Analyst

Check for updates

Evaluation of grade and invasiveness of bladder urothelial carcinoma using infrared imaging and machine learning†

Monika Kujdowicz, [ID] [a,b] David Perez-Guaita, [ID] *[c] Piotr Chlosta, [ID] [d] Krzysztof Okon [ID] [a] and Kamilla Malek [ID] *[b]

Urothelial bladder carcinoma (BC) is primarily diagnosed with a subjective examination of biopsies by histopathologists, but accurate diagnosis remains time-consuming and of low diagnostic accuracy, especially for low grade non-invasive BC. We propose a novel approach for high-throughput BC evaluation by combining infrared (IR) microscopy of bladder sections with machine learning (partial least squares-discriminant analysis) to provide an automated prediction of the presence of cancer, invasiveness and grade. Cystoscopic biopsies from 50 patients with clinical suspicion of BC were histologically examined to assign grades and stages. Adjacent tissue cross-sections were IR imaged to provide hyperspectral datasets and cluster analysis segregated IR images to extract the average spectra of epithelial and subepithelial tissues. Discriminant models, which were validated using repeated random sampling double cross-validation, showed sensitivities (AUROC) *ca.* 85% (0.85) for the identification of cancer in epithelium and subepithelium. The diagnosis of non-invasive and invasive cases showed sensitivity values around 80% (0.84–0.85) and 76% (0.73–0.80), respectively, while the identification of low and high grade BC showed higher sensitivity values 87–88% (0.91–0.92). Finally, models for the discrimination between cancers with different invasiveness and grades showed more modest AUROC values (0.67–0.72). This proves the high potential of IR imaging in the development of ancillary platforms to screen bladder biopsies.

Introduction

Bladder urothelial carcinoma (BC) constitutes 3% of cancer diagnoses worldwide. The available diagnostic methods undoubtedly need improvement.[1–5] The current diagnostic standard for bladder tumours (T) includes cystoscopy with transurethral biopsy and hematoxylin–eosin (HE) examination, urine cytology, radiological examinations, and molecular tests.[6–9] The implementation of appropriate patient treatment should include morphological verification and recognition of its type, stage, and grade. The most common bladder cancer is urothelial bladder carcinoma, which constitutes 90% of cases

in this localization. Histological examination is particularly difficult for small biopsies, especially if they are thermally damaged during their preparation. Moreover, unequivocal tissue morphology requires additional immunohistochemistry (IHC). The reproducibility of the histological recognition of BC features is approximately 60% while the reproducibility of its invasiveness amounts to 75% due to intra- and inter-observer variations in the subjective judgment of pathologists.[4,5] Misdiagnosis also results from the histological similarity of some tumours or poor cell differentiation in the tissue.

In the last decade, Fourier Transform Infrared (FTIR) microspectroscopy has been proposed as an alternative technique for the diagnosis of cancer in tissue. The IR spectrum shows absorption bands from different biochemical components including, proteins, lipids, carbohydrates, and DNA. Changes in the chemical composition of the tissue produced by the cancerous cells can be represented in the IR spectrum.[10–13] The main advantage of this technique is that it can be fast and easily automatable, without requiring staining and the inspection of a trained pathologist.[14] Infrared (IR) spectral histopathology offers information at a microscopic scale. Identified chemical changes appearing in cancers can indicate novel biomarkers and then serve as ancillary and diagnostic tools. IR

[a] Department of Pathomorphology, Faculty of Medicine, Jagiellonian University Medical College, Krakow, Grzegorzecka 16, 31-531, Poland
[b] Faculty of Chemistry, Jagiellonian University in Krakow, Krakow, Gronostajowa 2, 30-387, Poland. E-mail: kamilla.malek@uj.edu.pl
[c] Department of Analytical Chemistry, University of Valencia, 50 Dr. Moliner Street, Research Building, 46100 Burjassot, Valencia, Spain.
E-mail: david.perez-guaita@uv.es
[d] Department of Urology, Medical Faculty, Jagiellonian University Medical College, Krakow, Jakubowskiego 2, 30-688, Poland
†Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2an01583h

data sets are mathematically analysed using well-developed algorithms for unsupervised image segmentation in the first step and then for blind classification.[14] So far, IR spectroscopy and microscopy have been used for BC diagnosis by analysing urothelial cell lines,[15,16] urine sediment,[17–19] cytology,[20] and also bladder tissues.[21,22] A few studies showed differences in IR spectra between BC subtypes but from only nine patients or between control and BC when the bladder was probed by fiber-optics.[22,23] Other modalities have been also tested for the detection of BC, including Raman[24,25] and Coherent anti-Stokes Raman spectroscopy (CARS)[24] spectroscopy as well as Second-harmonic imaging (SHG) imaging.[25] A report by Demos *et al.* has demonstrated the detection of bladder carcinoma from autofluorescence signal in Raman spectra from fresh bladder tissues but without the recognition of the cancer aggressiveness.[24] Whilst Raman, CARS, and SHG imaging perfectly have differentiated urothelial cells in urine for the classification of high grade BC from control patients.[25]

The most important predictive factor in BC diagnosis is staging referring to the degree of tumour infiltration into the bladder and metastases. Non-invasive BC (Tni here) includes *in situ* and papillary BC (Tis and Ta, respectively). Invasive stages are denoted as follows, T1 – tumour present in subepithelial tissue only, T2 – tumour invading the muscularis propria, T3 – tumour invasion to perivesical connective tissue, and T4 – tumour invasion to other organs.[6,26,27] A grading system was also introduced by the International Society of Urological Pathology and WHO to express the low (LG) and high (HG) malignant potential of urothelial carcinoma.[6,28] The currently used WHO system aims at assigning tumours to different prognostic groups, which then allows for appropriate treatment through transurothelial resection of tumour or cystectomy.[6] A significant number of patients are nowadays diagnosed with non-invasive BC (51%), then localized (T1, 34%), and regional BC (T2 and T3, 7%). Approximately 5% of cases are urothelial BC with metastasis (T4).[1] The survival rate is *ca.* 95% for Tni and is lower for higher stages, for example, 69.5 and 36.3% for T1 and T2, respectively.[1]

