Metabolomics (2012) 8:S111–S121
DOI 10.1007/s11306-011-0310-7

ORIGINAL ARTICLE

Fluorescence spectroscopy as a potential metabonomic tool for early detection of colorectal cancer

Anders Juul Lawaetz · Rasmus Bro · Maja Kamstrup-Nielsen · Ib Jarle Christensen · Lars N. Jørgensen · Hans J. Nielsen

Received: 12 January 2011 / Accepted: 6 April 2011 / Published online: 21 April 2011

Abstract Fluorescence spectroscopy Excitation Emission Matrix (EEM) measurements were applied on human blood plasma samples from a case control study on colorectal cancer. Samples were collected before large bowel endoscopy and included patients with colorectal cancer or with adenomas, and from individuals with other non malignant findings or no findings ($N = 308$). The objective of the study was to explore the possibilities for applying fluorescence spectroscopy as a tool for detection of colorectal cancer. Parallel Factor Analysis (PARAFAC) was applied to decompose the fluorescence EEMs into estimates of the underlying fluorophores in the sample. Both the pooled score matrix from PARAFAC, holding the relative concentrations of the derived components, and the raw unfolded spectra were used as basis for discrimination models between cancer and the various controls. Both methods gave test set validated sensitivity and specificity values around 0.75 between cancer and controls, and poor discriminations between the various controls. The PARAFAC solution gave better options for analyzing the chemical mechanisms behind the discrimination, and revealed a blue shift in tryptophan emission in the cancer patients, a result that supports previous findings. The present findings show how fluorescence spectroscopy and chemometrics can help in cancer diagnostics, and with PARAFAC fluorescence spectroscopy can be a potential metabonomic tool.

Keywords Fluorescence spectroscopy · Colorectal cancer · Chemometrics · PARAFAC · Metabonomics

A. J. Lawaetz (✉) · R. Bro · M. Kamstrup-Nielsen
Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark
e-mail: ajla@life.ku.dk

I. J. Christensen
Finsen Laboratory, Copenhagen, Denmark

L. N. Jørgensen
Department of Surgery, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark

H. J. Nielsen
Department of Surgical Gastroenterology, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

1 Introduction

The idea of using autofluorescence measurements of blood to discriminate people with cancer from non-cancer was first presented by Leiner, Wolbeis and co-workers in the 1980s. They considered the fluorescence excitation emission matrix (EEM) of a diluted blood serum sample as a base for pattern recognition to monitor the health status of a person. The hypothesis was that, due to the high sensitivity of fluorescence spectroscopy, it would be possible to observe even small deviations in the fluorescence spectrum from "normal" healthy subjects to diseased subjects (Leiner et al. 1983, 1986; Wolfbeis and Leiner 1985). This hypothesis actually fits well into the present theories of metabonomic based diagnostics. Metabonomic based diagnostics explores metabolites in a biological system and its response to a stress situation such as disease. Metabonomics is often based on non-targeted quantitative and qualitative measurements using nuclear magnetic resonance spectroscopy (NMR) or chromatography [liquid (LC) or gas (GC)] combined with mass spectroscopy (MS) (Nordström and Lewensohn 2010; Zhang et al. 2007). In the present study we explore the possibilities for

introducing fluorescence spectroscopy of blood plasma samples as an alternative metabonomic tool for detection of cancer.

Other publications have followed up on the work from Leiner and co-workers or applied other strategies in using autofluorescence on blood to detect cancer (Hubmann et al. 1990; Kalaivani et al. 2008; Leineret al. 1986; Madhuri et al. 1997, 1999, 2003; Masilamani et al. 2004; Nørgaard et al. 2007; Uppal et al. 2005; Xu et al. 1988). Different approaches have been used; some use extracts or controlled fractions of the plasma, whereas others use the plasma or serum merely diluted or with no sample treatment at all. The studies by Madhuri et al. (1999, 2003) and by Masilamani et al. (2004) use an acetone extract of blood plasma in order to reduce spectral interference in their attempt to measure emission from porphyrins. The results from these studies show elevated levels of porphyrins in cancer patients compared to healthy patients. In the present study we will therefore also have a focus on emission from porphyrins.

Common for almost all of the previous studies was the use of only few or single specific wavelength pairs as opposed to the whole spectral approach combined with chemometrics used in the present study. Only the study from Nørgaard et al. (2007) applied chemometrics in their data analysis, and they got promising results on serum samples from breast cancer patients. The use of chemometrics allows us to use the whole spectrum instead of focusing on single wavelength pairs. Multivariate data analysis/chemometrics is a cornerstone in metabonomics used to extract important information from the complex data output, and hereby hopefully identify specific metabolites with discriminatory or predictive ability (biomarkers) that can be used e.g. for a diagnostic purpose (Ragazzi et al. 2006; Ward et al. 2006). The lack of methods to extract the useful information from the EEMs was exactly a problem for Leiner and co-workers and hence, despite the rather complex EEM measurements, the outcome of their analysis was a simple ratio between two wavelength pairs. In the present study, we apply chemometrics on the fluorescence spectra to discriminate between blood plasma samples from colorectal cancer (CRC) patients and healthy individuals. We apply two different methods of data analysis; one which has been applied previously using the raw spectra as input to the classification model, and one where we extract underlying chemical information from the spectra by Parallel Factor Analysis (PARAFAC) (see materials and methods for a description of PARAFAC). The combination of fluorescence spectroscopy and PARAFAC has not previously been applied in a diagnostic test approach. The combination of PARAFAC and three-way fluorescence data (the EEMs) is especially fruitful, as the parameters of the PARAFAC model can be seen as estimates of the relative concentrations (scores) and the emission and excitation spectra (loadings) of the fluorophores in the sample (Andersen and Bro 2003; Bro 1997). As for conventional NMR and LC–MS this chemical identification opens for fluorescence spectroscopy as a metabonomic tool.

Fluorescence spectroscopy is widely applied in biomarker research though almost solely in the field of labeled fluorescence, where designed fluorescence probes are used to detect the presence of specific biomarkers (Hamdan 2007). In autofluorescence or intrinsic fluorescence, naturally occurring fluorophores are measured with or without minimal sample preparation (Lakowicz 2006). The number of fluorophores in a blood sample is limited compared to the number of compounds detectable by MS and NMR, though among the fluorophores, biologically important compounds are found. In blood for example, the amino acids tryptophan, tyrosine and phenylalanine and also some cofactors and flavonoids NAP, NAD(P)H, FAD are among the fluorophores (Wolfbeis and Leiner 1985). Compared to MS and NMR, fluorescence spectroscopy is highly sensitive and can thus measure concentrations down to parts per billion (Lakowicz 2006). The fluorescent signal from a fluorophore is dependent on the surrounding environment. For example, tryptophan groups in different proteins or on different positions in the same protein can have different excitation and emission maxima, and can thus be distinguished from each other (Abugo et al. 2000). In fact Leiner et al. (1986) showed a difference in the fluorescence from the amino acid tryptophan in human serum from healthy individuals and patients with gynaecological malignancies.

