were carried out. Although a maximum of 32 scans were used for the SNR testing on the polystyrene film, expecting additional influential factors in tissue (such as heterogeneity, uneven thickness scattering, etc.), 64 scans were used for tissue measurements resulting in a feasible 5 min measurement time for a single FPA tile of 128x128 pixels or 704x704 μm^2 of tissue (note that the background measurement is an additional 15 min, but does not need to be repeated if multiple sample tiles are required to cover the sample). To obtain similar signal-to-noise with the high-magnification mode, analysis of the envelope results indicate that twice the scans would be needed. However this would double the time required per tile. Furthermore, it requires 25 high-magnification tiles to cover the same area as a single conventional magnification tile. Therefore in order to find a compromise between the signal-to-noise and the time taken, 64 scans were opted for, which is identical to the conventional set-up. It therefore takes 125 min plus a single 15 min background using high-magnification to cover the same area as a single conventional magnification tile. Data analysed thereafter on these images showed acceptable segmentation of the pathological features as discussed later, indicating that the signal-to-noise is still sufficient.

In order to test the applicability of the novel high-magnification optics for characterizing the histological features of biological samples, IR imaging was performed on colon tissues and then compared to conventional IR imaging. The important criteria was to validate if histological features on a smaller scale, such as individual cells from the whole tissue, could be directly visualized and thereby characterised together with the other histological features. For this, IR images were obtained and multivariate analysis based on *k*-means clustering was individually performed on all IR images. Shown in figure 2 is a comparison of the cluster images between the conventional and the high-magnification set-ups of the same normal colon tissue. The cluster image obtained with the conventional imaging (top panel, second column) show a segmentation for the glandular (clusters 1, 8, 5, 9 and 11) and connective tissue regions (4, 7, 2, 6 and 10) in comparison to the reference HE image (top panel, first column). This is in concordance with other studies including our own (4, 6, 29) using similar scale pixel resolutions. The cluster heterogeneity was represented in the corresponding

dendrogram (top panel, third column). Furthermore, the glandular clusters could be attributed to nuclear region (cluster 1), cytoplasmic region (clusters 8) and mucinous regions along the outside opening of the glands (clusters 5, 9 and 11). Interestingly, cluster 11 which contains mucinous features, could also be seen in the cytoplasmic regions which indicated mucin producing goblet cell; however this is less evident due to the lower resolution of the pixels. Until recently, this was the limit for observing small scale histological features using the bench top IR imaging systems.

Following this, the cluster image (bottom panel, second column) obtained from the same region of the tissue measured using the high-magnification set-up was then analyzed. In comparison to the reference HE image (bottom panel, first column) and the cluster image from conventional magnification (top panel, second column) similar partitioning could be observed for glandular region (clusters 1, 3, 4, 7, 8 and 9) and connective tissue regions (clusters 2, 5, 6, 11 and 10) that are characterised by an improved contrast between classes. Likewise, the glandular structures could be attributed to nuclear (cluster 3) and cytoplasmic regions (clusters 1 and 4). However, the most remarkable observation with the new high-magnification set-up are the clearly distinguishable cellular features localized in the glandular area (clusters 1, 8 and 9), dispersed across the gland. An enlarged version of the cluster map obtained using the high-magnification set-up is shown in figure 3 for more detailed information. The fact that similar cluster attribution was obtained for these classes and the classes along the outside of the glands which correspond to the secreted mucous, indicate that the clusters 1, 8 and 9 are the mucin producing goblet cells.

To further investigate this, a second benign and a larger colonic gland in this case with tubular adenoma was also measured and analyzed using both the conventional and high-magnification set-ups as shown in figure 4. Once again using the conventional modality, histological features conferring to glandular (clusters 2, 4, 7 and 3) and connective tissue (clusters 6, 8, 1 and 5) regions were discernible in the cluster image (top panel, second column) in comparison to the HE reference image (top panel, first column). From the glandular structures cluster 3 could be attributed to the nuclear region and 7, 2 and 4 to the cytoplasmic regions. Cluster 4 could be attributed to the mucin filled lumen of the glands. No further information could be exploited from this image.

Since the field of view was different between the conventional and the high-magnification images and in order to have a better comparison between the two, similar regions of interest was selected in the

IR image of the conventional modality and compared to the high-magnification image. Similar analysis was carried out on the selected region of interest independently. As shown in figure 5, globally the conventional image show certain similarities to the high-magnification image separating the glandular region from the connective tissue region and separating the mucin rich lumen region. However, details pertaining to the cellular features could not be resolved. Further, increasing the cluster numbers provided no further details as the image tend to become more pixelated (data not shown) as the pixels are averaged over larger area.

On the other hand, some significant details of the histology were observed using the high-magnification imaging (bottom panel, second column) in comparison to the HE reference image (bottom panel, first column) and also the cluster image from the conventional magnification. An enlarged version of the cluster map obtained using the high-magnification set-up is shown in figure 6 for more detailed information. Cluster analysis (N=17 clusters; bottom panel, second column) was able to segment the gland into detailed histological features where clusters were attributed to connective tissue regions (clusters 10, 11, 8, 5, 9 and 17) and remaining clusters to the glandular region. Within the glandular regions, clusters 7 and 13 could be attributed to the secreted mucous in the central lumen of the gland.

A patterned localization of cellular features were observed in this gland (clusters 3 and 12) which based on the corresponding dendrogram (figure 4, bottom panel, third column), are grouped close to extracellular mucin in the central lumen of the gland (clusters 7 and 13). This strongly suggests that they are goblet cell features. This holds true as usually goblet cells are the mucin secreting organelles in the colon filled with mucin up to most of its volume. Furthermore as shown in the cluster map (figure 4, bottom panel, second column), it was also observed that firstly the goblet cell features were attributed to two distinct clusters (clusters 3 and 12), and secondly as can be observed in the corresponding dendrogram (figure 4, bottom panel, third column) that these two classes grouped distinctly with two different cluster groups (cluster 3 with clusters 1, 4 and 14 and cluster 12 with clusters 7 and 13) probably indicating differences in these cell types. Interestingly, when compared with the clusters 7 and 13 originating from the extracellular mucin secreted into the lumen of the gland, and which could be considered as a purer form of mucin features, cluster 12 is more similar than cluster 3. To corroborate with this, when the spectral features were observed (figure 7), although

both the cluster groups show glycoprotein profile, the centroids of the cluster group 7, 12, 13 showing higher relative intensities than the cluster group 1, 3, 4, 14 as indicated by the spectral peaks in figure 7.

