J. Biophotonics 6, No. 1, 73–87 (2013) DOI 10.1002/jbio.201200126

Journal of BIOPHOTONICS

FULL ARTICLE

FTIR microspectroscopy of selected rare diverse sub-variants of carcinoma of the urinary bladder

Caryn Hughes[1,2], *Junaid Iqbal-Wahid*[3], *Michael Brown*[2], *Jonathan H. Shanks*[4], *Amanda Eustace*[3], *Helen Denley*[5], *Peter J. Hoskin*[6], *Catharine West*[3], *Noel W. Clarke*[2,7,8], *and Peter Gardner*[*,1]

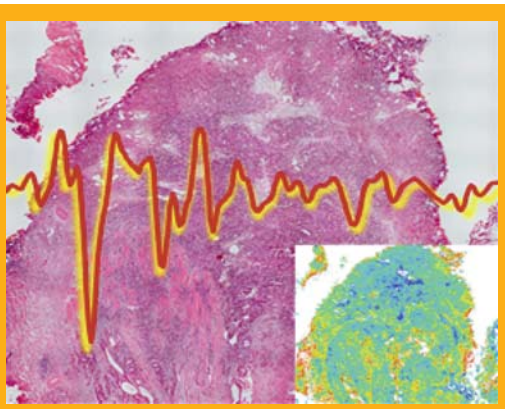
[1] Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK
[2] Genito Urinary Cancer Research Group, Institute for Cancer Sciences, Paterson Institute for Cancer Research, The University of Manchester, Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, Manchester, M20 4BX, UK
[3] Translational Radiobiology Group, Institute for Cancer Sciences, Paterson Institute for Cancer Research, The University of Manchester, Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, Manchester, UK, M20 4BX, UK
[4] Department of Histopathology, The Christie NHS Foundation Trust, Manchester, UK, M20 4BX, UK
[5] Histopathology, Central Manchester University Hospitals NHS Foundation Trust, Manchester Royal Infirmary, Oxford Road, Manchester, M13 9WL, UK
[6] Cancer Centre, Mount Vernon Hospital, Rickmansworth Road, Northwood, Middlesex, HA6 2RN, UK
[7] Department of Urology, The Christie NHS Foundation Trust, Manchester, UK, M20 4BX, UK
[8] Department of Urology, Salford Royal Hospital NHS Foundation Trust, Stott Lane, Salford, M6 8HD

Received 3 July 2012, revised 1 October 2012, accepted 1 October 2012
Published online 1 November 2012

Key words: FTIR, Infrared microspectroscopy, principal component analysis, cluster analysis, support vector machines, urothelial carcinoma, transitional cell carcinoma, lipoid, clear cell

Supporting information for this article is available free of charge under http://dx.doi.org/10.1002/jbio.201200126

Urothelial carcinomas of the bladder are a heterogeneous group of tumours, although some histological sub-variants are rare and sparsely reported in the literature. Diagnosis of sub-variants from conventional urothelial carcinoma can be challenging, as they may mimic the morphology of other malignancies or benign tumours and therefore their distinction is important. For the first time, the spectral pathology of some of these sub-variants has been documented by infrared microspectroscopy and an attempt made to profile their biochemistry. It is important not only to identify and separate the cancer-associated epithelial tissue spectra from common tissue features such as stroma or blood, but also to detect the signatures of tumour sub-variants. As shown, their spectroscopic signals can change dramatically as a consequence of differentiation. Example cases are discussed and compared with histological evaluations.



* Corresponding author: e-mail: peter.gardner@manchester.ac.uk

1. Introduction

Urothelial carcinoma, or transitional cell carcinoma (TCC), accounts for 85–90% of all epithelial tumours of the bladder and urinary collecting system. It is however, a very heterogeneous tumour, with a pronounced ability for divergent differentiation. It is common to find multiple patterns within a single tumour which can simultaneously display such variety as sarcomatoid, small-cell, squamous and glandular differentiation. The histological variants can have different diagnoses and prognoses that impact on tumour management [1]. For example, the presence of developed micro-papillary morphology in otherwise conventional urothelial carcinoma is associated with advanced tumour stage and poor prognosis [2]. The development of an objective method for identifying divergent histologies in TCC has potential to enhance diagnostic capabilities.

For unsupervised analysis of FTIR images, multivariate imaging techniques including principal component analysis (PCA), K-means clustering and hierarchical cluster analysis (HCA), have shown proficiency in the identification of spectral classes of tissue types that can be directly compared to H & E stained tissue sections. This has been demonstrated for a number of tissue types including carcinoma of the brain [3], cervix [4, 5], colon [6], kidney [7], lymph node [8, 9], salivary gland [10] and skin [11, 12].

Unsupervised cluster techniques provide a non-subjective method to identify diverse tissue structures and subsequent pathological analysis by a qualified reviewer can provide regulation of the results. Once tissue structures have been identified, however, it is often desirable to create a predictive model for future analysis of new patient samples. There are many examples where supervised classification algorithms such as linear discriminant analysis (LDA) and artificial neural networks (ANNs) have been applied to spectral classification of biological cells and tissues [13–17]. Support vector machines (SVMs) are a more recent breed of classification algorithm. Although there are relatively fewer examples in the application of SVMs in spectral pathology, the reports are impressive [18–20]. SVM theory is originally designed for binary classification, whereby a decision boundary either side of a hyperplane separates two classes of training data. Ideally the boundary is maximised to create the largest possible distance between the separating hyperplane and the two different class observations either side of it. Two classes of test data are then mapped into the same space and are class predicted based on which side of the decision boundary they fall on. For multi-class situations, a one-versus-rest (OVR) approach is used whereby the data trained to a set of multiple binary classifications. SVMs are suited to classification where the number of features is large with respect to the number of training instances, as the model complexity of an SVM is unaffected by the number of features in the training data. This makes SVMs an ideal method for spectroscopists where the data is often multidimensional. To complete the optimisation of class training, a global minimum is reached. This is an advantage over similar classification tools such as ANNs, that can potentially cease training at a local minima [21].

Within the vibrational spectroscopy community, Raman spectroscopy has been applied to study bladder cells and tissues [15, 22–27], yet there are very few reports of characterisation by infrared spectroscopy [28, 29]. Here we report for the first time a pilot investigation of identified sub-variants of TCC that have been preliminarily characterised by Fourier transform infrared (FTIR) microspectroscopy and chemometric analysis. We outline the case for the utility of spectral pathology as a useful tool in clinical diagnosis of challenging cases.

2. Experimental

2.1 Sample preparation

A retrospective pilot study was carried out in six patients with histologically proven high grade muscle invasive urothelial carcinoma of the bladder (stage *p*T2, Grade 3). Tissue samples were obtained by transurethral resection: all tissue was collected with full ethical approval and informed patient consent. One block per cm of tumour diameter was formalin-fixed and paraffin-embedded (FFPE). Pathological staging utilised the (AJCC/UICC classifications), grading was according to Royal College of Pathologists (RCPath) guidelines (WHO 1973). Tissue sections 4 μm thick were floated onto calcium fluoride substrates (76 × 26 × 1 mm). Paraffin was removed following standard histology protocols; slides were dipped 10 times and placed in xylene for 15 minutes each in three xylene troughs in succession before being dipped 10 times in 100% ethanol and air dried.

Tumour biopsies, displayed a variety of morphological variants within the tissue; some of which are uncommon, with only a small number of cases previously cited [30]. Supplementary Table 1 describes the observed variants found within the experimental samples and their known clinical and pathological significance [31]. The samples chosen for study (listed in Supplementary Table 2) were all from male patients diagnosed with the same grade and stage of TCC.

2.2 Infrared imaging

Transmission mode FTIR imaging spectroscopy was performed on a Varian 670-IR spectrometer coupled with a Varian 620-IR imaging microscope (Agilent Technologies, CA) equipped with a 128 × 128 pixel liquid nitrogen-cooled Mercury-Cadmium-Telluride (MCT) focal planar array detector. Spectra were collected in the 950–3800 cm^{-1} range, at a spectral resolution of 4 cm^{-1}, with the co-addition of 128 scans for sample spectra, and 256 scans for the background spectra. At a pixel resolution of ~5.5 μm, a tissue sampling area of ~700 × 700 μm was captured per hyperspectral image.

2.3 Histological staining

Haematoxylin and eosin (H & E) staining was performed directly onto the infrared substrate tissues. The slides were dipped 10 times in succession from 100%, 90% and 70% ethanol before being placed in water. The slides were then placed in a trough containing Gill's haematoxylin for 3 minutes, washed for 1 minute in running water before blueing in alkaline water (3 dips). The slides were then washed in running water for another minute, placed in a trough filled with eosin for one minute, before finally washing in water for 1 minute. The samples were dehydrated through ethanol (70%, 90%, 100%) to xylene and mounted with a cover slip using Pertex (DCM).

