

Journal of

[www. biophotonics-journal.org](http://www.biophotonics-journal.org)

# BIOPHOTONICS

 **WILEY-VCH**

**REPRINT**

FULL ARTICLE

# It's in your blood: spectral biomarker candidates for urinary bladder cancer from automated FTIR spectroscopy

Julian Ollesch<sup>1</sup>, Margot Heinze<sup>1</sup>, H. Michael Heise<sup>1</sup>, Thomas Behrens<sup>2</sup>, Thomas Brüning<sup>2</sup>, and Klaus Gerwert<sup>\*,1</sup>

<sup>1</sup> Ruhr-Universität Bochum, Department of Biophysics ND04, Universitätsstraße 150, 44780 Bochum, Germany

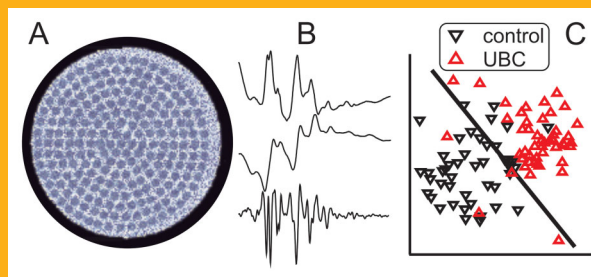
<sup>2</sup> Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-Universität Bochum (IPA), Bürkle-de-la-Camp Platz 1, 44789 Bochum, Germany

Received 15 October 2013, revised 18 November 2013, accepted 29 November 2013

Published online 7 January 2014

**Key words:** high throughput, FTIR spectroscopy, dry film blood analysis, cancer biomarker, urinary bladder cancer

Blood samples of urinary bladder cancer (UBC) patients and patients with urinary tract infection were analysed with advanced automated high throughput Fourier transform infrared (HT-FTIR)-spectroscopy. Thin dried film samples were robotically prepared on multi-well titer plates (MTP) for absorbance measurements in transmission mode. Within the absorbance, 1<sup>st</sup> and 2<sup>nd</sup> derivative spectra of serum and two plasma preparations, discriminative patterns were identified and validated using bioinformatic tools. The optimal spectral resolution for data acquisition was determined. An accurate discrimination of the patient groups was achieved with three different independent spectral variable sets. The HT-FTIR blood test may support future clinical diagnostics.



Dry robotically prepared blood sample films (A) were analysed with automated HT-FTIR spectroscopy (B) to identify and validate spectroscopic biomarker candidates for urinary bladder cancer (UBC) (C).

## 1. Introduction

Urinary bladder cancer (UBC) is one of the current major health burdens worldwide [1, 2]. As major risk factors, smoking and occupational exposure to toxins as e.g. aromatic amines and polycyclic aromatic hydrocarbons (PAH) were identified. Complementarily, infectious diseases as for example schistosomiasis contribute to the risk in developing countries [3, 4]. Men are roughly three times more affected than

female patients. When detected and treated at an early stage, a five-year survival rate of more than 70% can be achieved [5]. But, recurrent cancer is a grave issue [5]. UBC patients are required to undergo repeated cystoscopies in tight intervals of approximately two months. Cystoscopy itself is a painful procedure bearing the risks of bleeding, subsequent infection or inflammation, bladder perforation, or urethral stricture.

\* Corresponding author: e-mail: gerwert@bph.rub.de, Phone: +00 49 234 32 24461, Fax: +00 49 234 32 14238

Therefore, a reliable detection based on a simple, minimally invasive procedure would relieve the patient from strain, and at the same time would increase the chance for a successful therapy.

A reliable diagnostic test based on blood analysis would comply with this demand. Compared with cystoscopy, drawing a blood sample definitely requires less time and resources with less impact on the patient. Therefore, if a specific blood based test could confirm or avert an initial UBC suspicion, the number of more invasive examinations performed on the patient could be reduced.

Several potential blood-based biomarker candidates for UBC have been previously described [6–14]. Thus, an analysis of the complex marker candidates in their entirety using Fourier-transform infrared (FTIR) spectroscopy of blood samples in combination with disease pattern recognition (DPR) was demonstrated as a reasonable alternative [15].

The use of FTIR spectroscopy as a high throughput technology in combination with 96 well multi well titer plates (MTP) has been established [16–19], although the sample throughput as compared with, e.g., fluorescence based techniques is comparably low. Contrastingly, the FTIR absorbance spectrum of a biofluid sample reflects the individual biochemical patient status [20, 21]. No additional chemistry such as labelling or the introduction of markers is required for the simultaneous acquisition of the complex proteome, lipidome, and metabolome data. With specific bioinformatics, precise and multiparametric quantitative clinical chemistry assays for a single sample were demonstrated, and disease specific spectral band patterns had been identified and validated [19–32].

An improved and automated technology for the DPR approach was recently developed [15]. A possible user impact during analysis of the sample was ruled out by a robotic sample preparation. Thereby artefacts, usually encountered with dry film preparations, were excluded. Absorbance spectra were recorded by an automated high-throughput (HT)-FTIR system, thus achieving extreme spectral reproducibility. User interference – often described by e.g. “visual inspection” or “manual baseline correction” – was avoided by the implementation of automated bioinformatics routines without manual user input. Strict and efficient Monte Carlo cross-validation (MCCV) schemes were introduced for validation.

Here, we present improved results of our still ongoing study on the identification and validation of spectral biomarker candidates obtained from UBC patients *versus* patients with urinary tract infection, which are our clinically relevant control group for this study.

Eligible patients were referred to our collaborating clinics by urologists outside of our study for

transurethral resection (TUR) of urinary bladder tissue. TUR is performed during cystoscopy, causes discomfort for the patient, and bears the risks of bleeding, subsequent inflammation, thrombosis, embolism, bladder perforation or urethral stricture. Three to four days of stationary hospitalization are usually required for this medical procedure. In roughly a third of our study participants, a urinary tract infection, which could be treated with antibiotics, gave rise to the initial suspected diagnosis of cancer. For these patients, a highly specific negative blood test would have been a cost-efficient method to avoid a more invasive follow-up and to preserve hospital capacities. In particular, UBC patients under therapy, who undergo repeated cystoscopy to screen for recurrent bladder cancer, would greatly benefit from a less invasive method.

In the first report on our study, we presented an accurate and sensitive, but yet unspecific procedure to discriminate between UBC and control patients based on infrared spectra of blood serum and two plasma preparations [15]. Now, with a larger patient cohort available, a statistically significantly increased accuracy, sensitivity, and, in particular, an increased specificity of the patient class discrimination were achieved with the three blood samples, which were collected from each patient and processed as described before: serum, ethylene diamine tetraacetic acid (EDTA), and citrate stabilized plasma. This protocol takes into account expected differences between samples of induced coagulation *versus* coagulation prevention with two chemicals, which may mask specific spectral absorption bands. Absorbance sample spectra were combined with the respective 1<sup>st</sup> and 2<sup>nd</sup> derivative spectra to form one cumulative data vector, which also contains the expected subtle band shifts as possible indicators of the patient's health status.

As described in the earlier report, particular care was applied to select discriminative features from the spectral vectors. In addition to the previously applied iterative random forest (RF) algorithm, which exploited the RF intrinsic *Gini*-importance of features in repeated calculations on data subsets [15] (see below), a syn-entropy analysis method to identify the features of maximum relevance and minimum redundancy (MRMR) [33, 34] was applied.

