Developing Raman Spectroscopy as a Diagnostic Tool for Label-free Antigen Detection

Aaran T. Lewis [a*], Riana Gaifulina [a], Naomi J Guppy [b], Martin Isabelle [c], Jennifer Dorney [d], Gavin R. Lloyd [c], Manuel Rodriguez-Justo [a], Catherine Kendall [c], Nicholas Stone [d], Geraint M. Thomas [a]

* Correspondence to: Aaran T. Lewis, Department of Cell and Developmental Biology, University College London, London, UK.

E-mail: aaran.lewis@ucl.ac.uk

[a] Department of Cell and Developmental Biology, University College London, London, UK

[b] UCL-Advanced Diagnostics, University College London, London, UK

[c] Biophotonics Research Unit, Gloucester Royal Hospitals NHS Foundation Trust, Gloucestershire, UK

[d] School of Physics and Astronomy, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK

Abstract

For several decades, a multitude of studies have documented the ability of Raman spectroscopy (RS) to differentiate between tissue types and identify pathological changes to tissues in a range of diseases. Furthermore, spectroscopists have illustrated that the technique is capable of detecting disease-specific alterations to tissue before morphological changes become apparent to the pathologist. This study draws comparisons between the information that is obtainable using RS alongside immunohistochemistry (IHC); since histological examination is the current GOLD standard for diagnosing a wide range of diseases. Here, Raman spectral maps were generated using formalin-fixed, paraffin-embedded colonic tissue sections from healthy patients and spectral signatures from principal components analysis (PCA) were compared with several IHC markers to confirm the validity of their localizations. PCA loadings identified a number of signatures that could be assigned to muscle, DNA and mucin glycoproteins and their distributions were confirmed with antibodies raised against anti-Desmin, anti-Ki67 and anti-MUC2 respectively. The comparison confirms that there is excellent correlation between RS and the IHC markers used, demonstrating that the technique is capable of detecting compositional changes in tissue in a label-free manner, eliminating the need for antibodies.

Keywords: Raman Spectroscopy; Immunohistochemistry; Tissue; Diagnosis; classification

1 Introduction

Raman spectroscopy (RS) is a vibrational spectroscopic technique that relies on inelastic scattering of light by polarizable molecules. Raman scatter occurs when photons of a defined energy impinge on an analyte, leading to the emission of lower and higher energy photons, which are detected and provide information on the biochemical composition of cells, tissues and biofluids[1]. The method offers chemical detection, analysis and interpretation with minimal or no sample preparation as Raman is an optical

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jbio.201700028

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

technique[2]. RS has been used to analyze biological samples for decades, but with recent advances in lasers, detectors and fibre optics along with improvements in spatial resolution and scan speed it is now feasible to implement Raman in a routine clinical setting for *in vivo* disease classification and diagnosis[3]. The fundamental principle underpinning these developments is that disease-related molecular changes in cells, tissues or biofluids can be reliably and reproducibly detected by RS and used as characteristic fingerprints[1]. Importantly, multivariate calibration and classification models can be applied to mine large spectral datasets in order to obtain quantitative and objective diagnoses in a range of diseases[4, 5]. RS therefore has the potential to be used for diagnostics, prognostics and as a tool for evaluating response to treatment.

Colorectal cancer (CRC) is one of the most prevalent cancers in the developed world[6], but is also one of the most easily curable if detected at an early-stage[7]. Genetic mutations cause CRCs and the disease is broadly categorized into hereditary and sporadic types, which have differing genetic causes[8]. In most cases the disease develops slowly and a number of CRCs remain asymptomatic for many years prior to diagnosis. Fecal occult blood testing is widely used to screen for CRC, but although the technique is cheap and inexpensive it is error-prone in detecting CRCs[9]. Therefore patients with abnormal screening results are examined by colonoscopy and CRCs are routinely detected by visual inspection of suspicious lesions and polyps that are seen during the procedure. Suspicious tissues are removed during colonoscopy, stained with chemical dyes, and later inspected by a pathologist for any sign of cancer. This method can provide a basic diagnosis of cancer presence and stage, however simple visual inspection provides no information on the origin of the disease or disease-associated genetic markers. This information is obtained by analysis of known cancer markers, for example genetic mutations must be uncovered using next generation sequencing and aberrations at the protein level with immunohistochemistry (IHC) respectively.

Two of the most widely used groups of molecular markers are carcinoembryonic antigens and carbohydrate antigens[9]. The markers have been associated with cancer progression and prognosis[10, 11]. Microsatellite instability has also been linked with a large fraction of colorectal cancers and these are identified using immunohistochemistry (IHC). IHC is widely used in pathology to detect disease-related tissue antigens via antigen-antibody binding. In colon cancers, the technique is used to detect MLH1 and MSH2, which are key proteins involved in DNA mismatch repair and a loss of these proteins is indicative of microsatellite instability[9]. IHC also plays an important role in the detection of mucinous tumors in colon, with most advanced colorectal tumors abnormally secreting mucins, linking mucinous tumors with poor survival rates[12]. However, these types of genetic analyses are not routinely performed due to their cost. Furthermore, IHC is a very powerful tool in CRC diagnosis because it provides pathologists with additional (spatial) morphological information which can be used in phenotyping tumors. IHC relies on the ability of target-specific antibodies to bind antigens on tissue sections. For detection purposes the antibody probes are commonly linked to reporter molecules. These alert the pathologist to the antibody binding locations on the tissue sample by their direct UV or visible emission/absorption spectra or capacity to deposit easily visible stains (chromogens). Although extremely useful and well established in the field, IHC methods suffer a number of limitations when applied to tissue specimens. Firstly, the technique relies upon the availability of an antibody that specifically binds to an antigen of known interest. Once the appropriate antibody has been selected, staining is only reliable when samples are promptly fixed and reliably processed in order to maintain antigen viability[13]. The size of the tissue section available for staining is also of concern as small specimens are unsuitable for staining as the number of tumor cells may be below the detection limit and there may also be insufficient material available to confirm a positive result[9]. Staining is normally completed over a period of hours and protocols commonly implement a number of washing, blocking and incubation steps. The complex, multi-step nature of IHC staining also means that the reliability of results is governed by the experience of the operator performing staining, features of any automation used and quality of the antibodies; all of these features may contribute to variance in the

This article is protected by copyright. All rights reserved.