Where required, serial sections were floated onto standard histology slides, dewaxed and stained for periodic acid-Schiff (PAS); the samples were placed in periodic acid reagent (1%) for 5 minutes, then rinsed under tap water followed by distilled water. The samples were then placed in Schiff reagent for 15 minutes before counter-staining with Gill's haematoxylin for 1 minute, and then dipped twice in alkaline water before being dehydrated and coverslipped. A section of liver was stained with PAS as a control (Supp. Note 1a). Additionally, as PAS stains positively for mucin as well as glycogen, a control serial section was treated with diastase (a solution containing amylase) for 30 minutes prior to PAS staining. Glycogen presence is evidenced by loss of staining after enzyme treatment when compared to the untreated diastase sections (Supp. Note 1b).

All photomicrographs were captured using a Nikon Eclipse 90i microscope equipped with colour camera using either a 10× objective or 40× objective with a total pixel resolution of 0.85 or 0.27 μm/pixel respectively. To separate the PAS stain from the counterstain, a colour deconvolution plug-in implementing stain separation using Ruifrok and Johnston's method [32] was used with ImageJ (National Institutes of Health) (Supp. Note 2).

2.4 Spectral pre-processing

A PCA-based noise reduction algorithm was used for each image (containing 16384 spectra) by retaining 30 principal components. The level of noise removal was carefully monitored by observation of the difference spectrum (noise reduced spectrum extracted from the raw spectrum) (See Supp. Note 3 for further explanation). The noise-reduced data was then corrected with 20 iterations of the RMieS-EMSC scattering correction algorithm using a Matrigel spectrum as the Refs. [33–35].

The images were quality-controlled by integration of the amide I for threshold removal of blank regions of substrate. A second threshold extraction was based upon total intensity of absorbance, where particularly high intensity signals from red blood cells and artefacts as a consequence of thermal ablation were removed.

2.4 Data analysis

The spectral range was cut to 1000–1760 cm^{-1} and 2800–3000 cm^{-1} to remove spectral regions of no biological interest, reducing subsequent computational processing time. The spectra were transformed to the first derivative, with 13 point Salvitzky-Golay smoothing. Mean-centred, vector normalised (vnmc) PCA was performed for each hyperspectral image, retaining the scores for 10 PCs for dimension reduction [36–39]. K-means cluster analysis was performed (with three replicates to avoid local minima convergence) on the principal component scores using a cosine distance metric to define the image by an appropriate number of partitioned clusters. The computational time for cluster analysis on the PC scores was calculated at a fraction of a second per image. Subsequent pseudo-colour cluster maps were constructed, assigning a specific colour to each spectral cluster and compared to the respective H & E stain for definitive histopathological evaluation. Using the defined cluster classes as a mask, the corresponding original spectra were extracted and vector normalised for subsequent evaluation. Table 1 lists the numbers of patients, infrared images and extracted spectra per sub-variant class.

175 random spectra were transformed to the second derivative with 7 point Salvitzky-Golay smoothing and analysed by vector-normalised, mean-centred principal component analysis initially to establish discrimination between the classes of compiled spectra from different infrared images.

Finally, a radial basis function (RBF) support vector machine (SVM) was used for classification of the different spectral types [40]. 10,000 vector-normalised (non-derivative) spectra were each extracted

Table 1 Urothelial sub-variants and corresponding extracted spectra.

Sub-variant	#Spectra	#Sampled IR images	#Patient cases used
Conventional TCC	11776	3	2
Micro-papillary	19796	2	1
Stroma	14757	4	3
Lymphocyte-rich (epithelial)	19671	2	1
Clear cell level 1	12748	4	2
Clear cell level 2	20064	4	2
Lipoid	307	2	1

from the database of tissue structures previously identified as conventional TCC (class 1), stroma (class 2), micro-papillary (class 3), lymphocyte rich (class 4) and clear cell (class 5) spectra respectively. (The lipoid class was omitted from the classification due to a low number of spectral observations.) The dataset was randomly split into 5000 training spectra and 5,000 testing spectra per class. Each instance in the training set contained the class label and 251 features. For an accurate prediction of test data, suitable values for the RBF kernel parameters C and γ were determined by 3-fold cross-validation. This involved dividing the training set into 3 subsets of equal size. In sequential order, two subsets were used for training and one subset for testing to quantify the cross-validation accuracy as a percentage of data which correctly classified. This was repeated in a loop for different parameter selections to find the best validation accuracy. The best parameters, yielding a cross-validation accuracy of 98.088% were determined as C (362.0387), γ (90.5097). All computations were performed in Matlab (Mathworks Inc.).

3. Results and discussion

3.1 Cluster analysis of infrared images

The pseudo-colour k-means cluster maps derived from principal component scores, demonstrated a high level of pathological agreement with their subsequent H & E micrographs. The following examples detail a number of sub-variant morphologies of urothelial carcinoma, some of which have been rarely reported in the literature.

3.1.1 Micro-papillary sub-variant

The presence of developed micro-papillary morphology in otherwise conventional urothelial carcinoma is associated with advanced tumour stage and poor prognosis [2]. Micro-papillary structure presents with filliform projections that have secondary or tertiary hierarchical branching. Tumour cell growth is usually represented as small nests surrounded by lacunae resembling vascular-lymphatic or retraction spaces [30]. These spaces may be lined focally by flattened spindled cells, or devoid of lining. There are currently no established imaging techniques to reliably diagnose some types of deeply invasive urothelial carcinoma of urinary bladder, particularly the micropapillary variant. Diagnosis of this variant by pathological findings are therefore important [41]. The most problematic pathologic issue, however, are cases that present with metastatic disease where the primary lesion is unknown. Establishing the anatomic site of origin is challenging as several organs may give rise to carcinomas with an identical micropapillary appearance. Carcinomas with micro-papillary features have been described in the breast, urinary bladder, lung, and ovary and are said to be morphologically identical in the breast, urinary bladder, and lung [42].

Figure 1a displays H & E regions from one patient sample that displayed typical micro-papillary morphology. Chemical heterogeneity was explored in the micro-papillary regions of interest as shown in the two infrared images of total absorbance (Figure 1b). A two- class unsupervised cluster analysis suggested regions of chemical distinction (Figure 1c). A notable difference was observed in the ratio of the lipid $\nu_{as}(CH_3)$ at $\sim 2960\ cm^{-1}$ versus $\nu_{as}(CH_2)$ at $\sim 2920\ cm^{-1}$. This was indicative of an increase in short-chain lipids within certain regions of each IR image (the K-means class coloured black in Figure 1c) [43]. In both IR image A and IR image B (Figure 1c), the areas of greater lipid contribution (coloured black) directly corresponded to the H & E photomicrograph regions with more extensive retraction spaces (For clarity, a superimposition can be found in Supp. Note 4). An increase in absorbance of short chain lipid signals i.e. CH_2 stretches, suggests a change in lipid composition. This change may be due to either a breakdown of molecules to form longer-chain lipid bi-products, or an increase in the assembly of lipid-based building blocks. This is possibly due to greater disruption of the surrounding stromal extracellular matrix (ECM). Disorganisation of the ECM deregulates the behaviour of stromal cells and facilitates tumour-associated angiogenesis and inflammation [44]. Unsurpirsingly, large protein differences were also observable by eye in the Amide I and II bands (Figure 1d). These changes may be indicative of the complex network of enzymes such as proteases and sulfatases that are involved in activation of various signalling pathways during matrix re-organisation and tumour progression [45].

