Introduction

The latest generation of Fourier transform infrared (FT-IR) spectrometers, that incorporate large liquid nitrogen cooled focal plane array (FPA) detectors within an infrared microscope system, have no doubt accelerated the development of the field of spectral histopathology. Significant improvements in data acquisition, processing, and classification times, in part due to the increased field of view (FOV) of these FPA based devices, have enabled spectroscopic investigations that now include clinically relevant patient populations. Nevertheless, despite these marked improvements when compared to linear detector array (LDA) or point detector based systems, data collection times from tissue micro-array (TMA) cores or whole tissue sections are still in the order of hours or days. All FT-IR based microscope systems are inevitably limited by a tradeoff between spatial resolution, signal-to-noise ratio, field of view and acquisition time. Traditionally with a conventional benchtop FT-IR based system, to achieve true diffraction-limited resolution, high magnification objectives with the largest possible numerical aperture (N.A.), or use of upstream optics to increase the overall system magnification are required that provide a much-reduced field of view (ca. 150 × 150 µm) and extended sample dwell times to achieve an acceptable signal-to-noise ratio for tissue classification purposes. More recently, however, the application of high resolution imaging for clinical diagnostics has shown clear advantages for the identification of small tissue structures that are essential for detecting early forms of disease. Identification of very small changes in the biochemical components of glandular, endothelial and myoepithelial cells, or intra-lobular stroma could be key for early pre-malignant changes to be identified and tracked.

The recent development of the Spero® microscope (Daylight Solutions Inc., San Diego, USA), a laser-based mid-infrared microscope, provides the capability to perform diffraction-limited imaging across the molecular fingerprint region (900–1800 cm\(^{-1}\)), at high N.A. (0.7) and with a much enhanced FOV (650 × 650 µm). This equates to a ~20× enhancement in the FOV, making high resolution imaging of tissue samples more viable in practical timeframes. Currently, the two major bottlenecks predominant in IR chemical imaging are long data collection times and spectral hypercube file size. The Spero QCL infrared microscope can address both issues by harnessing the power and flexibility of the tunable laser source. If diagnostic spectral features are located at a reduced number of frequencies across the full spectrum, which is often the case, the laser can be programmed to target these frequencies alone, reducing acquisition time and data file size proportionately.

This contribution presents data acquired from biopsied colorectal tissue specimens as an example case study of this new capability. In addition to the acquisition of full band infrared spectra, sparse frequency datasets were recorded from...
10 infrared bands previously identified as salient for tissue segmentation in the literature. This was conducted to create rapid chemical maps. Such images do not provide diagnostic contrast (i.e. the segmentation of cancerous cells), but could serve as methods for more rapid global screening of samples. As in traditional histology, identification of abnormal tissue architecture could potentially be visualized using these more rudimentary chemical images, prior to moving on to more rigorous analyses, including full-band infrared analysis, or by using new laser microdissection techniques for downstream proteomic and genomic testing of targeted regions.

Materials and methods

Sample preparation

Colorectal tissue sections were cut from archival tissue banks of formalin fixed paraffin embedded (FFPE) tissue blocks (SDPathology, San Diego, USA). Two parallel sections were cut from each sample for analysis. The first section was cut at 10 µm thickness and floated onto a CaF₂ disc (25 mm × 2 mm) for infrared microscopy. The second and adjacent section was cut at 5 µm thickness and floated onto a regular glass slide for brightfield visible microscopy. Both sections were subsequently de-paraffinised using standard protocols. The section mounted onto glass was further H&E stained and cover slipped before conventional brightfield imaging at 20× magnification using a whole slide imaging system (Hamamatsu, Japan). The section mounted onto CaF₂ disc remained stain-free and stored in a desiccator prior to infrared microscopy.

QCL-based infrared microscopy

QCL-based infrared microscopy was performed using a Spero microscope (Daylight Solutions Inc., San Diego, USA). Mosaic images were recorded from the tissue samples using the 12.5×, 0.7 N.A., compound refractive infrared objective of the microscope. This optic has a field of view 650 µm × 650 µm (480 × 480 pixels), and sample referenced pixel size of 1.35 µm.

Mercury cadmium telluride (MCT) focal plane arrays (FPAs) have been the choice for FT-IR based infrared imaging and have been employed in some QCL microscope prototypes. These cameras can operate over 2–12 µm when cryogenically cooled with frame rates exceeding 1 kHz but suffer from poor linear dynamic range (ca. 100 : 1) and long-term reliability. Room temperature operated microbolometer cameras based on VO₂ or a-Si have intrinsic responsivity over the entire mid-IR band and can exhibit excellent linear dynamic ranges (>1000 : 1). However, commercially available microbolometers are spectrally limited to 8–14 µm in order to suppress water vapor noise in thermal imaging applications. The Spero microscope uses a custom uncooled 480 × 480 microbolometer VO₂ FPA having a 30 Hz frame rate and excellent responsivity over the 5–14 µm molecular fingerprint band.