To further investigate this, principal component analysis (PCA) was carried out on the ensemble of spectra from these two cluster groups of interest namely cluster 12 labelled as G1 (Goblet 1) and cluster 3 labelled as G2 (Goblet 2). As displayed in figure 8, the PCA score plot shows a clear separation of these two groups indicating differences between them and also confirming the cluster analysis results. When the PCA loading was observed (figure 8), most of the mucin peaks (1039 cm^{-1} , 1074 cm^{-1} , 1125 cm^{-1} and 1317 cm^{-1} ; highlighted in green) (7, 29) were associated with G1 (positive peaks in the loading and positive scores in the score plot) compared to G2. Although both these classes are goblet cells with in the same gland, only some of the cells (G1) appear to show predominant secretion of mucins. All the above results (differential segmentation of goblet cells and predominant mucin secretion in some of the goblet cells) appear to indicate probable molecular changes in the gland which were not observed in the other benign tissues (SI 2 and SI 3). It could be hypothesized that these changes could possibly be attributed to predominant or selective secretion of mucins by some goblet cells or simply due to the different degree of maturation of these cells as they develop from the bottom to the top region of the crypt. However, this differential segmentation was not observed along the length of the crypts in other benign colon tissues analyzed in this study as shown in SI 2 and SI 3 although a patterned organization of goblet cells was observed. Further planned studies are required to understand if such observations have any association with samples of tubular adenoma as in this case. Nevertheless, the high-magnification IR imaging does seem to offer some important applications that could be of pathological importance.

In several studies including our previous work, colon tissues have been characterised using IR spectroscopic imaging (4, 6, 7 and 29). Due to the limit of the detector pixel size to $6.25 \times 6.25\ \mu\text{m}^2$ or $5.5 \times 5.5\ \mu\text{m}^2$ in most of the commercially available benchtop instruments, only the gross histological features were resolved. Recently, with improvements to various components of the IR imaging systems, biological tissues are being looked into in an unprecedented highly resolved scale. Some of the first studies demonstrated the feasibility to obtain improved image quality using smaller pixel resolutions of around $1 \times 1\ \mu\text{m}^2$ using the conventional Globar[®], QCL and synchrotron sources in

combination with the FPA detectors (20, 24). Following the availability of commercial benchtop IR imaging system with an improved optical set-up, high-resolution imaging has been possible where measurements on cells and tissues were carried out (34, 35, and 36) with promising outcomes.

In this study, IR imaging of colon tissues using the high-magnification set-up in combination with multivariate analysis has been carried out. For the first time, we are able to observe a patterned localization of cells in the colonic crypt directly without any staining procedures together with a very detailed segmentation of the colon tissue. These observations supported by the spectral characteristics indicate that they are goblet cells. Usually goblet cells are present all around the colon crypts together with other types of epithelial cells and are not easily recognized using the routine HE staining procedures. In addition, due to the limit in the spatial resolution of the benchtop instruments, it was not possible to identify these cell types morpho-spectroscopically in earlier studies. This study on the one hand demonstrated the higher resolution that can be achieved using conventional sources on a bench top instrument using improved optics and on the other hand provides important histological information that could be of clinical significance. From histopathological point of view, mucins are implicated in disorders such as cancer and inflammatory bowel disorders such as uncreative colitis (26, 27). Direct identification of the mucin producing goblet cells could be of interest to pathologists to characterise their number, size, localization and the type of mucins they produce. In addition to this, in cancers especially adenocarcinomas, mucins are also known to alter both qualitatively and quantitatively as a result of the deregulation of expression of mucin core proteins during the transformation of tumour cells (27). If mucin associated changes can be characterised using such technology at cellular level, it could provide vital information related to the characteristics of cancers being investigated. Previous IR spectroscopy studies on colon tissues were undertaken to identify the abnormalities in colon epithelial regions (4-6) and the secreted mucin (7) as potential indicators of cancer. These were carried out using the conventional IR set ups. The ability of novel technologies to enable cellular scale characterization of tissues could further enhance such types of work where cancer and inflammatory signatures can be localised back to the respective cellular regions of the tissue. Furthermore, the fact that this technology could be carried out directly on paraffinized tissues without any chemical de-paraffinization procedures, which are time consuming and known to be toxic, holds important prospects for rapid applications and retrospective studies in clinical field.

Finally, characterization of a tumoral sample using the two modalities was carried out (figure 9). In comparison to the HE images, the clusters 1, 2 and 11 could be attributed to the tumor in the conventional magnification set-up and the clusters 1 and 7 could be attributed in the high-magnification set-up. The comparison between the conventional and the high-magnification set-ups show quite similar partitioning of the tumor from its surrounding stroma in both the modalities. As the structural organization of colonic glands is lost in a cancerous tissue, morphologically no striking difference was observed between these two modalities. Although at present, histological information is being obtained in detail using the high-magnification imaging, its discrimination capabilities between tumoral and non-tumoral tissues have not been tested. Accurate discrimination could be of significant importance in complicated cases such as those in which tumors are surrounded by complex network of stroma or other non-tumoral tissue, in order to isolate purer tumor pixels especially where they will be detrimental in developing training algorithms. This also has important implications in delimiting the tumor margins and also shedding light on the tumor-stroma boundaries.

Conclusion:

The novel IR high-magnification spectroscopic imaging was able to provide much more detailed information of the colon histological features in comparison to the conventional imaging set-up. In this study, the goblet cell features were resolved demonstrating the capability of this novel imaging modality in segmenting cellular features from other tissue features. Additionally, the high-magnification imaging together with multivariate analyses was firstly able to show a differential segmentation of goblet cells and secondly demonstrate a predominant mucin secretion within some of the goblet cells. Further tests will be undertaken to explore the relative performance of high- versus conventional-magnification IR imaging for discriminating pathology from normal tissues in the colon. Further work will be required to evaluate the merits of high-magnification IR imaging for discriminating pathology from normal tissues in the colon. It may be the case that the increased spectral heterogeneity from each tissue type, which is advantageous when looking at individual tissue sections, may be detrimental for spectral prediction of pathology.