Since clinical classification of BC is not trivial, this study evaluates FTIR microspectroscopic as digital pathology to discriminate BC stage and grade in a large cohort of biopsies from 50 patients. IR imaging identified not only the tumour region as

in HE but also other tissue types as confirmed by the certified pathologists. Next, we examined each of them to select promising candidates for further discrimination analysis of the clinical importance. Finally, we determined Receiver Operating Characteristic (ROC) curves to assess accuracy. Our FTIR-based models are directly related to clinical histopathology procedures.

Experimental

The study was accepted by the First Local Ethics Committee of the Jagiellonian University Medical College in Krakow – UJ CM (no. 1072.6120.100.2018). Patient consent was not necessary because we used only the remaining sample after standard diagnostics. Permission to use human tissues for research was obtained with the supervision of clinically-qualified staff in accordance with Polish law.

Sample collection and clinical classification

Clinical characteristics of fifty patients are listed in Table 1. Bladder excisions were collected during cystoscopy according to a standard operating procedure in the Department of Urology of UJ CM after standard clinical diagnosis of suspicion of bladder cancer. The inclusion criterion was adult, and the exclusion criteria were pregnancy, history of radiotherapy in the pelvic region, and bladder cancer other than the urothelial type. Biopsy samples were fixed with 4% buffered formalin and embedded in paraffin in the Department of Pathomorphology, UJ CM. Adjacent cross-sections were placed on CaF_2 and glass slides for FTIR imaging and staining, respectively. Sections were prepared at a thickness of 7 and 3.5 μm, respectively. Expression of glucose transporter 1 protein (Glut-1) was determined according to a standard IHC protocol employed in UJ CM (1 : 200 rabbit polyclonal antibody; Bioassay Technology Laboratory, Shanghai, China). Photographic documentation was collected by using an Olympus BX53 light microscope equipped with an Olympus DP27 digital camera. A histopathological assessment was performed according to guidelines of WHO 2017 and the International Society of Urological Pathology.[6] appropriate.

Patients were assigned to five groups: (1) normal urothelium (N), (2) potentially malignant changes (PM), (3) BC

Table 1 Clinical characteristics of 50 patients with pathological classification

	N	PM	Tni	T1	≥T2
Number of patients (female : male)	9 (3 : 6)	8 (2 : 6)	13 (7 : 6)	13 (2 : 11)	7 (1 : 6)
Average age (SD)	63.6 (10.2)	71.1 (12.5)	67.5 (11.4)	70,8 (6.4)	69.3 (7.2)
LG/HG BC	NA	NA	12/1	6/7	0/7
Papillary morphology	NA	2	8	4	1
Number of patients with Glut-1 membranous expression in the epithelium	0	2	9	11	3
Number of excisions	18	16	33	29	18

Abbreviations: N – normal; PM – potentially malignant group, including dysplasia, hyperplasia, and PUNLMP (Papillary Urothelial Neoplasm of Low Malignant Potential); Tni – non-invasive BC, including BC *in situ* and papillary urothelial carcinoma (Ta); T1 – BC invading subepithelial (lamina propria) tissue; ≥T2 – BC invading deep bladder wall and adjacent structures, it includes as follows: invasion of muscularis propria (T2, N: 5), perivesical tissue (T3, N: 0) and adjacent organs (T4, N: 2); LG/HG BC – low- and high-grade bladder carcinoma; Glut-1 – IHC results for expression of glucose transporter; NA – not applicable.

without invasion (Tni) and BC infiltrating deeper tissues (Tinv), including (4) invasion to subepithelial connective tissue (T1), and (5) invasion of muscularis propria or perivascular tissue (T2 and higher stages; ≥T2).[6,26] We additionally assessed LG and HG BC *c.f.* Table 1. The PM group included dysplasia, hyperplasia and papillary urothelial neoplasm of low malignant potential (PUNLMP), and it is the most heterogeneous group in this work and we grouped all cases with any doubts due to their unequivocal nature, morphology, and treatment guidelines.[6,26,28,29–32]

FTIR imagining

Tissue sections deposited on CaF_2 substrates were deparaffinized in xylene and ethanol baths before IR imaging. An Agilent 670-IR FTIR spectrometer coupled with a 620-IR microscope and a focal plane array (FPA) detector was employed (Santa Clara, CA, USA). The detector consisted of a matrix of 16 384 pixels, arranged in a 128 × 128 grid format. IR images were acquired by illumination of samples through a Cassegrain 15× (NA: 0.62) lens giving a projected FPA pixel size of 5.5 µm × 5.5 µm (Standard Definition (SD) FTIR imaging). An imaged area was adjusted to the size of the tissue cross-section (approx. 4900 µm × 3500 µm) and one IR image was acquired from this area (Fig. 1). We applied also high spatial sampling with 5× enhancing optics with a projected FPA pixel size of 1.1 µm × 1.1 µm (High Definition (HD) FTIR imaging). All transmission FTIR spectra were recorded by co-adding 32 (SD) and 128 (HD) scans in the range of 3700–900 cm^{-1} with a spectral resolution of 4 cm^{-1}.