Accepted Article

data and create significant bias[13]. Furthermore, since markers need to be pre-selected and demonstrated to correlate with disease the technique is not suitable for identifying general abnormalities.

We are focusing on FFPE tissue blocks since these are the standard sample format utilized by pathology departments for IHC and other diagnostic tests. When working with FFPE issues, insufficient paraffin removal can lead to a number of known issues in both Raman and IHC. Firstly, paraffin peaks appear across the biological spectral region, showing distinctive Raman contributions around 1063 cm^{-1}, 1133 cm^{-1}, 1296 cm^{-1}, and 1441 cm^{-1}[14]. Peaks in these regions can obscure a multitude of tissue peaks, negatively impacting feature detection[15]. For IHC paraffin not only physically blocks epitopes on the tissue surface, preventing antibody binding, but it can also trap excess chromogens, giving a false positive or high background signal[16]. It is therefore essential to implement thorough and reproducible deparafination procedures during tissue processing in order to minimize artifacts caused by paraffin contribution during analysis.

RS has been applied to study tissue and cells for several decades and has proven that tissue, cell and disease-specific spectral signatures are attainable. Spectra are generated using Raman scatter from components of the genome, proteome and metabolome and are displayed as collections of overlapping peaks, which can be mined for the molecular information they contain. The ability to collect all of this data simultaneously, in a single spectrum, means that abnormalities can be detected rapidly through spectral differences. For clinical application, spectral changes must be correctly assigned to the disease under investigation and this is routinely accomplished through classification algorithms. However, one major drawback to this methodology is the current failure of spectroscopists to validate their algorithms by reference to proven and accepted detection methods and therefore adoption of RS by clinicians has proved problematic. To validate the capabilities of RS to detect pathological changes here we correlate spectral features with clinically validated antibody localization and show that RS has the potential to, at least, complement current IHC methods in the CRC pathology workflow. Novel studies using surface enhanced RS alongside conventional IHC methods are now being implemented. Deemed immuno-RS, the method employs surface-enhanced Raman scattering from an immuno-Raman label in order to selectively detect disease-specific antigens with high sensitivity. The protocol has been previously described for PSA detection in prostate tissue with promising initial results[17]. Additionally, Raman microscopy has more recently been compared against IHC data in comparing EGFR mutation status in lung adenocarcinoma[18]. The study correlated IHC-associated molecular changes with Raman spectral data and built an algorithm capable of diagnosing E19del and L858R mutations with an accuracy of 87.8%. RS could overcome a number of current limitations in IHC workflows as it is capable of providing simultaneous detection of multiple antigens and their spatial distributions in the same tissue specimen, removing the need for multiple biopsies and antibody panels. The technique is also label-free and does not require expensive antibodies and dyes to detect antigens like in conventional IHC, reducing the cost of analysis.

These studies have been made possible through advances in instrumentation and substrate development. Recent substrate developments have highlighted stainless steel as suitable, low-cost material for tissue analysis, substantially reducing the costs associated with Raman analysis of tissue[19]. Newly characterized substrates also exhibit improved signal to noise ratios over commonly used and expensive calcium fluoride, indicating that tissue spectra can be obtained at greater speeds [20]. Reduction in mapping costs and improvements in mapping speed now means that RS has reached the stage where it has the potential to be integrated into the pathology workflow. RS is objective and offers a very reliable diagnosis as it provides a biochemical fingerprint of the tissue under investigation and does not rely upon subjective assessment. Previous studies have combined RS with basic hematoxylin and eosin (H&E) staining and white light microscopy images [21-23], while this paper is the first of its kind to investigate the tissue morphological information that can be obtained in a direct comparison between

This article is protected by copyright. All rights reserved.

Accepted Article

RS and IHC. Results clearly show the co-localization of spectral signatures with IHC staining and demonstrate the multiplexing capabilities of RS.

2 Materials and Methods

2.1 Tissue Preparation

Formalin fixed paraffin embedded human colon tissue blocks were obtained from the UCL Biobank (REC 15/YH/0311). For each tissue block 8 µm tissue sections were prepared for RS on stainless steel slides, described previously by our group[19]. All samples for RS were deparaffinated with xylene with four, five-minute baths prior to rehydration and downstream analysis. Dewaxed samples were stored at room temperature and were Raman-mapped within 24 hours of rehydration.

For each sample prepared for RS contiguous 3 µm tissue sections were also cut for IHC and H&E staining. H&E staining was carried out using the appropriate standard protocol and IHC staining, deparaffination and antigen retrieval was performed on a Leica Bond III auto-immunostaining platform, using Leica Bond Polymer Refine Detection (Leica DS9800) with DAB chromogen. The fully-automated system was run under optimal conditions for each antibody, as determined by positive controls. Primary antibodies were applied for 15 minutes at room temperature. Ki-67 antibody (Dako mouse clone MIB-1, M7240) was used at a dilution of 1/120 and epitope retrieval performed in Leica ER2 solution for 20 minutes at 99°C. Desmin antibody (Leica mouse clone DE-R-11, PA0032) was used as per manufacturer's protocol and antigen retrieval performed in Leica ER2 solution for 20 minutes at 99°C. Anti-MUC2 (Leica mouse clone Ccp58, NCL-MUC-2) was applied at a dilution of 1/200 and epitopes retrieved in Leica ER2 solution for 30 minutes at 99°C. Collagen III antibody (GeneTex mouse clone FH-7A, GTX26310) was applied at a dilution of 1/500 and epitope retrieval was carried out in Leica ER1 solution for 30 minutes at 99°C. Finally, anti-CD20 (Dako mouse clone L26, M0755) staining was performed at a dilution of 1/100 and epitope retrieval performed in ER1 solution for 20 minutes, also at 99°C.

