

Analyst

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

PAPER

Discrimination and Classification of Liver Cancer Cells and Proliferation States by Raman Spectroscopic Imaging

T. Tolstik^{a,*,#}, C. Marquardt^{a,#}, C. Matthäus^b, N. Bergner^b, C. Bielecki^a, C. Krafft^b, A. Stallmach^a, J. Popp^{b,c}

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

Discrimination of nodular lesions in cirrhotic liver is a challenge in the histopathologic diagnostics. For this reason, there is an urgent need for new detection methods to improve the accuracy of the diagnosis of liver cancer. Raman imaging allows to determine the spatial distribution of a variety of molecules in cells or tissue label-free and to correlate this molecular information with the morphological structures at the same sample location. This study reports investigations of two liver cancer cell lines, - HepG2 and SK-Hep1, - as well as HepG2 cells in different cellular growth phases using Raman micro-spectroscopic imaging. Spectral data of all cells were recorded as a color-coded image and subsequently analyzed by hierarchical cluster and principal component analysis. A support vector machine-based classification algorithm reliably predicts previously unknown cancer cells and cell cycle phases. By including selectively the Raman spectra of the cytoplasmic lipids in the classifier, the accuracy has been improved. The main spectral differences that were found in the comparative analysis can be attributed to a higher expression of unsaturated fatty acids in the hepatocellular carcinoma cells and during the proliferation phase. This corresponds to the already examined de novo lipogenesis in cells of liver cancer.

1. Introduction

A range of optical techniques such as infrared, fluorescence, Raman and CARS spectroscopy have been widely used in recent years and proven to be reliable methods for cancer diagnostics.¹⁻⁹ Raman microscopic imaging is an innovative technology, which combines the spatial resolution of microscopic imaging with the highly specific spectroscopic information of classical Raman spectroscopy. It allows to excite the vibrational and rotational states of molecules in biological samples in a label-free manner at submicron resolution. Hence, the biochemical composition of the sample can be examined at the subcellular level. Raman spectra of a biological sample are complex and result in a biochemical fingerprint, containing information about the chemical structure of proteins, nucleic acids, lipids and carbohydrates. Changes in these chemical structures in cells detected by Raman imaging can therefore help to differentiate and classify malignant tumors. For the evaluation of the spectral information, different multivariate unsupervised (e.g. clustering) and supervised (e.g. classification) methods have been employed.^{1,10-32} Raman imaging has already successfully been applied for the identification of cell lines from the same and different origins,^{1,10-17,28-31} stem cells,^{18,19,32} the proliferative status of cells²⁰ and preliminary stages of cell apoptosis.²¹⁻²³

The histopathological evaluation of biopsies is a crucial step to distinguish unclear nodular lesions in cirrhotic liver. With an occurrence rate of more than 85%, the hepatocellular carcinoma (HCC) is the most prevalent malignant primary liver tumor worldwide.³³ This second most lethal cancer with a 5-year

survival rate of 8.9%³⁴ occurs with great frequency in Asia and Africa and is increasingly found in Europe. By an earlier diagnosis and an appropriate treatment the survival rate of patients with this malignant tumor can be improved. Despite a number of common staining techniques, morphological features have not proven to be clinically significant for HCC diagnosis³⁵. Therefore, it is often difficult for the pathologist to diagnose the underlying disease just by cytohistologic criteria according to the International Consensus Group for Hepatocellular Neoplasia. Consequently, new detection methods are needed to reinforce the diagnosis of HCC.

In order to support the diagnosis of liver cancer, the aim of this research study was to characterize and differentiate two types of liver cancer cells by Raman spectroscopy and to prove its reproducibility. As an in vitro model system the HepG2 cell line (human cell line that was derived from liver tissue of a patient with well differentiated hepatocellular carcinoma) and the SK-Hep1 cell line (received from ascitic fluid of a patient with adenocarcinoma of the liver) were chosen for investigations by Raman imaging. Furthermore, we determined the Raman spectra of hepatocellular carcinoma cells (HepG2) in the exponential and plateau phase of cell growth. These two approaches were performed to detect Raman signatures and to generate classification models for two different types of liver cancer cells and also metabolic changes in liver cancer cells during a high cell division rate. The cell type and the proliferation behaviour are essential distinguishing features of malignant tumors. The results suggest that Raman spectroscopy is a particularly suitable method to easily identify molecular changes in liver cancer.