The patient prediction was performed using a classifying linear discriminant analysis (LDA) and an advanced ensemble random forest classifier. The LDA required only relatively low computational power compared with the RF algorithm. The RF is reliable but computationally intensive due to the construction of the classifier. It is a collection of decision trees built from random selections of spectral features [35]. Internal cross-validation leads to a collection of correctly classifying decision trees calculated from randomly selected data variables. A ma-

majority vote of the included trees is interpreted as the classifier prediction. For the spectral features included in the RF, the so-called *Gini* importance was calculated, which is useful to exclude uninformative features from classification [15, 18, 36, 37]. Furthermore, RFs were used as ensemble classifiers of multiple RFs predicting validation datasets based on majority votes of the combined RFs.

The earlier report was based on a class-unbalanced study cohort of 89 UBC patients *versus* 46 controls. The distortive effects of unbalanced training data were now evaluated and removed. From the total patient population, even sets of definite UBC patients *versus* definite non-cancer control patients were randomly assembled. Thus, the validation results were largely more balanced considering sensitivity and specificity. As a result, the accuracy of class prediction was significantly improved.

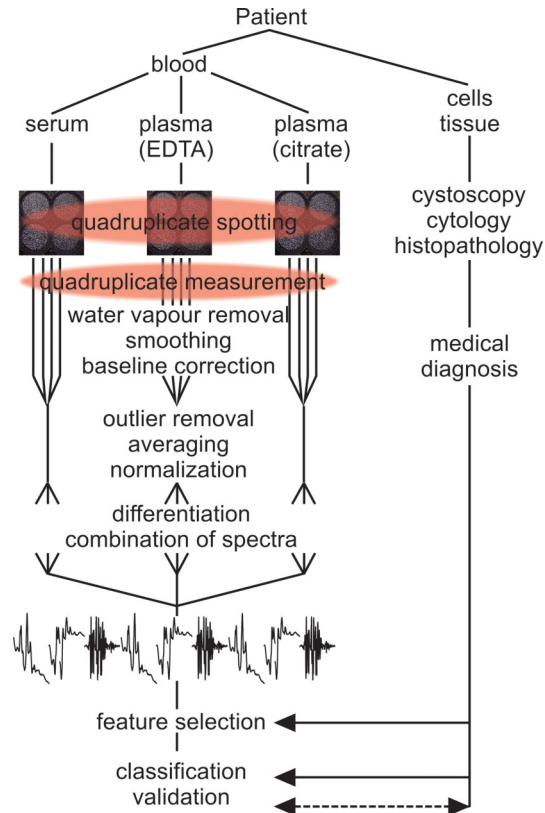
The sample throughput rate was increased by a reduction of the spectral resolution during the measurement. Thereby, the minimum resolution necessary for accurate class distinction was determined as  $4\text{ cm}^{-1}$ , thus doubling the previously reported sample throughput, and reducing the number of total spectral variables per patient by a factor of two.

## 2. Experimental

The study reported here is a continuation of our study on UBC with all experimental procedures described there [15], whereas differences to the previous protocols are discussed in detail below. The workflow (Figure 1) comprised quadruplicate spotting of blood preparations, HT-FTIR-measurement, spectral preprocessing, feature selection, and the validation of two different classifiers (Figure 1).

### 2.1 Patient population

Strictly defined standard operating procedures (SOP) were developed with the PURE Scientific Epidemiological Study Centre of the IPA (Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-Universität Bochum, Germany, member of the research initiative PURE, Protein research Unit Ruhr within Europe) according to the rules of Good Epidemiological Practice. Following these protocols, epidemiologic data were collected of patients, who were fully informed about the study and gave their written consent. Blood samples were collected and processed to serum, EDTA-, and tri-sodium citrate stabilized plasma with clinical routine equipment (BD Biosciences, Heidelberg, Germany) obeying



**Figure 1** Workflow of sample analysis (adapted from [15]): With each biofluid sample, four wells of a 384 well MTP were robotically coated with a thin film. From the recorded four absorbance spectra, spectra containing artefacts were removed individually. Outlier removal, averaging and normalization resulted in a representative absorbance spectrum of each sample. After differentiation, spectra were combined to a synthetic and patient-unique, sequentially arranged vector, consisting of the respective absorbance, 1<sup>st</sup> and 2<sup>nd</sup> derivative spectrum of serum, EDTA and citrate plasma. Using these data, classification relevant variables were identified and validated based on the medical diagnosis.

strict SOPs, as reported previously [15]. The samples were shock frozen within less than 30 min for plasma, and less than 50 min after sampling for serum. All samples were stored at  $-80\text{ }^{\circ}\text{C}$  until experimental use. Using the clinical chemistry data of the participating patients alone, no UBC specific signatures could be identified. The established diagnosis by combination of cytology, cystoscopy and histopathology served as gold standard for the DPR approach. This study complies with the applicable ethical guidelines and was approved by the Ethical Committee of the Ruhr-Universität Bochum (Ethics vote 3674-10, Ethical Committee of the Ruhr-Universität Bochum, Bochum, Germany).

From the total pool of participating patients, four datasets were randomly assembled (Table 1). Set (i)

**Table 1** Datasets randomly selected from the total patient pool. Sets (i) and (ii) were analysed with a spectral resolution of  $8\text{ cm}^{-1}$ , (iii) and (iv) with  $4\text{ cm}^{-1}$ .

| dataset             | (i)         |             | (ii)        |             | (iii)       |             | (iv)        |             |
|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                     | m           | f           | m           | f           | m           | f           | m           | f           |
| patients            | 222         | 286         | 129         | 166         | 74          | 100         | 78          | 100         |
| av. age $\pm\sigma$ | $71 \pm 10$ | $72 \pm 10$ | $72 \pm 12$ | $71 \pm 10$ | $71 \pm 11$ | $72 \pm 11$ | $71 \pm 11$ | $73 \pm 12$ |
| controls            | 61          | 81          | 63          | 83          | 36          | 50          | 38          | 50          |
| UBC G2+             | 161         | 205         | 66          | 83          | 38          | 50          | 40          | 50          |
| recurrent           |             | 64*         |             | 31**        |             | 14          |             | 0           |

\* 23 missing information if recurrent

\*\* 7 missing information if recurrent

comprised 286 patients, 81 diagnosed non-UBC cases, and 205 positive UBC with a tumour grade of G2, G3 or G4 (WHO 1973), termed in the following “UBC G2+”. Thus, papilloma and low-malignancy grade G1, which are unlikely to secrete similar biomarker amounts as occurring with the advanced grades, were excluded. At least 64 recurrent cancer cases were included. Of 23 UBC G2+ patients, no information about recurrence was available. The spectral dataset was acquired with  $8\text{ cm}^{-1}$  spectral resolution.

Patient set (ii) was a balanced selection of 166 patients (i). It consisted of 83 UBC G2+ patients (34 G2, 30 G3, 19 G4) and 83 control patients with urocystitis or urethral infection who were free of cancer based on the clinical and pathological diagnosis. At least 31 recurrent cancer cases were included in the UBC G2+ group; of 7 patients the recurrence status was unclear. The spectral resolution was  $8\text{ cm}^{-1}$ .

For patient set (iii), blood samples of 50 patients with urocystitis or urethral infection *versus* 50 UBC G2+ patients were analyzed with  $4\text{ cm}^{-1}$  spectral resolution (Figure 2). The UBC G2+ group (22 G2, 17 G3, 11 G4) included 14 recurrent cancer cases.

Finally, patient set (iv) consisted of 50 control patients with urocystitis or urethral infection *versus* 50 definite first-time UBC G2+ patients (17 G2, 20 G3, 13 G4). The blood samples were analyzed with  $4\text{ cm}^{-1}$  spectral resolution (Figure 3).