2.2 Raman Measurement

A Renishaw RA800 series benchtop Raman system configured for pathology use was used for all measurements. The instrument is configured with a 1500 lines per mm diffraction grating, 785 nm laser excitation wavelength, a 50x, NA 0.75 objective and a motorized XYZ stage. The system is capable of imaging in transmitted and reflected white light mode for sample identification on both transparent and opaque substrates. StreamLine™ Raman imaging, a fast mapping method, was used for spectral collection. The spectral resolution was 2.5 cm^{-1} and data was collected between 150-2100 cm^{-1}. Fully-automated spectral alignment and calibration routines were in operation to ensure data reproducibility. Spectral collection was controlled by WiRE 4.0 software (Renishaw PLC). Raman maps were collected at an acquisition time of 0.3 seconds per pixel at 1.05 µm step size. Spectra were collected with a laser power of 230 mW measured at the sample.

Pre-processing of raw spectra included cosmic ray removal and baseline correction using WiRE 4.0. Principal component analysis (PCA), also performed in WiRE 4.0 software, was then used for spectral data reduction and visualization. PCA was applied to the entire spectral region and spectra were mean centered and normalized by standard normal variate normalization. Spectral data was reduced to ten key components that explained the majority of the variance in the data. These components were used to plot pseudo-color images, which were created by applying a color weighting to the loadings of principal components from the spectral datasets. The resulting image displays the relative intensity of each component and its spatial distribution across the mapped region, and therefore in the sampled tissue.

3 Results and Discussion

Normal colon sections with easily distinguishable tissue morphology were selected for analysis in this

This article is protected by copyright. All rights reserved.

Accepted Article

proof of concept study (Figure 1). Areas covering longitudinal crypt sections were selected for Raman mapping in contiguous sections since these contained discernable tissue features that could be easily compared to Raman data in the initial stages. The outermost structure of the colon consists of the lumen, which forms the tubular space inside the colon. Next, the mucosal layers makeup the inner lining of the colonic tissue and consists of the epithelium, a layer of connective tissue called the *lamina propria* and a layer of thin muscle, the *muscularis mucosa*. The colonic crypts are found within the mucosal layer in normal, healthy colon and contain a base of stem cells and are also lined with mucus secreting cells, amongst other cell types. Tissue maps were generated using Raman spectroscopy and corresponding tissue sections stained with the antibody panel; anti-Ki-67, anti-desmin and anti-MUC2. Antibody stained tissue appears brown due to the DAB reaction and cell nuclei are counterstained blue/violet with hematoxylin, which binds to lysine residues localized to nuclear histones.

Raman spectroscopy provides a biological 'fingerprint' of the specimen under investigation between 450 – 2000 cm^{-1} and specific regions of the spectrum can be used to confirm the presence of various tissue components. Most notably, bond vibrations attributed to proteins at 1500 –1700 cm^{-1}, carbohydrates at 470 –1200 cm^{-1} and DNA at 980, 1080 and 1240 cm^{-1} are all present in the spectrum[24, 25]. Additionally, higher frequency modes due to lipid-associated bond vibrations can be found at 2700 – 3500 cm^{-1} [25].

PCA is an unsupervised, multivariate analysis technique that reduces the variance of the spectral dataset into a set of components that explain the variance within the data set. Individual components can be more easily visualised in space and represent a combination of loadings derived from a mixture of spectral features that best describe each component. The majority of the variance in the spectral dataset is explained by the first component, and the fraction explained then decreases sequentially with each subsequent component. For us the first three components explained the majority of the variance in the dataset and were deemed most histologically relevant.

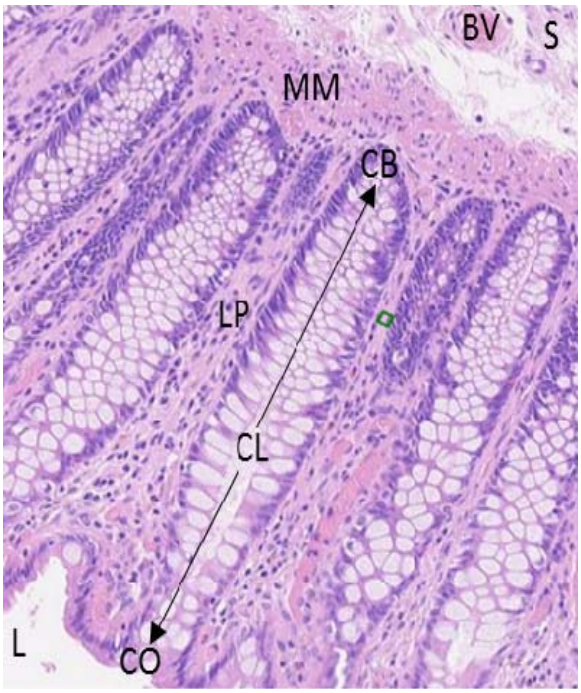

Figure 1 H&E stained longitudinal section of a normal colon tissue sample outlining major tissue structures. The colonic lumen (L) is visible in the bottom left and a whole colonic crypt, from crypt opening (CO) to crypt base (CB), through crypt lumen (CL), is identified between arrows. The lamina propria (LP), a layer of connective tissue surrounding the crypts can also be identified. The submucosa (S) is orientated to the top right corner of the image and connects the mucosa to the muscularis propria. The muscularis mucosa (MM) is a thin layer of muscle separating crypt bases and the submucosa. Blood vessel (BV).