Acknowledgments:

This work is part of the MINERVA project (<http://www.minerva.eu>) which is supported by the European Union's Seventh Framework Programme (FP7) for research, technological development and demonstration under grant agreement no 317803. The authors would like to thank Jim Wood and Tim Cook at Gloucestershire Hospitals NHS Trust for collecting the tissue samples used here. Mustafa Kansiz is acknowledged for helpful discussions.

References:

1. F. L. Martin, J. G. Kelly, V. Llabjani, P. L. Martin-Hirsch, I. Patel, J. Trevisan, N. J. Fullwood, M. J. Walsh, *Nat. Protoc.*, 2010, **5**, 1748-1760.
2. A. Akalin, X. Mu, M. A. Kon, A. I. Ergin, S. H. Remiszewski, C. M. Thompson, D. J. Raz, M. Diem, *Lab. Invest.*, 2015, **95**, 406-421.
3. A. Salman, G. Sebbag, S. Argov, S. Mordechai, R. K. Sahu, *J. Biomed. Opt.*, 2015, **20 (7)**, 075007, (1-10).
4. J. Nallala, M. D. Diebold, C. Gobinet, O. Bouche, G. D. Sockalingum, O. Piot, M. Manfait, *Analyst*, 2014, **139**, 4005-4015.
5. A. Kallenbach-Thieltges, F. Großeruschkamp, A. Mosig, M. Diem, A. Tannapfel, K. Gerwert, *J. Biophotonics*, 2013, **6**, 88-100.
6. J. Nallala, O. Piot, M. D. Diebold, C. Gobinet, O. Bouché, M. Manfait, G. D. Sockalingum, *Cytometry, Part A*, 2013, **83**, 294-300.
7. A. Travo, O. Piot, R. Wolthuis, C. Gobinet, M. Manfait, J. Bara, M. E. Forgue-Lafitte, P. Jeannesson, *Histopath*, 2010, **56**, 921-931.
8. J. D. Pallua, C. Pezzei, B. Zelger, G. Schaefer, L. K. Bittner, V. A. Huck-Pezzei, S. A. Schoenbichler, H. Hahn, A. Kloss-Brandstaetter, F. Kloss, G. K. Bonna, C. W. Huck, *Analyst*, 2012, **137**, 3965-3974.
9. F. Le Naour, C. Sandt, C. Peng, N. Trcera, F. Chiappini, A. M. Flank, C. Guettier, P. Dumas, *Anal. Chem.*, 2012, **84**, 10260-10266.

10. J. T. Kwak, S. M. Hewitt, S. Sinha, R. Bhargava, *BMC Cancer*, 2011, **11**(62), 1-16.
11. E. Gazi, M. Baker, J. Dwyer, N. P. Lockyer, P. Gardner, J. H. Shanks, R. S. Reeve, C. A. Hart, N. W. Clarke, M. D. Brown, *Eur. J. Urology*, 2006, **50**, 750-761.
12. D. C. Fernandez, R. Bhargava, S. M. Hewitt, I. W. Levin, *Nat. Biotechnol.*, 2005, **23**, 469-474.
13. R. Baker, K. D. Rogers, N. Shepherd, N. Stone, *B. J. Cancer*, 2010, **103**, 1034-1039.
14. H. Fabian, N. A. Thi, M. Eiden, P. Lasch, J. Schmitt, D. Naumann, *Biochim. Biophys. Acta*, 2006, **1758**, 874-882.
15. A. Benard, C. Desmedt, M. Smolina, P. Szternfeld, M. Verdonck, G. Rouas, N. Kheddoumi, F. Rothe, D. Larsimont, C. Sotiriou, E. Goormaghtigh, *Analyst*, 2014, **139**, 1044-56.
16. B. Bird, M. Romeo, N. Laver, M. Diem, *J. Biophotonics*, 2009, **2**, 37-46.
17. B. Bird, M. Miljkovic, M. J. Romeo, J. Smith, N. Stone, M. W. George, M. Diem, *BMC Clinical Path.*, 2008, **8** (8), 1-14.
18. W. Steller, J. Einkenkel, L. C. Horn, U. D. Braumann, H. Binder, R. Salzer, C. Krafft, *Anal. Bioanal. Chem.*, 2006, **384**, 45-154.
19. P. Lasch, W. Haensch, E. N. Lewis, L. H. Kidder, D. Naumann, *Appl. Spectros.*, 2002, **56** (1), 1-9.
20. K. Yeh, S. Kenkel, J. N. Liu, R. Bhargava, *Anal. Chem.*, 2015, **87** (1), 485-493.
21. N. Wald, E. Goormaghtigh, *Analyst*, 2015, **140**, 2144-2155.
22. F. Großerueschkamp, A. Kallenbach-Thieltges, T. Behrens, T. Brüning, M. Altmayer, G. Stamatis, D. Theegarten, K. Gerwert, *Analyst*, 2015, **140**, 2114-2120.
23. P. Bassan, A. Sachdeva, J. H. Shanks, M. D. Brown, N. W. Clarke, P. Gardner, *Analyst*, 2013, **138**, 7066-7069.
24. R. K. Reddy, M. J. Walsh, M. V. Schulmericha, P. S. Carney, R. Bhargava. *Appl Spectrosc.*, 2013, **67**(1), 93-105.
25. C. J. Hirschmugl, K. M. Gough, *Appl. Spectrosc.*, 2012, **66**, 475-491.
26. Y. S. Kim, S. B. Ho., *Curr. Gastroenterol. Rep.*, 2010, **12**, 319-330.