## 2.2 High-throughput FTIR spectroscopy

Automated HT-FTIR-measurements (Vertex 70v FTIR spectrometer, HTS-XT extension, Twister robotic plate feeder, Bruker Optics GmbH, Ettlingen, Germany) of robotically spotted blood serum and plasma (50 nl each) in concentric circles of 217

200 pl spots per well (instrumentTwo, M2 Automation GmbH, Berlin, Germany) on 384 well silicon MTPs (Bruker) were performed with extreme reproducibility as described [15].

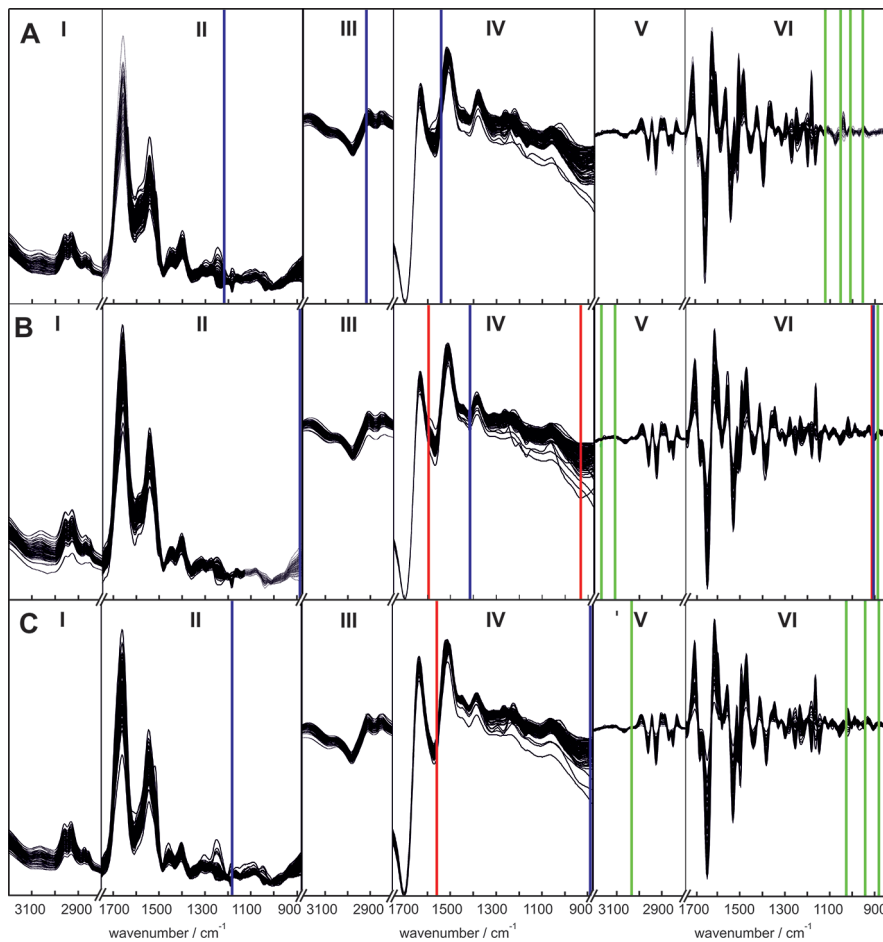
In FTIR spectroscopy, interferograms are recorded, averaged and converted to spectra by Fourier transformation. The spectral resolution is defined by the length of the recorded interferogram, which is proportional to the instrument scan time [38]. To reduce the measurement time and dimensionality of the dataset, the spectral resolution of the data acquisition was reduced from the earlier reported  $2\text{ cm}^{-1}$  to  $4\text{ cm}^{-1}$  and  $8\text{ cm}^{-1}$ , respectively. In theory, a reduction of measurement time by the respective factors of two and four was expected. All further instrument parameters remained unchanged.

## 2.3 Data preprocessing

The absorbance spectra of a sample were collected in quadruplicate in transmission mode. Trace spectral contributions of atmospheric water vapour were removed by scaled subtraction. Remaining high frequency noise was filtered out by means of a Gaussian low pass filter. Outlier removal, averaging, adaptive iteratively penalized least squares (airPLS) baseline correction, derivation and spectral combination were performed as reported [15].

Considering the spectral resolution of the dataset, the noise filter was adjusted to  $6\text{ cm}^{-1}$  when applied to the  $4\text{ cm}^{-1}$  resolution spectra, and to  $8\text{ cm}^{-1}$  with  $8\text{ cm}^{-1}$  resolution data. The 1<sup>st</sup> and 2<sup>nd</sup> derivation of the  $4\text{ cm}^{-1}$  resolution absorbance spectra was calculated with Fourier transformation and low pass filtering at  $6\text{ cm}^{-1}$  and  $8\text{ cm}^{-1}$ , respectively [15]. Spectra with  $8\text{ cm}^{-1}$  resolution were derivated with  $10\text{ cm}^{-1}$  and  $12\text{ cm}^{-1}$  filtering.





**Figure 2** Spectral overview and selected features of dataset (iii), 50 UBC G2+ patients including 14 recurrent cases, 50 controls. Spectra of serum (**A**), of EDTA stabilized plasma (**B**), and sodium citrate stabilized plasma (**C**) are shown, divided in the spectral regions C–H stretching absorbance (I), absorbance fingerprint (II), 1<sup>st</sup> derivative of the C–H stretching absorbance region (III), 1<sup>st</sup> derivative of the fingerprint absorbance region (IV), 2<sup>nd</sup> derivative of the C–H stretching absorbance (V), and the 2<sup>nd</sup> derivative of the fingerprint absorbance region (VI). The respective regions (I–II, III–IV, and V–VI) were scaled for optimal display. Green: 15 features reported [15], blue: MRMR algorithm results on this dataset, red: RF-algorithm results on this dataset (compare with Table 4 and Figure 4D).

For each patient, a representative spectral vector was assembled from all three blood preparations, as documented before [15]. The absorbance spectrum of serum was concatenated with its 1<sup>st</sup> and 2<sup>nd</sup> derivative, followed by the corresponding data of EDTA and citrate stabilized plasmas (Figure 1). This resulted in a 11,493 dimension vector of wavenumber-intensity pairs with a datapoint spacing of 1 cm<sup>-1</sup> (@ 2 cm<sup>-1</sup> resolution). Hence, we report vectors reduced to 5751 and 2871 features with 2 cm<sup>-1</sup> and 4 cm<sup>-1</sup> spacing recorded with 4 cm<sup>-1</sup> and 8 cm<sup>-1</sup> instrumental resolution, respectively.