Anti-desmin binds to desmin, a filament protein localized to muscle fibers that is responsible for connecting myofibrils to one another and to the plasma membrane. We observe intense anti-desmin staining localized to the *muscularis mucosae,* a thin layer of muscle that can be seen running along the crypt base. Staining of the plasma membrane on stromal cells, likely pericytes[26], outlining the boundaries of crypt structures can also be seen[27]. Fundamentally, the corresponding Raman map created using the PC1 loading shows good correlation to the staining density obtained with anti-desman IHC. Again, the *muscularis mucosae* has the highest Raman signal and signals outlining the plasma membranes of individual cells are also visible. The major spectral contributions used to generate the PC1 red pseudo-color image can be attributed to amino acid scatter, which have easily

This article is protected by copyright. All rights reserved.

Accepted Article

distinguishable spectra, even in complex biological samples[28]. Intense peaks corresponding to proline at 852 cm^{-1}, phenylalanine at 1004 cm^{-1} and the CH_2 bending mode of proteins at 1449 cm^{-1} all indicate that protein scatter is the major contributor to PC1[29].

Anti-Ki67 stains a large nuclear protein that is expressed during all active phases of the cell cycle and is an established marker of cell proliferation. We would therefore expect the distribution of this stain to be localized to cell nuclei and to be of highest intensity towards the crypt base as this is where the cells are documented to be proliferating most rapidly. Figures 2 and 3 verifies this observation; the most intense anti-Ki67 staining is localized to crypt bases, with minor staining also found along the epithelium where cells are also dividing. The pseudo-color Raman comparison (created from the PC2 loading in the Principal Component Analysis) arises from DNA/RNA signals. Peaks characteristic of DNA have previously been assigned at 782 cm^{-1}, 1100 cm^{-1}, 1335 cm^{-1}, 1573 cm^{-1} and 1662 cm^{-1} [29] and corresponding peaks in the loadings of PC2 could be found at 780 cm^{-1} and 1105 cm^{-1} in our spectra. Even though the major peak in the loading identified around 556 cm^{-1} could not be assigned to DNA, the IHC/RS comparison for the Ki67 marker shows good correlation with DNA since the greatest scattering intensity is also mapped to crypt bases, where cells are actively proliferating. The same image also shows nucleotide scatter along crypt epithelium where cells are also growing, although to a lesser extent than at the crypt base. Variation in cellular DNA content is associated with different phases of the cell cycle[30]. RS is capable of identifying nucleotide signals and a major component identified in the spectral dataset could be related to nucleotide scatter; it is this component that was used to generate the pseudo-color (green) image for comparison against the Ki67 stain.

MUC2 is the major mucin localized to the colon, more specifically to colonic crypts and the lumen. It is likely that some mucus remains trapped within the lumen of the crypt structures and within goblet cells following sample processing. Therefore, anti-MUC2 was used to check for the presence of the glycoprotein. It is useful to be able to identify the localization of mucins in different pathologies since abnormal expression of these highly-glycosylated proteins can signal abnormalities. As expected, the antibody intensely stained the lumen of colonic crypts and the mucin secretory cells that line them. Since mucins are highly decorated with oligosaccharides, the major spectral component used to explore the distribution of mucins was based on glycan scatter in the tissue spectra. The blue distribution in the pseudo-color RS image (PC3 loading) also indicates the presence of mucin in the secretory cells lining the crypt epithelium and in the crypt lumen. Alongside several protein peaks identified in the loadings of PC3, a peak at 440 cm^{-1} and a considerable scatter in the low wavenumber region is seen to dominate PC3. Raman scatter in the low wavenumber region (450 – 650 cm^{-1}) can be attributed to δ(C-O-C) found in glycosidic linkages of carbohydrates[31], which are found attached to mucins. Furthermore, since it is known that mucins are decorated with sulfate groups and that they form disulfide bond stabilized networks via cysteine resides, then S-S stretching modes arising at 498 cm^{-1}, 512 cm^{-1} and 529 cm^{-1}[32] are most likely contributing to the loadings on PC3.

This article is protected by copyright. All rights reserved.