27. M. A. Hollingsworth, B. J. Swanson, *Nat. Rev. Cancer*, 2004, **4**, 45-60.
28. A. Kohler, C. Kirschner, A. Oust, H. Martens, *Appl. Spectrosc.*, 2005, **59**, 707-716.
29. J. Nallala, C. Gobinet, M. D. Diebold, V. Untereiner, O. Bouché, M. Manfait, G. D. Sockalingum, O. Piot, *J. Biomed. Opt.*, 2012, **17**, 1-12.
30. E. Ly, O. Piot, R. Wolthuis, A. Durlach, P. Bernard, M. Manfait, *Analyst*, 2008, **133**, 197-205.
31. R. Wolthuis, A. Travo, C. Nicolet, A. Neuville, M. P. Gaub, D. Guenot, E. Ly, M. Manfait, P. Jeannesson, O. Piot, *Anal. Chem.*, 2008, **80**, 8461-8469.
32. J. Nallala, G. R. Lloyd, N. Stone, *Analyst*, 2015, **140**, 2369-2375.
33. D. Sebiskveradze, V. Vrabie, C. Gobinet, A. Durlach, P. Bernard, E. Ly, M. Manfait, P. Jeannesson, O. Piot, *Lab. Invest.*, 2011, **91**, 799-811.
34. C. Hughes, A. Henderson, M. Kansiz, K.M. Dorling, M. J. Hernandez, M.D. Brown, N.W. Clarke, P. Gardner, *Analyst*, 2015, **140**, 2080-2085.
35. C. R. Findlay, R. Wiens, M. Rak, J. Sedlmair, C. J. Hirschmugl, J. Morrison, C. J. Mundy, M. Kansiz, K. M. Gough, *Analyst*, 2015, **140**, 2493-2503.
36. L. S. Leslie, T. P. Wrobel, D. Mayerich, S. Bindra, R. Emmadi, R. Bhargava, *PLoS ONE*, 2015, **10** (6), 1-17.

Figures:

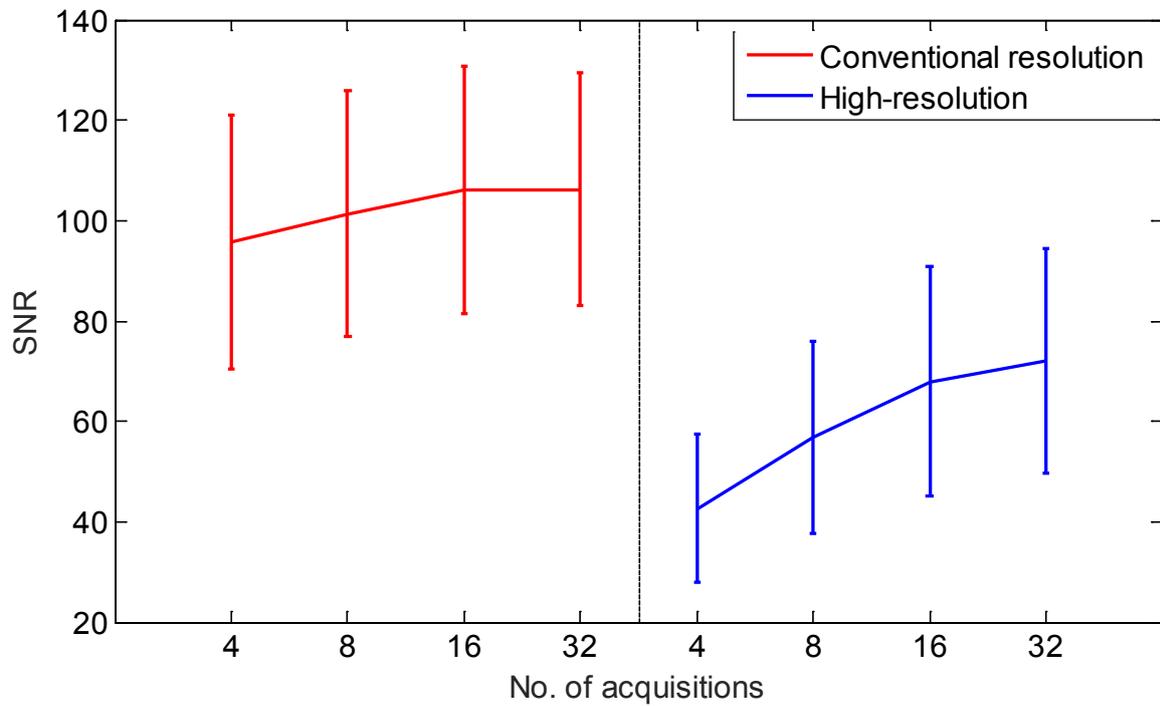


Figure 1: Comparison of the signal to noise (SNR) ratios between the conventional and the high-resolution imaging setups at different number of scans. Plots indicate median signal to noise ratio over all the pixels of the IR image of a polystyrene film with error bars of ± 1 median absolute deviation from the median.