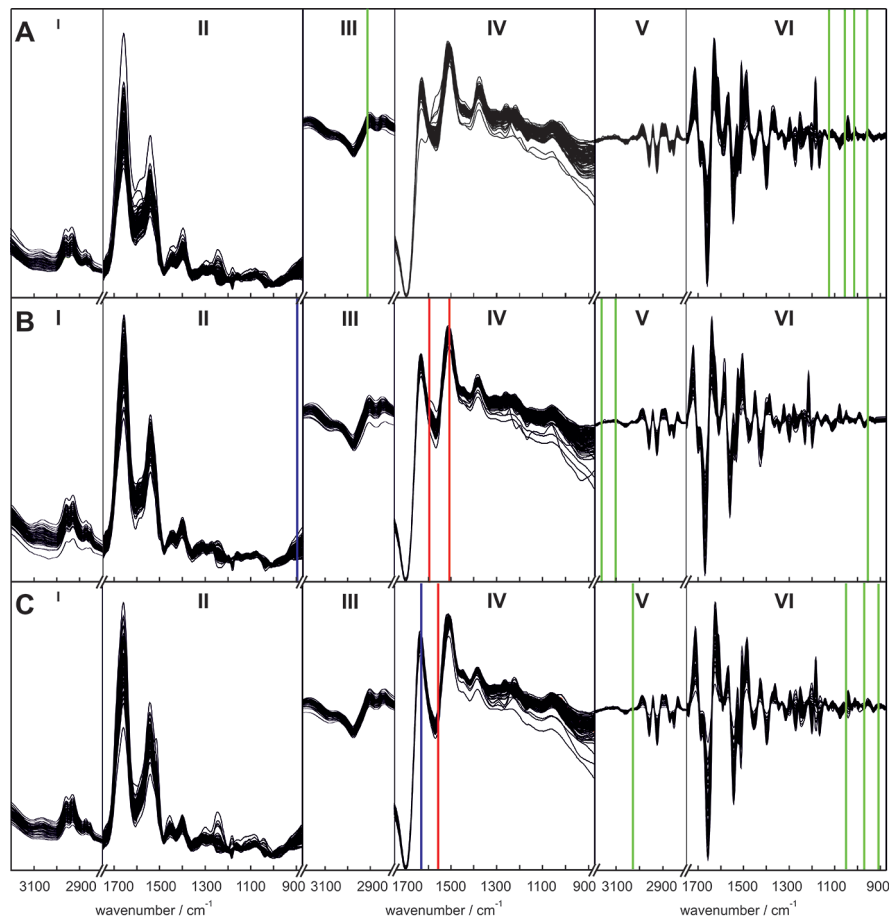
## 2.4 Feature selection

The term “feature selection” comprises a dimensionality reduction of the classification problem in a way that redundant and uncorrelated information is removed and only the most discriminative data are preserved. The algorithms applied here were shown to perform well in nonlinear multivariate classification problems [15, 34, 37].

For the comparison with previous findings, the set of fifteen wavenumber-intensity pairs identified previously as the set of optimum classification [15] was evaluated for classification performance on the actual datasets. If the spectral resolution of the datasets did not match, nearest neighbours to the previously found wavenumbers were selected to represent the corresponding spectral band.

A second feature set specific to the respective spectral data was determined by a maximum relevance, minimum redundancy (MRMR) approach [33, 34]. This algorithm identifies spectral variables depending on the discriminative power and the redundancy of information. It was performed with the algorithm as published and can be downloaded from [http://www.mathworks.com/matlab\\_central/fileexchange/14916](http://www.mathworks.com/matlab_central/fileexchange/14916) (September 10, 2013). Each dataset was analysed for the 100 most discriminative features. Stepping down with the MRMR ranking from the single top ranked feature to the bottom hundred ones, the highest-ranked feature set performing with highest average accuracy in 1000 independent leave-one-third-out MCCV with LDA classifiers was identified as the MRMR selection result, requiring only relatively low processing power.

**Figure 3** Spectral overview and selected features of dataset (iv), 50 UBC G2+ patients excluding recurrent cases, 50 controls. Spectra of serum (A), of EDTA stabilized plasma (B), and sodium citrate stabilized plasma (C) are shown, divided in the spectral regions C–H stretching absorbance (I), absorbance fingerprint (II), 1<sup>st</sup> derivative of the C–H stretching absorbance region (III), 1<sup>st</sup> derivative of the fingerprint absorbance region (IV), 2<sup>nd</sup> derivative of the C–H stretching absorbance (V), and the 2<sup>nd</sup> derivative of the fingerprint absorbance region (VI). The respective regions (I–II, III–IV, and V–VI) were scaled for optimal display. Green: 15 features reported [15], blue: MRMR algorithm results on this dataset, red: RF-algorithm results on this dataset (compare with Table 5 and Figure 5D).



With regard to processing power, the iterative wrapper algorithm for random forest based feature selection that we successfully applied before [15], is vastly more demanding. Briefly, a random forest can be used to determine the *Gini*-importance of a spectral feature for correct classification [18, 36, 37, 39]. The selection process was repeatedly performed on MC derived data subsets comprising of 80% of the total dataset, resulting in a selection frequency map of each identified feature. For each subset, the cumulative *Gini*-importance of all spectral features was determined from 192 random forests, the 20% least important features excluded from the dataset, and the next 192 RFs were calculated obeying strict leave-one-third-out MCCV procedures. This procedure was repeated until only 5 features were left. Based on the average accuracy determined on each 192 MCCVs, the best predicting set was registered into a pool of selected features. This pool was analysed by stepping down in search of a minimum selection frequency threshold. For each threshold, the identified feature sets were individually validated for optimum average accuracy in a 1000 fold LDA leave-one-third-out MCCV to determine the average accuracy. The best performing feature set determined the threshold, which is given in the according Tables 2–5.

The identified classification-characteristic features were checked for agreement with spectral contributions of the silicon substrate and the sample additives, as e.g. citrate, without apparent overlap.

**Table 2** Classifier evaluation of spectral marker candidates of unbalanced dataset (i), 8 cm<sup>-1</sup> resolution, 205 UBC G2+ versus 81 control patients (set: feature set, #f: number of features, cf: classifier, acc: accuracy/%, sens: sensitivity/%, spec: specificity/%).

| set*    | #f | cf    | acc     | sens    | spec   |
|---------|----|-------|---------|---------|--------|
| RF [15] | 15 | LDA   | 50 ± 2  | 98 ± 2  | 2 ± 2  |
| RF [15] | 15 | RF*** | 51 ± 1  | 98 ± 2  | 4 ± 4  |
| RF**    | 49 | LDA   | 54 ± 3  | 89 ± 5  | 19 ± 6 |
| RF**    | 49 | RF*** | 54 ± 2  | 97 ± 2  | 11 ± 4 |
| MRMR    | 24 | LDA   | 22 ± 26 | 40 ± 47 | 3 ± 5  |
| MRMR    | 24 | RF*** | 51 ± 2  | 97 ± 2  | 4 ± 4  |

\* datasets as published in [15] or individually calculated on the dataset with RF or MRMR algorithm.

\*\* features were selected in ≥9/45 selection cycles

\*\*\* average values of 20 MCCV steps with ensembles of 1001 RFs

**Table 3** Classifier evaluation of spectral marker candidates of balanced dataset (ii), 8 cm<sup>-1</sup> spectral resolution, 83 UBC G2+ including recurrent cancer versus 83 control patients (for legend see Table 2).

| set*    | #f | cf  | acc    | sens    | spec    |
|---------|----|-----|--------|---------|---------|
| RF [15] | 15 | LDA | 55 ± 6 | 55 ± 10 | 55 ± 10 |
| RF [15] | 15 | RF  | 56 ± 5 | 53 ± 11 | 60 ± 10 |
| RF**    | 4  | LDA | 66 ± 6 | 67 ± 9  | 65 ± 9  |
| RF**    | 4  | RF  | 67 ± 6 | 66 ± 10 | 68 ± 10 |
| MRMR    | 3  | LDA | 67 ± 5 | 70 ± 8  | 63 ± 9  |
| MRMR    | 3  | RF  | 68 ± 5 | 71 ± 9  | 65 ± 9  |

\* datasets as published in [15] or individually calculated on the dataset with RF or MRMR algorithm.