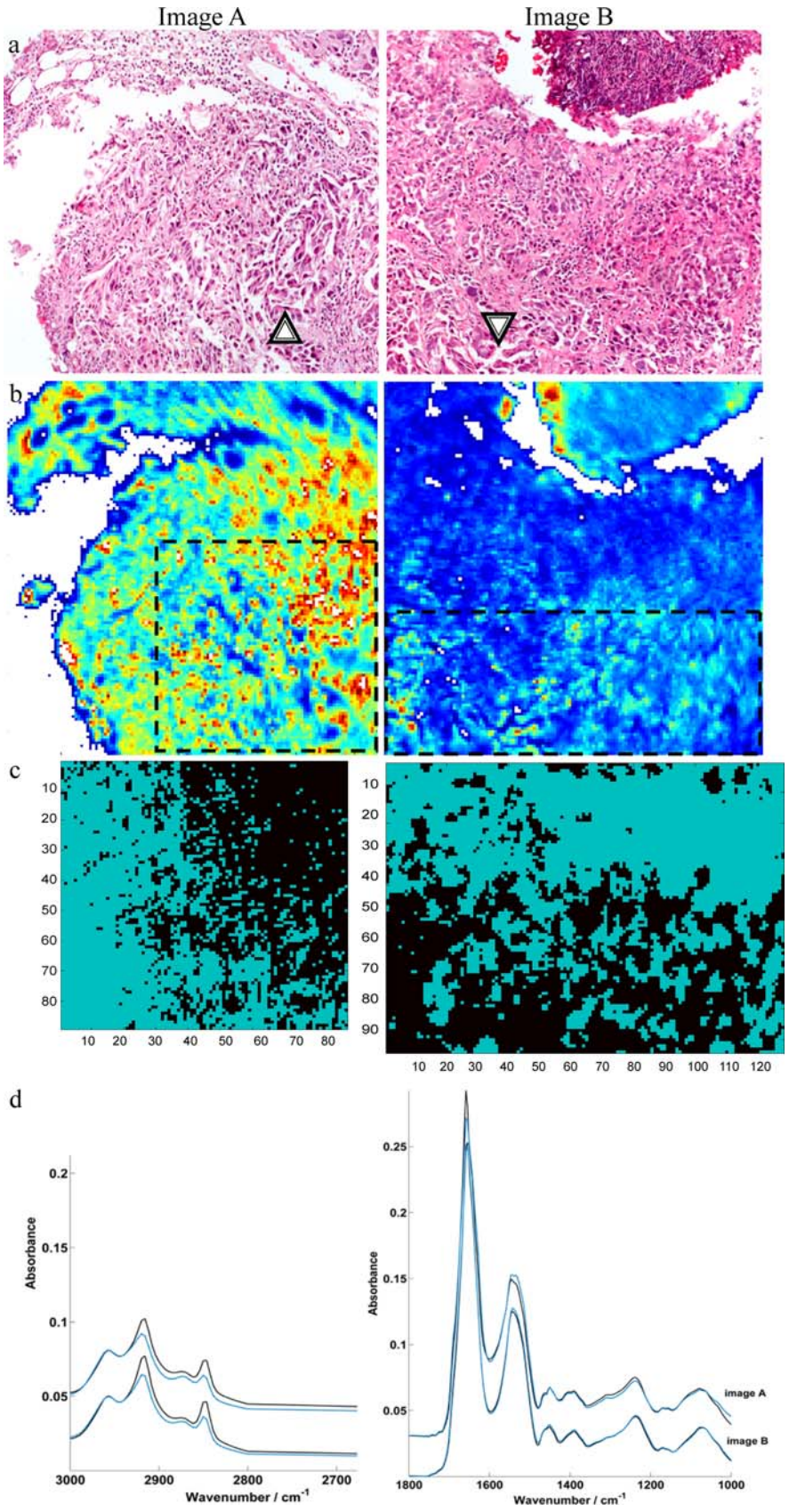


Figure 1 (online color at: www.biophotonics-journal.org) Micro-papillary regions of TCC. Images display: (**a**) H & E photomicrographs where the black arrows indicate regions with more extensive retraction spaces (**b**) infrared total intensity of absorbance maps of the regions of interest; dashed lines indicate the areas of cluster analysis for image *A* (440 μm × 440 μm) and image *B* (495 μm × 660 μm) (**c**) corresponding 2-class *K*-means classification matrix and (**d**) mean representative spectra.

Histologically, retraction spaces, such as the prominent pattern found in the urothelial micro-papillary variant, are frequently observed and can be difficult to differentiate from indicative retraction spaces in lymphovascular invasion [46]. Although there is no defined correlation, true lymphovascular invasion, has in fact reported to be present in most TCC cases presenting with the micro-papillary subvariant [47]. More convincingly, the presence of micro-papillary features or extensive retraction spaces in breast carcinoma has been shown to predict the presence of nodal metastasis. It is hypothesised that the characteristic clear spaces are related to altered tumour-stromal interactions, which might have an important role in lymphatic tumour spread and tumour progression [48].

3.1.2 *Lymphocyte-rich TCC*

Lymphoepithelioma-like carcinoma (LELC) in the urinary tract is a rare malignancy, named for its resemblance to nasopharyngeal undifferentiated carci-

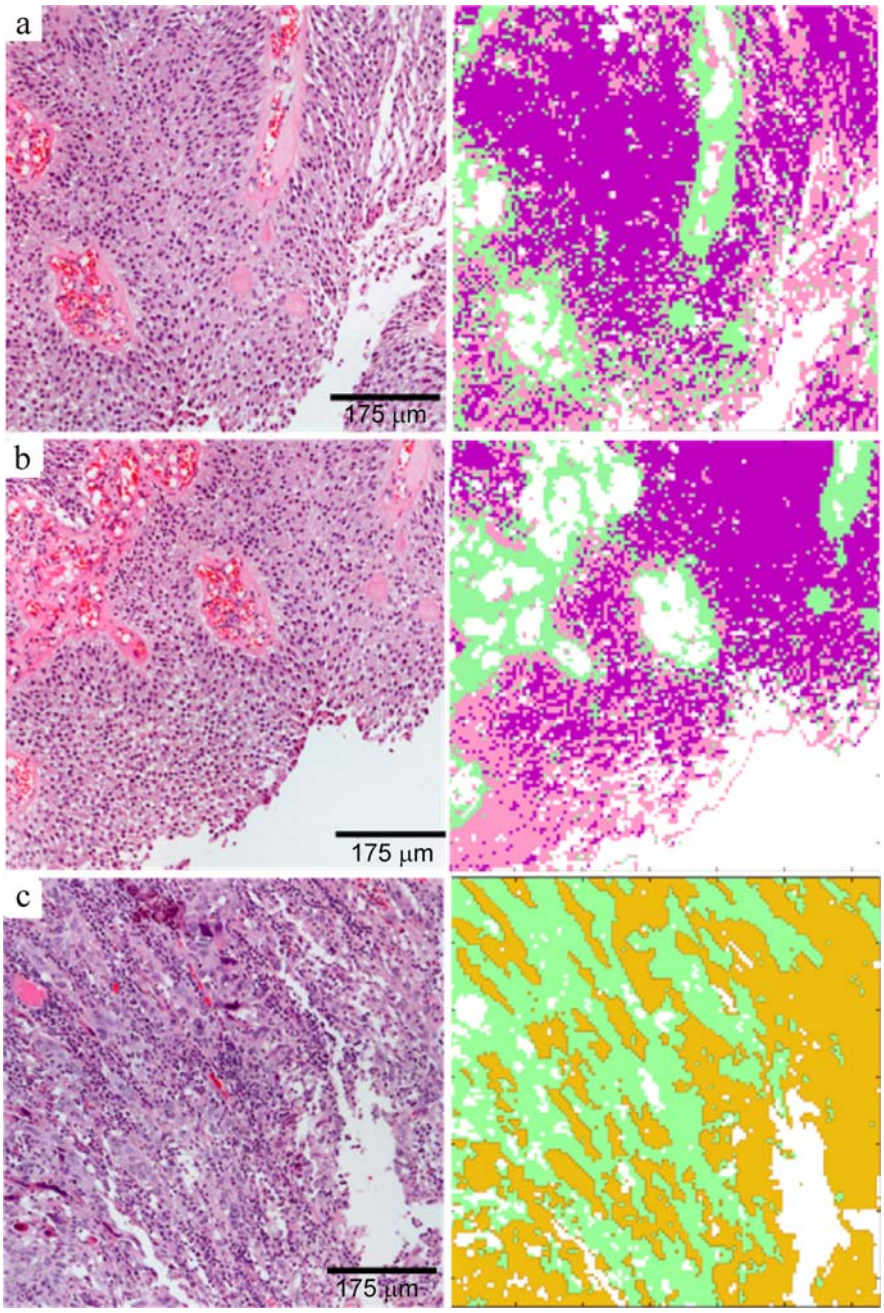


Figure 2 (online color at: www.biophotonics-journal.org) Lymphocyte-rich regions of TCC: Photomicrographs of the H & E stained tissue after IR acquisition (left) and the *K*-means pseudo-colour map associated with each infrared image (right). Lymphocyte-rich epithelium is classified (magenta/pink) and stroma (green) for two regions of patient Sample 2 (**a**, **b**). (**c**) displays TCC with prominent lymphoid stroma (green) and typical TCC (yellow) from patient Sample 3.

