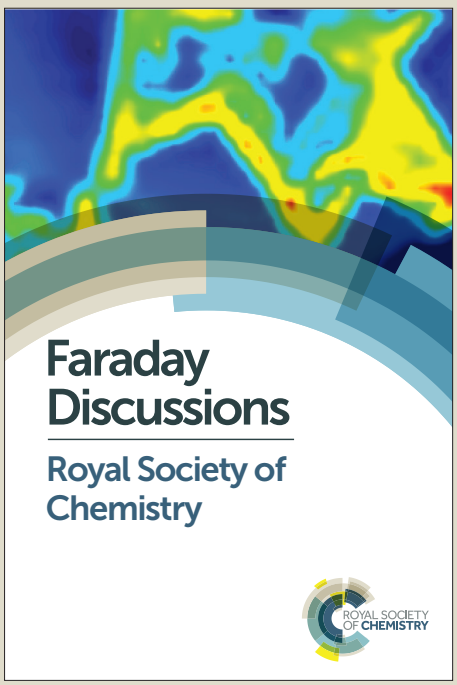
Faraday Discussions

Accepted Manuscript



This article can be cited before page numbers have been issued, to do this please use: M. Isabelle, J. Dorney, A. Lewis, G. R. Lloyd, O. Old, N. Shepherd, M. Rodriguez-Justo, H. Barr, K. Lau, I. Bell, S. Ohrel, G. Thomas, N. Stone and C. Kendall, *Faraday Discuss.*, 2015, DOI: 10.1039/C5FD00183H.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

View Article Online
DOI: 10.1039/C5FD00183H

Faraday Discussions

ARTICLE

Multi-Centre Raman Spectral Mapping of Oesophageal Cancer Tissues: a study to assess system transferability

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

M. Isabelle[a], J. Dorney[b], A. Lewis[c], G.R. Lloyd[a], O.Old[a], N.Shepherd[a], M. Rodriguez-Justo[c], H. Barr[a], K. Lau[d], I. Bell[d], S. Ohrel[d], G. Thomas[c], N. Stone[b], and C. Kendall[a]

The potential for Raman spectroscopy to provide early and improved diagnosis on a wide range of tissue and biopsy samples in situ is well documented. The standard histopathology diagnostic methods of reviewing H&E and/or immunohistochemical (IHC) stained tissue sections provides valuable clinical information, but requires both logistics (review, analysis and interpretation by an expert) and costly processing and reagents. Vibrational spectroscopy offers a complimentary diagnostic tool providing specific and multiplexed information relating to molecular structure and composition, but is not yet used to a significant extent in a clinical setting. One of the challenges for clinical implementation is that each Raman spectrometer system will have different characteristics and therefore spectra are not readily compatible between systems. This is essential for clinical implementation where classification models are used to compare measured biochemical or tissue spectra against a library training dataset.

In this study, we demonstrate the development and validation of a classification model to discriminate between adenocarcinoma (AC) and non-cancerous intraepithelial metaplasia (IM) oesophageal tissue samples, measured on three different Raman instruments across three different locations. Spectra were corrected using system transfer spectral correction algorithms including wavenumber shift (offset) correction, instrument response correction and baseline removal. The results from this study indicate that the combined correction methods do minimize the instrument and sample quality variations within and between the instrument sites. However more tissue samples of varying pathology states and greater tissue area coverage (per sample) are needed to properly assess the ability of Raman spectroscopy and system transferability algorithms over multiple instrument sites.

Introduction

Oesophageal cancer is the fifth biggest cancer killer, with rising incidence rates[1]. Once patients are found to have symptoms such as trouble swallowing the disease is advanced and usually incurable (15% of patients survive 5 years[1]). Individuals with a precursor condition called Barrett's oesophagus (caused by reflux disease), have an increased risk of developing cancer[2,3]. These patients can be diagnosed at an early stage using endoscopic surveillance, which includes biopsy. Most gastrointestinal cancers occur on the mucosal

[a] *Biophotonics Research Unit and Pathology Department, Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK.*
[b.] *University of Exeter, Exeter, UK*
[c.] *University College London, London, UK*
[d.] *Renishaw Plc*
† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

ARTICLE
View Article Online
Faraday Discussions
DOI: 10.1039/C5FD00183H

surface; however dysplastic changes in the oesophagus are invisible to the naked eye during endoscopy[4]. Therefore patients with Barrett's oesophagus are subjected to routine endoscopy with collection of biopsy samples for analysis by histopathology. However the gold standard histopathology is not able to accurately and reproducibly identify those patients with early disease or those who may be harbouring more advanced disease in un-sampled areas of the oesophagus[5]; therefore improved and objective diagnostic tools are needed[6].

Raman spectroscopy (RS) makes use of biochemical or molecular analysis of target tissues rather than an individual's assessment of cellular and tissue appearance[7], therefore RS provides an objective, non-destructive, label free, non-invasive, rapid and economical method for diagnosis. RS has the potential to be more automated and adaptable to routine screening and provides a spectroscopic signature of all molecular constituents of the tissue sample, which facilitates the analysis and use of subtle molecular changes to classify different pathology and tissue states. Due to these pathology and tissue-specific spectroscopic signatures, RS has been used as a tool in histopathology[8–14], and cytology[15–18]. It is based on inelastic light scattering and occurs when the biological sample is illuminated by monochromatic laser light. The interactions between the incident photons and molecules in the sample result in scattering of the light to produce a characteristic biomolecular fingerprint of the sample[19].

Kendall et al demonstrated the ability of RS to discriminate between 8 pathological groups in oesophageal cancer: normal squamous, three subtypes of Barrett's oesophagus (intestinal metaplasia, cardiac and fundic type mucosa), high-grade dysplasia, adenocarcinoma, squamous dysplasia and squamous cell carcinoma with sensitivities between 73% and 100% and specificities of 90-100%[8]. Hutchings et al and Shetty et al have demonstrated the advanced use of chemometrics (Principal Component Analysis score maps) to generate biochemical maps. These maps were used to identify cellular constituents associated with malignancy [20], and to discriminate between different pathology and tissue types[21].

Development of a miniature confocal fibre optic probe for Raman spectroscopic measurements (Day et al 2009[22]) enabled the use of fibre optic Raman probe spectral measurements ex-vivo in which Almond et al (Almond et al 2012[23]) performed Raman spectroscopic measurements on tissue samples from 28 patients using a custom-built fibre-optic Raman probe, in conjunction with multivariate classification models, to differentiate between benign and neoplastic oesophageal cancer and pre-cancer. The Raman probe system was able to differentiate between normal squamous, Barrett's oesophagus and neoplasia with sensitivities of (83% to 86%) and specificities of (89% to 99%). In another study by the same group (Almond et al 2014[24]), using tissue from 62 patients, the authors demonstrated that the Raman probe system, in conjunction Principal component fed linear discriminant analysis (PC-LDA), was able to achieve a sensitivity of 86% and a specificity of 88% for detecting high-grade dysplasia and adenocarcinoma. An additional study using a Raman fibre optic probe was conducted by Bergholt and co-workers who used, in clinic during endoscopic examination, multimodal image-guided Raman endoscopy technique for real-time in vivo diagnosis of cancer in the oesophagus.[25] 75 oesophageal tissue sites from 27 patients were measured and using a LDA diagnostic model was able to achieve an accuracy of 96%, sensitivity of 97% and specificity of 95% in the in vivo diagnosis of oesophageal cancer.