2. Materials and methods

2.1 Cell cultivation

HepG2 and SK-Hep1 cells were cultivated in RPMI 1640 liquid medium with 20 mM HEPES, stable glutamine (FG 1235; Biochrom AG, Germany), 10% fetal bovine serum (DE14-801F; Lonza, Belgium) and 1% penicillin-streptomycin (15140; Gibco®, Life Technologies GmbH, Germany). Both cell types were maintained in a 5% CO₂ incubator at 37°C in cell culture flasks (658170; Greiner Bio-One GmbH, Germany). Every two days the medium was changed until approximately 50% confluence was reached. After that the cells were detached with Trypsin/EDTA solution (L2143; Biochrom AG, Germany) and transferred onto CaF₂ slides at a concentration of 60000 cells/ml. The Raman measurements were performed on CaF₂ slides in order to avoid autofluorescence background from regular glass slides.

2.2 Cell preparation of HepG2 and SK-Hep1

HepG2 and SK-Hep1 cells were incubated for 24 hours, washed twice with phosphate buffered saline (PBS) (0689; Lonza, Belgium) and fixed with 10% neutral buffered formalin solution (HT501128; Sigma-Aldrich, USA) for 20 minutes. One more washing step was performed after the fixation process. Finally, the slides were stored in PBS buffer at +4°C until the Raman measurements were done.

2.3 Preparation of exponential and plateau growth phase of HepG2 cells

To measure HepG2 cells with Raman spectroscopy in the exponential phase, the cells were cultivated in fresh medium for 24h on CaF₂ slides. In order to achieve maximum cell proliferation, this cultivation process was performed until a monolayer culture with a confluence of 50% was reached. For the Raman measurement of HepG2 cells in the plateau phase, the monolayer culture was cultivated for 96h until 100% confluence was obtained. Throughout the period of cell cultivation the medium was not exchanged, which leads to an inadequate nutrition of the cells and thus to a growth arrest (plateau phase). Afterwards, the samples were fixed with neutral buffered formalin solution for 20 minutes. Subsequently a washing step was performed and the samples were stored in PBS buffer at +4°C.

2.4 Raman image acquisition

Raman spectra were acquired using a confocal Raman microscope (WITec, Ulm, Germany, Model CRM 2000) equipped with an air-cooled argon ion laser (LASOS Lasertechnik GmbH, Jena, Germany) with a laser power of ca. 5 - 10 mW at the sample (grating 600 g/mm, BLZ=500 nm). An excitation wavelength of 488 nm was chosen for our measurements because Raman scattering intensity is directly proportional to λ^{-4} (λ = laser wavelength), and therefore a blue laser with a shorter wavelength of 488 nm results in an increase in scattering intensity in compare with the longer wavelength. The raw Raman spectra acquired with 488 nm laser (Fig. 1) showed high signal to noise ratio with an acceptable fluorescent signal. The laser was coupled to a microscope by a single mode optical fiber. The Raman scattered light was detected by a back-

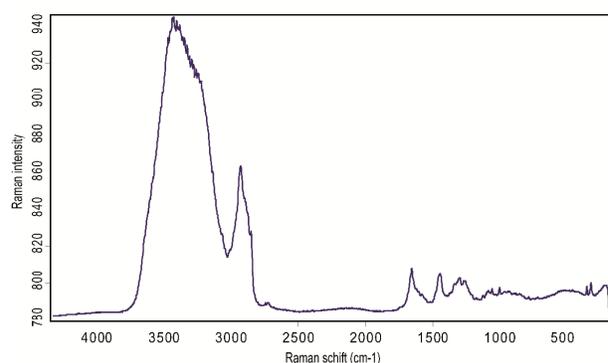


Fig. 1 Raw Raman spectra in the spectral region 4200-200 cm⁻¹ of one HepG2 cells measured in water before any preprocessing steps.

illuminated deep-depletion CCD camera operating at -65°C. The system was pre-calibrated to the 520.7 cm⁻¹ spectral line of silicon and the laser light was focused on the sample with a 60x/NA=1.0 water immersion objective (Nikon NIR Apo, Tokyo, Japan). All samples were prepared in four repeats and the Raman measurements were performed in PBS solution. Raman spectral images of the cells were detected in the spectral region of 4100-200 cm⁻¹ with a step size of 0.5 μ m and an integration time for each spectrum of 1 second. Subsequently after the measurement the cells were stained with haematoxylin and eosin (HE) for morphological confirmation using a classical HE-staining protocol.