Currently, a full data cube containing 226 spectral images takes 5 minutes or about 9000 individual camera frames. Future improvements in instrumentation could allow spectral data collections rates to increase to about one spectral image per 1–2 camera frames collected.

As all molecules respond to the exciting infrared radiation and produce relatively complicated infrared spectra, the response observed for a single cell or an area of tissue is a complex superposition of all spectral features of all biomolecules in the sample. Although infrared spectroscopy is usually referred to as a “fingerprint” spectroscopic technique, which implies that every molecule known exhibits a distinct spectrum that identifies it, the superposition of such fingerprints leads to relatively broad spectral features that need to be decoded, or de-convolved, to enable an interpretation or diagnosis. Nevertheless, it is typical that small but reproducible changes in the spectral features can be identified between cell/tissue types. These changes can be identified by use of unsupervised methods of multivariate image analysis or mined and ranked using annotated spectral libraries and feature selection algorithms based on statistical measures such as the t-test. In general, a smaller number of salient spectral features can be used to train robust supervised algorithms that provide optimal classification.

In this study, all infrared images were recorded in duplicate from the full fingerprint region between 900–1800 cm⁻¹, and from a reduced set of 10 sparsely distributed salient frequencies within the same region, located at 940, 1044, 1180, 1300, 1336, 1360, 1480, 1544, 1656 and 1800 cm⁻¹ respectively. These frequencies were selected after consultation of several rigorous FT-IR based studies performed by 3 independent research groups and are further summarized in Table 1. Each recorded spectral hypercube took 5 minutes for a full broadband spectrum and 0.9 minutes for a sparse frequency data collection. The data was recorded at 8 cm⁻¹ spectral resolution with a 4 cm⁻¹ data interval, and ratio’d against a background from a clean area of the substrate. Spectral hypercubes 480 × 480 × 256 (102 MB) or 480 × 480 × 10 (5 MB) were thus created.

Data pre-processing

Standard pre-processing techniques for mid-infrared spectroscopy were used. These have been reported at great length previously. Raw data sets were imported into the multisensor imaging software package ImageLab. The data processing included:

- Noise reduction via the maximum noise fraction transform.
- Spectral quality test to remove pixel spectra from regions not occupied by tissue (pixels describing an amide I intensity at 1656 cm⁻¹ <0.05 au were rejected).
- Full band spectra were converted to second derivatives (Savitzky-Golay algorithm, 9-point smooth) to provide better resolved peaks and eliminate background slopes.
- All spectral vectors were scaled to a zero mean and a standard deviation of 1.0 (standard normal variate normalisation) to reduce the influence of intensity changes caused by differences in cellular density and tissue thickness.
Table 1  List of spectral descriptors useful to differentiate different colon tissue classes

<table>
<thead>
<tr>
<th>Spectral descriptor type</th>
<th>Peak 1 band position (cm⁻¹)</th>
<th>Vibrational mode/assignment</th>
<th>Peak 2 band position (cm⁻¹)</th>
<th>Molecular assignment</th>
<th>Baseline positions (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>1656</td>
<td>Combination of:</td>
<td>1544</td>
<td>Combination of:</td>
<td>1480, 1800</td>
</tr>
<tr>
<td>Peak height ratio</td>
<td></td>
<td>C=O stretching (mainly)</td>
<td></td>
<td>N-H bending</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-H bending</td>
<td></td>
<td>C-N stretching</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Amide I of proteins</td>
<td></td>
<td>Amide II of proteins</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1656 cm⁻¹ position is correlated with α-helix structures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ratio highlights gross protein compositions changes i.e. changes in secondary structure/amino acids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collagen</td>
<td>1336</td>
<td>CH_2 wagging</td>
<td></td>
<td></td>
<td>1300, 1360</td>
</tr>
<tr>
<td>Peak height</td>
<td></td>
<td>Amino acid proline side chains</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucin</td>
<td>1044</td>
<td>Abundant in connective tissues</td>
<td></td>
<td></td>
<td>940, 1180</td>
</tr>
<tr>
<td>Peak height</td>
<td></td>
<td>COH deformation</td>
<td></td>
<td>Glycosylated protein mucin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abundant in mucus membranes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data analysis

Sparse frequency absorbance datasets were used to create chemical images as follows:

- Protein map: peak height ratio of the amide I (1656 cm⁻¹)/amide II (1544 cm⁻¹) bands. The ratio was linear baseline corrected using intensity values recorded at 1480 and 1800 cm⁻¹. The ratio image produced was mapped against a white–blue (min–max) colour palette.