Data acquisition and pre-treatment

Pre-processing and chemometric analysis of collected FTIR images was performed using CytoSpec (ver. 2.00.01),[33] MATLAB (R2021a, MathWorks, Natick, MA, USA), Unscrambler X (v. 10.5, Camo, Montclair, NJ, USA), and Origin 9.1 (ver. 2020b, OriginLab, OriginLab Corporation, Northampton, MA, USA) software.

Subsequently, the removal of water vapour lines, quality test (based on signal-to-noise ratio in the regions of 1700–1600 and 1800–1900 cm^{-1}, respectively), PCA-based denoising (10 PCs), calculation of second derivative spectra (Savitzky–Golay; 13 points) and vector normalization (spectral region: 1780–1000 cm^{-1}) were performed before segmenting images (CytoSpec, MATLAB). Unsupervised hierarchical cluster analysis (UHCA) was executed in the 1780–1000 cm^{-1} region using the second derivative FTIR spectra. Spectral distances were calculated as D-values while individual clusters were extracted according to a Ward algorithm. The number of UHCA classes was adjusted according to HE tissue morphology. In that way, IR images were clustered into classes of tissue structures (UHCA maps) and their mean FTIR spectra were used for building models and their verification. In some cases, more than one class had to be assigned to epithelial and subepithelial tissues due to the complexity of tissue organisation.

IR images after pre-processing described above were also used to construct the distribution of carbohydrates by calculating integral intensity of the IR spectra in the region of 1200–1000 cm^{-1} (CytoSpec).
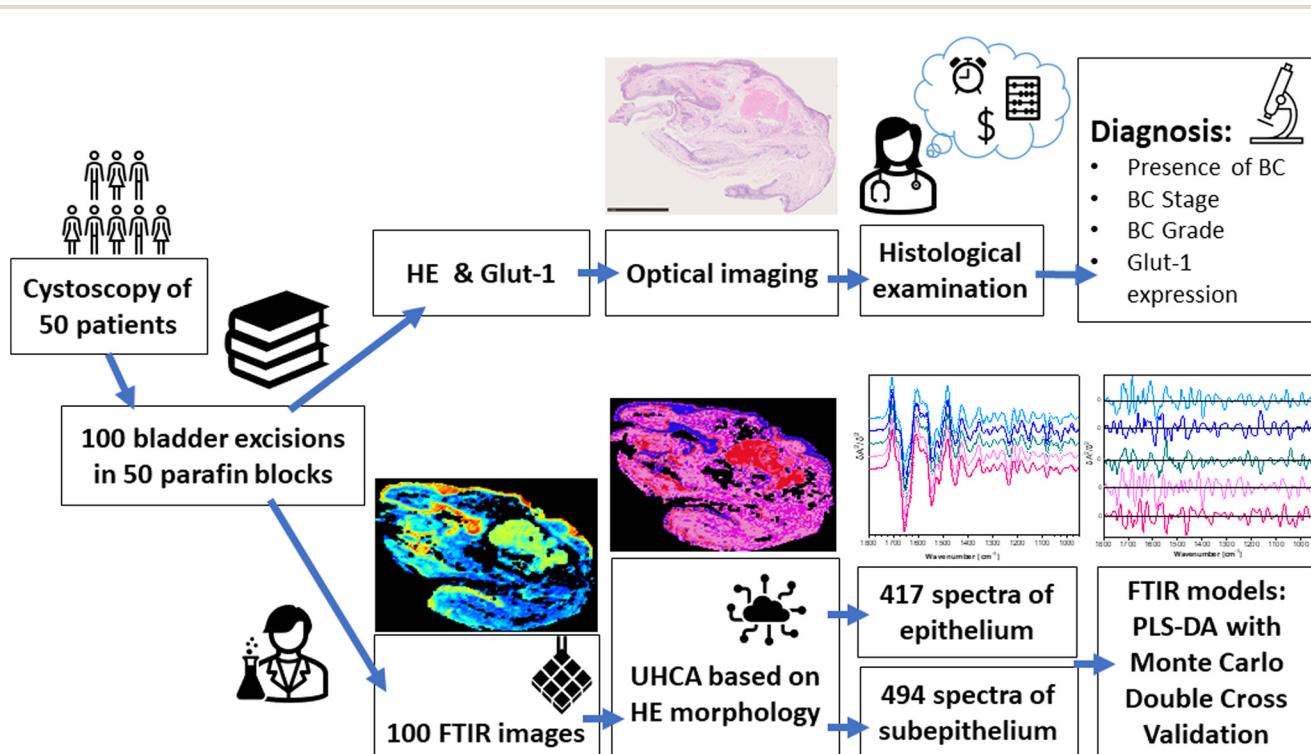
Fig. 1 The data collection and analysis scheme.

Mean second derivative UHCA spectra were then used to train a diagnostic classification model. In total, we obtained 417 and 494 spectra extracted from epithelial and subepithelial tissues, respectively. Datasets included 62/81 (N), 44/61 (PM), 124/140 (Tni), 94/122 (T1), 83/84 (≥T2), 167/192 (LG), and 134/159 (HG) spectra of epithelium/subepithelium, respectively.

Model building and testing

Standard Normal Variate and mean centring were performed before modelling in the 950–1800 cm^{-1} region. Repeated random sampling double cross-validation (DCV) was used to (1) optimize the model and (2) provide a realistic estimation of classification errors when the model is applied to external patients. In the outer loop, the spectra were randomly split between calibration (66%) and test (34%) keeping all the spectra of a patient either as a calibration or as a test set. Then, PLS-DA models were created and optimized with the calibration dataset using the leave-one patient-out to select the optimal number of latent variables (from a maximum of 10). The selected model was next used to predict the corresponding test set and figures of merit of classification, *i.e.* AUROC (Area Under the Receiver Operating Characteristic), Specificity, Sensitivity, and Accuracy. This process was repeated 100 times with different calibration and test sets until obtaining a distribution of the different classification parameters, which was employed to study the generalization capabilities of the model. On the other hand, to study the regression vector, a model was constructed with the entire data set and validated by cross-validation of the leave-one patient out. The model was validated using a permutation test as in[34] to ensure the significance of the spectral markers obtained. The classifications considered were BC *versus* normal, invasive BC *versus* non-invasive BC, and HG BC *versus* LG BC. A schematic of our approach is presented in Fig. 1.