\*\* features were selected in  $\geq 45/50$  selection cycles

**Table 4** Classifier evaluation of spectral marker candidates of balanced dataset (iii), 4 cm<sup>-1</sup> resolution, 50 UBC G2+ including 14 recurrent cases versus 50 control patients (for legend see Table 2).

| set*    | #f | cf  | acc    | sens    | spec    |
|---------|----|-----|--------|---------|---------|
| RF [15] | 15 | LDA | 75 ± 7 | 75 ± 11 | 75 ± 11 |
| RF [15] | 15 | RF  | 84 ± 5 | 82 ± 9  | 86 ± 9  |
| RF**    | 6  | LDA | 88 ± 5 | 93 ± 6  | 83 ± 9  |
| RF**    | 6  | RF  | 89 ± 5 | 91 ± 8  | 88 ± 9  |
| MRMR    | 7  | LDA | 89 ± 5 | 93 ± 6  | 84 ± 9  |
| MRMR    | 7  | RF  | 92 ± 5 | 93 ± 6  | 92 ± 8  |

\* datasets as published in [15] or individually calculated on the dataset with RF or MRMR algorithm.

\*\* features were selected in  $\geq 26/50$  selection cycles

**Table 5** Classifier evaluation of spectral marker candidates of balanced dataset (iv), 4 cm<sup>-1</sup> resolution, 50 UBC G2+ excluding recurrent cancer versus 50 control patients (for legend see Table 2).

| set*    | #f | cf  | acc    | sens    | spec    |
|---------|----|-----|--------|---------|---------|
| RF [15] | 15 | LDA | 73 ± 7 | 72 ± 11 | 74 ± 10 |
| RF [15] | 15 | RF  | 80 ± 8 | 78 ± 10 | 81 ± 11 |
| RF**    | 6  | LDA | 85 ± 5 | 90 ± 7  | 80 ± 9  |
| RF**    | 6  | RF  | 88 ± 4 | 87 ± 8  | 88 ± 8  |
| MRMR    | 2  | LDA | 88 ± 3 | 94 ± 6  | 83 ± 9  |
| MRMR    | 2  | RF  | 90 ± 4 | 93 ± 5  | 86 ± 8  |

\* datasets as published in [15] or individually calculated on the dataset with RF or MRMR algorithm.

\*\* features were selected in  $\geq 37/50$  selection cycles

## 2.5 Classification

Two classifiers with different processing power requirements were applied. The classifying linear dis-

criminant analysis (LDA), which requires only relatively little computing resources was performed using the Matlab provided routine ‘classify’, with *a priori* class membership estimation and a linear discriminant function [40, 41].

Second, a complex ensemble random forest classifier requiring advanced computational resources was applied [15]. In brief, a prediction is not achieved directly by the majority vote of the trees in a single random forest; rather, an ensemble of 1001 random forests was used for prediction based on the majority vote of the included random forests.

For all validation procedures, a strict leave-one-third-out MCCV scheme was obeyed, in which classifiers were trained on a randomly selected 2/3 of the dataset to predict the left-out 1/3 subjects.

In line with common practice, the accuracy was defined as percentage of correct classifications, whereas sensitivity reflects the percentage of true positive (UBC G2+) predictions among all cancer positive predicted patients, and specificity provides the percentage of true negative (non-UBC control) patients among all negative predicted patients.

## 2.6 Bioinformatics environment

Calculations of the random forest routines were performed within the Matlab environment, version 2012a and version 2013a with the R-project based [42] Matlab port (downloadable from <http://code.google.com/p/ran-domforest-matlab/>, January 30, 2013) on a High-Performance Computing Server Supermicro SYS-5086B with 8x Intel® Xeon® Westmere EX (E7-8837, 2.66 GHz, 8-Core), 512 GB RAM. The linear discriminant analysis (LDA) was performed with the internal Matlab function (‘classify’). Final cross-validation and MRMR feature selection were performed on office PCs equipped with Intel Core2Quad CPU Q9650@3.0 GHz, 8 GB RAM running Matlab 2012a, and Intel Core i7-3770 CPU @ 3.40 GHz, 8 GB RAM running Matlab 2013a.

## 3. Results and discussion

The applied preparation procedures, spotting and measurement setup were shown to generate highly reproducible spectra of bodyfluids [15]. Spectral marker candidate bands for the discrimination of UBC from control patients with urocystitis and urinary tract infection were identified, based on subtle spectral differences. With two classification systems, the feature set showed a high sensitivity of 93%, but only a low specificity of 46% for the disease discri-



mination of the then available data of only 135 study participants [15].

For further validation, we proposed the investigation of a larger patient cohort with a balanced class distribution. The following results were obtained along these lines.

To analyse large datasets in due time requires high throughput capabilities. Originally, we reported our system set-up operating with a spectral resolution of  $2\text{ cm}^{-1}$ . To increase the sample throughput, the scanning time per sample had to be reduced. Thus, reducing the instrumental resolution by half would halve the required interferogram in length. Therefore, the spectral resolution was limited to  $8\text{ cm}^{-1}$  with a theoretical gain of a fourfold sample throughput. A full 384 MTP was entirely scanned within 7 h as compared with 21 h at a resolution of  $2\text{ cm}^{-1}$  and the further reported parameters [15]. At  $4\text{ cm}^{-1}$  resolution, the scanning time was still reduced to 12 h, roughly doubling the original sample throughput. Thereby, it became apparent that the remaining procedures of data acquisition, interferogram processing, and mechanics contribute *a fortiori* to the measurement time with increased speed of the spectral acquisition.

The resolution reduction was accompanied by a reduction in the number of spectral variables of the dataset from 11,493 wavenumber-intensity pairs per patient to 5,751 ( $4\text{ cm}^{-1}$ ) and 2,871 ( $8\text{ cm}^{-1}$ ), which further reduced the computer time required for calculation.

The continued patient recruitment allowed us to select among the participants UBC cases of reliably diagnosed, unambiguously manifested cancer of grades G2, G3, and G4. Early and pre-cancer-states like papilloma, which may be too small to secrete detectable amounts of biomarker molecules into the blood, were excluded to assess the principal potential of FTIR spectroscopy to discriminate UBC from a urinary tract infection. Thus, the first patient group analysed consisted of (i) 81 control and 205 UBC G2+ patients, including 64 recurrent cancer cases, a total of 286 patients.

Cross-validation showed that on this dataset, distinguishing controls from UBC patients was possible with high sensitivity, but a rather poor specificity resulting in an average accuracy of 47% with both classifiers and all three feature sets (Table 2).

Considering the again existing imbalance of UBC and control patients, both classifiers may have given overdue weighting to the UBC class, resulting in the apparent low specificity. Thus, a balanced control and UBC patient set was selected for the analysis: dataset (ii) consisted of 166 patients, 83 controls versus 83 UBC G2+ patients.