There is clearly a clinical need for these types of validated multivariate disease-specific classification models, developed to classify and discriminate pathology classes. However due to instrument artefacts and environmental differences between locations (mentioned below), a successful classification model constructed using spectra from one instrument system may not have the ability to accurately predict spectra measured on another second instrument.

Physical and chemical constitution (e.g. viscosity, particle size, and surface texture) of the tissue samples and standards can impact this transfer process. The instrument response function can affect signal intensity values (these differences/changes could be due to part replacement, aging of parts and sources, and gradual change in quantum efficiency of the detector over time) and wavenumber registration shifts. Ideally these factors can be prevented by pre-calibrating prior to sample measurement using calibration standards. Lastly, instrument environment can affect the validity of the classification model on the different instruments. Absolute temperature changes, both short and long term, may possibly induce instrument changes due to thermal expansion, leading to misalignment of optical components and the shifting of spectral peaks along the wavelength axis[26]. There are a number of standardization methods to attempt to correct the intensity and wavenumber shifts including single wavelength standardization (SWS), direct standardization (DS) and piecewise direct standardization (PDS)[26]. These methods involve determining the instrumental differences from a subset of samples measured on the primary site instrument and regressing against the same subset measured on the secondary site instrument. This determines a spectral response that can be used to correct the model differences for future measurements.

The ability to implement models created with data collected on a number of machines at various times that can classify data measured on different instruments at different sites is critical for large scale implementation of Raman spectroscopy in the clinic. In this work we investigate the use of calibration standards (wavenumber offset correction, instrument response correction and baseline fluorescence correction) to develop transferable classification models capable of discriminating between intraepithelial metaplasia (IM) and adenocarcinoma (AC) pathologies in the oesophagus tissue using Raman spectroscopic mapping. Stone et al[27] first attempted system transferability between two similar Raman systems (at the same site) by determining the energy transfer function of each system from a calibrated tungsten-filament lamp and a secondary NPL (National Physical Laboratory) standard and then correcting the measured oesophageal tissue spectra by multiplying this by the transfer function. The authors noted that by calibrating for energy sensitivity in each of the systems, there were very few differences in the mean tissue spectra between the two systems. This current study, however, is the first paper that has attempted to build a classification model, from mapped Raman tissue spectra, and validate in three different geographical instrument sites with the intention of creating a transferable diagnostic tool suitable for

Journal Name
View Article Online
ARTICLE
DOI: 10.1039/C5FD00183H

clinical application. Additionally it is the first paper to document the use of Raman map spectra (compared to commonly point-measured spectra) collected from oesophageal tissue samples to develop and validate a pathology classification model.

Materials and methods

Tissue collection and preparation

All tissue used in this work was obtained from patients undergoing scheduled endoscopy (for Barrett's surveillance or dysplasia) or from patients undergoing surgery for oesophageal cancer. These procedures were performed under local (endoscopic resection) or general (oesophageal resection) anaesthetic in accordance with an approved ethical proposal [Gloucestershire Local Research Ethics Committee]. At all times the General Medical Council (GMC) Guidelines on good clinical practice were followed.

20 oesophageal samples were collected in total, 10 IM and 10 AC. Informed written-consent was provided, prior to collection of the tissue. Each specimen was placed in a labelled cryovial (Thermo Scientific, Waltham, MA, USA) and the specimens were immediately snap-frozen in liquid nitrogen, before storage at -80°C.

For every specimen, tissue sections were cut and stained with H&E (Haematoxylin and Eosin) and reviewed by a consultant histopathologist. Acetate paper on which the specimen was mounted in the operating theatre acted not only to orientate the sample but also allowed the specimens to adhere to the microtome chuck without the need for cutting agents such as Optimal Cutting Temperature compound (OCT).

A freezing microtome was used to obtain a 5-6 micron section which was mounted on a plain glass slide for H&E staining, while three 8 micron thick tissue sections were mounted on three (one for each instrument site) circular 20mm x 1mm Raman-grade Calcium Fluoride (CaF2) discs (Crystran, Poole, UK) and stored at -80°C in 20mm round coin capsules (Leuchtturm, Geesthacht, Germany). Before being used in all of the Raman experiments samples were allowed to passively thaw at 23°C for up to 15 minutes prior to measurement.

Raman spectral tissue measurement

All three sites (Biophotonics Research Unit, Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK; University of Exeter, Exeter, UK; and University College London, London, UK) used a Renishaw RA800 series benchtop Raman system configured for pathology use; 785nm laser excitation with a X50 objective, and motorised XYZ stage. The system is equipped with transmitted and reflected white light imaging for sample location. StreamLine™ Raman imaging, a fast mapping method, was used for data collection. The spectral resolution was 2.5 cm^{-1} with a range of 150-2100 cm^{-1}. The Raman spatial resolution is ~1 um. The system comes with fully-automated alignment and calibration routines to ensure optimal data reproducibility and transferability. An acquisition time of 60s/line was used to collect a 50 × 50 pixel Raman image of two selected regions of interest from the tissue section (determined by reviewing the H&E slide).

The spectrometers were regularly calibrated and checked using an additional set of external standards including silicon, green glass and polystyrene, to automatically calibrate the wavenumber axis and to ensure optimal signal.

Data Processing and Analysis

Streamline™ maps were loaded into Matlab R2014b (The Mathworks Inc., Natick, Massachusetts, USA) for data pre-processing. The spectral range was cropped to the fingerprint region (400–1800 cm^{-1}).

Three spectral correction transfer methods were applied and tested to determine their influences on model classification performance: instrument response correction, wavenumber offset correction and EMSC fluorescence baseline correction.

Instrument response correction

Instrument response correction was achieved by comparing the ratio of the daily measured and calibrated standard spectra from each of the instrument system measured from a piece of standardized green glass (three pieces from the same source were used, one on each of the three systems) to provide an instrument response profile that was then used to correct spectra obtained from the tissue specimens.

Wavenumber off-set correction

Spectra were wavenumber off-set corrected using the phenylalanine peak of the average spectrum of each map. Peak fitting a pseudo-Voigt profile was used to correct the phenylalanine position to 1003cm^{-1}.

Extended Multiplicative Scatter correction (EMSC) Fluorescence Baseline Correction

Using Extended Multiplicative Scatter Correction (EMSC), a least squares modelling procedure, in a manner already described in detail in previous publications[28,29] including an investigation for electronic removal of paraffin peaks from infrared spectra[30]. EMSC was used to account for scaling and artefacts using a polynomial baseline (3rd order), non-tissue measured instrument spectrum for each of the instrument systems (measuring without CaF2 substrate or tissue sample present) and a representative mean tissue spectrum for each instrument spectral dataset. In brief, EMSC is adapted in order to neutralize any background (e.g. fluorescence) using a least squares modelling procedure. This involves an interference matrix introduced into the EMSC algorithm composed of the non-tissue instrument spectrum and the mean tissue spectrum for each instrument site. This function removes baseline fluorescence arising from any tissue fluorophores and instrument optics and normalises the dataset to reduce the effect of total signal variation (e.g. due to variable focussing and thickness of the sample).

Cosmic ray and fluorescence saturation removal

It is crucial to either remove or replace sharp spikes due to cosmic rays and saturated spectra due to fluorescence in order to prevent them unduly influencing the data analysis. These

ARTICLE
View Article Online
Faraday Discussions
DOI: 10.1039/C5FD00183H

artefacts were filtered from the spectral images by applying a 5 × 5 moving window two-dimensional median filter to each wavenumber. For the 5 × 5 window used, the centre pixel was replaced with the median value of all pixels in the window for the chosen wavenumber.