Accepted Article

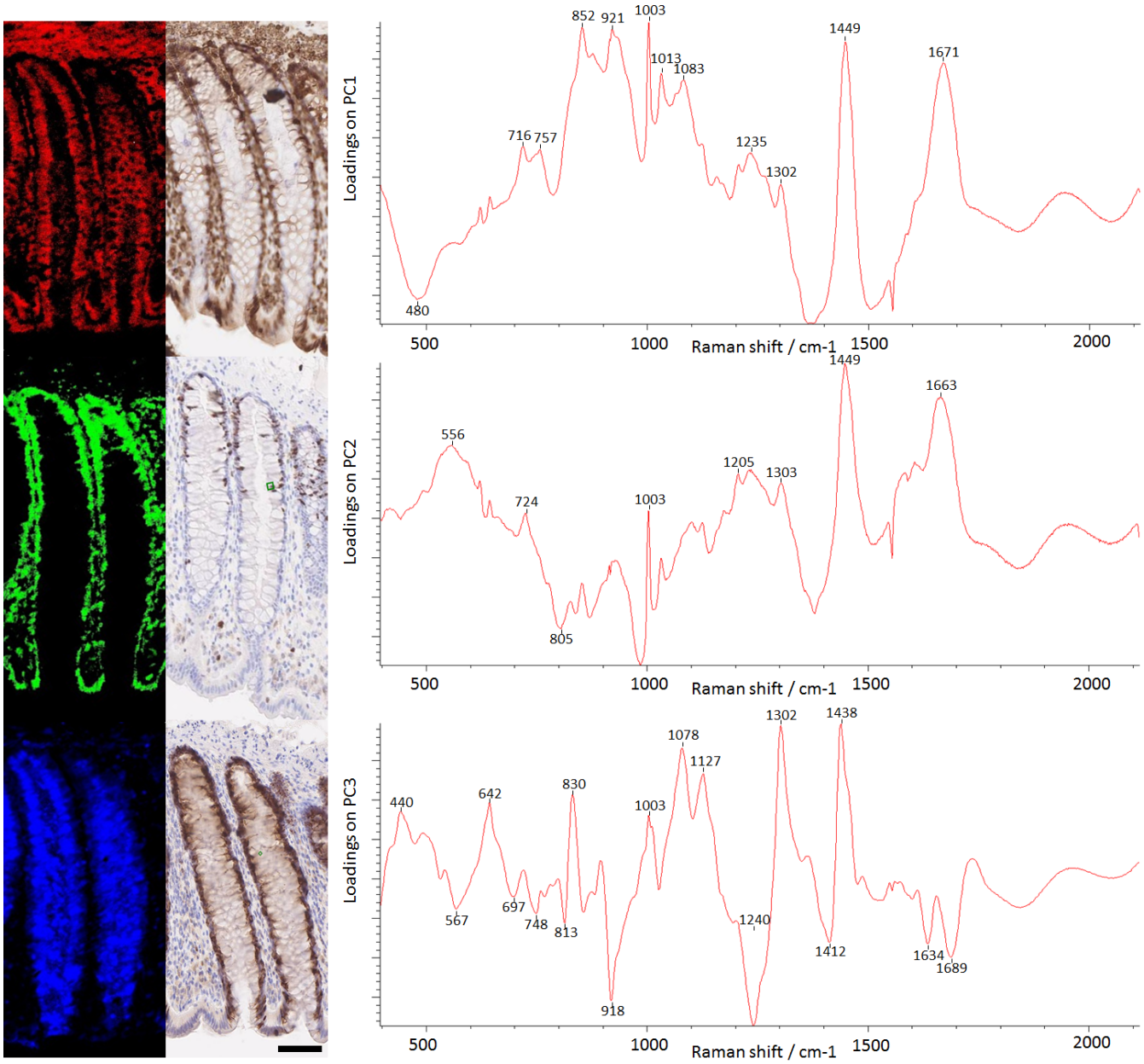

Figure 2 Alternating Raman maps (left) and IHC images (right) of colon section showing crypt structures. Red Raman map indicates the distribution of muscle fibers and is correlated to anti-desmin, a muscle-specific antibody. Green Raman map highlights the distribution of DNA and is correlated to anti-Ki67 staining for proliferating cells. The blue Raman map shows the distribution of mucin glycoproteins in colonic crypts and this correlates to anti-muc6 staining in the IHC image. Scale bar is 50 µM. Right hand panels display the loadings of the first three principal components identified by PCA with major peaks labelled. Percentage variance explained by individual components: PC1 = 40.0; PC2 = 21.5; PC3 = 5.4.

This article is protected by copyright. All rights reserved.