On the previously identified fifteen spectral features, LDA and RF classifier achieved an accuracy of  $55 \pm 6\%$  and  $56 \pm 5\%$ , respectively. A *t*-test

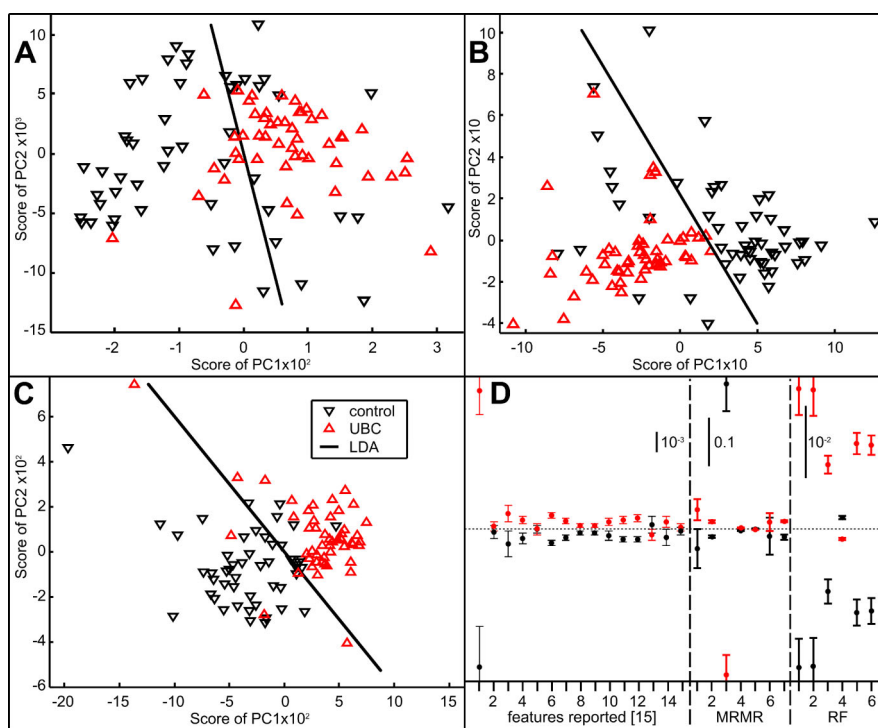
( $p < 0.001$ ) indicated an already existing significance over ambiguity, but still, this result is far from practical applicability. Nevertheless, the increased specificities of  $55 \pm 10\%$  and  $60 \pm 10\%$  are already noteworthy (Table 3).

Four features were identified specifically for this dataset by the RF algorithm in  $\geq 45/50$  selection cycles. LDA and RF classifier resulted in an accuracy of 66 and  $67 \pm 6\%$ , respectively. Sensitivities of  $67 \pm 9\%$  (LDA) and  $66 \pm 10\%$  (RF) were achieved. The specificity accounted for values of  $65 \pm 9\%$  (LDA) and  $68 \pm 10\%$  (RF), the highest achieved values on dataset (ii) (Table 3).

On three features identified by the MRMR algorithm, both classifiers also performed comparably well. The LDA achieved an accuracy of  $67 \pm 5\%$ , a sensitivity of  $70 \pm 8\%$ , and a specificity of  $63 \pm 9\%$ . The RF led to values of  $68 \pm 5\%$ ,  $71 \pm 9\%$ , and  $65 \pm 9\%$ , respectively. These accuracies on MRMR features differ significantly ( $p < 0.001$ ) from an ambiguous classification as achieved with the unbalanced dataset (i) (Tables 2, 3).

To evaluate whether discriminative features were obscured by the low spectral resolution, a balanced dataset (iii) of 50 control patients, 50 UBC G2+ patients including 14 recurrent cancer cases was acquired with  $4\text{ cm}^{-1}$  spectral resolution (Figure 2). On this dataset, a clear distinction was already achievable with both the LDA (accuracy of  $75 \pm 7\%$ ) and the RF classifier (accuracy of  $84 \pm 5\%$ ) on the previously reported set of 15 features (Figure 2, green lines). Remarkably, with sensitivities of  $75 \pm 11\%$  (LDA) and  $82 \pm 9\%$  (RF), and specificities of  $75 \pm 11\%$  (LDA) and  $86 \pm 9\%$  (RF), both classifiers outperformed the predictors we reported previously (acc.  $66 \pm 8\%$ , spec.  $45 \pm 14\%$  LDA, acc.  $68 \pm 7\%$ , spec.  $46 \pm 18\%$  RF [15]) especially with regard to specificity. These improvements were statistically significant ( $p < 0.001$ , Table 4), and the features enabled an accurate dataset separation (Figure 4A, D).

The performance of two entirely different classifiers indicates that an accurate, sensitive and specific discrimination of UBC patients from controls on infrared absorption spectra of blood can be achieved with the previously identified 15 spectral features. Remarkably, these were identified using an unbalanced dataset of 135 patients (89 UBC, 46 controls) which included 38 UBC G1 stages, and two prostate carcinoma cases within the controls. It was argued, that the poor specificity of the presented classifiers was most likely due to the mismatched class sizes. Here, we present evidence that training the applied LDA and ensemble RF classifiers with class-unbalanced data resulted in a generally unintended preference of prediction, so that sensitivity outbalanced specificity. Therefore, our data gives evidence to train these predictors exclusively on class-balanced datasets to avoid distortions in class membership prediction.



**Figure 4** Using three differently selected feature sets, 50 control and 50 UBC G2+ patients including 14 recurrent cases were well separable by LDA. For illustrative purposes, the dimensionality of the classification problem was reduced by PCA. The LDA discriminative function separating patients based on scores of the first two principal components is shown for (A) 15 features determined previously [15], (B) seven features determined on this dataset with the MRMR algorithm, and (C) six features determined on this dataset with the repeated RF algorithm in  $\geq 26/50$  selection cycles (Table 4). (D) The class-averaged, centred intensities and standard error of mean at the determined vibrational biomarker candidates of control (black) and UBC G2+ (red) patients illustrate the spectral separability.

On a six feature set determined individually for the dataset by the RF algorithm (threshold 26/50 cycles) (Figure 2, red lines), both classifiers performed comparably well with a respective accuracy of  $88 \pm 5\%$  and  $89 \pm 5\%$  (Table 4, Figure 4 C, D). The MRMR algorithm identified a set of seven best discriminating features (Figure 2, blue lines). Using these, the RF classifier performed with a higher accuracy of  $92 \pm 5\%$  versus  $89 \pm 5\%$  of the LDA (Table 4, Figure 4B, D).

The validation results of all classifiers on all three feature sets indicate a fair to superb class separability in datasets (ii) and (iii) (Figure 4). A prediction performance increase of 20 (LDA) and 30 percent units (RF) on the fixed 15 feature set, along with a comparable performance improvement using individually calculated feature sets, indicates that in dataset (ii) with only  $8 \text{ cm}^{-1}$  resolution insufficiently discriminating spectral features were present. For the analysis with the applied HT-FTIR-methodology, a spectral resolution of at least  $4 \text{ cm}^{-1}$  is required for an accurate UBC G2+ prediction.

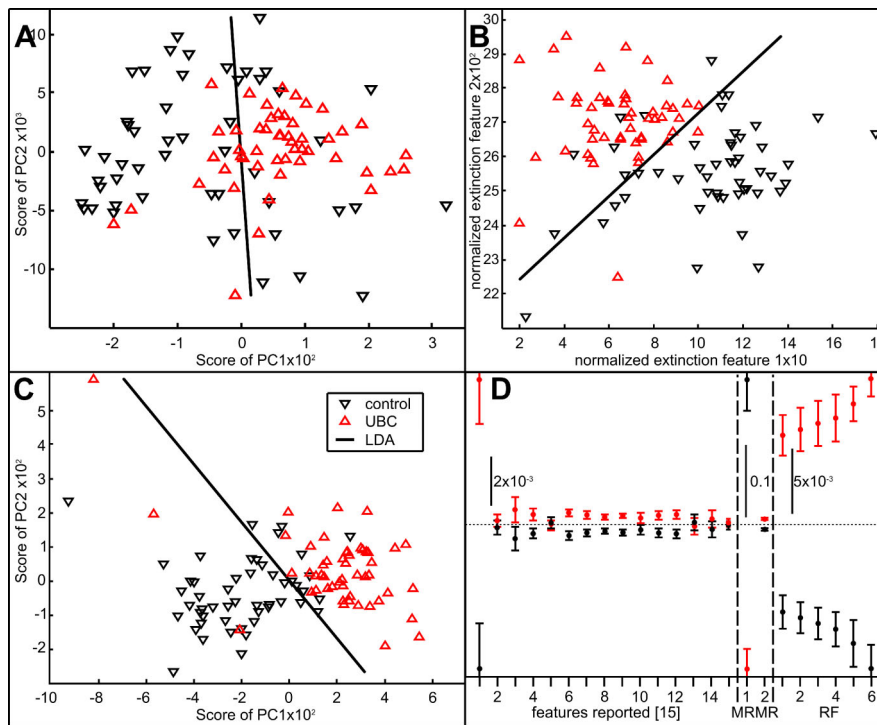
Recurrent cancer is a serious issue with UBC [5]. Therefore, nearly half of the patients with confirmed UBC recruited in our study suffer from a recurrent