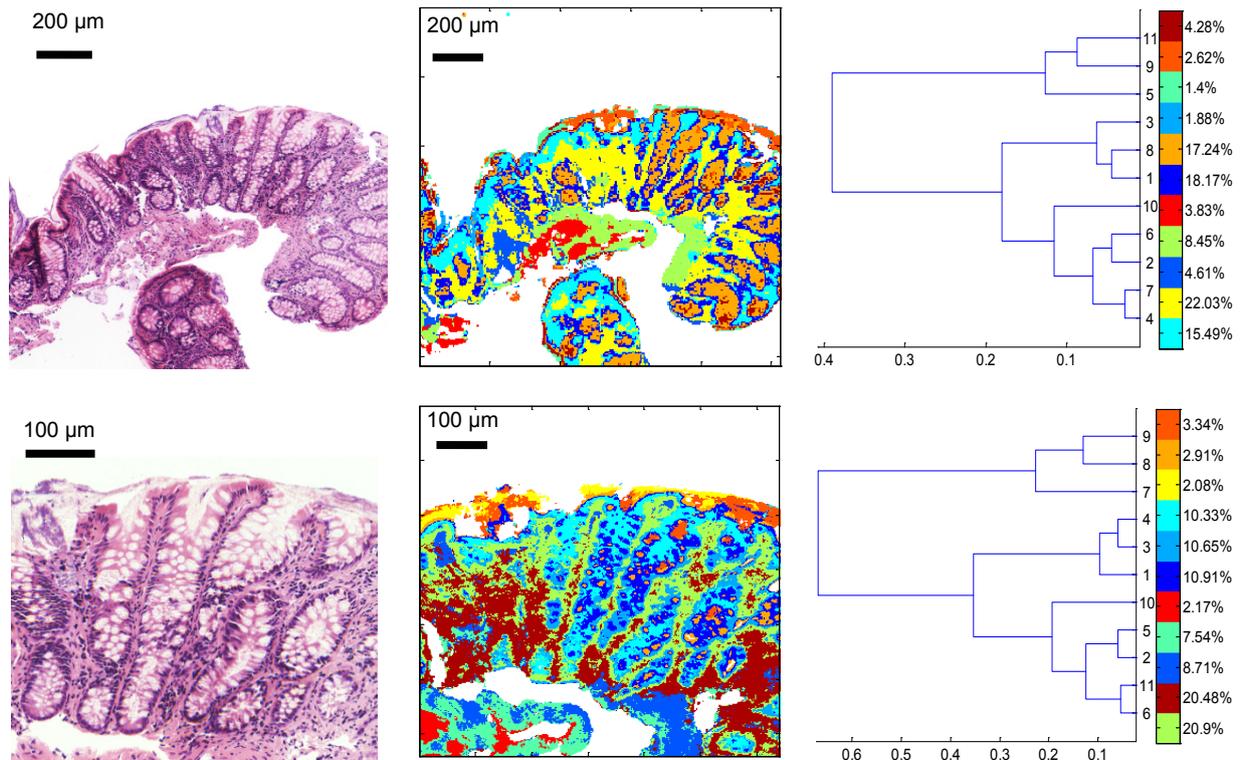


Figure 2: Comparison of infrared spectral-histopathological characteristics of a normal colon tissue measured using the conventional (top row) and the high-magnification (bottom row) infrared imaging set ups independently. Top row: Cluster analysis of a normal colon tissue (second column) using 11 cluster groups, measured using the conventional imaging set up ($5.5 \times 5.5 \mu\text{m}^2$), in comparison to the reference HE stained image (left column). The corresponding dendrogram (third column) representing the heterogeneity of the clusters is also shown. Bottom row: Cluster analysis of the same normal colon tissue (second column) using 11 cluster groups, measured using the high-magnification imaging set up ($1.1 \times 1.1 \mu\text{m}^2$), in comparison to the reference HE stained image (left column). The corresponding dendrogram (third column) representing the heterogeneity of the clusters is also shown.

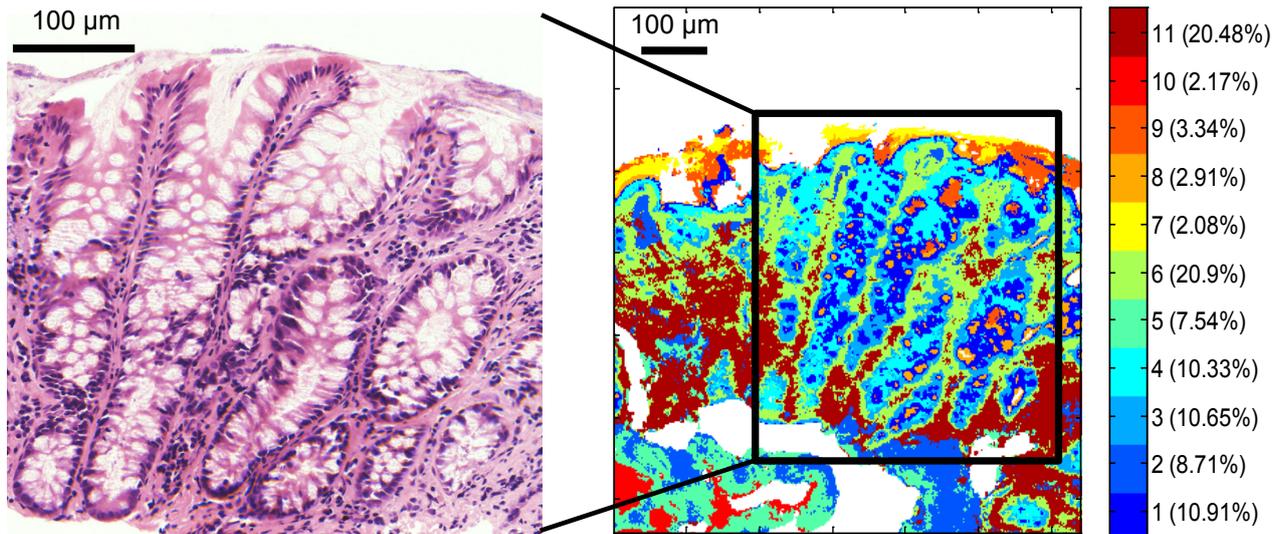


Figure 3: An enlarged version of cluster analysis of the normal colon tissue measured using the high-magnification imaging set up ($1.1 \times 1.1 \mu\text{m}^2$) in comparison to the reference HE stained image. Cluster analysis show visibly more apparent cellular features (clusters 1, 8 and 9).