Results and discussion

Morphological and IHC assessment

The inner layer of the urinary bladder wall is lined with transitional epithelium called urothelium. The deeper structure – subepithelial connective tissue includes lamina propria and muscularis propria. The outer structure – adventitia is composed of fat and fibrous tissue and leads to adjacent organs. Herein, taking into account the morphology of typical excisions, which rarely contain muscularis propria, the segmentation of SD IR images gave classes attributed to the epithelium (blue shades of UHCA classes) and subepithelium (pink shades of UHCA classes), see Fig. 2A and D. The epithelium contains both normal urothelium and epithelium subjected to the neoplastic process.

The subepithelial class with lamina propria includes fibrous tissue, blood and lymph vessels, muscularis mucosae, and muscularis propria (Fig. 2C and F). Muscularis propria
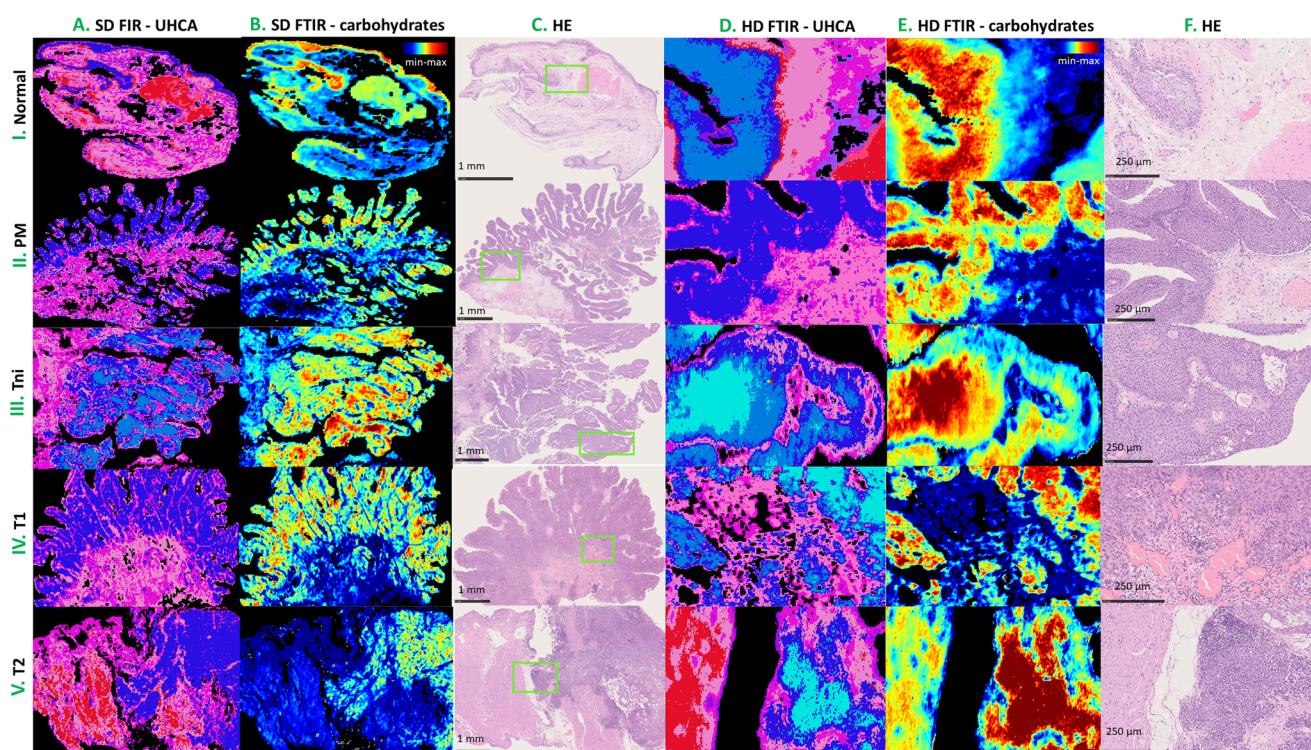

Fig. 2 IR imaging results for normal (N), potential malignant (PM) and BC groups (Tni LG, T1 HG and ≥T2 HG). (A and D) UHCA segmentation, (B and E) distribution of carbohydrates, (C and F) HE microphotographs (magnification 2× and 20×, respectively). Colour code in (A and D): blue – epithelial tissue (light and dark blue correspond to a high- and low-content of carbohydrates, respectively); pink– subepithelial fibrous tissue; red – subepithelial muscular tissue. The green rectangle in (C) marks HD FTIR image.

might be cumbersome in identification in HE images, even by the trained pathologist, since their muscle fibres are often irregular, with multivarious directions and are embedded in connective fibrous tissue. The latter is also heterogeneous.[35] Clustering all IR spectra into the epithelium and subepithelium was important to answer the question of whether the spectra of the deeper tissue, which is not BC, could be useful in spectroscopic BC diagnosis. Pathologists cannot recognize BC based only on changes in subepithelial tissue morphology without BC cells.