- Collagen map: peak height of collagen CH2 side chain vibrations (1336 cm⁻¹). The measurement was linear baseline corrected using intensity values recorded at 1300 and 1360 cm⁻¹. The intensity image produced was mapped against a white–red (min–max) colour palette.

- Mucin map: peak height of the strongest mucin glycosylation band (1044 cm⁻¹). The measurement was linear baseline corrected using intensity values recorded at 1180 and 940 cm⁻¹. The intensity image produced was mapped against a white–green (min–max) colour palette.

- RGB image: each chemical map mentioned above was merged into a single red, green, blue colour image.

These chemical maps or stains were generated in an effort to describe the potential of sparse frequency collection protocols to provide meaningful and easily interpretable chemical images for rapid inspection/screening of samples.

Full band (900–1800 cm⁻¹) second derivative datasets were analysed using the unsupervised algorithm, k-means clustering, a non-hierarchical iterative method that obtains “hard” class membership for each spectrum. A total of 6 clusters was chosen for analyses. False colour images were constructed, where a pixel’s cluster membership is defined by a corresponding colour in the image. Pixels rejected by the quality test are provided a black colour. These datasets were collected to provide further evidence that QCL-based infrared spectral imaging data can produce high fidelity images and spectra sufficient for rigorous multivariate analyses and allay concerns from early reports of potentially problematic coherence effects.25

Results and discussion

Due to the longer wavelengths of mid-infrared light in the biological fingerprint region (λ = 5.5–11 μm) the lateral resolution of IR based microscopes is reduced compared to visible light microscopes by an order of magnitude. However, this inherent drawback is offset by the rich chemical information that can be visualised by this non-destructive, label free technique. The performance of the Spero microscope’s 12.5× optic is demonstrated in Fig. 1(a)–(d), whereby the contrast of a USAF 1951 resolution test pattern was used to quantify spatial resolution in reflection geometry. Fig. 1(a) displays a reflectance image of groups 6 and 7 of the target recorded at 1555 cm⁻¹ (6.4 μm). Fig. 1(b) displays a line profile plot across group 6 elements that has patterns of spatial frequencies down to 4.38 μm (widths of each group of elements are displayed in red). Spatial resolution is governed by the diffraction limit. A microscope can be considered diffraction limited if able to distinguish objects separated by 0.61λ/N.A. The line plot clearly indicates that features of 5 microns can be resolved, meeting the Rayleigh criterion of 26% image contrast. Fig. 1(c) shows an image captured by the microscope from a microtome tissue section. The greyscale IR chemical image was acquired from a healthy colorectal tissue section (3 × 3 image tile mosaic) using the amide 1 (1656 cm⁻¹) vibration of proteins. Fig. 1(d) displays a brightfield image captured from the parallel H&E stained section. By direct comparison of the images it is clear the small diagnostic features of the colorectal tissue can be adequately resolved with this optic, including microstructures such as blood vessel walls and the colonocytes that surround the goblet cells of the crypts (a reference is provided by the coloured arrows in the stained image).

Fig. 2(a)–(c) display chemical images from the same tissue region shown in Fig. 1. These were calculated using the sparse frequency absorbance datasets that comprise 10 salient IR frequencies (45MB in size, ca. 20× smaller than a full band...
dataset of a mosaic the same size). Fig. 2(a) shows a chemical image based on the gross protein composition of the tissue and was calculated by ratio of the two most prominent bands of proteins, the amide I and amide II vibrations. After normalisation of the spectral vectors, a ratio of the protein bands yields greater contrast between tissue structures than using a single peak height alone. The colonocytes and blood cells are provided the greatest contrast in this chemical map. Fig. 2(b) shows a chemical image based on the presence of collagen and was calculated using the CH$_2$ side chain vibrations of collagen. The submucosa, being comprised mostly of collagen, has the highest contrast when using this chemical map. The greatest contrast for the goblet cells and lumen of the crypts was achieved using this chemical map, which again makes biochemical sense since they both comprise high concentrations of mucos- and glyco-proteins. The final image in Fig. 2(c) shows a chemical image based on the presence of mucin and was calculated using a characteristic glycosylation absorbance band in the low wavenumber part of the IR spectrum. The greatest contrast for the goblet cells and lumen of the crypts was achieved using this chemical map, which again makes biochemical sense since they both comprise high concentrations of muco- and glyco-proteins. The final image in Fig. 2(d) is an RGB fused image of the chemical maps shown in Fig. 2(a)–(c) respectively. By direct comparison to the bright-field image captured from the parallel H&E stained section in Fig. 1(d), the RGB image provides clear contrast of the major colorectal tissue structures, including the mucosa, submucosa and muscularis externa. Such chemical based images could help identify discrete regions within large areas of tissue where the mucosa no longer provides normal architecture, i.e. the appearance of clusters of enlarged crypts (aberrant crypts) and other structural hallmarks of epithelial dysplasia and neoplasia. The combination of these RGB type chemical images with morphometric descriptors used in digital pathology analysis may be worth consideration. In the example presented, colour channels were selected with colon tissue in mind, however these can be adjusted to suit the application. For example, lipid associated diseases (e.g. glioblastoma multiforme) could be visualised using the lipid ester band at 1740 cm$^{-1}$, whereas disease associated changes in cellular metabolism (e.g. hypoxia driven changes) could be tracked using one of the characteristic glycogen peaks at 1030, 1080 or 1152 cm$^{-1}$ respectively. Such simplified chemical represen-
mean cluster spectra. This is the first report, to the authors
section. Fig. 3(b) displays the
the brightfield image captured from the parallel H&E stained
mosaic composed 1.65 million pixel spectra. Fig. 3(a) displays
same mosaic size). After pre-processing and quality check the
in size,
diseased colorectal tissue section (3 × 3 mosaic image, 918 MB
lysis of a full band second derivative dataset acquired from a
(cols can be su
still required to ascertain whether frequency targeting proto-
investigations into clinically relevant patient populations is
been pondered whether all the frequencies that are routinely
collected in a broadband FT-IR measurement are required for
infrared spectral staining/diagnostics.1,4,26,27 Subsequent proof
concept studies performed using QCL microscopy techno-
ality have shown great promise.13–15 However, more exhaustive
investigations into clinically relevant patient populations is
still required to ascertain whether frequency targeting proto-
can be sufficiently robust.