In the practical data acquisition, fluorescence spectroscopy has some advantages compared to both traditional metabonomic techniques. Sample preparation is limited to a minimum of only diluting the sample, and the time of acquisition can be down to few minutes, depending on the spectral area covered and the integration time. A spectrofluorometer can be small and compact compared to MS and NMR, and the price is often much lower. Compared to standard diagnostic tools such as X-ray, MR and CT scanning, fluorescence spectroscopy is very cheap, but at the present stage not a viable alternative. Compared to targeted methods for single biomarkers based on immunochemical tests the onetime investment in fluorescence spectroscopy is, like in MS and NMR, relatively high, but the running costs are much lower, and fluorescence spectroscopy is faster and easy to use.

Some drawbacks of fluorescence spectroscopy are the instrument dependent results that call for spectral correction before they are globally comparable (DeRose and Resch-Genger 2010). The fluorescence intensity is also highly dependent on the overall absorbance of the sample. At low concentrations of fluorophores (and/or low absorbance), the

linear relation between concentration and intensity known from Lambert-Beers law is also valid in fluorescence spectroscopy. At higher concentrations/high absorbance this relation is broken. This phenomenon is called concentration quenching or the inner filter effect (Lakowicz 2006). Blood plasma is highly absorbent, and thus precautions must be taken to avoid or reduce inner filter effects. In the present study the samples are both diluted and undiluted. For the undiluted samples the pathway of the exciting light is reduced to reduce absorbance.

Colorectal cancer is one of the most frequent malignant diseases for both women and men in the western world. In Denmark in 2008, 4194 cases of CRC were diagnosed, which accounted for more than 12% of all malignant diseases (The Danish Cancer Society 2010; The Danish National Board of Health 2010). The 5-year survival rate of CRC patients is approximately 50%, only ovarian, lung, and pancreas cancers have lower rates (UK, national statistics, 2010). The low rate is primarily due to high recurrence frequencies in some patients undergoing intended curative resection and disseminated disease at the time of diagnosis in other patients. At present fecal occult blood test (FOBT) combined with subsequent colonoscopy in those with positive tests is the method of choice for early detection of colorectal cancer. In recent years national screening programs based on FOBT have been introduced in several countries. The FOBT has been criticized for limited compliance rates, which reduce the advantage of the test, and therefore new, improved screening modalities with high compliance rates are urgently needed (Jenkinson and Steele 2010). The only accepted serum biomarker for CRC is carcinoembryonic antigen (CEA), but with sensitivity and specificity values of 0.34/0.93, this is only accepted for prognosis after detection. Other biomarkers have been suggested with similar or better performance, for example free DNA (Flamini et al. 2006) and plasma lysophosphatidylcholine levels (Zhao et al. 2007). None of these biomarkers have yet been clinically accepted. In search for alternative methods with improved detection rates, and/or better compliance rates in screening for CRC, a metabonomic approach with broad unbiased search for changes in the metabolic profile is a possible solution. Interesting results have been published by Ward et al. (2006) by use of MALDI MS. The present paper will explore whether a solution with fluorescence spectroscopy could be an interesting approach.

2 Materials and methods

2.1 Samples

Human plasma samples (sodium citrate anticoagulant) from 308 individuals were used for the experiment. The samples are a part of a larger sample set from a multicentre cross sectional study conducted at six Danish hospitals of patients undergoing large bowel endoscopy due to symptoms associated with CRC (Nielsen et al. 2008). The present sample set is designed as a case control study with one case group (verified CRC) and three different control groups. The three control groups are (1) healthy subjects with no findings at endoscopy, (2) subjects with other, non malignant findings and (3) subjects with pathologically verified adenomas (Lomholt et al. 2009). Each of the groups, case and controls, consisted of samples from 77 individuals. Additional control samples, standardized pooled human citrate plasma, were purchased from 3H-Biomedical AB, Sweden.

2.2 Sample handling and data acquisition

Before measurements, the samples were defrosted on wet ice (0°C) for app. one hour, or until thawed, and each sample was divided in four aliquots of 200 µL to 1 mL for different analytical methods. The divided samples were immediately refrozen at −80°C. The standardized plasma samples were received in 50 mL aliquots, and stored at −80°C. Before use they were thawed at 0°C and divided into aliquots of 300 µL, and refrozen at −80°C. For fluorescence measurements, the samples were defrosted on wet ice (0°C) for app. 40 min.

The samples were measured both undiluted and in a hundred fold dilution in Phosphate Buffered Saline (PBS) (pH 7.4). The diluted samples were prepared immediately after the samples were thawed, and then stored on wet ice (0°C) until measured (app. 20 min). The non diluted fractions of the samples were measured as fast as possible after thawing. Fluorescence spectra were acquired on an FS920 spectrometer (Edinburgh Instruments) with double monochromators and a red sensitive photomultiplier (R928P, Hamamatsu) in a cooled detector house. The EEMs were acquired for the samples using the following settings. Diluted and undiluted samples were measured with excitation from 250 to 450 nm with a 5 nm increment, and emission from 300 to 600 nm with a 1 nm increment. Integration time was 0.05 s. This spectral area consists of light in both the ultra violet and visual area. The ultra violet area is dominated by excitation and emission from the aromatic aminoacids tyrosine and tryptophan hence the fluorescence from proteins. The visual area covers among other things excitation and emission from vitamins and cofactors (for example riboflavin and NAD(P)H) (Wolfbeis and Leiner 1985). In an attempt to capture emission from porphyrins, additional EEMs were acquired from the undiluted samples with excitation wavelengths from 385 to 425 nm with a 5 nm increment and emission wavelengths from 585 to 680 nm with a 1 nm increment, and an

integration time of 0.2 s. Every day a spectrum of the PBS used for dilution was measured with the same settings as the diluted samples. Excitation and emission slit widths were set at 4 nm for all measurements. The fluorescence data were corrected for the wavelength dependent excitation intensity by an internal reference detector in the spectrometer. Likewise the spectra were corrected for instrument dependent emission spectral biases by a correction factor supplied with the instrument. Total time spent for measuring all three EEMs was app. 40 min.