The normal epithelium is composed of 5 to 7 cellular layers that often form invaginations called Von Brunn nests. The most important morphological features of BC are increased nuclear to cytoplasmic ratio, altered cell stratification, an increased thickness of the epithelium, mitoses, irregular nuclei shape, and hyperchromasia.[6] Papillary structures are features of bladder cancer, hyperplasia, and PUNLMP. The exemplary HE images show the complexity of the histological features, see Fig. 2.

Analysing massive hyperspectral data brings to us some valuable observations. One of them regards the usefulness of standard and high definition IR imaging for mirroring tissue structure. Applying higher spatial resolution provides better discrimination of tissue structures and their borders than SD (Fig. 2). The irregular contours show the invasion of carcinoma by small nests or cells, whereas the smooth ones result from the tangential arrangement of structures. In particular, HD FTIR images reveal irregular patches of different classes in invading BC, contrary to smooth epithelium borders in PM and Tni cases. From a histopathological point of view, the smoother the epithelium borders are, the lower the possibility of BC invasion is expected. In turn, SD FTIR microscopy is a more efficient tool than HD because of the shorter data collection time and thus imaging of the whole excision is possible. Bearing in mind that SD FTIR imaging has a 5-fold lover spatial resolution, we can calculate that it is 200 times faster in the measurement of the same area than HD FTIR. The question appears whether the whole sample imaging is required to bring valuable information about BC. If early BC foci are to be detected and the heterogeneity of advanced BC needs to be investigated, the answer is that the SD FTIR approach is more robust than HD FTIR imaging (Fig. 2).

The IHC detection of glucose transporter 1 (Glut-1) can support the proper recognition of malignant epithelium in the bladder. Membranous positive Glut-1 expression around single cells co-occurs with BC morphological changes and is associated with an increased glucose uptake assisting the proliferation and survival of cancer cells. Its overexpression is also associated with tumour hypoxia and implicates a worse prognosis and chemotherapy-resistant neoplasm. We compared Glut-1 immunostaining patterns in the investigated cases with well-detected carbohydrates in the 1000–1200 cm^{-1} region of FTIR spectra (Fig. 3, detailed band assignments in Table S1†).[36] BC cases have a membranous positive Glut-1 reaction with a patchy/focal pattern (Fig. 3). We observed there a cytoplasmic reaction also, but it is not useful for diagnosis.
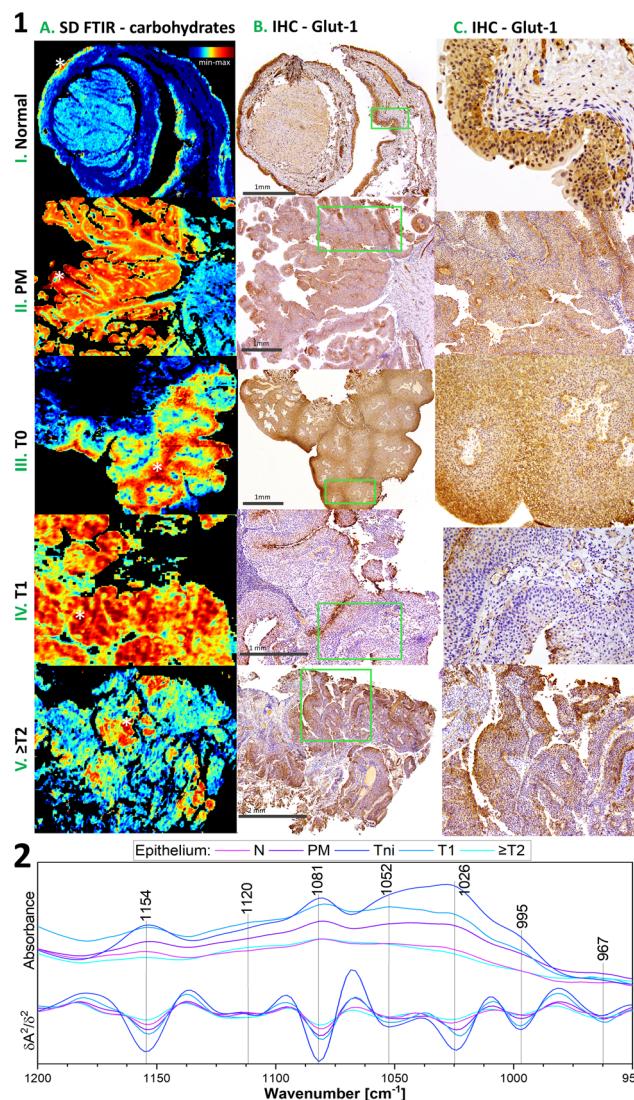

Fig. 3 Carbohydrates distribution and Glut-1 expression. (1) Comparison of IR distribution of carbohydrates and IHC staining. (A) IR images for integration in the 1000–1200 cm^{-1} region, (B and C) Glut-1 IHC images; the darker colour is observed, the higher is the presence of transporter density (magnification 2× and 20×, respectively). (2) FTIR spectra from pixels marked with a white star in (1.A). I−V: normal urothelium (N), potentially malignant urothelium (PM), and BC groups (T0 LG, T1 HG, and T2 HG).

The majority of the positive Glut-1 area in Tni is similar to those of the highest carbohydrate level in the IR images in contrast to Tinv cases (Fig. 3.1A–C).

This finding results from the fact that the IR-based level of carbohydrates does not illustrate the distribution of the glucose – protein transporter complex only. The IR profile in the 1000–1200 cm^{-1} region shows variation in the signal shape in normal epithelium and BC cases (Fig. 3.2). Furthermore, the accumulation of carbohydrates in some cases is more sensitive to early neoplasia than the expression of Glut-1 (Fig. S1†). This implicates that the IR distribution patterns of carbohydrates have power as a label-free glycomic tool. The

correlation of carbohydrate FTIR-based distribution and IHC-based Glut-1 expression at different stages of BC remains enigmatic, because of the similar location of the high carbohydrate level and membranous Glut-1 expression in T0 but not at higher stages accompanied by changes in the IR band shape (Fig. 3). In addition, we demonstrate that the IR map for the carbohydrate distribution localised small foci of early BC, even though immunohistochemistry does not indicate them (Fig. S1†).