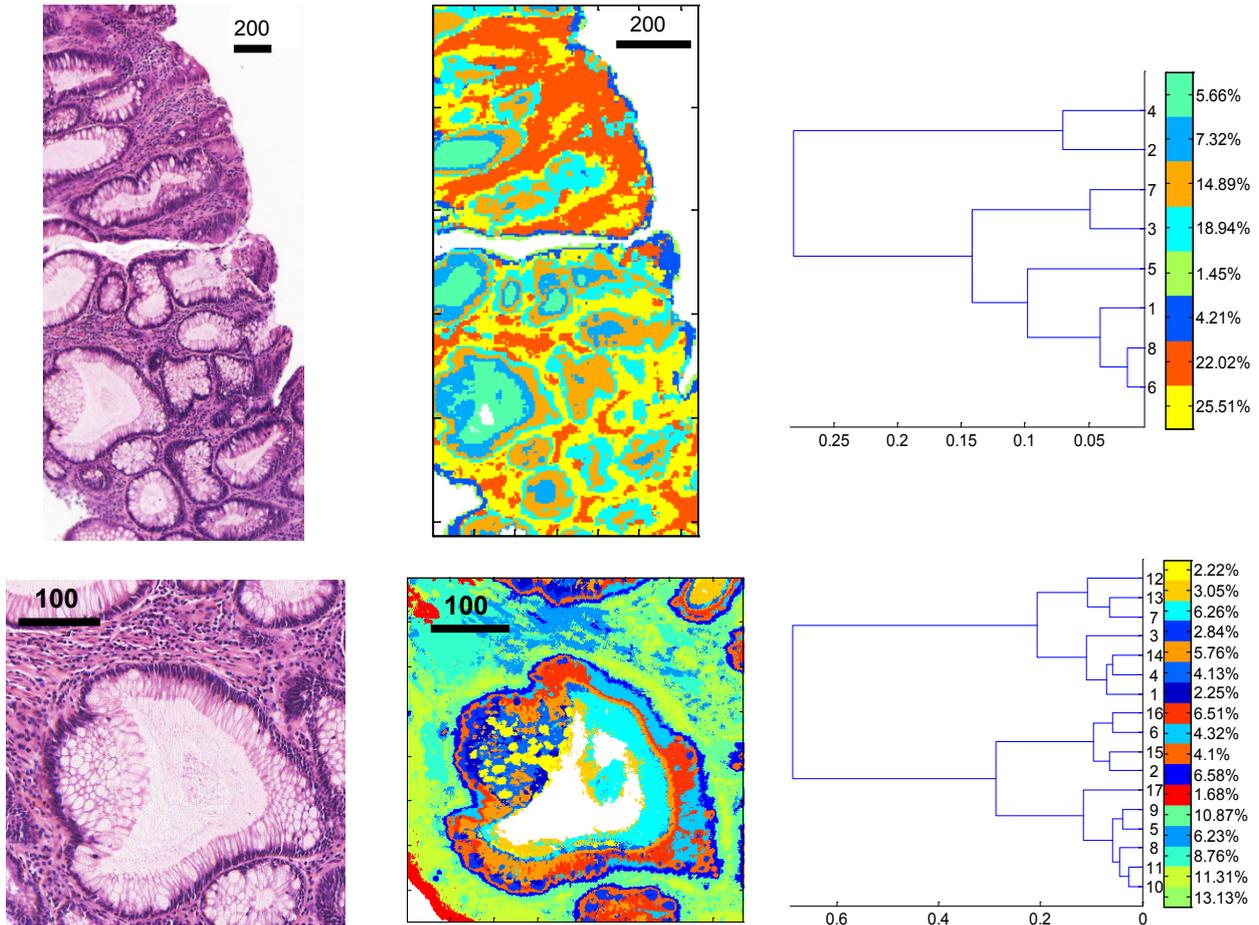


Figure 4: Comparison of infrared spectral-histopathological characteristics of a colon tissue with tubular adenoma measured using the conventional (top row) and the high-magnification (bottom row) infrared imaging set ups independently. Top row: Cluster analysis of a colon tissue (second column) with tubular adenoma using 8 cluster groups, measured using the conventional imaging set up ($5.5 \times 5.5 \mu\text{m}^2$), in comparison to the reference HE stained image (left column). The corresponding dendrogram (third column) representing the heterogeneity of the clusters is also shown. Bottom row: Cluster analysis of the same colon tissue (second column) using 17 cluster groups, measured using the high-magnification imaging set up ($1.1 \times 1.1 \mu\text{m}^2$), in comparison to the reference HE stained image (left column). The corresponding dendrogram (third column) representing the heterogeneity of the clusters is also shown.

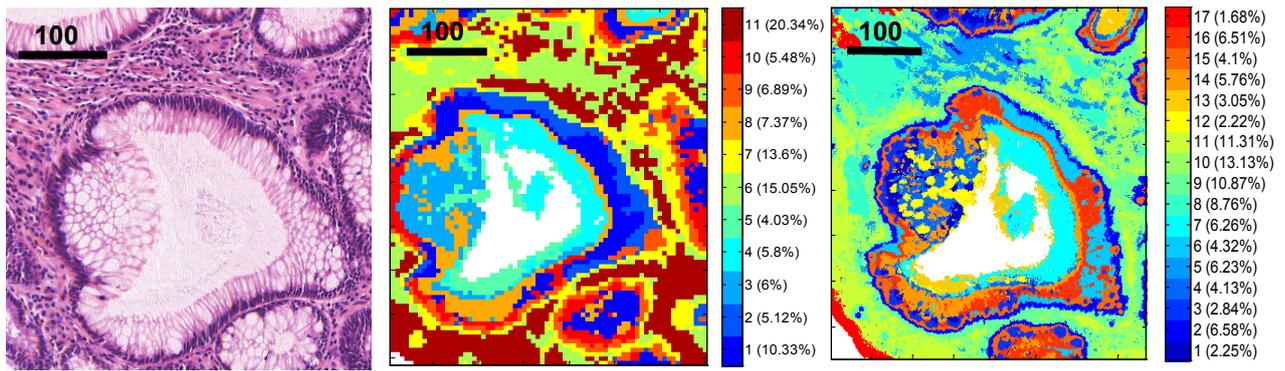


Figure 5: Comparison of infrared spectral-histopathological characteristics of a colon tissue with tubular adenoma measured using the conventional and the high-magnification infrared imaging set ups independently over the same region of interest. Cluster analysis of a colon tissue with tubular adenoma (using 11 and 17 cluster groups respectively) measured using the conventional imaging set up at $5.5 \times 5.5 \mu\text{m}$ (middle column) and the high-magnification imaging set up at $1.1 \times 1.1 \mu\text{m}^2$ (right column) in comparison to the reference HE image (left column).