Diluted samples were measured in a 10 × 10 mm quartz cuvette. To reduce inner filter effect in the undiluted samples, these were measured in a 2 × 10 mm quartz cuvette with the 2 mm in the emission direction.

An external cooling system was mounted on the spectrometer keeping the measurement temperature constant at 15°C. To monitor the performance of the fluorescence instrument, a standard plasma sample was measured every day. All spectra were saved as ASCII and exported to Matlab® by an in-house routine. The raw spectra are available for download at http://www.models.life.ku.dk/.

2.3 Data analysis

Some samples were discarded due to either obviously erroneous measurements, or too little sample material. From the three different EEMs acquired, the numbers of samples ready for data analysis were then 301, 295 and 300 from low wavelength undiluted, high wavelength undiluted and diluted, respectively. Before the actual data analysis, the data were subjected to certain signal processing steps meant to appropriately handle and minimize the influence from non-relevant artifacts. When measuring fluorescence EEMs, non-chemical phenomena such as Rayleigh scatter and second order fluorescence may be present (Lakowicz 2006). These were removed and replaced with missing data and zeros using in-house software (Andersen and Bro 2003). For the diluted samples, a background spectrum of the solute PBS, measured the same day as the sample, was subtracted from each sample in order to remove possible Raman scatter (McKnight et al. 2001). All samples were intensity calibrated by normalizing to the integrated area of the water Raman peak of a sealed water sample measured each day prior to the measurements. This converts the scale into Raman units and allows comparison of intensity of samples measured on other fluorescence spectrometers (Lawaetz and Stedmon 2009).

A data reduction/decomposition of the fluorescence EEMs to less complex features was performed using the multi-way decomposition method called PARAFAC. A set of fluorescence EEMs can be seen as a three-way data array ($I \times J \times K$), where I is the number of samples measured (objects), J the number of emission wavelengths, and K the number of excitation wavelengths. Just as PCA is decomposing a two-way data matrix, a three-way data structure can be decomposed by PARAFAC into a number of latent PARAFAC components, by minimizing the sum of squared residuals e in the PARAFAC model (equation below).

$$X_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}$$

a_{if} is the ith element of the score vector, b_{jf} the jth element of the loading vector of the emission mode and c_{kf} the kth element of the loading vector for the excitation mode, for the fth PARAFAC component. If the correct number of PARAFAC components is used to decompose data with an approximately true trilinear structure and an appropriate signal to noise value, the solution from the PARAFAC model will give estimates of the true underlying profiles of the variables. This makes PARAFAC perfect for fluorescence spectroscopy when applied on EEMs. The loadings and scores can be treated as estimates of the excitation and emission spectra, and relative concentrations of the fluorophores in the samples respectively (Andersen and Bro 2003; Bro 1997).

PARAFAC models were fitted applying nonnegativity constraints on all parameters in the model (Andersen and Bro 2003); hence the estimated parameters were found in such a way that they would not be negative. Models were validated by split-half analysis (Harshman and DeSarbo 1984) combined with trained judgment of the loadings. PARAFAC models were fitted separately to each of the three sets of EEMs. The score matrices from the PARAFAC analyses were pooled to one matrix with 19 variables, which were subjected to further data analysis. PCA was fitted to get a preliminary overview of the data. Classification models were built using PLS-DA, a PLS regression with the pooled PARAFAC scores as independent X variables and a dummy matrix as the dependent Y variable with ones for samples belonging to the class, and zeros for samples not belonging to the class (Wold et al. 2001). Forward selection was applied for variable selection. For all classifications, the data sets were divided into training and test sets (10–30% in test set). The training sets were used for model building, and the test samples were used for validating the models. During model building of the training sets, the models were cross validated with 10% of the samples randomly removed in each segment and averaging over ten repetitions for each cross-validation run. The test sets for subsequent model validation were randomly selected from the data with the same relative number of samples removed from each class.

As an alternative to building classification models on the three combined PARAFAC score matrices, classification was tried directly with the raw spectra as the independent

variables. Variable selection was applied using Interval PLS (iPLS) (Nørgaard et al. 2000). Before the direct classification the three-way array of EEMs were unfolded to a two-way matrix.

All data analyses were performed in Matlab R2010® (The Mathworks Inc.) and chemometric analyses were performed in PLS_Toolbox v.5.8.2 (Eigenvector Research, Inc).

3 Results and discussion

Spectra from the three setups are seen in Fig. 1. Comparing the spectra from one undiluted sample and a sample diluted 100 times (leftmost and rightmost spectra respectively in the figure) the effect of dilution is clear. In both the raw undiluted sample and in the diluted, the major peak is in the region with fluorescence from the aromatic amino acids tryptophan and tyrosine (phenylalanine is also among the fluorescing amino acids, but it has excitation/emission maximum outside the measured area). For the undiluted sample there are two distinct peaks in that area, whereas in the diluted sample there is only one distinct peak. Furthermore in the undiluted sample a distinct peak is seen with emission maximum at a higher wavelength. The complex peak structure indicates that it is a mixture of several peaks, which could reflect analytes such as NAD(P)H, FAD, Riboflavin etc. (Wolfbeis and Leiner 1985). This peak structure is not apparently visible in the diluted sample.

It is also worth noticing that the intensity of the diluted sample is higher than the raw. This shows that even though the raw sample is measured in a micro cuvette, it still suffers from inner filter effect. Though it was also observed that the dilution in PBS buffer had an effect besides the reduced inner filter effect, a slight blue shift was observed in emission following excitation at 295 nm in the diluted samples. This might be explained by a slight change in the configuration of the proteins, which can change the emission profile.

The high wavelength area of the undiluted samples was measured separately in order to capture possible fluorescence from porphyrins. In the diluted samples this area gave no signal and was therefore not measured. In Fig. 1, middle plot, the high wavelength area primarily shows the descending tail of a peak with maximum outside the measured area, but a closer inspection of the EEM reveals a little bump at app. 405/610 nm which is in accordance with literature values of porphyrin fluorescence (Madhuri et al. 2003).

In order to monitor the performance of the fluorescence spectrometer, a standard plasma sample was measured every day. The standard deviation among these standard samples was up to five times lower than the standard deviation for the real samples, indicating good performance of the instrument and consistent sample handling, and at the same time revealing a large biological variation among the real samples.