Accepted Article

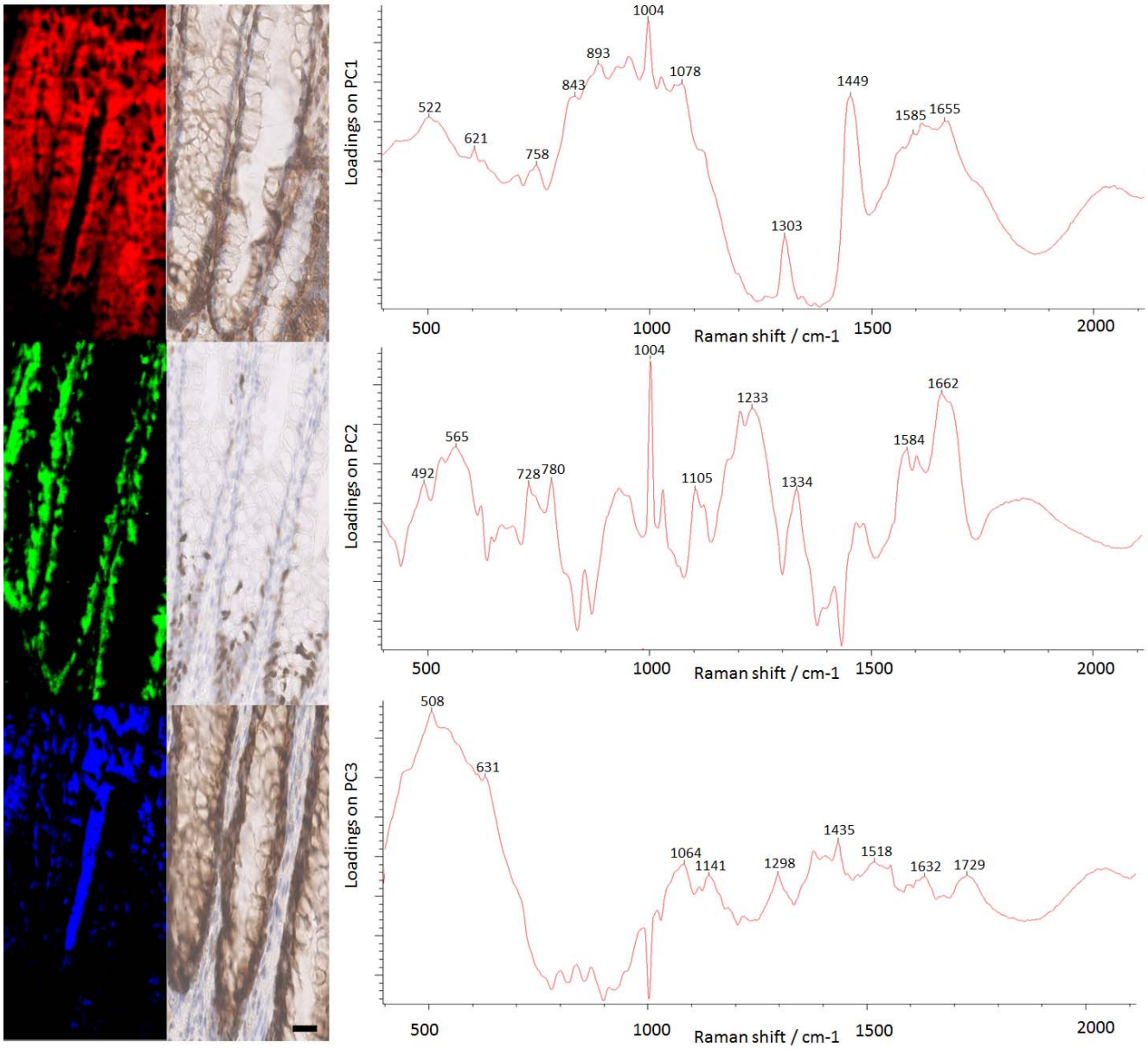

Figure 3 Alternating Raman maps (left) and IHC images (right) of colon section showing crypt structures. Red Raman map indicates the distribution of muscle fibers and is correlated to anti-desmin, a muscle-specific antibody. Green Raman map highlights the distribution of DNA and is correlated to anti-Ki67 staining for proliferating cells. The blue Raman map shows the distribution of mucin glycoproteins in colonic crypts and this correlates to anti-muc6 staining in the IHC image. Scale bar is 50 µM. Right hand panels display the loadings of the first three principal components identified by PCA with major peaks labelled. Percentage variance explained by individual components: PC1 = 61.7; PC2 = 7.3; PC3 = 3.6.

This article is protected by copyright. All rights reserved.

Accepted Article

There is good correlation between the spatial distributions of components in both tissue specimens studied. This was to be expected since the first set of components are most likely to map to protein and DNA, making these molecular species particularly easy to discriminate by PCA[33]. Although the distribution of protein and DNA are easily identified via this method, it is likely that other components that scatter less intensely will require more sophisticated multivariate analysis methods in order to identify their presence and distribution in relation to IHC markers. Additionally, although the first three components distinguish between protein, DNA and mucin, it should be noted that the loadings of components between tissue samples are not entirely comparable, which is particularly true for PC3. It is likely that spectral contamination due to fluorescence and paraffin retention are the cause of these inconsistencies in the loadings. Paraffin retention can contaminate Raman spectra across the whole biological region, masking peaks of interest and detrimentally contributing to PC loadings; particularly those associated with discriminating lipids[29, 33]. Additionally, mucin-related signals will be highly influenced by paraffin since we have observed that it is commonly retained in mucus and that the vigorous deparaffination required to remove trapped paraffin sometimes also removes the non-attached mucus layer. It is therefore important that future studies implement careful yet vigorous methods to remove paraffin from sections prior to analysis so that these signals do not interfere with the identification of low abundance species or alter the native lipid and mucin component of the sample.

To highlight the capabilities of Raman in multiplexing, we mapped the distributions of multiple components onto a single tissue map. The first five components from figure 2 were superimposed onto figure 4, with additional components that localized to lipids and vasculature (identified by characteristic lipid-associated Raman scatter attributed to C=C stretch at 1654 cm^{-1} [34] in spectral loadings). In the multiplexed, pseudo-color image, individually colored components clearly indicate the distributions of various morphological features and tissue and cellular subtypes and these can all be identified from spectra collected during a single experimental run. The distribution of muscle tissue, DNA, mucus glycoproteins and lipids can all be seen and there is an obvious greater wealth of information in the Raman