Principal Component Analysis (PCA)

Prior to creating and validating a classification model, PCA was performed on the complete tissue dataset prior to and after implementing the correction algorithms. This determined whether the correction methods were able to minimize instrument system differences created as a result of different instrument environment and instrument response function as well as potential sample quality differences.

PCA reduces the spectral data to a smaller number of components that describe the major variations in the dataset, in this case instrument, sample quality and pathology variations.

Partial Least Squares followed by Discriminant Analysis (PLS-DA)

A total of 219,367 individual Raman spectra were collected, across all three systems (intraepithelial metaplasia = 114,456; adenocarcinoma = 104,911).

The data matrix was mean centred prior to the application of Partial Least Squares Discriminant Analysis (PLS-DA). PLSDA reduces the spectral data to a smaller number of components that describe the major variations in the dataset, while simultaneously correlating the captured variance with the consultant pathology classification obtained for each tissue sample[31].

The number of components used for discriminant analysis was determined by increasing the number included until the ability of the algorithm to accurately predict the pathology group of spectra in the second leave-some-out test dataset no longer increased.[31]

Model validation using leave-one-map-out (LOMOCV) cross-validation

To fully assess the suitability of Raman spectroscopy for distinguishing between the two pathology groups (intraepithelial metaplasia and adenocarcinoma) and to test the system transferability performance across the three instrument sites, the dataset was split into training and validation sets by leaving one map out of the tissue spectra

The training dataset was used to generate a statistical model and the validation dataset was projected onto the model to assess its performance.

Results and discussion

In order to assess the ability of the wavenumber offset correction (WOC) algorithm on the tissue spectral data and whether it will adequately correct for any potential wavenumber shift at any of the instrument sites, the polymer spectra recorded every day, prior to tissue measurements, was analysed prior and post WOC. Figure 1 shows the mean spectra of the polymer data calculated from a 5s time-acquired 50 x 50 pixel map. In figure 1a, it is clear that there is a slight shift in the ~1033cm^{-1} peak (and therefore rest of the spectrum) between each of the sites. Three polymer peaks were initially selected across the fingerprint range and a 2nd order polynomial correction, including an offset, was applied to correct for wavenumber shifts between sites. However, on inspection of the polynomial coefficients, it was determined that the values were effectively zero for all terms apart from the offset. This indicated that it was only necessary to apply an offset correction to the tissue data set. For the actual tissue dataset, this offset was corrected using the phenylalanine 1003cm^{-1} tissue peak. Without correction of these slight shifts in the tissue dataset, it could negatively affect the ability of our algorithm to accurately differentiate between pathology groups because the tissue peaks accounting for pathology state differences will be masked by these spectral shifts. By applying the pseudo-Voigt profile peak-fitting algorithm to the data, figure 1b demonstrates that it is possible to correct for these spectral shifts and align the spectral data.

Figure 1 – Plot of mean spectra of polymer standard (a) **before** and (b) **after** corrections for offset wavenumber correction. Right panels represent zoom enhanced area of the spectra for the 1033cm^{-1} peak

Using WOC in addition to response correction and EMSC on the complete tissue data set from each of the instrument sites, figure 2 demonstrates the effectiveness of these combined correction methods to reduce the variation in the standard deviation at all of the sites. Site 1 still seems to show increased variation in comparison to the other sites. Although similar tissue areas are selected to be measured from the H&E by the site 1 user, it is possible that each user at the other two instrument sites could be measuring a slightly different area from site 1 and therefore could account for the difference in the calculated standard deviation.

Figure 2 – Plot of mean spectra of pathologies (by instrument site) with 0.5STD before and after offset wavenumber, fluorescence baseline correction and normalisation applied

However the combined correction methods (as depicted in the flowchart in figure 3) are helping to correct for any sample thickness differences as well as variations in baseline, caused by fluorophores in tissue and any instrument optical fluorescence. In addition the instrument response function is corrected by comparing the ratio of the daily green glass standard measurement and the calibrated standard spectra to decipher the instrument response profile, which is then applied to the same day collected map tissue spectra.

Figure 3 – Steps applied of correction and processing of tissue data

Journal Name

View Article Online
ARTICLE
DOI: 10.1039/C5FD00183H

In order to illustrate the impact of the correction methods to minimise the instrument site variation within the entire dataset, PCA was applied to the mean-centred tissue dataset and the first two PCA components are plotted in figure 4. PC1 and PC2 account for the greatest variance in the tissue data independent of any pathology supervision or prior knowledge. It is clear in figure 4a, that there is potentially an instrument and sample quality variation between and within the instrument sites. The individual sites are separating and site 3 is seems to have 3 clusters of data which could be due to either pathology or sample quality differences.

After applying the correction algorithms in figure 4b which will reduce the variation due to sample thickness and quality, fluorescence baselines as well as spectral offset and instrument transfer function, it is clear that between and within the individual instrument sites, that these differences are reduced and the data seems to be more tightly clustered. However there still appears to be some instrument site variation even after the correction has been applied. Due to the small sample size (10 adenocarcinoma and 10 intraepithelial metaplasia tissue samples) as well as small region of area actually mapped on each tissue section (two 50x50 mini-maps at ~1.1µm spatial resolution), we do not expect there to be a great variance within the pathology states in the existing limited tissue dataset. Therefore, in this current tissue dataset, it is assumed that any instrument and sample quality variation would become more prominent. We believe that these remaining site variations could be due to two different factors: 1) instrument and 2) sampling area.

Correction for instrument differences (i.e. wavenumber offset correction, instrument response and background correction) have been applied, therefore the remaining differences seen in the figure 4b between the sites could be either due deficiencies in the correction methods or remaining instrument differences which we have not addressed. For example, the signal and noise levels were higher in site 1 (due to higher laser power at objective) which would have increased both the Raman peak signals from tissue as well the instrument noise, enough to separate some of site 1 data from site 2 & 3 datasets, and demonstrated in the PC plot. Additionally, point spread function (PSF) could also play a part in instrument differences as it is already a challenge to maintain a constant PSF imaging performance over the full detector of one detector, never mind three different detectors on the three different systems.

The second source of difference is possibly due to sampling differences between the sites. After further analysis of the PC plot, it was discovered that a majority of the data points having the most variation were in the IM pathology. After checking with the areas selected from each of the tissue sections from all the sites, it was concluded that there was a small difference in how the IM tissue sampling areas were selected between sites 1 and sites 2 & 3. The Barrett's epithelium in IM, compared with adenocarcinoma, contains a number of tissue features: goblet cells and intervening non-goblet columnar cells in crypts with irregular, budding and distortion. IM areas selected by sites 2 & 3 concentrated mostly on dense areas of goblet cells showing dysplasia while site 1 selected areas included more tissue feature coverage: the goblet cells, non-goblet columnar cells in the crypts as well as some of the luminal surface. We believe it is the inclusion of a broader region of tissue features at site 1 measurements compared to site 2 & 3 that is responsible for the increase in data variation. For future measurements, there will be more agreement of the sampling area and features being measured. These considerations will hopefully help reduce some of the variation between the measurements sites.