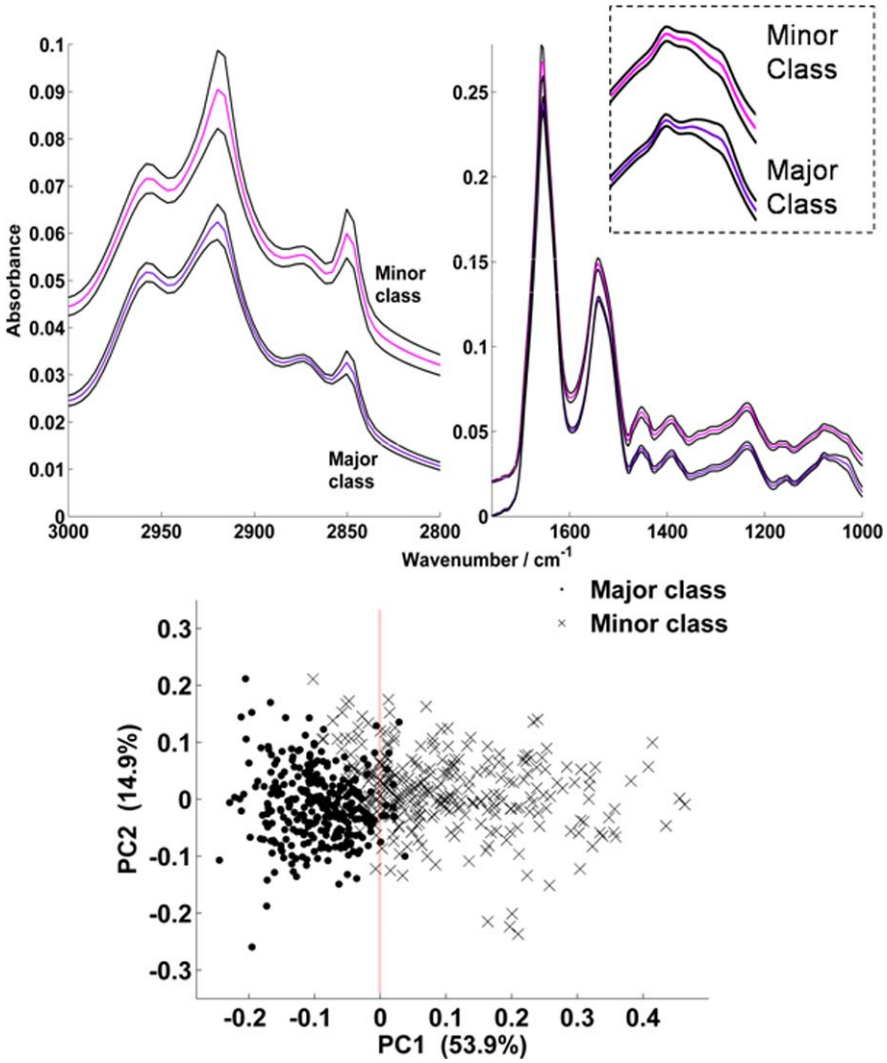


Figure 3 (online color at: www.biophotonics-journal.org) Mean representative spectra ± standard deviation (black lines) for the two classes of lymphocyte-associated tissue spectra; majority spectra (magenta) and minority (pink) in the *K*-means pseduo-colour maps in Figure 2a, b (top). PCA score plot of major and minor data points from 300 randomly selected second derivative spectra per class (below).

noma or lymphoepithelioma. Investigation of immunohistochemical and molecular characteristics of bladder LELC is limited [49]. LELC typically consists of high-grade tumour cells with a syncytial appearance, arranged in sheets or anastomosing nests with a prominent polyclonal lymphoplasmacytic infiltrate [1, 49]. Histological analysis of the lymphocyte-rich tissue in this study, however, found urothelial carcinoma with lymphocyte infiltration rather than lymphoepithelioma-like carcinoma due to a lack of syncytial growth pattern in the H & E photomicrographs. The presence of dense, tumour-infiltrating lymphocytes is said to be a highly significant predictor of favourable prognosis in invasive TCC [50]. Figure 2a and b display captured regions of lymphocyte-rich invasion in the epithelium of TCC (coloured magenta – major class/coloured pink – minor class) and stroma (green) by unsupervised cluster analysis. Figure 2c displays a different example, where the lymphocytes are restricted to stromal regions around otherwise conventional TCC. In both cases the lymphocyte-rich epithelial tissue spectra displayed pronounced symmetric $\nu_s(1080\ \mathrm{cm}^{-1})$ and asymmetric $\nu_{as}(1240\ \mathrm{cm}^{-1})$ phosphodiester stretches (Figure 3), notable spectral features reported in FTIR studies of lymphocyte cells [51, 52]. There were observable differences between the two lymphocyte classes (Figure 3). The majority spectra displayed a more pronounced shoulder in the carbohydrate region associated with glycogen at ~1030 cm^{-1}. It is possible that an increased rate of glycogen consumption in the minor class may be indicative of greater metabolic activity [53]. The minority class of spectra (pink) exhibit larger variance, particularly in the lipid regions at higher wavenumbers. This variance was reflected in PCA space when 300 randomly selected 2nd derivative spectra per class were PC analysed; the minority data points projecting more extensively in PC1 (Figure 3). Loadings for PC1 can be found in Supp. Note 5 where differences are noted for the symmetric and anti-symmetric stretching of CH_2, protein differences in the amide I

and II bands and the C–O stretch associated with glycogen. It is hypothesized that the spectral differences within the minority tissue are possibly due to the disruption of the surrounding matrix, similar to the observations in the micro-papillary variant, as larger matrix-associated spacing is observed in the minority spectral regions in the H & E.

3.1.3 *Clear cell sub-variant*

Clear cell change can be seen in some TCCs, although it is uncommon. Histological findings consist of tumour cells with abundant clear cytoplasm occurring in the bladder or upper urinary tract. An alternative situation where clear cells feature, is in clear cell adenocarcinoma of the urinary tract [54]. This is somewhat different from TCC with clear cell change and different again from clear cell renal carcinoma. The clear cell variant of TCC has high levels of cytoplasmic glycogen, resulting in clear appearance. Its distinction from clear cell adenocarcinoma is usually without difficulty. The clear cell sub-variant is, however, difficult to identify in poorly differentiated carcinoma of the upper tract (renal pelvis/upper ureter) and to distinguish from high grade clear cell renal cell carcinoma [55, 56]. An objective approach for distinguishing the two histological variants would be particularly useful.

The H & E photomicrographs (Figure 4) show regions of TCC with mostly conventional appearance, however at higher magnifications, sub-regions were found to be potentially indicative of clear cell change (Supp. Note 6). Morphological changes in the H & E, however, were relatively subtle compared to the dramatic changes in the infrared signatures extracted from the unsupervised classification. Figure 5 displays the mean spectrum of extracted classes from the *k*-means classification matrix of (Figure 4c) where typical cells, stroma, presumable 'clear cell change' and extensive clear cell chemistry are classified as brown (dark yellow), green, blue and red respectively. A second example from this Patient Sample 5 (Figure 4c) can be found in Supp. Note 7. The blue class has been assigned as an emerging clear cell class as the mean spectrum displays somewhat transitional biochemistry between the conventional

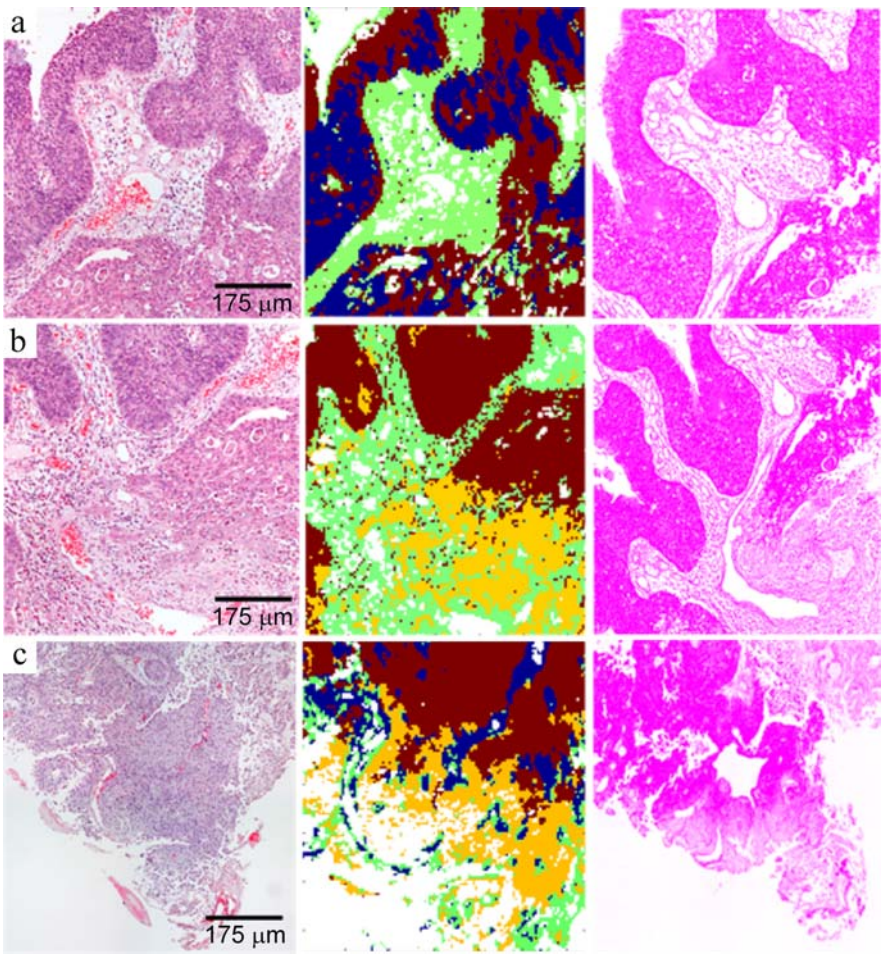


Figure 4 (online color at: www.biophotonics-journal.org) Photomicrographs of the H & E stained tissue after IR acquisition (left), the pseudo-colour *k*-means classification matrix (middle) and photomicrographs of serial sections stained for PAS (where the intensity of the pink staining is proportional to the proportion of glycogen) for each infrared image (right). Two regions shown are from Sample 5 (**a**, **b**) and one from Sample 6 (**c**). Clear cell are classified by blue and red; with red having the most developed glycogen signal in the spectra. Areas of conventional transitional cells and stroma are classified yellow and green respectively.