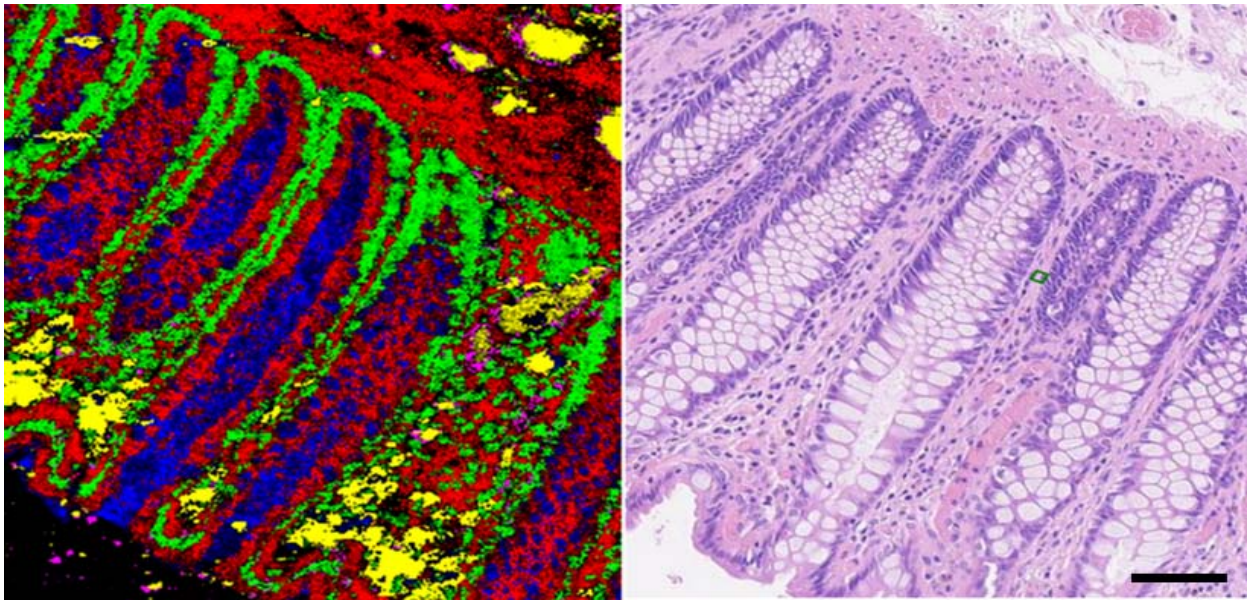

Figure 4 Raman false colour map (left) compared alongside a standard H&E stained section that reveals only cytoplasmic and nuclear distributions (right). Raman map shows several biochemical components on same section with each color representing the spatial distribution of a specific biochemical component, ranging from DNA (green) to muscle fibers (red), lipids (yellow) and mucin glycoproteins (blue) as detailed in Figure 2. Scale bar is 50 µM.

This article is protected by copyright. All rights reserved.

Accepted Article

map, at least comparable to that achievable with multi-antigen IHC compared to the corresponding conventional H&E section, which provides only basic morphological detail regarding the localization of cell nuclei and cytoplasmic regions. The colonic crypts can be easily visualized in the H&E stained section, but there is no indication of mucus presence and vasculature is more difficult to distinguish.

On the multiplexed image, as expected, PCA was able to distinguish musculature from vascular and connective tissue since muscle fibers are predominantly comprised of actin and myosin proteins. Both actin and myosin possess secondary structures rich in alpha helices, which give characteristic Raman scatter, and these are dissimilar to the spectral patterns obtained from collagen, which contains high concentrations of glycine in an ordered type-2 helix, hydroxyproline (an amino acid rare in other proteins) and a triple helical structure. The presence of blood vessels can be distinguished, albeit at low contrast, in the H&E stained section, however the yellow component attributed to lipids in the pseudo-color image cannot be seen. The co-localization of Ki-67 staining and DNA signal obtained using RS to cells lining the colonic crypts was also anticipated. Previous work has identified that proliferative cells have distinct Raman spectra based on the conformation of DNA during various phases of the cell cycle. Since the most active, proliferating cells are found adjacent to the crypt base (where the colonic stem cells reside) and lining the colonic epithelium, we would expect an overlap of signals in these tissue regions. All cell nuclei can be easily distinguished in the H&E image since DNA is densely stained violet with hematoxylin. There is good correlation between the well-established hematoxylin stain for DNA and the DNA Raman scatter, which indicates the reliability of the Raman signal; strong Raman and dye intensities can be seen in crypt bases and along the epithelium lining the crypts. Importantly all of these individual components have been identified in a single Raman map, using one tissue section, proving the multiplexing capabilities of RS and directly correlating them to the “ground truth” of known IHC and H&E markers. Routinely using RS in this manner would minimize the usage of expensive antibodies and reduce the number of tissue sections required for pathologists to reach a diagnosis, since multiple components could be identified in a single piece of tissue. Multiplexing would also save time and reduce the burden on healthcare staff. The non-destructive nature of RS also offers the possibility of returning the sample to the conventional downstream analysis pathway and investigation with conventional methods after the tissue has been mapped, which is not possible following IHC or H&E staining.

4 Conclusion

This proof-of-concept study has demonstrated that RS is capable of providing complimentary information to IHC for anatomical regions based on tissue-subtyping. The capabilities of RS to detect chemical signatures and to provide predictive models of disease using these signatures is already well documented. However, previous studies of this kind are rarely confirmed by secondary methods, whereas this study validates the use of RS in the detection of cell and tissue subtyping since the correct tissue signals could be mapped to clinically validated IHC markers. We show that mucins can be easily distinguished due to their rich glycosylation, nuclear regions by their DNA content and various other tissue types can be identified based on differences in amino acid content and structural conformation, as is the case with muscle fibers and collagen found in connective tissue. There is a vast and growing catalogue of non-validated RS tissue signatures already reported in the literature and, like our limited dataset, these should also be verified using complementary analytical methods, such as MALDI imaging and IHC. Without this link to the ground truth established by IHC or other molecular techniques the assignment of spectral signatures to discrete tissue and cellular subtypes through subjective inspection alone will remain speculative and loosely correlated. Future studies should also include spectra and corresponding markers from cell lines, in order to confirm the identification of cell type by RS.

This article is protected by copyright. All rights reserved.

Accepted Article

Acknowledgements

The authors would like to acknowledge Katherine Lau, Stephanie Ohrel and Ian Bell at Renishaw PLC for equipment and software support and the UCL Biobank for providing tissue for the study. We also thank UCL Advanced Diagnostics for IHC support. Samples of stainless steel slides can be obtained from GMHT on request. This work was funded by Innovate UK.