Figure 4 - Plot of PCA scores of PC1 vs PC2 of mean-centred tri-centre data to demonstrate the instrument site dependence of data and need for correction algorithms. (a) **before** (b) **after** correction

Table 1 documents the training performance of the site 1 tissue dataset after applying PLS-DA and the impact of applying the correction algorithms. The tissue data from site 1 was mean centred and then PLS was applied to reduce the spectral data to smaller number of components the major variation in the dataset relevant to the pathology states. Although the correction method improved the resulting specificity (99.5% vs 95.9%) in discriminating adenocarcinoma from intraepithelial metaplasia, the sensitivity reduced from 99.5% to 94.7% following correction, which indicates little impact on the training performance for a single site. However this is not greatly surprising as the dataset is from one instrument site and therefore the instrument transfer function corrections we applied will have minimal impact compared to multi-site analysis.

Table 1 - Table for training performance of PLS-DA on site 1 before and after correction applied to differentiate intraepithelial metaplasia (IM) from adenocarcinoma (AC)

In order to assess some of the spectral differences between the two pathology groups of interest (adenocarcinoma and intraepithelial metaplasia), the corrected tissue spectra from all three instrument sites were combined and the mean spectrum for each of the two pathology groups was calculated (figure 5). By analysing the spectral differences in these two spectra, adenocarcinoma has higher intensities in nucleic acids (cytosine and uracil at 783cm^{-1} and 840cm^{-1} O–P–O stretch) but reduced protein content (amino acid peaks at 754, 1002, 1121, 1603-1622cm^{-1}, and amide III at 1246cm^{-1}). Shetty et al[20] documented similar changes and attributed this to cells undergoing dysplastic or malignant change in carcinogenesis in oesophageal tissue. This observation is also in agreement with Stone et al[10] who observed an increase in nucleic acid content in increasing carcinogenesis in laryngeal cancer tissue samples using similar peak positions for reference as did Haka et al[32] in breast cancer tissue using Raman spectroscopy.

Figure 5 - Mean spectra for intraepithelial metaplasia (IM) vs adenocarcinoma (AC) for all sites

ARTICLE
View Article Online
Faraday Discussions
DOI: 10.1039/C5FD00183H

Finally, to assess the classification performance of measured tissue spectra from all three instrument sites with and without correction applied, leave one sample out cross validation (LOOCV) and using data from one instrument site as the training site (site 1) analysis was performed. Table 2 describes the classification performance achieved. Again, although the correction methods improved the sensitivity of the classification model to differentiate between adenocarcinoma and intraepithelial metaplasia (from 86% to 96% when site 2 was the test set and 73% to 79% when site 3 was the test set) the specificity was reduced (from 84% to 77% for site 2 as test set and 85% to 81% when site 3 was the test set). This suggests that although the correction method improves the classification sensitivity of the model (i.e. the ability of the model to correctly identify those samples that are adenocarcinoma correctly) there is a reduction in the specificity (ability of the model to correctly identify those samples that are not classified as adenocarcinomas).

Table 2 - Table of classification performance for LOOCV **with** and **without** all corrections methods applied to differentiate intraepithelial metaplasia (IM) from adenocarcinoma (AC)

Overall, the results together indicates that although the correction models do minimize the instrument and sample quality variations within and between the instrument sites, they do not currently improve the overall performance of the classification model when taking into account both the specificity and sensitivity performance. In order to further improve the classification performance of these models, potentially further standardisation procedures, calibrations, transfer corrections and modelling options need to be implemented in order to achieve similar results to the within-site model. Additionally this classification model was constructed from only 20 tissue samples over two pathology groups with limited sampling area from each tissue sample. A larger sample collection with more tissue samples of varying pathology states and greater tissue area coverage (per sample) is needed to properly assess the ability of Raman spectroscopy and system transferability algorithms over multiple instrument sites in oesophageal cancer diagnostics.

A few studies in the last 20 years have investigated models transferability in the field of spectroscopy[33–36]. Myles et al[37] published a study on a multivariate classification model constructed from near-infrared spectra measured in order to discriminate between green Arabica and Robusta coffee beans. This model was transferred to a second instrument and investigated for classification performance. In order to allow this classification model to perform optimally on this second system, multiple correction algorithms were investigated including: modified version of slope/bias correction (SBC), orthogonal signal correction (OSC), and model updating (MU). These corrections resulted in misclassification errors in the second instrument site between 5-10%. Anibal et al[38] assessed the performance of multivariate calibration transfer methods (such as Piecewise Direct Standardization (PDS)) used to discriminate between unadulterated samples and samples adulterated with four Sudan dyes. These were measured using UV-visible spectroscopy by different operators and varying the time in which the samples were analysed to represent the different experimental parameters/conditions that need correcting. By using PDS, the authors were able to determine transfer functions and apply them to improve the classification performance for the other measurement conditions to yield results comparable to the original measurement condition. Rodriguez and co-workers[39] used standardisation models on Raman spectra of a set of pharmaceutical chemical samples. These chemicals were measured on five-commercially available 785nm Raman spectrometers (different laboratory-based and portable platforms). As part of the standardisation process, the authors applied shift correction, intensity correction, and resolution matching and interpolation. The results indicated that the differences in Raman spectra from these chemical standards acquired on different spectrometers can be corrected, with satisfactory results, using their recommended standardisation protocol. They concluded that the most dramatic effects came from correcting for instrument-instrument variations in Raman shift axis and resolution matching.
All these previous studies were performed on commercial chemicals and food products. Human tissue samples measurements are subject to greater variations in sample quality, thickness, density, composition and preparation, which require a greater level of standardisation and system transfer spectral correction algorithms.

A larger study including both oesophageal and colon cancer tissue samples using the existing correction algorithms documented in this current study as well as a number of multivariate calibration transfer methods (such as Piecewise Direct Standardization (PDS) model updating (MU)) are currently being investigated within our collaboration.
Of interest to future clinical implementation of Raman spectroscopy as is the need to generate a classification model from data collected on a number of different machines and to then use the system transferability correction algorithms to predict pathology samples collected from a completely different set of sites and Raman spectroscopy systems. This current study used instruments of similar parameter designs however the true test of the implementation of this technology into clinic and worldwide is being able to implement this model on independent machine(s) of different design specifications. This would however require significant collaboration between an increased number of institutes, much larger sample sizes and more advanced multivariate analysis.

Conclusions

Journal Name

View Article Online
ARTICLE
DOI: 10.1039/C5FD00183H

Raman spectroscopy combined with advanced multivariate analysis is able to decipher a wealth of biochemical information from tissue samples of different pathology states. This information can be used to identify biomarkers associated with pre-malignant and carcinogenic change. This is important for non-invasive detection of disease as well as its classification and grading to help in patient diagnosis and prognosis.

Streamline™ Raman mapping and imaging in conjunction with rapid spectral analysis has significant potential to be able to collect more tissue spectra data from tissue samples increasing the amount of information available to spectroscopists to develop validated classification models.

In order to implement this technology in clinic and in general introduce Raman spectroscopy as a clinical tool for more global use in medical diagnostics, classification models constructed from spectra from one or more instruments must to be transferable between instruments and centres. Studies, such as this, help to investigate the potential of these models with their system transfer correction algorithms to transfer between different Raman spectroscopy systems at geographically different locations. There are a number of challenges ahead in order to develop the classification and optimal system transfer models, however collaborative studies such as this introduces the means and foundation for these further investigations.