tumour. In the following, we evaluated whether the spectral pattern of the blood samples was distorted by a pre-existing UBC history, using dataset (iv). This included 50 control patients, and 50 UBC G2+ without previous cancer history. The absorbance spectra were again recorded at  $4 \text{ cm}^{-1}$  resolution (Figure 3).

With the 15 previously identified features (Figure 3, green lines), the LDA reached an accuracy of  $73 \pm 7\%$  with  $72 \pm 11\%$  sensitivity and  $74 \pm 10\%$  specificity (Table 5, Figure 5A, D). Resulting in an accuracy of  $80 \pm 8\%$ , the RF classifier performed significantly ( $p < 0.001$ ) better, reaching a sensitivity and specificity of  $78 \pm 10\%$  and  $81 \pm 11\%$ , respectively.

Six features were identified with the RF algorithm specifically for this dataset (Figure 3, red lines). The LDA performed with  $85 \pm 5\%$  accuracy,  $90 \pm 7\%$  sensitivity and  $80 \pm 9\%$  specificity. The RF classifier reached an improved result with  $88 \pm 4\%$  accuracy,  $87 \pm 8\%$  sensitivity and  $88 \pm 8\%$  specificity (Table 5, Figure 5C, D).

The MRMR algorithm identified only two relevant features on this dataset (Figure 3, blue lines). We are well aware that a predictor based on such few features may be less robust against misclassifica-



**Figure 5** Using three differently selected feature sets, 50 control and 50 UBC G2+ patients without recurrent cancer were also well separable by LDA. For illustrative purpose, the dimensionality of the classification problem was reduced by PCA in feature sets (A) and (C). The LDA discriminative function is shown for (A) 15 features determined previously [15], (B) two features determined on this dataset with the MRMR algorithm, and (C) six features determined on this dataset with the repeated RF algorithm in  $\geq 37/50$  selection cycles (Table 5). (D) The class-averaged, centred intensities and standard error of mean at the determined vibrational biomarker candidates of control (black) and UBC G2+ (red) patients again illustrate the spectral separability.

tion of outlier patients. Spectral outliers of an individual sample, however, have been eliminated during preprocessing [15]. The RF classifier performed best with an accuracy of  $90 \pm 4\%$ , a sensitivity of  $93 \pm 5\%$  and a specificity of  $86 \pm 8\%$ . The LDA classifier led to similar results of  $88 \pm 3\%$  accuracy, a sensitivity of  $94 \pm 6\%$  and  $83 \pm 9\%$  specificity (Table 5, Figure 5B, D).

In total, using datasets (iii) and (iv), all classifiers validated comparably well with 85–92% accuracy on dataset specific features. Whether the UBC is recurrent appears not to affect the spectral prediction based upon a patient’s blood sample. The identified feature sets were obtained with distinct strategies from spectral datasets of different resolution. Thus, an overlap of wavenumber positions cannot be expected.

This finding could render an HT-FIR-blood analysis an attractive less invasive supplement to the diagnostics available for UBC patients in therapy. Currently, these patients are repeatedly examined by cystoscopy on a regular basis to monitor therapy progression. A spectroscopic blood test of high specificity would immediately and efficiently reduce the amount of stress put on the patient, reduce infection risks, and minimize hospital stays. For such an appli-

cation, the decay time of tumour-induced spectral patterns after therapy onset has to be validated.

Unfortunately, our current patient population is still short of a sufficient number of securely diagnosed recurrent UBC cases to evaluate the spectral separability from subjects with newly developed UBC. Judging from the small overlap of only two among 21 discriminative features identified specifically for datasets (iii) and (iv), a certain probability for the successful discrimination could be expected. However, we report from an on-going study, and this aspect will be evaluated in future.

The main findings about the existence of spectral biomarker candidates for UBC in blood preparations [15] have been confirmed. Even more, the classification results were fundamentally improved. The analysis provided validated results with a particularly enhanced specificity.

#### 4. Conclusion

Here, we demonstrated the principal existence of marker patterns for the discrimination of manifested

UBC from urinary tract infections in the FTIR absorbance spectra of blood samples from a risk collective. Finally, the discriminative power of the technique for the identification of other diseases in a broad screening approach remains to be shown with specifically defined patient groups of appropriate size.

The improved results of our HT-FTIR-spectroscopic approach to bodyfluid analysis [15] demonstrate its practical applicability. The accurate discrimination of UBC from control patients was shown on three balanced datasets with and without recurrent cancer cases. For each dataset, three combinations of spectrally discriminative features were identified and evaluated.

A significantly ( $p < 0.001$ ) better than ambiguous patient group separability was already obtained with a spectral resolution of  $8 \text{ cm}^{-1}$ , corresponding to a threefold increased sample throughput compared with our earlier study. With our data, optimum prediction quality was achieved with  $4 \text{ cm}^{-1}$  resolution datasets, still equivalent to a doubled sample throughput compared with our previously reported procedure [15].

Using  $4 \text{ cm}^{-1}$  spectrally resolved data, even the least discriminative set of spectral biomarker candidates resulted in an RF classification accuracy of  $80 \pm 8\%$ , with  $78 \pm 10\%$  sensitivity and  $81 \pm 11\%$  specificity. The sample preparation process and spectral measurement was strictly automated as far as reasonably achievable. Particularly, critical steps of the thin film preparation from fluid samples were performed by specialized robotics. Therewith, operator impact on data processing and evaluation was ruled out by automated procedures. Therefore, objective evidence for the existence of blood-borne spectral biomarkers from HT-FTIR spectroscopic analysis is given.

Further studies with a progressively increasing patient population will support the identification of the optimum spectral feature combinations and the most accurate classifier. Further prediction performance testing remains to be performed with even larger independent datasets.

**Acknowledgements** These studies were made available by support of PURE (Protein research Unit Ruhr within Europe), financed by the state of North Rhine-Westphalia. The participation of all members of the PURE consortium is acknowledged. Particularly for the organization and support of the clinical study with regard to bladder cancer, we thank Thomas Deix and Katharina Braun of the Urological Clinics (director Joachim Noldus) of the Marien-Hospital Herne for collaboration. The authors are grateful to all patients participating in the reported study.