On each of the three measured areas, a PARAFAC model was fitted. Due to the high complexity of the plasma matrix and the large biological variation in the samples, a large number of PARAFAC components was expected, which makes modelling more challenging. For the undiluted samples in the main spectral area (excitation from 250 to 450 nm, emission from 300 to 600 nm), ten PARAFAC components were chosen. To the spectra from the diluted samples, a model of six PARAFAC components was fitted. Only a reduced area of the spectra from the diluted samples was used, as the highest emission and excitation wavelengths did not contribute positively to the model. To the last selected area, the high wavelength area of the undiluted samples, a three component PARAFAC model was fitted. The number of PARAFAC components reflects the chemical rank of the system. For each component we get a set of loadings and scores, which are estimates of the excitation and emission profiles for the underlying chemical compounds. The excitation and emission loadings for the three models are seen in Fig. 2. Many of the components can be identified chemically but some are more difficult and even impossible to assign to specific chemical analytes. Despite the large number of PARAFAC components it is possible that some of these peaks reflect more than one chemical compound and the non-Gaussian peak shape of some of the loadings supports this.

In case of "just" making a model to discriminate between cancer and non cancer the issue would be to; objectively and in an unsupervised manner reflect the underlying variation, and then chemical assignment is of secondary concern. On the other hand if we at the same time want to gain knowledge about the reason for the discrimination and hereby move fluorescence spectroscopy into the world of metabonomics, chemical identification is an important parameter. A perfect PARAFAC model will give loadings which are estimates of the underlying excitation and emission spectra, and therefore we expected more unambiguous loadings with better options for chemical assignment. The reason for such non-ideal behaviour can be a low signal of some analytes, correlation between different compounds or non-linear behaviour due to quenching and similar phenomena. Given the relatively low number of samples and that some of the samples are not diluted, it is actually impressive that the PARAFAC models come out as chemically interpretable as they do. Still, we anticipate that the interpretability would be possible to improve if many more samples were included in

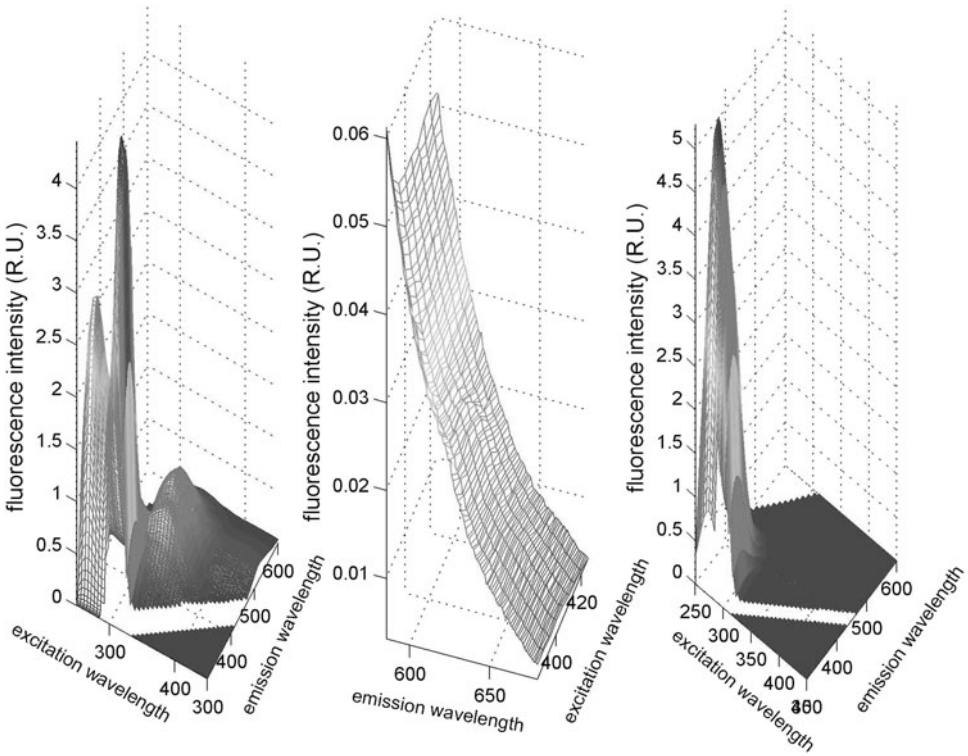

Fig. 1 Different EEMs recorded on one sample. *Left*: undiluted sample in main spectral area. *Middle*: undiluted sample in high wavelength area (notice the *axes* are different from the two other). *Right*: sample diluted 100 times in PBS

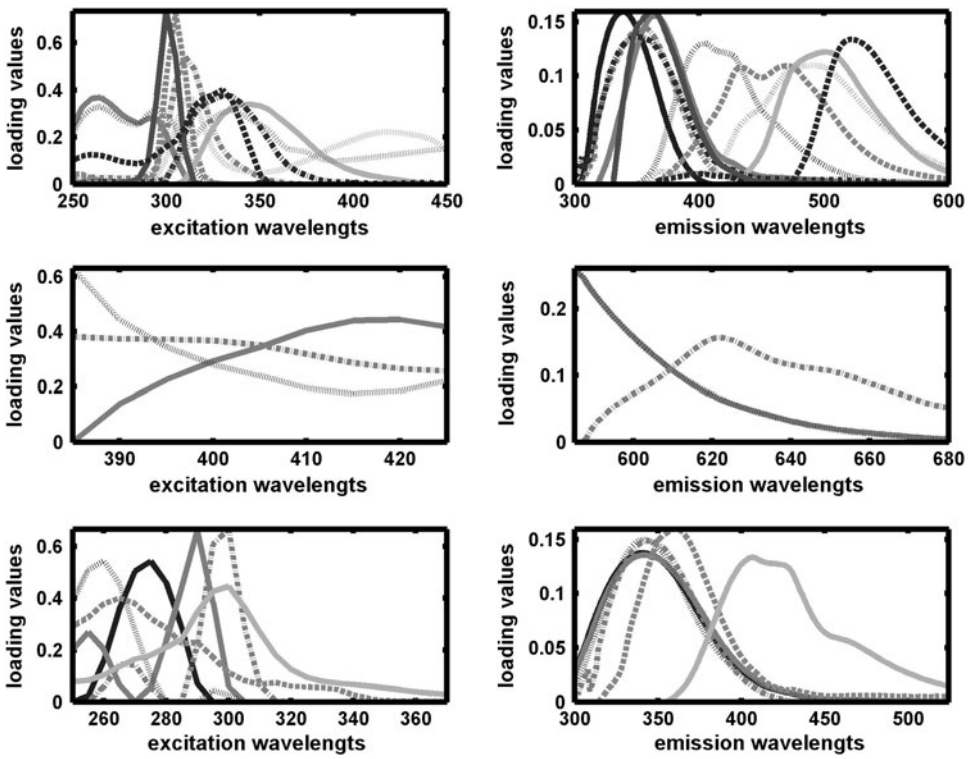

Fig. 2 PARAFAC excitation and emission loadings from the three datasets. *Upper*: undiluted main area. *Middle*: undiluted high wavelength area. *Lower*: diluted main area

the model and possibly also by using targeted standard addition of hypothesized analytes in the modelling phase.