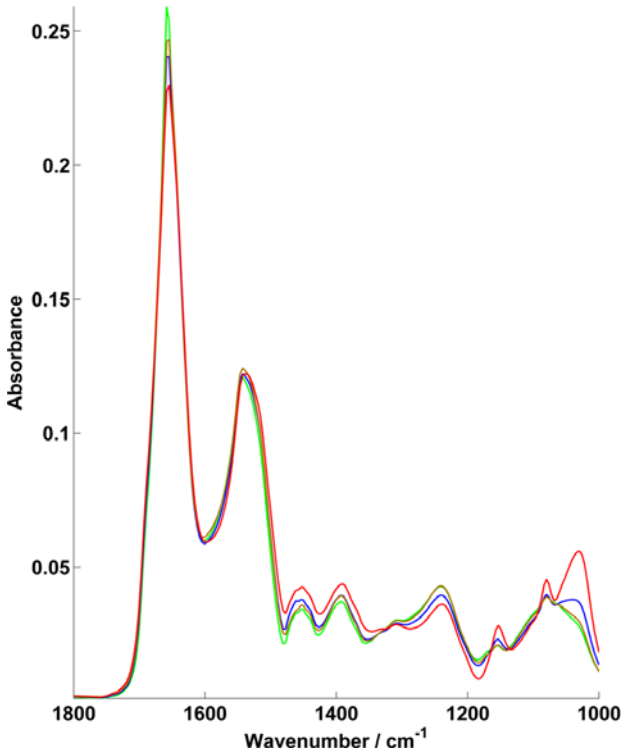


Figure 5 (online color at: www.biophotonics-journal.org) Mean representative spectrum for the classes of *k*-means image (Fig. 4c) illustrating a dramatic increase in carbohydrates in the clear cell region (red class). It is presumed that an intermediate signature, indicating progression to clear cell development is also detected (blue class). Conventional TCC and stroma-classed spectra are coloured brown (dark yellow) and green respectively.

urothelial cell and the clear cell sub-variant; a substantial increase in lipids yet with an increasing shoulder at ~1030 cm^{-1}, suggestive of glycogen accumulation. The serial sections were stained with PAS for glycogen presence, where deeper staining is synonymous with greater relative abundance. This gave confirmatory evidence of the clear cell phenotype and the associated high infrared signal of glycogen at ~1030 cm^{-1}

3.1.4 Lipoid sub-variant

Lipoid cells are large carcinoma cells with optically clear empty multi-vacuolated cells resembling lipoblasts. They are recognised by the WHO classification as relatively rare TCCs: very few cases have been reported [30, 57]. They may mimic a clear-cell appearance, but differ due to the multi-vacuolated appearance and negative PAS staining (Figure 6) [58]. The presence of lipid has not convincingly been proven by histochemistry, leaving the chemical contents a mystery; hence the term 'lipoid' variant. TCC with lipoid features is an important variant, as it can pose diagnostic difficulties, especially in limited biopsy samples in which it may be mistaken for a sarcomatoid carcinoma or secondary tumour. The TCC involved is usually high grade and invasive, however there is no known prognostic significance of this histological variant of bladder cancer. The example found in Patient Sample 6 presented with infiltrative growth and frank necrosis (Supp. Table 2).

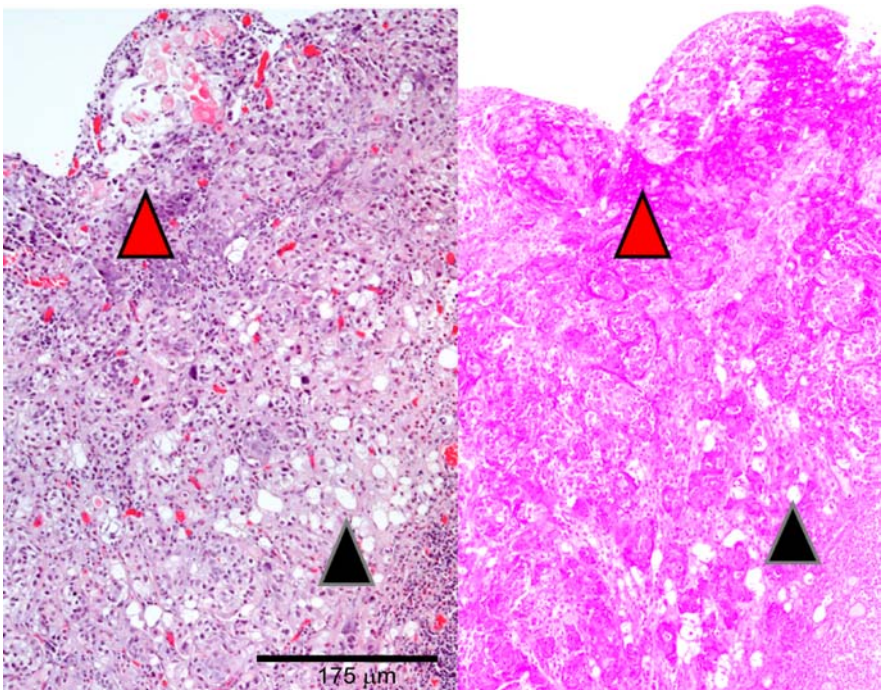


Figure 6 (online color at: www.biophotonics-journal.org) H & E (left) displaying a multi-differentiated TCC tumour with a smaller degree of clear cell regions (red arrow) and lipoid cell variants (black arrow). In a serial section, the glycogen-rich clear cells stain positively for PAS whereas the lipoid cells are PAS negative (right).

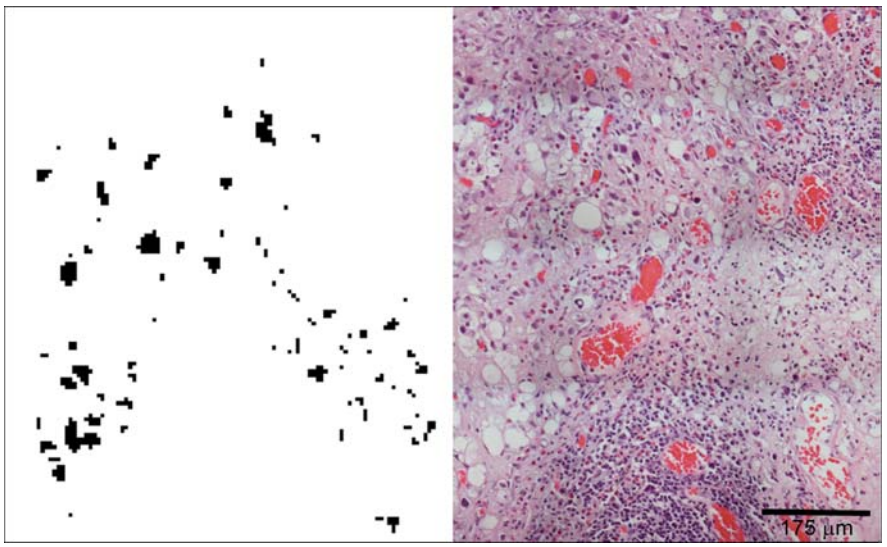


Figure 7 (online color at: www.biophotonics-journal.org) To assess the chemistry within the lipoid cells, a low-intensity of absorbance threshold was applied (left), in conjunction with the corresponding H & E (right), to obtain an accurate lipoid co-ordinate binary mask. The mask was then used to extract the lipoid-associated spectra (black class).

Spectra were extracted from these lipoid cells by using a total intensity of absorbance threshold due to the cells' uniquely low overall intensity (Figure 7) across two IR images. The lipoid spectra were found to be protein and carbohydrate-rich, although generally less abundant in lipids than the clear cell variant.

3.2 *Collective spectral analysis*

Mean spectra of each identified class are shown in Figure 8. The most distinctive signatures are unsurprisingly from the clear cell sub-variant, which is known to have high levels of glycogen and lipids. The lipoid cells also have a similar spectral profile. Mean 2nd derivative equivalent spectra were compared for each class (Figure 9). The clear cell derivatives have the strongest CH_2 signals whereas the lipoid cells dominate at the protein amide I band (Figure 9b). The clear cell and lipoid spectra display pronounced carbohydrate peaks at 1030, 1080 and 1150 cm^{-1} assigned to carbohydrate signatures (a wavenumber shift is observed for the clear cell derivative at 1026 cm^{-1}) (Figure 9c). The lipoid cells are also distinguishable at the anti-symmetric phosphodiester stretch at 1240 cm^{-1}

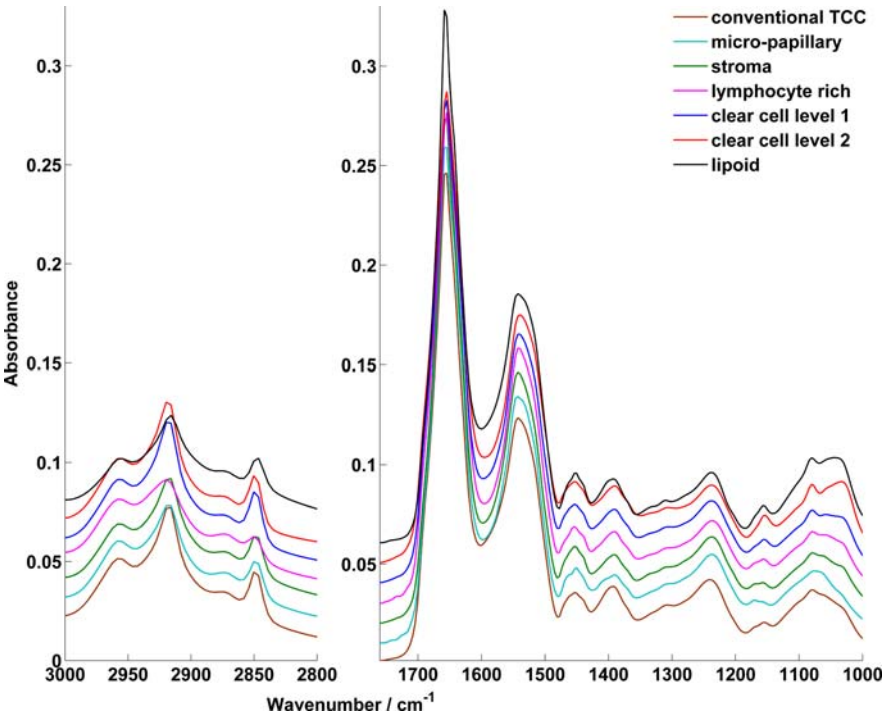


Figure 8 (online color at: www.biophotonics-journal.org) Spectra from Table 1 were combined and vector-normalised as a total dataset, then averaged to represent the mean class spectrum. (Displayed spectra are artificially super-imposed for clarity).