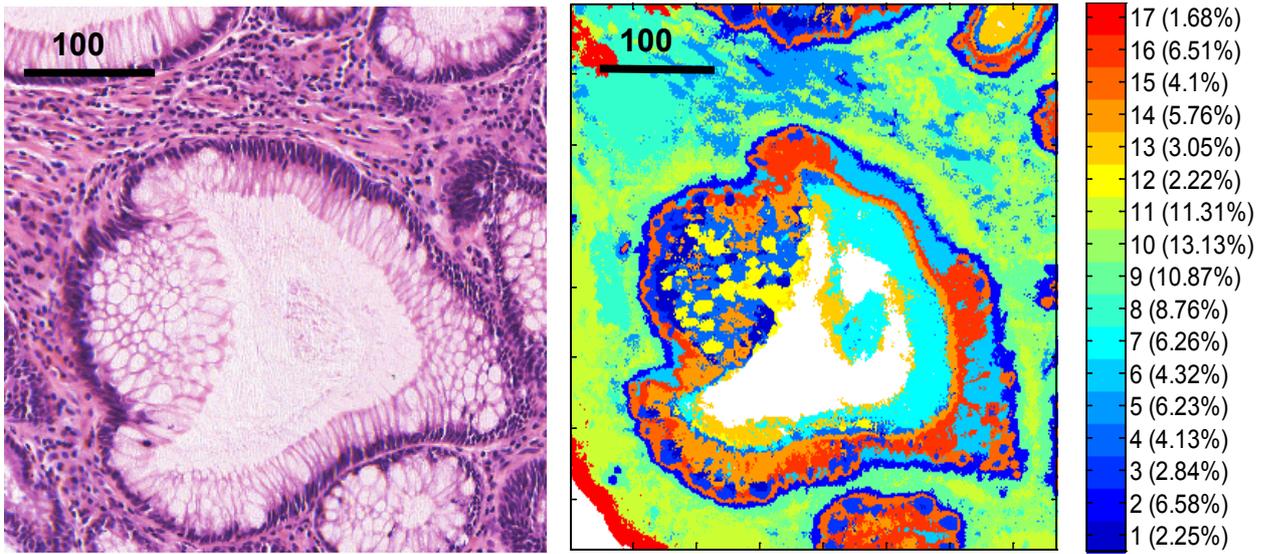


Figure 6: An enlarged version of the cluster analysis of the colon tissue with tubular adenoma measured using the high-magnification imaging set up ($1.1 \times 1.1 \mu\text{m}^2$) in comparison to the reference HE stained image. Cluster analysis show visibly more apparent cellular features (clusters 3 and 12) and the extra cellular mucin in the lumen of the gland (clusters 7 and 13).

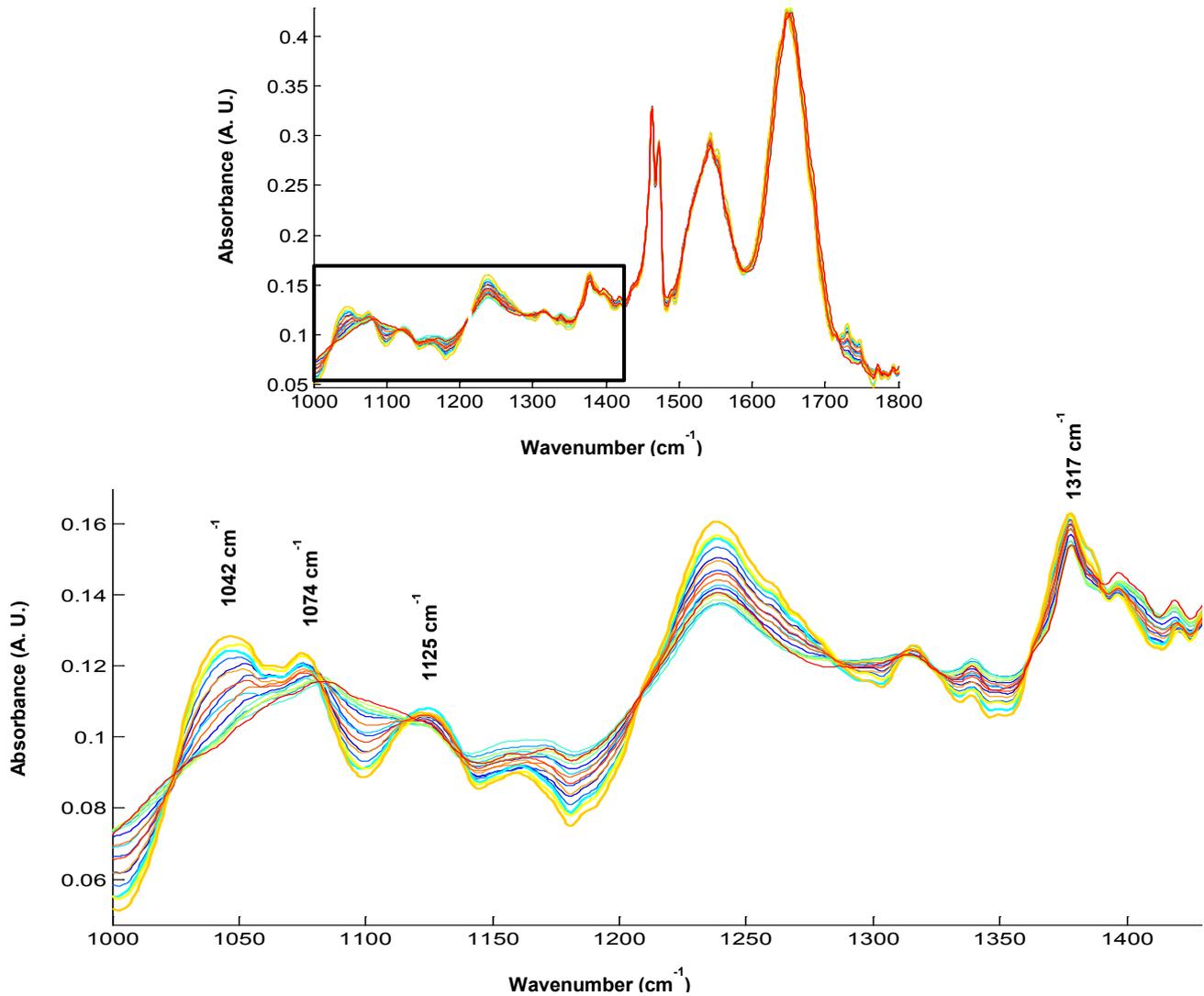


Figure 7: Spectral centroids from figure 5 (colon tissue with tubular adenoma). Top figure showing the 1000-1800 cm⁻¹ region and the bottom figure representing the highlighted clusters 7 (cyan), 12 (yellow) and 13 (orange) which show the typical glycoprotein profile peaks (mucin) in the selected region (1000-1450 cm⁻¹).