Qualified presumptions on the chemical origin of some of the loadings are made. In both the undiluted and the diluted samples, several loadings are seen with excitation maximum from 250 to 305 nm, and emission maximum from app. 330 to 350 nm. In this region, fluorescence from protein-bound tryptophan is strong. The emission from

tryptophan can shift when the polarity of the microenvironment changes, hence tryptophan which is bound to different proteins, or at the internal or external parts of a protein, can give rise to different emission maxima. In fact, literature values are reported for tryptophan emissions from 307 to 355 nm (Vivian and Callis 2001). This can explain the numerous peaks for tryptophan emission. Some of the excitation loadings fit well with excitation of tyrosine (app. 265 nm) whereas there is no emission loading supporting the presence of tyrosine emission (app. 300 nm). Energy transfer from excited state tyrosine to tryptophan is a known phenomenon and a reasonable explanation of the absent emission from tyrosine (Lakowicz 2006).

The peaks with maximum at higher wavelengths in both the undiluted and diluted samples can possibly be assigned to compounds such as NAD(P)H, FAD and FMN. In the model from the high wavelength region, it is worth noticing that the little, hardly visible "bump" in the pure spectra gives a clear component with excitation/emission maximum at 400/620 which is in agreement with literature values for porphyrins. There are two other components in this model. One has excitation maximum at 420 nm, but emission maximum outside the measured area, and the other has both excitation and emission maxima outside the measured area. The loadings are in agreement with some of the peaks in the undiluted "main" area (two rightmost peaks in Fig. 2 upper right), and could be tentatively assigned to compounds such as NADH or flavins.

The score matrices from the three PARAFAC models are "pooled" into one common score matrix. This matrix now contains all the quantitative information extracted from the fluorescence measurements. Thus we have reduced the complex spectra with several thousand variables to a matrix with 19 variables consisting of estimated relative concentrations of the underlying chemical compounds of the plasma samples. This matrix is now the input to a classification analysis. Note that absolutely no information about the health status of the patients has been used for building the PARAFAC models. This is important from a validation point of view, as it ensures that the matrix is simply an unbiased representation of the raw data.

3.1 Classification

The combined score matrix is used for building classification models. An initial exploratory PCA analysis of the score matrix explains 52% of the variation in the first three components and needs more than 12 components to explain 95% of the variation. The somewhat low explained variation is most likely due to the biological variation in the data and shows that the 19 PARAFAC scores are not overly redundant. No clear separation of cancer and control samples is found by the PCA analysis. There is thus no unsupervised direction in the variable space directly separating cancer from controls and hence the major part of the variation in the data is not related to the cancer/non cancer issue at all. Supplementary information such as age, gender, smoking habits, and co-morbidity could not explain further of this variation either. It is most likely just individual differences.

The score matrix with 19 variables was used as input to a PLS-DA classification model. During model building, some samples were removed as outliers based on evaluation of residuals and Hotellings T^2 (Jackson 1991). Classification models were built for all combinations of cancer and control and also control/control. Models are cross validated and the models are tested on a set of samples left out during model building. The huge biological variation from the raw data is still reflected in the extracted 19 variables in the score matrix. Therefore it makes sense to apply variable selection to select those variables of the 19 that reflect the variation relevant for discriminating cancer and non-cancer. We applied forward selection on the calibration data to find the optimal variables for classification. In the different models the number of variables was reduced from 19 variables to between five and 15 variables.

Results from the different models with sensitivity and specificity values for the cross validated and the tested models as well as area under the receiver operating characteristic (ROC) curve are seen in Table 1. A PLS-DA model with all the three control groups pooled to a common control versus the cancer patients gives an area under the ROC curve of 0.69 with optimal sensitivity and specificity values of 0.70 in the cross validated model, and similar values of 0.73 and 0.77 validated on new samples. Similar values are obtained on models with cancer vs. controls from the group of healthy individuals with no findings, and cancer vs. other non malignant findings. These models give areas under the ROC curves of 0.75 and 0.77, and sensitivity and specificity values between 0.73 and 0.80. In the models of cancer vs. adenomas, the area under the curve, sensitivity and specificity values are at the same level as the model with all controls. The results are to some extent surprising as one would expect it to be easier to discriminate between individuals with no findings and cancer, than between individuals with adenomas and cancer. Models of the different controls against each other give poor models with area under the curve values of 0.5–0.6. Even though they have different imbalances (adenomas or other non malignant findings), the controls are thus not much different from a fluorescence point of view. This result is important for future work of building better diagnosis models, as it underlines that the essential differences found in this study are related to cancer, non-cancer. In a different study on the same samples searching for differences in plasma levels of soluble urokinase

Table 1 PLS-DA models for classification of different classes based on the PARAFAC scores

Groups	Sensitivity CV	Specificity CV	AUC CV	Sensitivity predict	Specificity predict
Crc vs. no	0.68	0.84	0.75	0.73	0.77
Crc vs. onf	0.79	0.73	0.76	0.79	0.73
Crc vs. ade	0.73	0.74	0.77	0.92	0.63
Ade vs. no	0.57	0.55	0.50	0.45	0.43
Ade vs. onf	0.47	0.75	0.57	0.47	0.47
Onf vs. no	0.63	0.58	0.59	0.53	0.40
Crc vs. all controls	0.70	0.70	0.69	0.74	0.71

Crc cancer, *No* no findings, *Onf* other non malignant findings, *Ade* adenomas, *All* all three control groups, *CV* cross validated

plasminogen activator receptor (suPAR), the level of discrimination between cancer and other non malignant findings was better than between cancer and no findings. The discrimination between cancer and adenomas was less significant in this study (Lomholt et al. 2009).

The sensitivity and specificity values in Table 1 are found as the optimal value (maximizing the sum of the two). In diagnostic models, a high specificity value is often preferred as this reduces the number of false positives. For the models cancer vs. other non malignant findings and cancer vs. no findings we get sensitivity values of 0.48 and 0.43 at specificity values of 0.9. The result achieved by use of fluorescence spectroscopy and PARAFAC is thus comparable to the performance of the known biomarkers for CRC; CEA that has sensitivity and specificity values of 0.34 and 0.93.