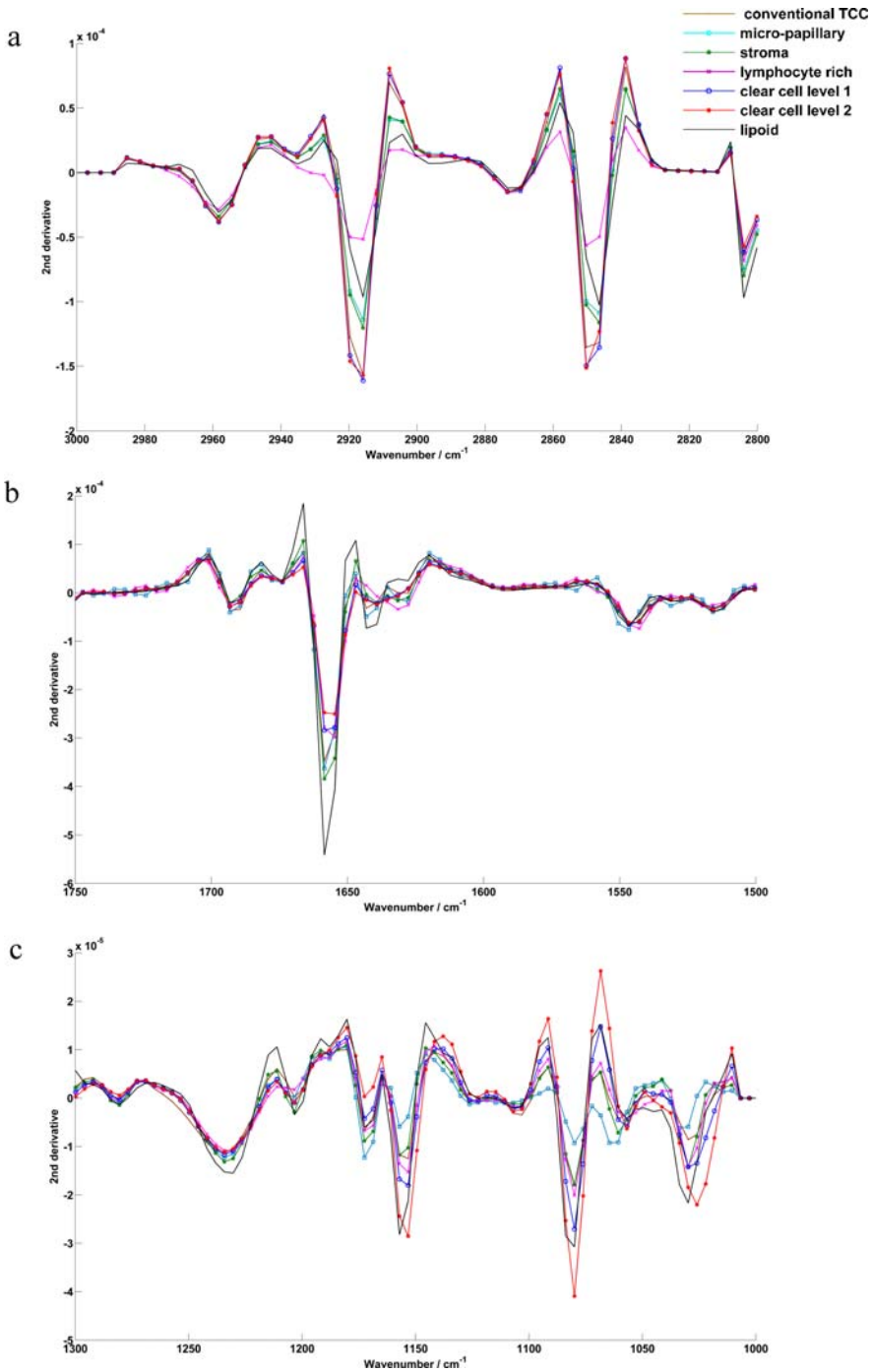


Figure 9 (online color at: www.biophotonics-journal.org) Mean representative 2nd derivative spectra per class (**a**) for 2800–3000 cm^{-1} (**b**) 1500–1750 cm^{-1} and 1000–1300 cm^{-1} (**c**) respectively.

3.2.1 Principal component analysis

A balanced sample set of 175 randomly chosen 2nd derivative spectra was selected per class for principal component analysis (Figure 10a, b) (Lipoid spectra were omitted due to low spectral observations). Each class displayed coherent clustering in PC space, with the exception of the micro-papillary variant (Figure 10a). This is most likely due to its highly aggressive phenotype that leads to disruption and replacement of the normal stromal matrix and so causes more transitional chemical variation (extensive stromal disruption found in the H & E images are highlighted by the black arrows in Figure 1a).

For clarity, PCA was repeated with the removal of the micro-papillary variant (Figure 10b). The PCA process was repeated several times to ensure that randomly extracted spectra gave an accurate representation of the full spectral population (Supp Note 8). PCA is an unsupervised chemometric techni-

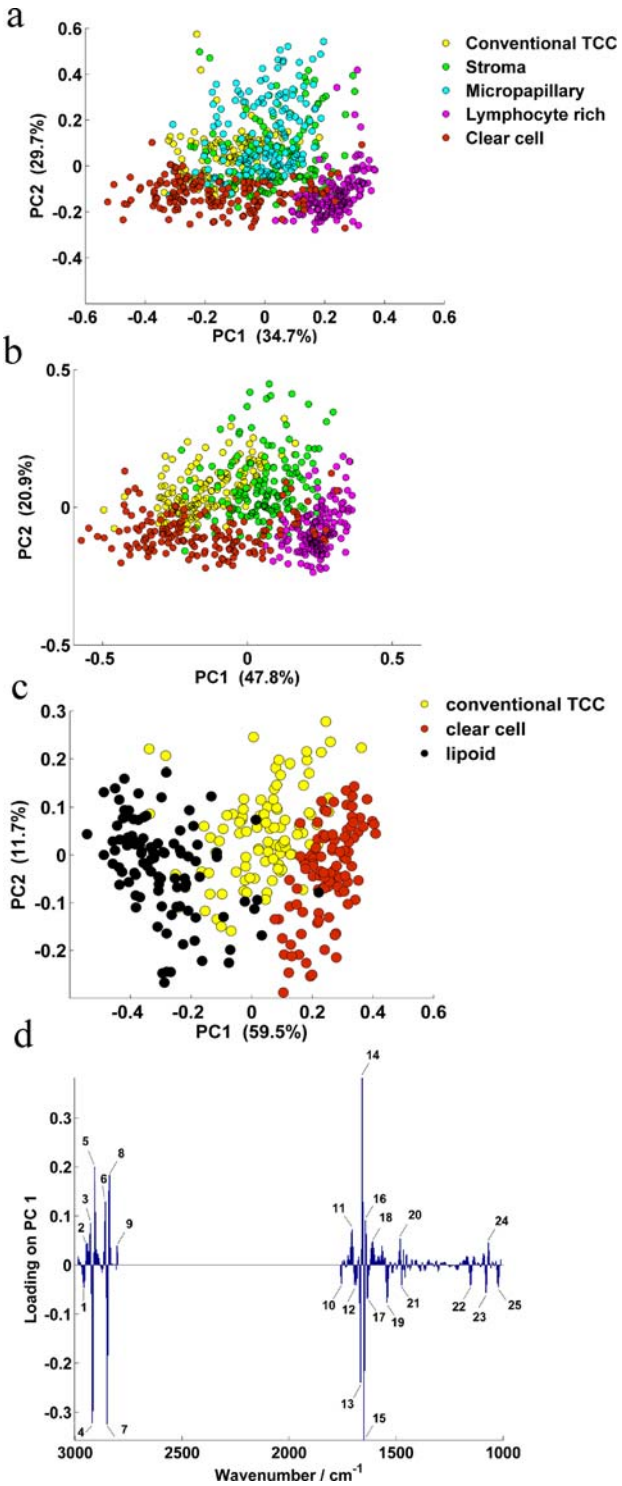


Figure 10 (online color at: www.biophotonics-journal.org) PCA score plot of 5 (**a**) and 4 (**b**) of the classed tissue types. PCA of clear cell versus the lipoid cell spectra against the conventional TCC (**c**). Loading plot of PC1 for the corresponding lipoid/clear cell/conventional TCC data PCA score plot (**d**). Notable spectral features (1–25) are tabulated in Supp. Table 3.

que with no *a priori* knowledge of class assignment; therefore it was encouraging to observe fairly coherent data group clusters, despite the fact that many classes were made up from spectra of different patient samples. This gave confidence that chemometric separation of class types were not simply due to a separation based on individual patient signatures due to the low sample numbers. The sub-variant clusters projected in opposite directions from the centroid of conventional TCC and stroma spectra; suggestive of a common shared biochemical origin with an eventual differentiation into independent classes.