References

[1] K. Kong, C. Kendall, N. Stone, I. Notingher *Adv Drug Deliv Rev*. **2015**, 89, 121-134.
[2] K. Eberhardt, C. Stiebing, C. Matthaus, M. Schmitt, J. Popp *Expert Rev Mol Diagn*. **2015**, 15, 773-787.
[3] J. C. C. Day, R. Bennett, B. Smith, C. Kendall, J. Hutchings, G. M. Meaden, C. Born, S. Yu, N. Stone *Phys Med Biol*. **2009**, 54, 7077-7087.
[4] S. K. Teh, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, Z. Huang *Brit J Cancer*. **2008**, 98, 457-465.
[5] G. R. Lloyd, J. Wood, C. Kendall, T. Cook, N. Shepherd, N. Stone *Vib Spectrosc*. **2012**, 60, 43-49.
[6] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal *CA Cancer J Clin*. **2015**, 65, 87-108.
[7] K. B. Greer, G. S. Cooper *Gastrointest Endosc*. **2012**, 76, 365-366.
[8] M. P. Legolvan, R. J. Taliano, M. B. Resnick *Hum Pathol*. **2012**, 43, 1157-1168.
[9] H. J. Kim, M. H. Yu, H. Kim, J. Byun, C. Lee *BMB Rep*. **2008**, 41, 685-692.
[10] M. J. Duffy *Clinical Chemistry*. **2001**, 47, 624-630.
[11] R. Shimono, M. Mori, K. Akazawa, Y. Adachi, K. Sgimachi *Am J Gastroenterol*. **1994**, 89, 101-105.
[12] V. N. Papadopoulos, A. Michalopoulos, S. Netta, G. Basdanis, D. Paramythiotis, A. Zatagias, P. Berovalis, N. Harlaftis *Tech Coloproctol*. **2004**, 8 Suppl 1, s123-125.
[13] M. Werner, A. Chott, A. Fabiano, H. Battifora *Am J Surg Pathol*. **2000**, 24, 1016-1019.
[14] E. O. Faolain, M. B. Hunter, J. M. Byrne, P. Kelehan, H. A. Lambkin, H. J. Byrne, F. M. Lyng *J Histochem Cytochem*. **2005**, 53, 121-129.
[15] P. Meksiarun, M. Ishigaki, V. A. Huck-Pezzei, C. W. Huck, K. Wongravee, H. Sato, Y. Ozaki *Sci Rep*. **2017**, 7, 44890.
[16] J. H. Williams, B. L. Mepham, D. H. Wright *J Clin Pathol*. **1997**, 50, 422-428.
[17] S. Schlücker, B. Küstner, A. Punge, R. Bonfig, A. Marx, P. Ströbel *Journal of Raman Spectroscopy*. **2006**, 37, 719-721.
[18] L. Wang, Z. Zhang, L. Huang, W. Li, Q. Lu, M. Wen, T. Guo, J. Fan, X. Wang, X. Zhang, J. Fang, X. Yan, Y. Ni, X. Li *Analyst*. **2014**, 139, 455-463.
[19] A. T. Lewis, R. Gaifulina, M. Isabelle, J. Dorney, M. L. Woods, G. R. Lloyd, K. Lau, M. Rodriguez-Justo, C. Kendall, N. Stone, G. M. Thomas *Journal of Raman Spectroscopy*. **2017**, 48, 119-125.
[20] L. T. Kerr, H. J. Byrne, B. M. Hennelly *Anal Methods-Uk*. **2015**, 7, 5041-5052.
[21] K. Kong, C. J. Rowlands, S. Varma, W. Perkins, I. H. Leach, A. A. Koloydenko, H. C. Williams, I. Notingher *Proc Natl Acad Sci U S A*. **2013**, 110, 15189-15194.
[22] F. L. Cals, T. C. Bakker Schut, J. A. Hardillo, R. J. Baatenburg de Jong, S. Koljenovic, G. J. Puppels *Lab Invest*. **2015**, 95, 1186-1196.
[23] O. Old, M. Isabelle, G. Lloyd, C. Kendall, H. Barr, N. Stone *Gut*. **2015**, 64, A485-A486.
[24] G. Clemens, J. R. Hands, K. M. Dorling, M. J. Baker *Analyst*. **2014**, 139, 4411-4444.
[25] Z. Movasaghi, S. Rehman, I. U. Rehman *Appl Spectrosc Rev*. **2007**, 42, 493-541.
[26] D. Verhoeven, N. Buyssens *Virchows Arch B Cell Pathol Incl Mol Pathol*. **1988**, 54, 263-272.
[27] A. Armulik, A. Abramsson, C. Betsholtz *Circ Res*. **2005**, 97, 512-523.
[28] G. Zhu, X. Zhu, Q. Fan, X. Wan *Spectrochim Acta A Mol Biomol Spectrosc*. **2011**, 78, 1187-1195.
[29] R. Gaifulina, A. T. Maher, C. Kendall, J. Nelson, M. Rodriguez-Justo, K. Lau, G. M. Thomas *Int J Exp Pathol*. **2016**, 97, 337-350.
[30] Z. Darzynkiewicz, H. D. Halicka, H. Zhao *Advances in experimental medicine and biology*. **2010**, 676, 137-147.
[31] H. Noothalapati, T. Sasaki, T. Kaino, M. Kawamukai, M. Ando, H. Hamaguchi, T. Yamamoto *Sci Rep-Uk*. **2016**, 6.
[32] P. Bazylewski, R. Divigalpitiya, G. Fanchini *Rsc Adv*. **2017**, 7, 2964-2970.
[33] P. Meksiarun, M. Ishigaki, V. A. C. Huck-Pezzei, C. W. Huck, K. Wongravee, H. Sato, Y. Ozaki *Sci Rep-Uk*. **2017**, 7.
[34] H. Wu, J. V. Volponi, A. E. Oliver, A. N. Parikh, B. A. Simmons, S. Singh *Proc Natl Acad Sci U S A*. **2011**, 108, 3809-3814.

This article is protected by copyright. All rights reserved.