The table above shows the results of the different classification models. The different models are based on different data, and thus use different variables for classification. A score and a loading plot for the classification model of cancer vs. other non malignant findings based on the PARAFAC scores are seen in Fig. 3. As expected from the sensitivity and specificity values, there is not a perfect separation between the two classes. However, there is a tendency towards separation along the diagonal from the second to forth quadrant in the score plot of the first vs. third PLS-DA component. From the loading plot we can see which variables are important for this separation. The loadings are likewise separated along a diagonal, with samples that are positively correlated to the "cancer direction" and samples negatively correlated to the "cancer direction" or positively correlated to the control samples; in this case the samples with other non malignant findings. A similar exercise can be done for all models.

Common for the models with cancer vs. one or all groups of controls is that the variables 1, 2, 8, 16 and 19 for several of the models are negatively correlated to the cancer direction, and likewise variables 6, 7 and 10 are positively correlated to the cancer direction. These variables are thus important in the discrimination between cancer and controls, though a model based on only those variables does not perform as well as models with more variables. The excitation and emission loadings from components seven and 10 which are positively correlated to cancer and likewise from components eight and 17 which are positively correlated to the controls are shown in Fig. 3 (lower plot). From the excitation and emission loadings these variables can most likely be assigned to tryptophan (variables 7 and 17) or tyrosine, with energy transfer to tryptophan (variables 1 and 4). They have pair wise similar excitation loadings, but the tryptophan emissions in the "cancer variables" are all shifted to shorter wavelengths (blue shift) compared to the "control variables". This confirms the findings from Leiner et al. (1986) who also experienced a blue shift in tryptophan emission in blood from cancer patients.

As opposed to what was expected, variable 3 (excitation/emission at 400/620), which corresponds to porphyrin, was not correlated to cancer. Several studies have shown elevated porphyrin levels in the blood from cancer patients (Madhuri et al. 2003; Masilamani et al. 2004; Xu et al. 1988). In this study all the subjects were included due to symptoms associated with CRC, and thus, even though three of four do not have cancer, some cellular biochemical imbalance might be expected, and therefore elevated levels could be expected in some of these controls. Additionally, the studies showing porphyrin to be important used acetone extracts of either blood plasma or cells, and not pure blood plasma as in the present study.

In the above models, PARAFAC scores were included from measurements on both diluted and undiluted samples, and as explained earlier there are some important effects of dilution. Fluorescence measurements on the undiluted samples may suffer from inner filter effect due to the high absorbance from the plasma samples. Diluting the samples induce physical/chemical changes in the plasma causing blue shift in the spectra. We found that variables from both the diluted and undiluted measurements were important for detecting cancer. Modelling only on scores from the diluted or undiluted samples gave similar but slightly worse

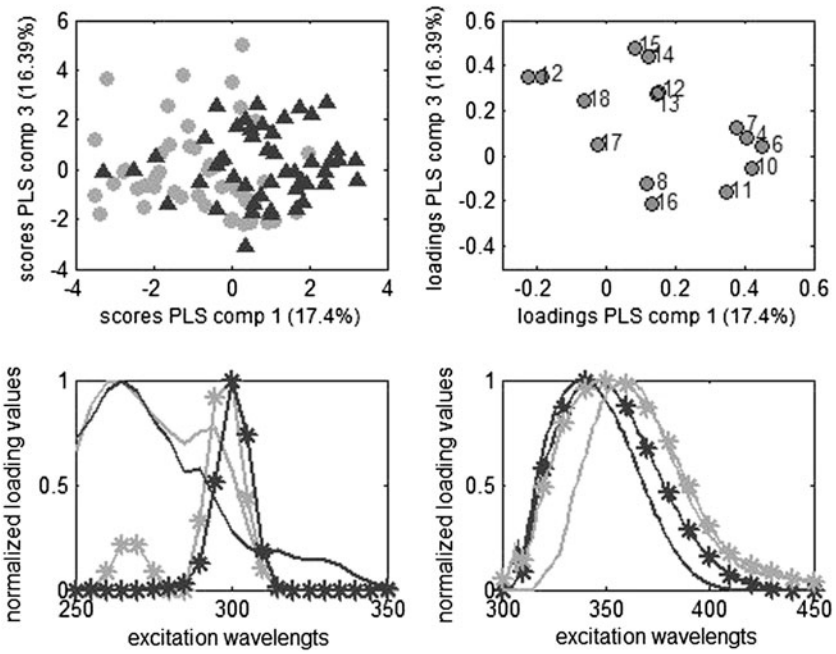

Fig. 3 *Upper left*: PLS-DA score plot of the first vs. third PLS-DA component from the model cancer vs. other non malignant findings on PARAFAC loadings. *Triangles* are cancers and *circles* are controls. *Upper right*: corresponding loading plot. *Lower*: selected PARAFAC excitation (*left*) and emission (*right*) loadings. *Dark gray line* (loading #7) and *dark grey* with *asterisk* (loading #10) are correlated with cancer, *light gray* (#8) and *light gray* with *asterisk* (#17) are correlated with control samples

models compared to the combination of scores from the diluted and undiluted samples, thus predictive power is gained by including both. From an analytical point of view, measuring only on the undiluted samples would be preferred as it makes the measurements faster and simpler to perform. Additionally there is a risk that the changes in sample matrix due to dilution could break some of the cancer specific correlations/interactions and thus make discrimination more difficult. A more thorough study addressing this could be interesting. In fact in analysis of the raw spectra (see below) better models were obtained using only the undiluted samples.

3.2 Classification on the raw data

A study similar to this on breast cancer by Nørgaard et al. (2007) applied discrimination only on the raw spectra. The authors did recommend applying more advanced techniques such as PARAFAC on the spectra but did not pursue this. Recall that we have used PARAFAC here, in order to provide more direct chemical information on how a possible classification can come about. Nevertheless, it is interesting to see whether we have gained anything from a quantitative point of view by applying PARAFAC on the data. Hence, classification models were built directly on the raw spectra as well. We have analyzed both diluted and undiluted samples individually and combined, and achieved similar results. However, the results from the undiluted measurements were slightly better than the alternative results, and are thus the only ones presented below. In Table 2 the results from the classifications based on the raw spectra are shown. Compared to the results based on the PARAFAC scores, these classification models perform equally well and these results are thus also comparable to the performance of CEA. Again the models on control vs. control perform worse than the cancer vs. control models. As for the models based on the PARAFAC scores we have applied variable selection on the models. Different variables are used for the models, but some of the same areas are represented in all four models.