**Author biographies** Please see Supporting Information online.

## References

- [1] D. M. Parkin, *Scand. J. Urol. Nephrol. Suppl.* **42**(s218), 12 (2008).
- [2] M. Ploeg, K. K. H. Aben, and L. A. Kiemeny, *World J. Urol.* **27**, 289 (2009).
- [3] P. Boffetta, *Scand. J. Urol. Nephrol. Suppl.* **42**(s218), 45 (2008).
- [4] S. M. Cohen, T. Shirai, and G. Steineck, *Scand. J. Urol. Nephrol. Suppl.* **205**, 105 (2000).
- [5] M. Adibi, R. Youssef, S. F. Shariat, Y. Lotan, C. G. Wood, A. I. Sagalowsky, R. Zigeuner, F. Montorsi, C. Bolenz, and V. Margulis, *Int. J. Urol. Off. J. Jpn. Urol. Assoc.* **19**, 1060 (2012).
- [6] J. L. Summers, J. S. Coon, R. M. Ward, W. H. Falor, A. W. Miller 3rd, and R. S. Weinstein, *Cancer Res.* **43**, 934 (1983).
- [7] S. Ramakumar, J. Bhuiyan, J. A. Besse, S. G. Roberts, P. C. Wollan, M. L. Blute, and D. J. O'Kane, *J. Urol.* **161**, 388 (1999).
- [8] I. Osman, *Clin. Cancer Res.* **12**, 3374 (2006).
- [9] J. Villanueva, *J. Clin. Invest.* **116**, 271 (2005).
- [10] L. C. Kompier, A. A. G. van Tilborg, and E. C. Zwarthoff, *Urol. Oncol.* **28**, 91 (2010).
- [11] C. J. Marsit, D. C. Koestler, B. C. Christensen, M. R. Karagas, E. A. Houseman, and K. T. Kelsey, *J. Clin. Oncol.* **29**, 1133 (2011).
- [12] N. Putluri, A. Shojaie, V. T. Vasu, S. K. Vareed, S. Nalluri, V. Putluri, G. S. Thangjam, K. Panzitt, C. T. Tallman, C. Butler, T. R. Sana, S. M. Fischer, G. Sica, D. J. Brat, H. Shi, G. S. Palapattu, Y. Lotan, A. Z. Weizer, M. K. Terris, S. F. Shariat, G. Michailidis, and A. Sreekumar, *Cancer Res.* **71**, 7376 (2011).
- [13] A. A. G. van Tilborg, L. C. Kompier, I. Lurkin, R. Poort, S. El Bouazzaoui, K. van der Keur, T. Zuijverloon, L. Dyrskjot, T. F. Orntoft, M. J. Roobol, and E. C. Zwarthoff, *PLoS One* **7**, e43345 (2012).
- [14] M. R. Karagas, A. S. Andrew, H. H. Nelson, Z. Li, T. Punshon, A. Schned, C. J. Marsit, J. S. Morris, J. H. Moore, A. L. Tyler, D. Gilbert-Diamond, M.-L. Guerinot, and K. T. Kelsey, *Hum. Genet.* **131**, 453 (2011).
- [15] J. Ollesch, S. L. Drees, H. M. Heise, T. Behrens, T. Brüning, and K. Gerwert, *The Analyst* **138**, 4092 (2013).
- [16] J. Moecks, G. Kocherscheidt, W. Koehler, and W. H. Petrich, in: *Proc SPIE*, edited by A. Mahadevan-Jansen, M. G. Sowa, G. J. Puppels, Z. Gryczynski, T. Vo-Dinh, and J. R. Lakowicz (San Jose, CA, 2004), pp. 117–123.
- [17] G. G. Harrigan, R. H. LaPlante, G. N. Cosma, G. Cockerell, R. Goodacre, J. F. Maddox, J. P. Luyendyk, P. E. Ganey, and R. A. Roth, *Toxicol. Lett.* **146**, 197 (2004).
- [18] B. H. Menze, W. Petrich, and F. A. Hamprecht, *Anal. Bioanal. Chem.* **387**, 1801 (2007).
- [19] W. Petrich, K. B. Lewandrowski, J. B. Muhlestein, M. E. H. Hammond, J. L. Januzzi, E. L. Lewandrowski, B. Dolenko, J. Früh, W. Köhler, and R. Mischler, *Analyst* **134**, 1092 (2009).



- [20] H.-U. Gremlich and B. Yan, *Infrared and Raman Spectroscopy of Biological Materials* (M. Dekker, New York, 2001).
- [21] M. Diem, J. M. Chalmers, and P. R. Griffiths, *Vibrational Spectroscopy for Medical Diagnosis* (John Wiley & Sons, Chichester, England; Hoboken, N.J., 2008).
- [22] P. Lasch and J. Kneipp, *Biomedical Vibrational Spectroscopy* (Wiley-Interscience, Hoboken, N.J., 2008).
- [23] G. Hoşafçı, O. Klein, G. Oremek, and W. Mänteles, *Anal. Bioanal. Chem.* **387**, 1815 (2007).
- [24] P. Lasch, J. Schmitt, M. Beekes, T. Udelhoven, M. Eiden, H. Fabian, W. Petrich, and D. Naumann, *Anal. Chem.* **75**, 6673–6678 (2003).
- [25] T. C. Martin, J. Moecks, A. Beloousov, S. Cawthraw, B. Dolenko, M. Eiden, J. Von Frese, W. Kohler, J. Schmitt, R. Somorjai, T. Udelhoven, S. Verzakov, and W. Petrich, *The Analyst* **129**, 897 (2004).
- [26] D. I. Ellis and R. Goodacre, *Analyst* **131**, 875 (2006).
- [27] M. Beekes, P. Lasch, and D. Naumann, *Vet. Microbiol.* **123**, 305 (2007).
- [28] G. Bellisola and C. Sorio, *Am. J. Cancer Res.* **2**, 1 (2012).
- [29] J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scott, and F. L. Martin, *The Analyst* **137**, 3202 (2012).
- [30] P. Carmona, M. Molina, M. Calero, F. Bermejo-Pareja, P. Martínez-Martín, and A. Toledano, *J. Alzheimers Dis.* **34**, 911 (2013).
- [31] K. Gajjar, J. Trevisan, G. Owens, P. J. Keating, N. J. Wood, H. F. Stringfellow, P. L. Martin-Hirsch, and F. L. Martin, *Analyst* **138**, 3917 (2013).
- [32] D. Sheng, Y. Wu, X. Wang, D. Huang, X. Chen, and X. Liu, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **116**, 365 (2013).
- [33] H. Peng, F. Long, and C. Ding, *Pattern Anal. Mach. Intell. IEEE Trans. On* **27**, 1226 (2005).
- [34] C. Ding and H. Peng, *J. Bioinform. Comput. Biol.* **3**, 185 (2005).
- [35] L. Breiman, *Mach. Learn.* **45**, 5 (2001).
- [36] T. Hastie, R. Tibshirani, M. B. Eisen, A. Alizadeh, R. Levy, L. Staudt, W. C. Chan, D. Botstein, and P. Brown, *Genome Biol.* **1**, RESEARCH0003 (2000).
- [37] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, *BMC Bioinformatics* **10**, 213 (2009).
- [38] P. R. Griffiths and J. A. De Haseth, *Fourier Transform Infrared Spectrometry* (Wiley-Interscience, Hoboken, N.J., 2007).
- [39] R. Genuer, J. M. Poggi, and C. Tuleau-Malot, *Pattern Recognit. Lett.* **31**, 2225 (2010).
- [40] G. A. F. Seber, *Multivariate Observations* (Wiley-Interscience, Hoboken, N.J., 2004).
- [41] W. J. Krzanowski, *Principles of Multivariate Analysis: a User's Perspective* (Oxford Univ. Pr., Oxford [u.a.], 2008).
- [42] A. Liaw and M. Wiener, *R News* **2**, 18 (2002).