Acknowledgements

This work was funded and supported by Innovate UK – 'SMART: Stratified Medicine through Advanced Raman Technologies'. We thank Renishaw Plc for system support and development. The histopathology technicians at Gloucestershire Hospitals NHS Trust in Gloucester, especially Joanne Motte, have worked extremely hard on additional sample processing and preparation.

References

1 Cancer Research UK, *Cancer Res. UK*, 2012.

2 H. Pohl and H. G. Welch, *J. Natl. Cancer Inst.*, 2005, **97**, 142–146.

3 J. Lagergren, R. Bergstrom, A. Lindgren and O. Nyren, *N Engl J Med*, 1999, **340**, 825–831.

4 Stephan M. Wildi and Michael B. Wallace, *Barrett's Esophagus and Esophageal Adenocarcinoma*, John Wiley & Sons, 2008.

5 B. J. Reid, R. C. Haggitt, C. E. Rubin, G. Roth, C. M. Surawicz, G. Van Belle, K. Lewin, W. M. Weinstein, D. a Antonioli and H. Goldman, *Hum. Pathol.*, 1988, **19**, 166–178.

6 C. L. Booth and K. S. Thompson, *J. Gastrointest. Oncol.*, 2012, **3**, 232–42.

7 L. M. Almond, J. Hutchings, N. Shepherd, H. Barr, N. Stone and C. Kendall, *J. Biophotonics*, 2011, **4**, 685–95.

8 C. Kendall, N. Stone, N. Shepherd, K. Geboes, B. Warren, R. Bennett and H. Barr, *J. Pathol.*, 2003, **200**, 602–609.

9 N. Stone, C. Kendall, N. Shepherd, P. Crow and H. Barr, *J. Raman Spectrosc.*, 2002, **33**, 564–573.

10 N. Stone, P. Stavroulaki, C. Kendall, M. Birchall and H. Barr, *Laryngoscope*, 2000, **110**, 1756–1763.

11 P. Crow, N. Stone, C. a Kendall, J. S. Uff, J. a M. Farmer, H. Barr and M. P. J. Wright, *Br. J. Cancer*, 2003, **89**, 106–108.

12 J. Smith, C. Kendall, a Sammon, J. Christie-Brown and N. Stone, *Technol. Cancer Res. Treat.*, 2003, **2**, 327–32.

13 R. Kast, G. Auner, S. Yurgelevic, B. Broadbent, A. Raghunathan, L. M. Poisson, T. Mikkelsen, M. L. Rosenblum and S. N. Kalkanis, *J. Neurooncol.*, 2015, **125**, 287–295.

14 F. M. Lyng, D. Traynor, I. R. M. Ramos, F. Bonnier and H. J. Byrne, *Anal. Bioanal. Chem.*, 2015, **407**, 8279–8289.

15 Z. Farhane, F. Bonnier, A. Casey, A. Maguire, L. O'Neill and H. J. Byrne, *Analyst*, 2015, **140**, 5908–19.

16 S. J. Harder, Q. Matthews, M. Isabelle, A. G. Brolo, J. J. Lum and A. Jirasek, *Appl. Spectrosc.*, 2015, **69**, 193–204.

17 T. Tolstik, C. Marquardt, C. Matthäus, N. Bergner, C. Bielecki, C. Krafft, A. Stallmach and J. Popp, *Analyst*, 2014, **139**, 6036–43.

18 Q. Matthews, A. Brolo, J. Lum, X. Duan and a Jirasek, *Phys. Med. Biol.*, 2010, **56**, 19–38.

19 C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker and N. Stone, *Analyst*, 2009, **134**, 1029–45.

20 G. Shetty, C. Kendall, N. Shepherd, N. Stone and H. Barr, *Br. J. Cancer*, 2006, **94**, 1460–4.

21 J. Hutchings, C. Kendall, B. Smith, N. Shepherd, H. Barr and N. Stone, *J. Biophotonics*, 2009, **2**, 91–103.

22 J. C. C. Day, R. Bennett, B. Smith, C. Kendall, J. Hutchings, G. M. Meaden, C. Born, S. Yu and N. Stone, *Phys. Med. Biol.*, 2009, **54**, 7077–87.

23 L. M. Almond, J. Hutchings, C. Kendall, J. C. C. Day, O. A. C. Stevens, G. R. Lloyd, N. A. Shepherd, H. Barr and N. Stone, *J. Biomed. Opt.*, 2012, **17**, 081421–1.

ARTICLE
View Article Online
Faraday Discussions
DOI: 10.1039/C5FD00183H

24 L. M. Almond, J. Hutchings, G. Lloyd, H. Barr, N. Shepherd, J. Day, O. Stevens, S. Sanders, M. Wadley, N. Stone and C. Kendall, *Gastrointest. Endosc.*, 2014, **79**, 37–45.

25 M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. So and Z. Huang, *Technol. Cancer Res. Treat.*, 2011, **10**, 103–12.

26 D. Ozdemir, M. Mosley and R. Williams, *Appl. Spectrosc.*, 1998, **52**, 1203–1209.

27 N. Stone, C. Kendall and H. Barr, in *Handbook of Vibrational Spectroscopy*, John Wiley & Sons, Ltd, 2006.

28 N. K. Afseth and A. Kohler, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 92–99.

29 H. Martens and E. Stark, *J. Pharm. Biomed. Anal.*, 1991, **9**, 625–635.

30 J. Nallala, G. R. Lloyd and N. Stone, *Analyst*, 2015, **140**, 2369–75.

31 G. R. Lloyd, L. E. Orr, J. Christie-Brown, K. McCarthy, S. Rose, M. Thomas and N. Stone, *Analyst*, 2013, **138**, 3900–3908.

32 A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari and M. S. Feld, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 12371–6.