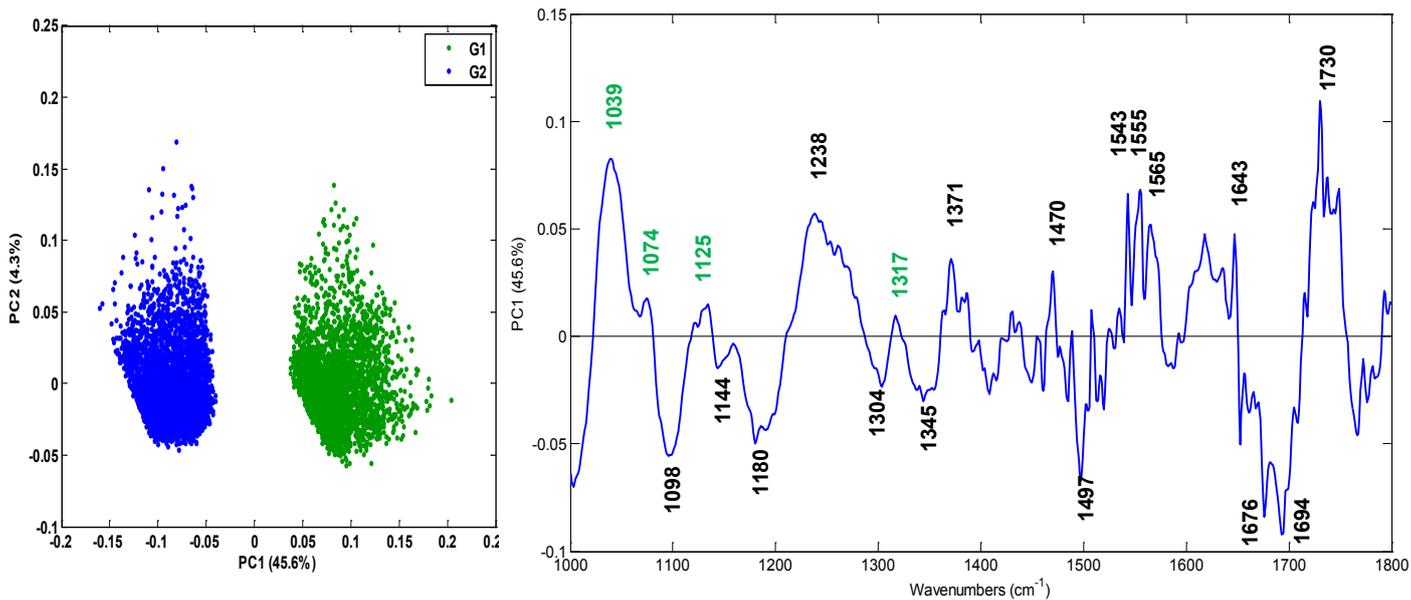


Figure 8: Principal component analysis (PCA) showing the score plot (left) and the loading (right) for the clusters 12 (G1) and 3 (G2) of fig 5. For the score plot, PC1 shows a clear separation with the highest explained variance. PC1 loading is plotted on the right which shows predominant mucin features (highlighted in green) separated from the other histological features.

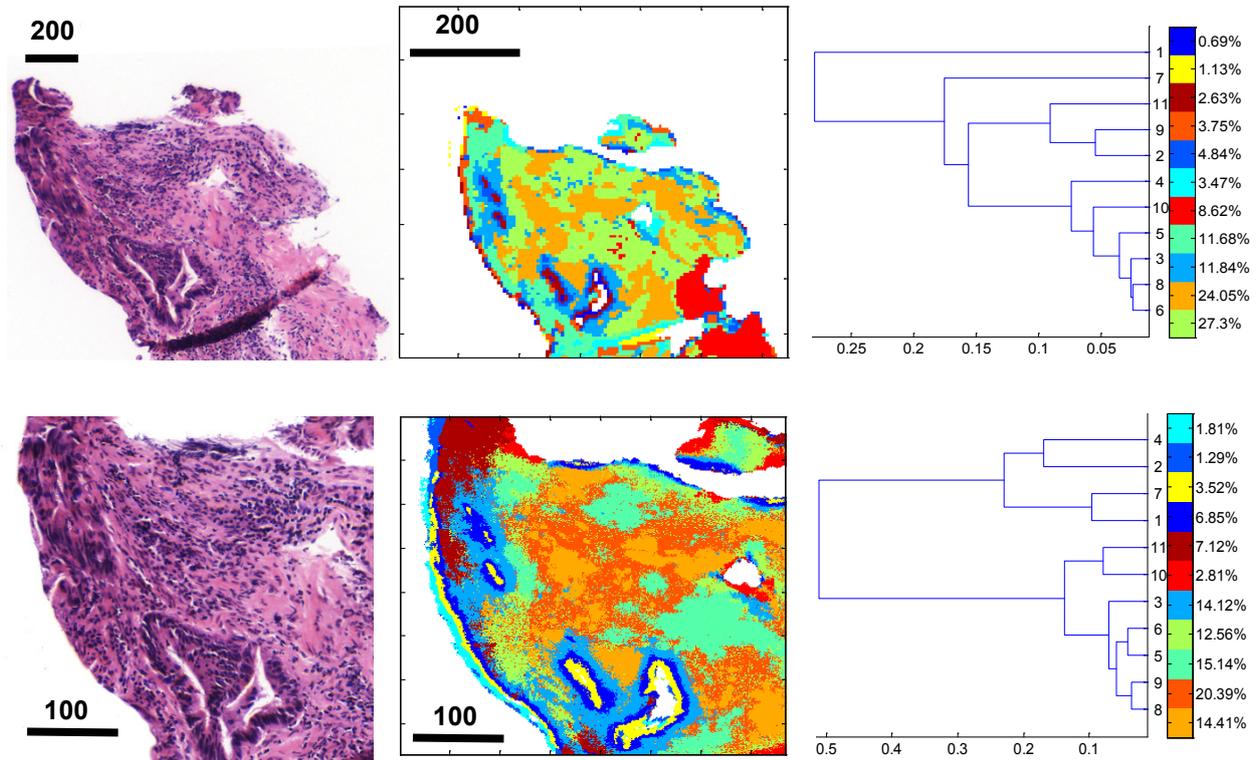


Figure 9: Comparison of infrared spectral-histopathological characteristics of a moderately differentiated colon adenocarcinoma (tumoral) measured using the conventional (top row) and the high-magnification (bottom row) infrared imaging set ups independently. Top row: Cluster analysis of a tumoral colon tissue (second column) using 11 cluster groups, measured using the normal imaging set up ($5.5 \times 5.5 \mu\text{m}^2$), in comparison to the reference HE stained image (left column). The corresponding dendrogram (third column) representing the heterogeneity of the clusters is also shown. Bottom row: Cluster analysis of the same tumoral colon tissue (second column) using 11 cluster groups, measured using the high-magnification imaging set up ($1.1 \times 1.1 \mu\text{m}^2$), in comparison to the reference HE stained image (left column). The corresponding dendrogram (third column) representing the heterogeneity of the clusters is also shown.