Fig. 3 displays results obtained from the multivariate ana-
ysis of a full band second derivative dataset acquired from a
diseased colorectal tissue section (3 × 3 mosaic image, 918 MB
in size, ca. 20× larger than a sparse 10 frequency dataset of the
same mosaic size). After pre-processing and quality check the
mosaic composed 1.65 million pixel spectra. Fig. 3(a) displays
the brightfield image captured from the parallel H&E stained
section. Fig. 3(b) displays the k-means clustering image (6 clus-
ters) calculated from the derivative infrared dataset. Fig. 3(c)
shows the absorbance spectrum calculated for the k-means cluster analysis. The absorbance data was baseline corrected
using an asymmetric least squares approach (Eilers algorithm,
5 iterations),28 and normalised for clarity before calculation of
mean cluster spectra. This is the first report, to the authors
knowledge, of rigorous analysis performed on full fingerprint
region second derivative spectra recorded using a QCL infrared
microscope. It has been reported that scattering effects, ampli-
ified by the coherent light source, can cause unwanted arte-
facts.23 Unsupervised methods of multivariate analyses are
very sensitive to these types of structures in the spectral and
spatial domain but were not observed as problematic in these
tissue studies. By observation of the mean cluster spectra, it
appears segmentation of the different structures within the
colonics are predominantly caused by intensity differ-
ences at several characteristic glycosylation bands located at
1044, 1076, 1120 and 1374 cm⁻¹ associated with mucin. A
strong lipid ester band at 1740 cm⁻¹ is also observed for cell
types of the crypts. The submucosa, in contrast, provides a
number of strong bands that can be directly attributed to the
structural protein collagen with bands located at 1204, 1236,
1280, 1336 (amide III), and 1452 cm⁻¹ respectively. Spectroscopic differences between the lamina propria and
adenocarcinoma are far more subtle, and are located at
nucleic acid-related vibrations at ca. 964, 1062, 1090 and
1236 cm⁻¹ respectively, with again very small differences in the
amide I and amide II band profiles.

Conclusions

The results presented in this contribution lend further
evidence to the great promise of QCL infrared microscopy. The
technology can provide high resolution, high throughput IR
chemical images useful for infrared based spectral pathology.
The efficiency of frequency targeting protocols will be subject
to continued debate within the spectroscopic community in
the coming years, especially when considering confounding
band lineshape distortions that can be caused by scattering,
which may necessitate collection of full band spectra for
robust correction. However, such concerns shall likely
dissipate as QCL imaging technology achieves even faster data
acquisition speeds. With a myriad of possible diagnostic
applications, it is reasonable to expect that for some cases,
sparse frequency protocols can realise higher throughput
without sacrificing accuracy. Proof of concept studies
performed on breast tissue microarrays13 and blood sera14
have shown great promise. However, further work needs to be
performed on significantly larger patient numbers to fully
understand and realise the potential of the technology.

Acknowledgements

The authors would like to acknowledge Dr Tim Robbins and
Dr Shirley Reidy at SDPathology for sourcing of archival tissue
blocks and pathology consult. We would in addition like to
thank Dr Hans Lohninger and Elisabeth Renner at Epina for
many fruitful conversations and support. We would also like to
acknowledge Dr Edeline Fotheringham for building and
calibrating the Spero microscope used in this work.
Notes and references