33 C. K. Mann and T. J. Vickers, *Appl. Spectrosc.*, 1999, **53**, 856–861.

34 H. Nguyen Quang, M. Jouan and N. Quy Dao, *Anal. Chim. Acta*, 1999, **379**, 159–167.

35 H. Swierenga, A. P. de Weijer, R. J. van Wijk and L. M. C. Buydens, *Chemom. Intell. Lab. Syst.*, 1999, **49**, 1–17.

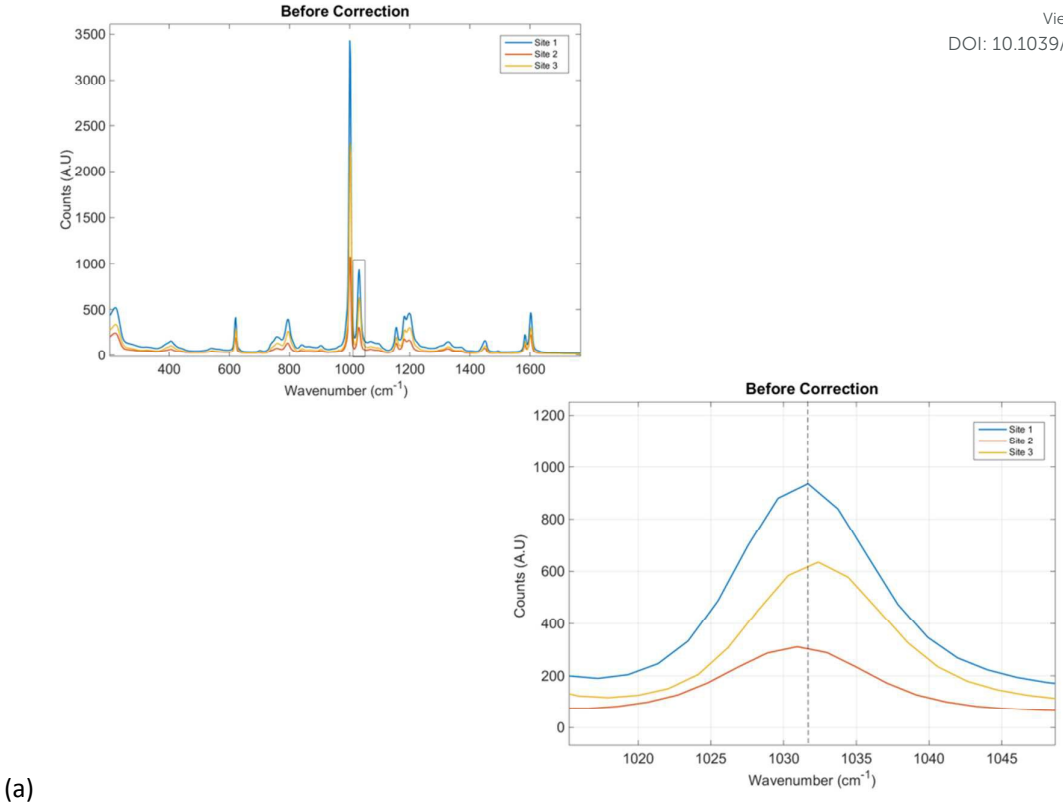
36 H. Swierenga, A. P. de Weijer and L. M. C. Buydens, *J. Chemom.*, 1999, **13**, 237–249.

37 A. J. Myles, T. A. Zimmerman and S. D. Brown, *Appl. Spectrosc.*, 2006, **60**, 1198–203.

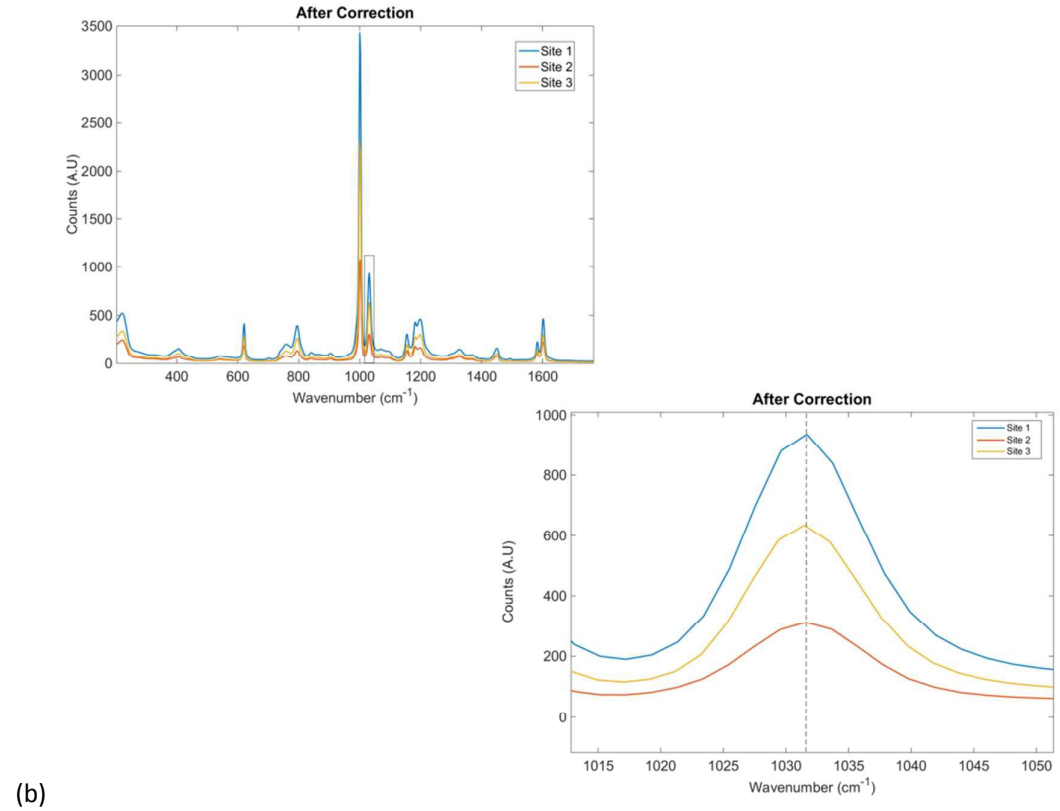
38 C. V Di Anibal, I. Ruisánchez, M. Fernández, R. Forteza, V. Cerdà and M. Pilar Callao, *Food Chem.*, 2012, **134**, 2326–31.

39 J. D. Rodriguez, B. J. Westenberger, L. F. Buhse and J. F. Kauffman, *Analyst*, 2011, **136**, 4232–40.

View Article Online
DOI: 10.1039/C5FD00183H



(a)



(b)

Figure 1 – Plot of mean spectra of polymer standard (a) **before** and (b) **after** corrections for offset wavenumber correction. Right panels represent zoom enhanced area of the spectra for the 1033cm^{-1} peak