The chemical differences between the lipoid and clear cell classes were explored by PCA with 100 spectra per class. Interestingly, the clear cell and lipoid cell spectra, which displayed similar spectral profiles, were anti-correlated in PC1, in respect to the conventional TCC. The loading plot for PC1 is shown in Figure 10. Key regions most influential in the class differentiation between the lipoid and clear cell predominantly stem from protein, lipid and carbohydrate contributions; namely vibrations from the anti-parallel β-sheet (*13*) and (C=O) stretching vibration of the amide I (*15*); stretching C–H and CH_2 (*7–11*); C–O stretching mode of the carbohydrates (*22*), symmetric stretching of the phosphodiester groups (*23*) and C–O stretching associated with glycogen (*25*). Full contributing wavenumber assignments (*1–25*) in the loadings for PC1 can be found in Supplementary Table 3.

3.2.2 Support Vector Machine (SVM) Classification

The extracted spectra of the highlighted subvariant tissue structures of TCC were compiled from a small number of cases. To illustrate the potential for future classification, the example database was evaluated for its predictive capabilites using a 50,000-strong spectral dataset.

An SVM with RBF kernel was trained with 3-fold cross-validation using a 25,000 spectal training set, comprised of 5,000 randomly selected spectra per class. The OVR approach was adopted whereby the decision boundary separates between class '*k*' versus the rest of the classes. The test set comprised of an equal number of respective class spectra (5,000). Figure 11 displays the binary confusion matrix summary where spectra have been classed as 1 (positive for the 'k'th class) or 0 for negative (the rest of classes). The resulting classification accuracy for classes 1–5 was 99.444 (conventional TCC), 98.496 (stroma), 99.972 (micro-papillary), 99.424 (lymphocyte rich) and 99.096% (clear cell) respectively, giving a total accuracy from the SVM of 98.36%. The absolute decision

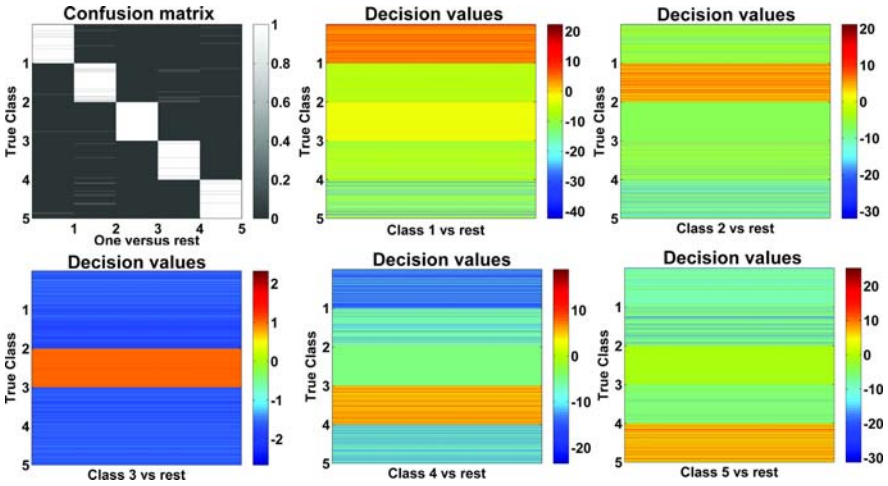


Figure 11 (online color at: www.biophotonics-journal.org) Confusion matrix of classified test spectra with an overall accuracy of 98.36%. The individual decision values for each class of 5000 spectra are also shown for conventional TCC (1), stroma (2), micro-papillary (3), lymphocyte rich (4) and clear cell (5) tissue spectra.

boundary values for each OVR classification are also shown in Figure 11. These values represent the distance from the separating hyperplane that was determined during training. The absolute decision values can be interpreted as a measure of confidence of the predicted classification; the larger the value, the more confident the classifier is in the decision [59]. Despite the chemical similarity indicated in the PCA score plot (Figure 10a), the classification performance of 'stroma versus rest' (class 2) and 'micro-papillary verus rest' (class 3) with the SVM was excellent.

4. Conclusion

Spectral pathology is an emerging field that has shown its diagnostic capability by providing biochemical qualification of clinical findings [60]. The technique may be particularly useful for TCC of the bladder to avoid diagnostic errors. This is because a variety of unusual architectural patterns of TCC can be mistaken for reactive processes or benign tumours, whilst others can mimic metastatic tumour from other sites [1]. There are also cases of poorly differentiated carcinomas with ambiguous morphology.

The application of spectral pathology via infrared microspectroscopy has demonstrated that tumour heterogeneity may be easily recognised and differentiated by chemical composition alone, with the use of staining required only as a validatory measure. Despite the low patient cases, this unique study has led to the spectral profiling of several tissue structures associated with TCC and its sub-variants. The clear cell variant, for example, can occur in several tissues and is generally known for its ability to accumulate high levels of glycogen [61–64]. This was reflected in its spectral phenotype with an atypically high absorbance in the infrared, a feature rarely observed in other tissue spectra.

The next step, beyond confirming the ability to identify and classify spectral signatures of the sub-variants studied with greater patient cases, would be to see whether clear cell bladder carcinoma can be differentiated by its spectroscopic signal from high-grade renal clear cell carcinoma of the upper tract. A similar aim would also be to discriminate between the micro-papillary phenotype that is morphologically identical in cancers from several organs [42]. If achievable, it would leave no question that spectral pathology could potentially have a critical role in the clinical diagnosis of rare and challenging cases.

The ability to localise spectral signals from cells using imaging capabilities has also led to an initial profiling of chemical constituents within the rare lipoid sub-variant cells. It is not clear whether the poor prognosis is due to the lipoid cells or merely a reflection of the association of this histology with high stage TCC [30]. The cells have revealed a protein and carbohydrate-rich spectral phenotype, rather than a lipid rich-composite; a significant finding that may help to unravel the mystery of the so-called lipoid cells [31, 58].

Acknowledgements The Engineering and Physical Sciences Research Council (EPSRC) for funding. LIBSVM Software freely available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. SVM guidance and codes written by Kittipat Kampa, Integrated Brain Imaging Center (IBIC), University of Washington, Seattle.

Author biographies Please see Supporting Information online.

References

[1] J. H. Shanks and K. A. Iczkowski, Histopathology **54**(7), 885–900 (2009).