IR imaging-based segmentation of urothelial tissues and classification model

SD FTIR imaging enables scanning of large areas, and our primary goal was to extract unique spectral profiles of epithelium and subepithelium tissues annotated in HE images (blue and pink classes in UHCA maps, Fig. 2A). Next, High Definition (HD) FTIR imaging was applied to the selected regions of the bladder wall to examine whether additional classes of a unique spectral signature could be discriminated (Fig. 2D).

Using spectra extracted from the UHCA analysis, a deep learning method was employed for the classification of bladder carcinoma. PLS-DA methodology is often used in spectral discrimination.[37] Here, the generalization power of the models was established by using repeated random sampling, where 100 splits of calibration and independent tests were considered. In all cases, the splits were performed considering the patients, *i.e.* FTIR spectra of the epithelial and subepithelial tissues from the same patients were always included either in calibration or test sets. Parameters of the PLS-DA models for various pairs of experimental groups show a high power to differentiate BC patients from healthy individuals (see Table 2 and Fig. S2†).

Here, FTIR spectra of both layers of the bladder wall can be used for the recognition of any BC case with accuracy above 85%. The attempts at the classification of the PM cases produced any model. Next, the cancer cases were classified to recognise bladder cancer invasiveness of urothelial cancer to deeper tissue layers (Tni *versus* Tinv) and grading (LG BC *versus* HG BC), *c.f.* Table 2. The AUROC, accuracy, sensitivity, and specificity for Tni and Tinv indicated the good performance of the test when IR spectra of the subepithelial tissue were considered. This also indicated significant spectral alternations in this bladder tissue due to the heterogeneous invasion of cancer cells. Discrimination attempts between the T1 and T2 spectra failed. The low and high grade groups that also included non and invasive urothelial carcinomas were well classified according to the IR signature of the epithelium. Moreover, the comprehensive investigation proved that the division of N *vs.* BC into the more complex model, for example, N *vs.* Tni and N *vs.* Tinv or N *vs.* LG BC and N *vs.* HG BC, is profitable as they present higher performance (Table 2). Averaged second derivative FTIR spectra are consistent with PLS-DA regression vectors calculated in the developed models (Fig. 4). FTIR spectra of detailed groups are collected in Fig. S3 and S4.†

The IR spectra displayed in Fig. 4, Fig. S3 and S4† show that high intensity of the bands of proteins (1652 and 1544 cm^{-1}) and nucleic acids (995 and 967 cm^{-1}) and low-intensity collagen features (1393, 1335, and 1200 cm^{-1}) indicate BC. For instance, unordered structures of proteins (1637 cm^{-1}) became pronounced in the HG BC tissue matrix compared to α-helical conformations and the collagen level. In turn, the glycogen spectral signature (1154, 1081, and 1026 cm^{-1}) does not show a clear linear relationship considering the malignant nature of the BC disease and the highest level is observed in the early LG BC. The invasive stages of BC are characterised by increased absorbances at 1637, 1393, and 1233 cm^{-1} and decreased ones at 1544 and 1200 cm^{-1} compared to non-invasive BC. This could be associated with epithelial to mesenchymal transformation. Moreover, the carcinogenesis caused the band shifts indicating changes in structures of biocomponents, *e.g.* 1543 *vs.* 1547 (proteins), 1028 *vs.* 1025 (carbohydrates), and 1120 *vs.* 1117 cm^{-1} (nucleic acids), in N and BC, respectively. The most dominant effect on cancer metabolism is the Warburg effect.[38] Cancer cells need more glucose, and this is accomplished by up-regulation of Glut-1. Glut-1 upregulation is in turn induced in early stages of many

Table 2 Results of the Monte Carlo double cross-validation

Classes	Tissue type	AUROC	Specificity	Sensitivity	Accuracy	p-Value[a]
N *vs.* BC	Epithelium	0.86 ± 0.10	0.67 ± 0.22	0.84 ± 0.10	0.80 ± 0.08	<0.001
N *vs.* BC	Subepithelium	0.87 ± 0.09	0.70 ± 0.26	0.85 ± 0.10	0.81 ± 0.07	<0.001
N *vs.* Tni	Epithelium	0.84 ± 0.11	0.79 ± 0.18	0.80 ± 0.15	0.75 ± 0.13	<0.001
N *vs.* Tni	Subepithelium	0.86 ± 0.10	0.80 ± 0.18	0.80 ± 0.14	0.81 ± 0.12	<0.001
N *vs.* Tinv	Epithelium	0.80 ± 0.10	0.73 ± 0.22	0.77 ± 0.12	0.74 ± 0.10	<0.001
N *vs.* Tinv	Subepithelium	0.73 ± 0.13	0.62 ± 0.27	0.76 ± 0.15	0.66 ± 0.08	<0.01
N *vs.* LG BC	Epithelium	0.91 ± 0.08	0.84 ± 0.18	0.87 ± 0.08	0.87 ± 0.08	<0.001
N *vs.* HG BC	Epithelium	0.92 ± 0.09	0.86 ± 0.16	0.88 ± 0.09	0.61 ± 0.01	<0.001
Tni *vs.* Tinv	Epithelium	0.67 ± 0.10	0.55 ± 0.17	0.68 ± 0.10	0.64 ± 0.07	<0.01
Tni *vs.* Tinv	Subepithelium	0.72 ± 0.14	0.64 ± 0.17	0.69 ± 10	0.68 ± 0.10	<0.01
HG BC *vs.* LG BC	Epithelium	0.69 ± 0.09	0.63 ± 0.13	0.65 ± 0.13	0.64 ± 0.07	<0.01