View Article Online
DOI: 10.1039/C5FD00183H

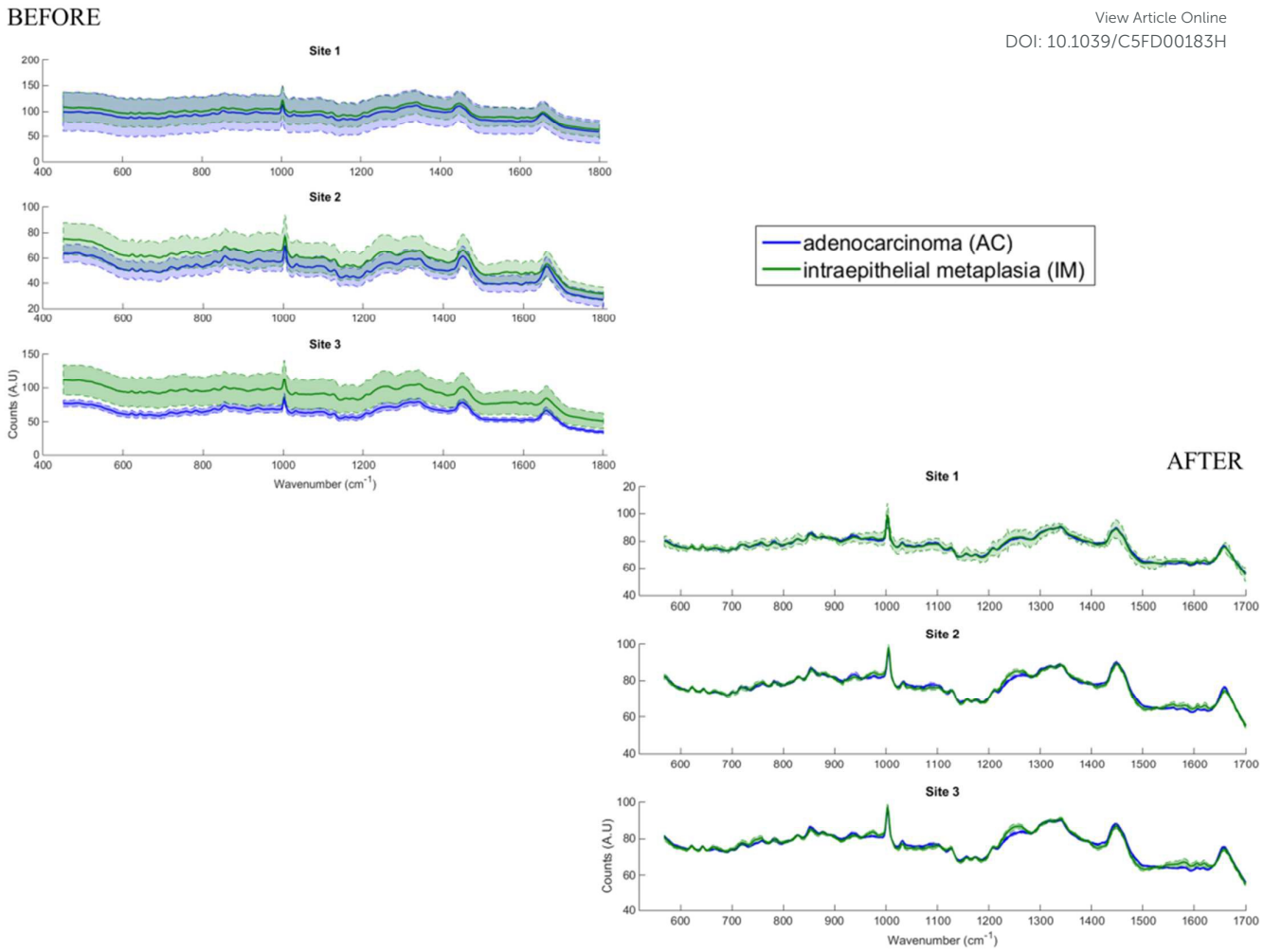


Figure 2 – Plot of mean spectra of pathologies (by instrument site) with ±0.5STD **before** and **after** offset wavenumber, instrument response correction, and fluorescence baseline correction and normalisation applied

View Article Online
DOI: 10.1039/C5FD00183H

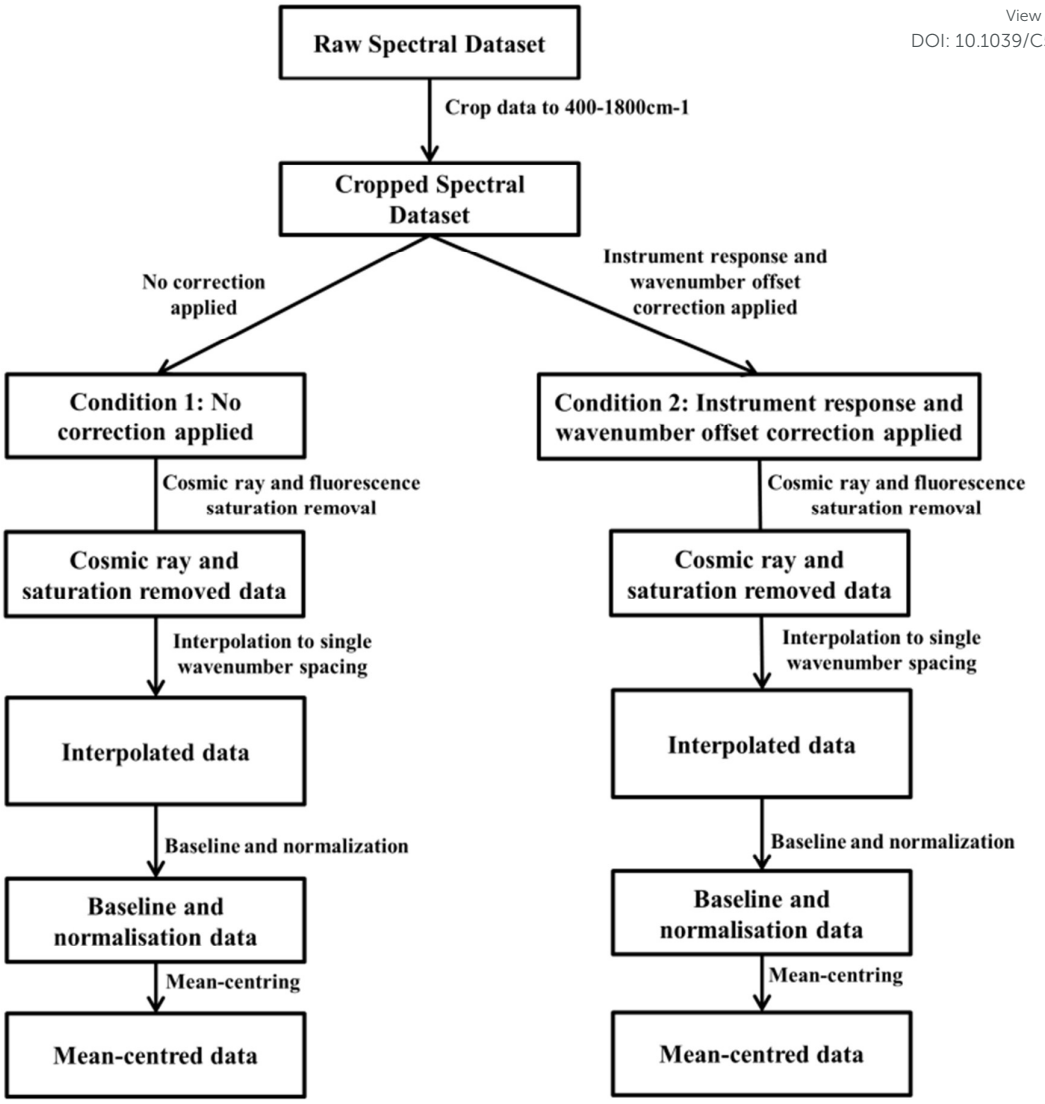
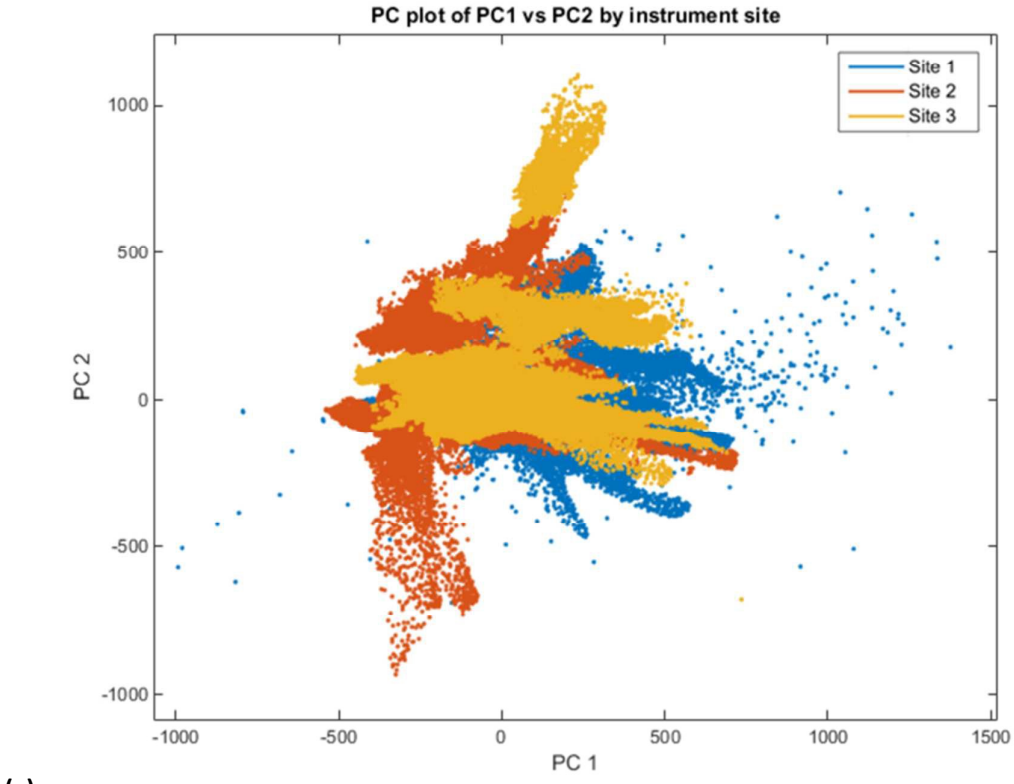
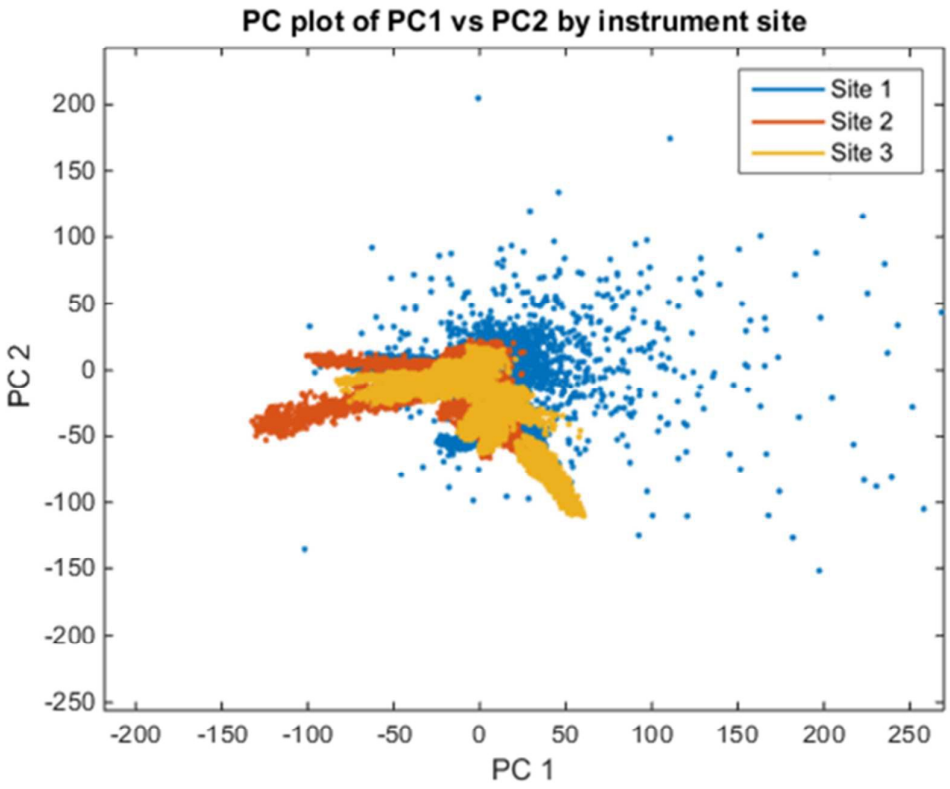