[2] M. B. Amin, J. Y. Ro, T. El-Sharkawy, K. M. Lee, P. Troncoso, E. G. Silva, N. G. Ordonez, and A. G Ayala, Am. J. Surg. Pathol. **18**(12), 1224–1232 (1994).
[3] C. Krafft, L. Shapoval, S. B. Sobottka, G. Schackert, and R. Salzer, Technol. Cancer Res. Treat. **5**(3), 291–298 (2006).
[4] B. R. Wood, K. R. Bambery, L. M. Miller, M. A. Quinn, L. Chiriboga, M. Diem, and D. McNaughton, Proc. SPIE **5651**, 78–84 (2005).
[5] B. R. Wood, K. R. Bambery, C. J. Evans, M. A. Quinn, and D. McNaughton , BMC Med. Imag. **6**, 12 (2006).
[6] P. Lasch, W. Haensch, D. Naumann, and M. Diem, Biochim Biophys Acta **1688**(2), 176–186 (2004).
[7] V. Sablinskas, G. Steiner, E. Koch, J. Ceponkus, M. Pucetaite, S. Strazdaite, V. Urboniene, and F. Jankevicius, Proc. SPIE **7902**, 790221 (2011).
[8] B. Bird, M. Miljkovic, M. J. Romeo, J. Smith, N. Stone, M. W. George, and M. Diem, BMC Clin. Pathol. **8**(1), 8 (2008).
[9] B. Bird, M. Romeo, N. Laver, and M. Diem, J. Biophotonics **2**(1/2), 37–46 (2009).
[10] C. Conti, P. Ferraris, M. Garavaglia, E. Giorgini, C. Rubini, S. Sabbatini, and G. Tosi, Microsc. Res. Tech. **72**(2), 67–75 (2009).
[11] S. F. Chew, B. R. Wood, C. Kanaan, J. Browning, D. MacGregor, I. D. Davis, J. Cebon, B. D. Tait, and D. McNaughton, Tissue Antigens **69**(Suppl. 1), 252–258 (2007).
[12] G. Tosi, C. Conti, E. Giorgini, P. Ferraris, M. G. Garavaglia, S. Sabbatini, S. Staibano, and C. Rubini, Analyst **135**(12), 3213–3219 (2010).
[13] R. Zendehdel, A. Masoudi-Nejad, and F. H. Shirazi, Iran. J. Pharm. Res. **11**(2), 401–410 (2012).
[14] C. Hughes, M. Liew, A. Sachdeva, P. Bassan, P. Dumas, C. A. Hart, M. D. Brown, N. W. Clarke, and P. Gardner, Analyst **135**(12), 3133–3141 (2010).
[15] T. J. Harvey, C. Hughes, A. D. Ward, E. C. Faria, A. Henderson, N. W. Clarke, M. D. Brown, R. D. Snook, and P. Gardner, J. Biophotonics **2**(1/2), 47–69 (2009).
[16] P. Lasch, M. Diem, W. Hänsch, and D. Naumann, J. Chemom. **20**(5), 209–220 (2006).
[17] S. Boydston-White, M. Romeo, T. Chernenko, A. Regina, M. Miljkoviæ, and M. Diem, Biochim. Biophys. Acta – Biomembranes **1758**(7), 908–914 (2006).
[18] M. Sattlecker, R. Baker, N. Stone, and C. Bessant, Chemom. Intell. Lab. Syst. **107**(2), 363–370 (2011).
[19] C. Cheng, J. Liu, C. Zhang, M. Cai, H. Wang, and W. Xiong, Appl. Spectrosc. Rev. **45**(2), 148–164 (2010).
[20] N. Bergner, B. F. M. Romeike, R. Reichart, R. Kalff, C. Krafft, and J. Popp, Progress in Biomedical Optics and Imaging Proc. SPIE, Munich (2011).
[21] Y. Liang, Q. S. Xu, H. D. Li, and D. S. Cao, Support Vector Machines and Their Application in Chemistry and Biotechnology (Taylor & Francis CRC Press, 2011).
[22] M .C. M. Grimbergen, C. F. P. Van Swol, R. O. P. Draga, P. Van Diest, R. M. Verdaasdonk, N. Stone, and J. H. L. R. Bosch, Proc. SPIE **7161**, 716114 (2009).
[23] A. Shapiro, O. N. Gofrit, G. Pizov, J. K. Cohen, and J. Maier, Eur. Urol. **59**(1), 106–112 (2011).
[24] V. Hanchanale, A. Rao, and S. Das, Indian J. Urol. **24**(4), 444–450 (2008).
[25] M. C. M. Grimbergen, C. F. P. van Swol, R. J. A. van Moorselaar, J. Uff, A. Mahadevan-Jansen, and N. Stone, J. Photochem. Photobiol. B **95**(3), 170–176 (2009).
[26] T. J. Harvey, E. C. Faria, A. Henderson, E. Gazi, A. D. Ward, N. W. Clarke, M. D. Brown, R. D. Snook, and P. Gardner, J. Biomed. Opt. **13**(6), 064004 (2008).
[27] L. Wang, J. H. Fan, Z. F. Guan, Y. Liu, J. Zeng, D. L. He, L. Q. Huang, X. Y. Wang, and H. L. Gong, Spectroscopy and Spectral Analysis, **32**(1), 123–126 (2012).
[28] N. A. Al-Muslet and E. E. Ali, Aust. J. Basic Appl. Sci. **5**(9), 1734–1739 (2011).
[29] N. A. Al-Muslet and E. E. Ali, J. Appl. Spectrosc. **79**(1), 139–142 (2012).
[30] P. Nigwekar and M. B. Amin, Adv. Anat. Pathol. **15**(4), 218–233 (2008).
[31] M. B. Amin, Modern Pathology **22**(Suppl. 2), S96–S118 (2009).
[32] A. C. Ruifrok and D. A. Johnston, Anal. Quant. Cytol. Histol. **23**(4), 291–299 (2001).
[33] P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas, and P. Gardner, Analyst **134**(8), 1586–1593 (2009).
[34] P. Bassan, A. Sachdeva, A. Kohler, C. Hughes, A. Henderson, J. Boyle, J. H. Shanks, M. Brown, N. W. Clarke, and P. Gardner, Analyst **137**(6), 1370–1377 (2012).
[35] P. Bassan, A. Kohler, H. Martens, J. Lee, H. J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke, and P. Gardner, Analyst **135**(2), 268–277 (2010).
[36] C. Ding and X. He, Proc. 21st Int. Conf. Machine Learning, Banff, Alberta, Canada, ACM (2004), 414–418.
[37] Y. B. Xie, Q. Liu, F. He, C. G. Guo, C. F. Wang, and P. Zhao, Chin. Med. J. **124**(16), 2517–2521 (2011).
[38] C. Ding and X. He, Proc. 21st Int. Conf. Machine Learning, Banff, Alberta, Canada, ACM, 2004.
[39] A. Beljebbar, N. Amharref, A. LeìveÌques, S. Dukic, L. Venteo, L. Schneider, M. Pluot, and M. Manfait, Anal. Chem. **80**(22), 8406–8415 (2008).
[40] C. C. Chang and C. J. Lin, ACM Transact. Intell. Syst. Technol. **2**(3), 1–27 (2011).
[41] A. M. Perepletchikov and A. V. Parwani, Pathol. Res. Pract. **205**(12), 807–810 (2009).
[42] P. Ramalingam, L. P. Middleton, P. Tamboli, P. Troncoso, E. G. Silva and A. G. Ayala, Ann. Diagn. Pathol. **7**(2), 112–119 (2003).
[43] P. G. Andrus, Technol. Cancer Res. Treat. **5**(2), 157–167 (2006).
[44] R. Jinka, R. Kapoor, P. G. Sistla, T. A. Raj, and G. Pande, Int. J. Cell Biol. **2012**, 219196 (2012).
[45] P. Lu, V. M. Weaver, and Z. Werb, J. Cell Biol. **196**(4), 395–406 (2012).
[46] G. Acs, K. L. Dumoff, L. J. Solin, T. Pasha, X. Xu, and P. J. Zhang, Am. J. Surg. Pathol. **31**(1), 129–140 (2007).

[47] A. M. Perepletchikov and A. V. Parwani, Pathol. Res. Pract. **205**(12), 807–810 (2009).
[48] G. Acs, G. Paragh, Z. Rakosy, C. Laronga, and P. J. Zhang, Mod. Pathol. **25**(2), 163–177 (2012).
[49] S. R. Williamson, S. Zhang, A. Lopez-Beltran, R. B. Shah, R. Montironi, P. H. Tan, M. Wang, L. A. Baldridge, G. T. MacLennan, and L. Cheng, Am. J. Surg. Pathol. **35**(4), 474–483 (2011).
[50] C. Jochems and J. Schlom, Exp. Biol. Med. (Maywood), **236**(5), 567–579 (2011).
[51] E. Benedetti, P. Vergamini, and G Spremolla, Microchim. Acta **94**(1), 139–141 (1988).
[52] B. R. Wood, M. A. Quinn, B. Tait, M. Ashdown, T. Hislop, M. Romeo, and D. McNaughton, Biospectroscopy **4**(2), 75–91 (1998).
[53] E. Gazi, J. Dwyer, P. Gardner, A. Ghanbari-Siahkali, A. P. Wade, J. Miyan, N. P. Lockyer, J. C. Vickerman, N. W. Clarke, and J. H. Shanks, J. Pathol. **201**(1), 99–108 (2003).
[54] B. Adams, R. Sawhney, A. Sheil, and U. B. Chaudhary, Am. J. Med. Sci. **333**(1), 63–65 (2007).
[55] D. Perez-Montiel, P. E. Wakely, O. Hes, M. Michal, and S. Suster, Mod. Pathol. **19**(4), 494–503 (2006).
[56] D. Perez-Montiel and S. Suster, Diagn. Histopathol. **14**(1), 48–54 (2008).
[57] T. Y. Chan, Urology, **65**(1), 214–215 (2005).
[58] X. Leroy, S. Gonzalez, L. Zini, and S. Aubert, Am. J. Surg. Pathol. **31**(5), 770–773 (2007).
[59] C. H. Yeang, S. Ramaswamy, P. Tamayo, S. Mukherjee, R. M. Rifkin, M. Angelo, M. Reich, E. Lander, J. Mesirov, and T. Golub, Bioinformatics **17**(Suppl. 1), S316–S322 (2001).
[60] E. Gazi, M. Baker, J. Dwyer, N. P. Lockyer, P. Gardner, J. H. Shanks, R. S. Reeve, C. A. Hart, N. W. Clarke, and M. D. Brown, Eur. Urol. **50**(4), 750–760; discussion 760–751 (2006).
[61] T. Kuo, Am. J. Surg. Pathol. **4**(6), 573–583 (1980).
[62] E. Dessy, M. Falleni, P. Braidotti, B. Del Curto, T. Panigalli, and G. G. Pietra, Arch. Pathol. Lab. Med. **125**(12), 1588–1590 (2001).
[63] D. P. Sarma, D. Olson, J. Olivella, T. Harbert, B. Wang, and S. Ortman, Patholog. Res. Int. **2011**, 386921 (2011).
[64] T. Terada, Breast J. **18**(3), 279–280 (2012).