N – normal; BC – all patients with BC; Tni - *in situ* and papillary urothelial carcinoma; Tinv – T1 (BC invading subepithelial tissue) and ≥T2 (BC invading muscularis propria and perivesical tissue); HG BC – high grade BC; LG BC – low grade BC; AUROC – area under the receiver operating characteristic. [a] Significance value of a permutation test.
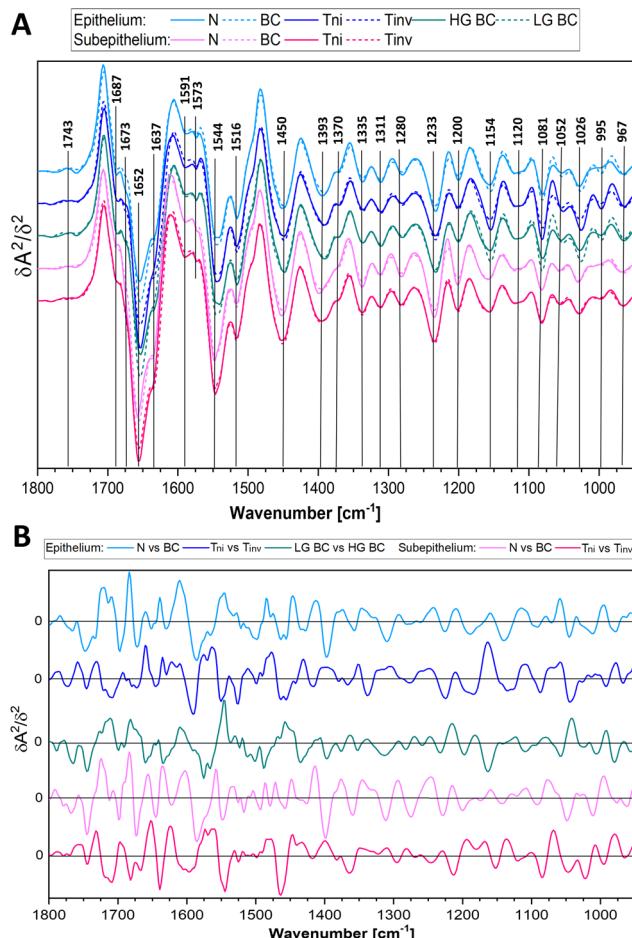
Fig. 4 (A) Averaged second derivative FTIR spectra of UHCA classes calculated from SD IR images. (B) Cumulative sums of PLS–DA regression vectors for the investigated cases (based on SD IR images). Legend: epithelium (n = 417): n(N) = 62, n(Tni) = 124, n(Tinv) = 177, n(lG) = 167, n(HG) = 134; subepithelium (n = 494): n(N) = 81, n(Tni) = 140, n(Tinv) = 206.

cancers.[39,40] Cancer cells need more glucose, and this is accomplished by up-regulation of Glut-1. We summarise the potential IR markers of the BC invasiveness and grading based on the most dominant vectors from the plots in Fig. 4B (vectors have been detailed in Table S2†). The epithelium and subepithelium IR features in the N *vs.* BC model show main differences in the range from 1700 to 1400 cm^{-1} (the protein region) whereas the Tni and Tinv groups exhibit several peaks in the entire fingerprint region. This comparison of the vector positions proves the fact that the IR profile of the studied cases is unique for each patient group.

The PLS-DA models reveal similar accuracy when epithelial and subepithelial spectra are used for the general BC classification (Table 2 and Fig. S2†). Whilst the better parameters of the assignment are obtained from the subepithelial IR spectra for the recognition of BC infiltration, the epithelial profiles are suitable for the grade definition. The performance (AUROC values) of epithelium and subepithelium-based classification for the N *vs.* Tni model is similar to the N *vs.*

BC groups (*ca.* 86) whereas for the N *vs.* Tinv is lower (70–80). In the case of grading, the accuracy of the epithelial model for N *vs.* LG BC and N *vs.* HG BC groups is higher than for N *vs.* BC (91, 92, and 86, respectively). Interestingly, we achieved similar classification accuracy for LG BC and HG BC for IR imaging of cytological samples.[20] This indicates the high power of this molecular tool for the prediction of BC grading. Considering the worst assignment of BC staging, we conclude that the invasion is often accompanied by inflammatory cells which could affect IR spectra of the tissue. In this work, the T1 group cannot be distinguished from T2 and higher stages because of too large variability between spectra of patients in the advanced BC epithelium and pronounced spectral changes in the subepithelial class for deep infiltration. In turn, potentially malignant cases (PM) studied here and assigned to the N or BC groups implicated a need for long prospective study with the observation of patients and the determination of their molecular and metabolomic features to confirm their correct classification. This is not surprising since very early neoplastic changes can stop, progress or regress.

Conclusions

FTIR imaging combined with clustering of the epithelium and subepithelium provides a powerful tool for the identification of BC and its invasion. This approach has detection and prognostic capability. Our work proves the possibility of BC recognition and its invasion based on IR spectra of the superficial and deep layers of the bladder wall with the specific features for grades and stages. The determined parameters of the classification are satisfactory and comparable to clinical standards. The added value of our findings is also the fact that the spectroscopic analysis can be limited to surface layers of the bladder wall reducing the time of data collection and analysis. Therefore, our work gives "the green light" to establish automatic spectral histopathology of bladder excisions supported by machine learning approaches.