Although it is possible to trace the original wavelengths behind the variables, these do not give the same intuitive information compared to the PARAFAC loadings. The scores and loadings for the model classifying cancer and other non malignant findings (Fig. 4) show a fairly good separation between the two groups in the first and fifth components. The loadings can be traced back to wavelengths around maxima for tryptophan, and the loading for the fifth component has a second derivative-like shape, which can be connected to the shift in the spectra from control to cancer that was shown above in the models based on PARAFAC scores. The results are thus similar, which was expected as it is originally the same data. Still, the extracted features by PARAFAC make the interpretation more straight forward and more comprehensive.

4 Conclusion

We have introduced excitation emission matrix fluorescence measurements on human blood plasma combined with multivariate data analysis as a potential alternative method to discriminate CRC patients from healthy controls, and controls with other cellular imbalances than cancer. With

Table 2 Results from the PLS-DA on the raw unfolded spectra

Groups	Sensitivity CV	Specificity CV	AUC CV	Sensitivity predict	Specificity predict
Crc vs. no	0.64	0.79	0.73	0.73	0.67
Crc vs. onf	0.73	0.79	0.75	0.73	0.73
Crc vs. ade	0.78	0.71	0.74	0.64	0.87
Ade vs. no	0.68	0.61	0.63	0.33	0.63
Ade vs. onf	0.84	0.34	0.55	0.70	0.33
Onf vs. no	0.45	0.82	0.62	0.20	0.82
Crc vs. all controls	0.69	0.7	0.73	0.67	0.83

Crc cancer, *No* no findings, *Onf* other non malignant findings, *Ade* adenomas, *All* all three control groups

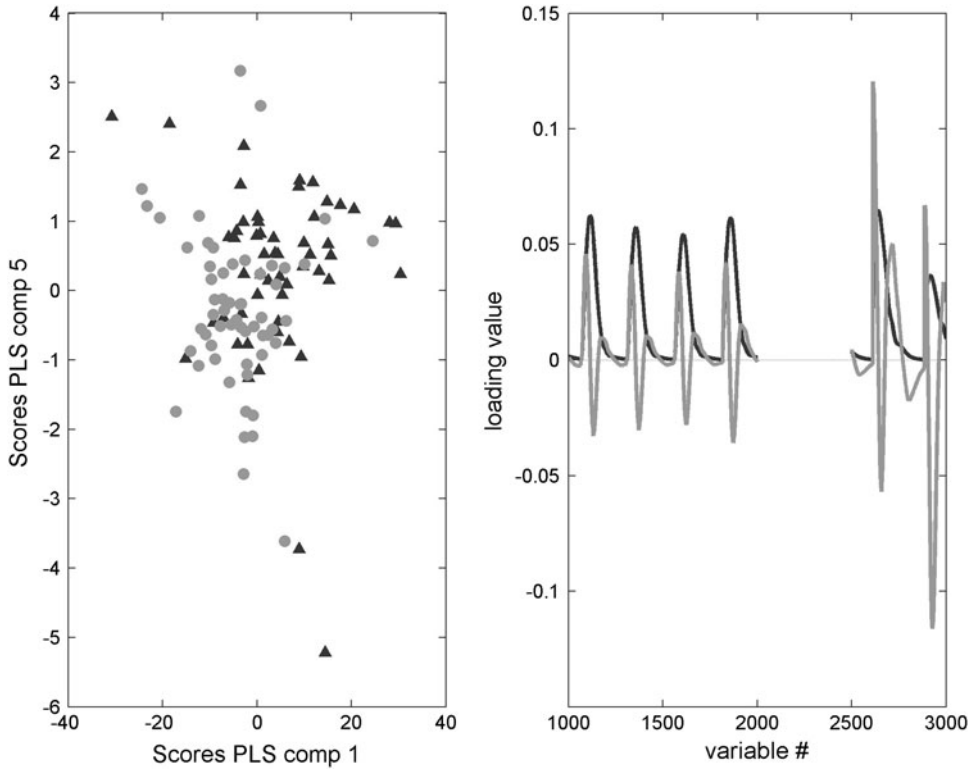

Fig. 4 *Left*: score plot of the first component vs. the fifth component for the PLS-DA model on cancer (*triangles*) vs. other non malignant findings (*circles*) on the raw spectra. *Right*: loadings from the first component (*dark gray*) and the fifth component (*light gray*)

sensitivity and specificity values of app 0.75 on a test set, the results are comparable to the known biomarker CEA. Previous studies with fluorescence spectroscopy have obtained similar results on other types of cancer but with a smaller number of samples. We obtained similar results in regards to discrimination whether we applied classification directly on the raw unfolded spectra or extracted estimates of the underlying fluorophores by use of PARAFAC. By the latter method, however, we obtained better conditions for a chemical interpretation/understanding of the results. We could see a blue shift in the tryptophan emission from cancer patients as one of the reasons for discrimination, a phenomenon described earlier in the literature. The use of PARAFAC on the fluorescence data to extract qualitative and quantitative chemical information from the human blood plasma samples, and base classification on this information is an example on how fluorescence spectroscopy can be used as a tool for metabonomic research. Compared to biomarker tests, fluorescence spectroscopy is an inexpensive alternative, and with minor sample preparation it is easy to perform the analysis. Further research is needed but we believe that there is room for fluorescence spectroscopy as metabonomic tool in cancer research.

Acknowledgments The VILLUM FOUNDATION is thanked for funding Anders Juul Lawaetz. Abdelrhani Mourhib is thanked for his laboratory assistance. Knud Nielsen, Randers Hospital, Søren Laurberg, Aarhus Hospital, Jesper Olsen, Glostrup Hospital and Hans B Rahr, Odense Hospital, are acknowledged for their contribution to the original protocol.

References

Abugo, O. O., Nair, R., & Lakowicz, J. R. (2000). Fluorescence properties of rhodamine 800 in whole blood and plasma. *Analytical Biochemistry, 279*, 142–150.

Andersen, C. M., & Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data 1. *Journal of Chemometrics, 17*, 200–215.

Bro, R. (1997). PARAFAC. Tutorial and applications 1. *Chemometrics and Intelligent Laboratory Systems, 38*, 149–171.

DeRose, P. C., & Resch-Genger, U. (2010). Recommendations for fluorescence instrument qualification: The new ASTM standard guide. *Analytical Chemistry, 82*, 2129–2133.

Flamini, E., Mercatali, L., Nanni, O., Calistri, D., Nunziatini, R., Zoli, W., et al. (2006). Free DNA and carcinoembryonic antigen serum levels: An important combination for diagnosis of colorectal cancer. *Clinical Cancer Research, 12*, 6985–6988.