Figure 3 – Steps applied of correction and processing of tissue data

View Article Online
DOI: 10.1039/C5FD00183H



(a)



(b)

Figure 4- Plot of PCA scores of PC1 vs PC2 of mean-centred tri-centre data to demonstrate the instrument site dependence of data and need for correction algorithms (a) **before** (b) **after**

View Article Online
DOI: 10.1039/C5FD00183H

	Site 1 as training set	
	Sensitivity (%)	Specificity (%)
Before any correction methods applied	99.5	95.9
After all correction methods applied	94.7	99.5

Table 1 – Training performance of PLS-DA on site 1 **before** and **after** correction applied to differentiate intraepithelial metaplasia (IM) from adenocarcinoma (AC)

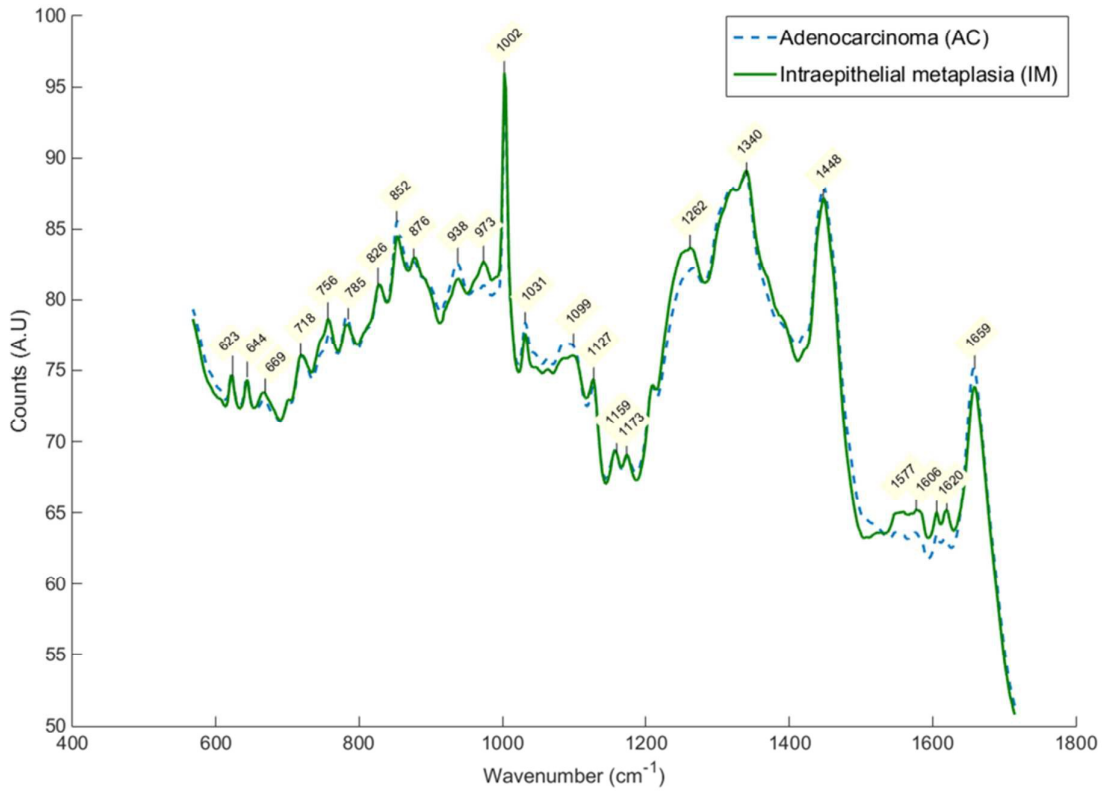


Figure 5 - Mean spectra for intraepithelial metaplasia (IM) vs adenocarcinoma (AC) for all sites

	Site 1 as training set (Site 2 as test set)		Site 1 as training set (Site 3 as test set)	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Before any correction methods applied	86.2	84.4	72.8	85.3
After all correction methods applied	95.7	76.9	78.8	80.8

Table 2 - Classification performance for LOOCV **with** and **without** all corrections methods applied to differentiate intraepithelial metaplasia (IM) from adenocarcinoma (AC)