Author contributions

Conceptualization, M. K., K. O., P. Ch., K. M. D. P.-G.; methodology, M. K., K. M., D. P.-G.; data collection, M. K.; histological analysis, M. K., K. O., spectral analysis and visualization, M. K, D. P.-G.; first draft preparation, M. K.; writing—review and editing, K. M., D. P.-G., K. O. and P. Ch.; supervision, K. M, K. O.; project administration, M. K., K. M., D. P.-G.; funding acquisition, M. K., K. M., D. P.-G.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 K. Saginala, *et al.*, *Prog. Clin. Biol. Res.*, 2020, **8**, 15.
2 D. F. I. Kurtycz, K. E. Sundling and G. A. Barkan, *Diagn. Cytopathol.*, 2020, **48**, 890–895.
3 S. Antoni, *et al.*, *Eur. Urol.*, 2017, **71**, 96–108.
4 V. Kvikstad, *et al.*, *Diagn. Pathol.*, 2019, **14**, 1–8.
5 E. Compérat, *et al.*, *Histopathology*, 2013, **63**, 756–766.
6 H. Moch, P. A. Humphrey, T. M. Ulbright and V. E. Reuter, *WHO Classification of Tumours of the Urinary System and Male Genital Organs*, International Agency for Research on Cancer, 2016.
7 L. G. Koss, G. Leopold and R. S. Hoda, *Koss's Cytology of the Urinary Tract with Histopathologic Correlations*, Springer, 2012.
8 R. Hartman and A. Kawashima, *Eur. J. Radiol.*, 2017, **97**, 119–130.
9 Y. Lotan, *et al.*, *Urol. Oncol.: Semin. Orig. Invest.*, 2017, **35**, 531.e15–531.e22.
10 M. Diem, *et al.*, *Faraday Discuss.*, 2016, **187**, 9–42.
11 S. Tiwari, *et al.*, *Sci. Adv.*, 2021, **7**, eabb8292.
12 H. J. Byrne, *et al.*, *Analyst*, 2015, **140**, 2066–2073.
13 K. Chrabaszcz, *et al.*, *Urol. Oncol.: Semin. Orig. Invest.*, 2018, **1864**, 3574–3584.
14 K. Y. Su and W. L. Lee, *Cancers*, 2020, **12**, 115.
15 M. Kujdowicz, *et al.*, *Cancers*, 2021, **13**, 123.
16 F. Santos, S. Magalhaes, M. C. Henriques, M. Fardilha and A. Nunes, Spectroscopic Features of Cancer Cells: FTIR Spectroscopy as a Tool for Early Diagnosis, *Curr. Metabolomics*, 2018, **6**, 103–111.
17 M. Kujdowicz, D. Perez-Guaita, P. Chłosta, K. Okon and K. Malek, *Microchem. J.*, 2021, **168**, 106460.
18 S. Gok, *et al.*, *J. Biophotonics*, 2016, **9**, 967–975.
19 S. Bensaid, *et al.*, *Expert Syst. Appl.*, 2017, **89**, 333–342.
20 M. Kujdowicz, *et al.*, *Cancers*, 2021, 13225734.
21 C. Hughes, *et al.*, *J. Biophotonics*, 2013, **6**, 73–87.
22 K. E. Witzke, *et al.*, *Am. J. Pathol.*, 2019, **189**, 619–631.
23 R. Bandzeviciute, J. Ceponkus, M. Velicka and V. Urboniene, *J. Mol. Struct.*, 2020, **1220**, 128724.
24 G. Demos, *et al.*, *J. Biomed. Opt.*, 2004, **9**, 767–771.
25 H. K. Yosef, *et al.*, *Anal. Chem.*, 2017, **89**, 6893–6899.
26 M. B. Amin, *et al.*, *AJCC Cancer Staging Manual*. Springer International Publishing, 2017.
27 C. C. Guo and B. Czerniak, *Arch. Pathol. Lab. Med.*, 2019, **143**, 695–704.
28 S. Bircan, O. Candir and T. A. Serel, *Urol. Int.*, 2004, **73**, 201–208.
29 J. K. Kim, *et al.*, *J. Cancer*, 2017, **8**, 2885–2891.
30 J. P. Maxwell, C. Wang, N. Wiebe, A. Yilmaz and K. Trpkov, *Diagn. Pathol.*, 2015, **10**, 3.
31 A. E. Hentschel, *et al.*, *Urol. Oncol.: Semin. Orig. Invest.*, 2020, **38**, 440–448.
32 H. Wijkström, *et al.*, *Scand. J. Urol. Nephrol., Suppl.*, 2000, 116–135, DOI: **10.1080/00365590050509878**.
33 P. Lasch, *A Matlab Based Application for Infrared Imaging*, Cytospec TM, 2001; Available online: **http://www.cytospec.com** (accessed on 3 January 2020).
34 D. Pérez-Guaita, J. Kuligowski, S. Garrigues, G. Quintás and B. R. Wood, *Analyst*, 2015, **140**, 2422–2427.
35 *Histology for Pathologists*, ed. S. E. Mills, Lippincott Williams & Wilkins, 2012.
36 E. Staniszewska-Slezak, *et al.*, *J. Biophotonics*, 2018, **11**, e201700044.
37 P. S. Gromski, *et al.*, *Anal. Chim. Acta*, 2015, **879**, 10–23.
38 R. J. DeBerardinis, J. J. Lum, G. Hatzivassiliou and C. B. Thompson, *Cell Metab.*, 2008, **7**, 11–20.
39 M. L. Macheda, S. Rogers and J. D. J. Best, *Cell. Physiol.*, 2005, **202**, 654–662.
40 H. Reis, *et al.*, *Oncol. Lett.*, 2011, **2**, 1149–1153.