Hamdan, M. H. (2007). *Cancer biomarkers*. Hoboken: John Wiley and sons.

Harshman, R. A., & DeSarbo, W. S. (1984). An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In H. G. Law, et al. (Eds.), *Research methods for multimode data analysis* (pp. 602–642). New York: Praeger.

Hubmann, M. R., Leiner, M. J. P., & Schaur, R. J. (1990). Ultraviolet fluorescence of human sera.1. Sources of characteristic differences in the ultraviolet fluorescence-spectra of sera from normal and cancer-bearing humans 1. *Clinical Chemistry, 36*, 1880–1883.

Jackson, J. E. (1991). *Operations with group data*. Hoboken: John Wiley & Sons, Inc.

Jenkinson, F., & Steele, R. J. C. (2010). Colorectal cancer screening—methodology. *Surgeon-Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland, 8*, 164–171.

Kalaivani, R., Masilamani, V., Sivaji, K., Elangovan, M., Selvaraj, V., Balamurugan, S. G., et al. (2008). Fluorescence spectra of blood components for breast cancer diagnosis. *Photomedicine and Laser Surgery, 26*, 251–256.

Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*. New York: Springer.

Lawaetz, A. J., & Stedmon, C. A. (2009). Fluorescence intensity calibration using the Raman scatter peak of water. *Applied Spectroscopy, 63*, 936–940.

Leiner, M. J., Schaur, R. J., Desoye, G., & Wolfbeis, O. S. (1986). Fluorescence topography in biology. III: Characteristic deviations of tryptophan fluorescence in sera of patients with gynecological tumors. *Clinical Chemistry, 32*, 1974–1978.

Leiner, M., Schaur, R. J., Wolfbeis, O. S., & Tillian, H. M. (1983). Fluorescence topography in biology. 2. Visible fluorescence topograms of rat sera and cluster-analysis of fluorescence parameters of sera of Yoshida ascites hepatoma-bearing rats. *IRCS Medical Science-Biochemistry, 11*, 841–842.

Lomholt, A. F., Hoyer-Hansen, G., Nielsen, H. J., & Christensen, I. J. (2009). Intact and cleaved forms of the urokinase receptor enhance discrimination of cancer from non-malignant conditions in patients presenting with symptoms related to colorectal cancer. *British Journal of Cancer, 101*, 992–997.

Madhuri, S., Aruna, P., Summiya Bibi, M. I., Gowri, V. S., Koteeswaran, D., Schaur, R. J., et al. (1997). Ultraviolet fluorescence spectroscopy of blood plasma in the discrimination of cancer from normal. *Proceedings of SPIE, 2982*, 41–45.

Madhuri, S., Suchitra, S., Aruna, P., Srinivasan, T. G., & Ganesan, S. (1999). Native fluorescence characteristics of blood plasma of normal and liver diseased subjects. *Medical Science Research, 27*, 635–639.

Madhuri, S., Vengadesan, N., Aruna, P., Koteeswaran, D., Venkatesan, P., & Ganesan, S. (2003). Native fluorescence spectroscopy of blood plasma in the characterization of oral malignancy. *Photochemistry and Photobiology, 78*, 197–204.

Masilamani, V., Al-Zhrani, K., Al-Salhi, M., Al-Diab, A., & Al-Ageily, M. (2004). Cancer diagnosis by autofluorescence of blood components. *Journal of Luminescence, 109*, 143–154.

McKnight, D. M., Boyer, E. W., Westerhoff, P. K., Doran, P. T., Kulbe, T., & Andersen, D. T. (2001). Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography, 46*, 38–48.

Nielsen, H. J., Brunner, N., Frederiksen, C., Lomholt, A. F., King, D., Jorgensen, L. N., et al. (2008). Plasma tissue inhibitor of metalloproteinases-1 (TIMP-1): a novel biological marker in the detection of primary colorectal cancer. Protocol outlines of the Danish-Australian endoscopy study group on colorectal cancer detection. *Scandinavian Journal of Gastroenterology, 43*, 242–248.

Nordström, A., & Lewensohn, R. (2010). Metabolomics: Moving to the Clinic. *Journal of Neuroimmune Pharmacology, 5*, 4–17.

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy, 54*, 413–419.

Nørgaard, L., Soletormos, G., Harrit, N., Albrechtsen, M., Olsen, O., Nielsen, D., et al. (2007). Fluorescence spectroscopy and chemometrics for classification of breast cancer samples—a feasibility study using extended canonical variates analysis. *Journal of Chemometrics, 21*, 451–458.

Ragazzi, E., Pucciarelli, S., Seraglia, R., Molin, L., Agostini, M., Lise, M., et al. (2006). Multivariate analysis approach to the plasma protein profile of patients with advanced colorectal cancer. *Journal of Mass Spectrometry, 41*, 1546–1553.

The Danish Cancer Society. (2010). http://www.cancer.dk.

The Danish National Board of Health. (2010). *National screening for tyk- og endetarmskræft*. The Danish National Board of Health, 1.

Uppal, A., Ghosh, N., Datta, A., & Gupta, P. K. (2005). Fluorimetric estimation of the concentration of NADH from human blood samples 1. *Biotechnology and Applied Biochemistry, 41*, 43–47.

Vivian, J. T., & Callis, P. R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal, 80*, 2093–2109.

Ward, D. G., Suggett, N., Cheng, Y., Wei, W., Johnson, H., Billingham, L. J., et al. (2006). Identification of serum biomarkers for colon cancer by proteomic analysis. *British Journal of Cancer, 94*, 1898–1905.

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*, 109–130.

Wolfbeis, O. S., & Leiner, M. (1985). Mapping of the total fluorescence of human-blood serum as a new method for its characterization. *Analytica Chimica Acta, 167*, 203–215.

Xu, X. R., Meng, J. W., Hou, S. G., Ma, H. P., & Wang, D. S. (1988). The characteristic fluorescence of the serum of cancer-patients. *Journal of Luminescence, 40–1*, 219–220.

Zhang, Xuewu, Li, Lin, Wei, Dong, Yap, Yeeleng, & Chen, Feng. (2007). Moving cancer diagnostics from bench to bedside. *Trends in Biotechnology, 25*, 166–173.

Zhao, Z., Xiao, Y., Elson, P., Tan, H., Plummer, S. J., Berk, M., et al. (2007). Plasma Lysophosphatidylcholine Levels: Potential Biomarkers for Colorectal Cancer. *Journal of Clinical Oncology, 25*, 2696–2701.