2.5 Data analysis

For the statistical analysis the complete spectral data set of all cells, which were recorded by the WITec software (WITec GmbH, Germany), were imported and calculated with the software packages CytoSpec (CytoSpec Inc., USA), OPUS (Bruker Optik GmbH, Germany) and MatLab (MathWorks Inc., USA) including the PLS_Toolbox (Eigenvector Research Inc., USA).

For preprocessing, baseline correction and vector normalization in the spectral region of 3100-600 cm⁻¹ were chosen³⁶, which have been demonstrated to be appropriate and are routinely used for Raman data analysis.^{12,37,38} A hierarchical cluster analysis (HCA), which was calculated with a Euclidean distance measure, clustered with the Ward's algorithm and filtered with the Generalized Least Squares Weighting, was applied to the complete data set of the cell lines, in order to split and identify new groups with a high level of similarity in the data (inter-individual cell variability). The relations between the data were illustrated in a dendrogram. For additional visualization of the data set, as a Raman image, the CH stretching intensities of the detected spectra at each measuring point were plotted. These color-coded Raman intensity maps were correlated with the underlying cytological compartments (cell organelles) and the differences in the data were displayed using the HCA. The HCA is grouping similar spectral information into different clusters, showing the spatial distribution of these clustered spectra corresponding to components of the cell (e.g. cell nucleus, cytoplasm, lipid storage organelles) as a Raman image. All spectra of each cell were grouped into five clusters including the

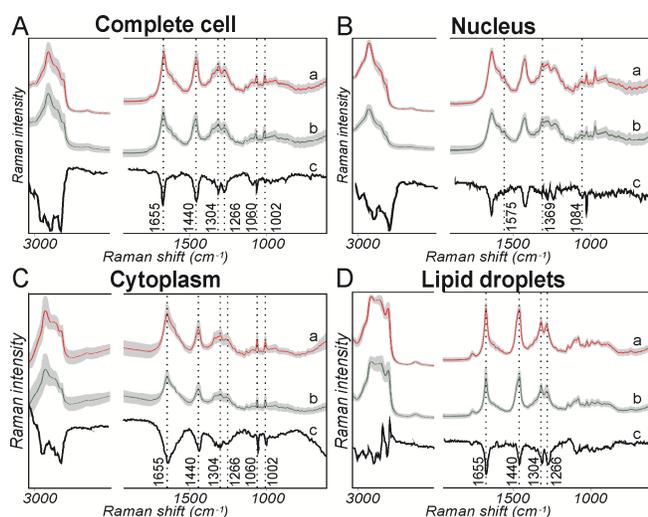


Fig. 2 Average spectra in the spectral region 3100-2800 cm^{-1} (x3) and 1800-600 cm^{-1} (x5) of (a) HepG2 cells, (b) SK-Hep1 cells, and (c) the corresponding difference spectra for (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets. Distinguishing Raman bands (in cm^{-1}) are indicated by dashed lines. Shaded areas of the spectra a and b represents the standard deviations.

spectra of nucleus, two different components of cytoplasm, cytoplasmic lipids and background spectra (intra-individual cell variability). The surrounding cluster of the cell background was removed from the data set. An average spectrum of each cell cluster and a difference spectrum were calculated with the OPUS software. Finally, the averaged data were imported into MatLab.

Principal component analysis (PCA) was applied to normalized and mean-centred Raman spectra. More than 99% of the data variance was represented by the first 20 principal components (PCs) and only noise was found in the higher PCs. In order to recognize a predictive pattern in the spectra, the first 20 PCs of a training data set were used to develop a supervised learning model using the support vector machine (SVM) algorithm, which is capable to classify future samples of an unknown validation set. Besides other classification algorithms, the SVM have already been successfully applied in various Raman imaging studies^{12,25,28,38}. To estimate the performance of this predictive classification model a 10-fold cross-validation was performed.

3. Results

3.1 Unsupervised analysis and classification of HepG2 and SK-Hep1 cells

By analysis of the Raman image data sets, acquired from 100 cells of HepG2 and SK-Hep1 cell lines, spectral differences based on the cytology have been identified. Images of the HE-staining and of the Raman measurement of the very same cell in combination with the corresponding average spectra for each cell line in the spectral region of 3100-500 cm^{-1} are presented in Fig. 2. The Raman spectra show the typical characteristics of cells, such as CH stretching intensities (3020-2800 cm^{-1}), amide I band (1680-1620 cm^{-1}), CH_2 deformation band (1440 cm^{-1}) and the phenylalanine band (1002 cm^{-1}). By subtraction of the

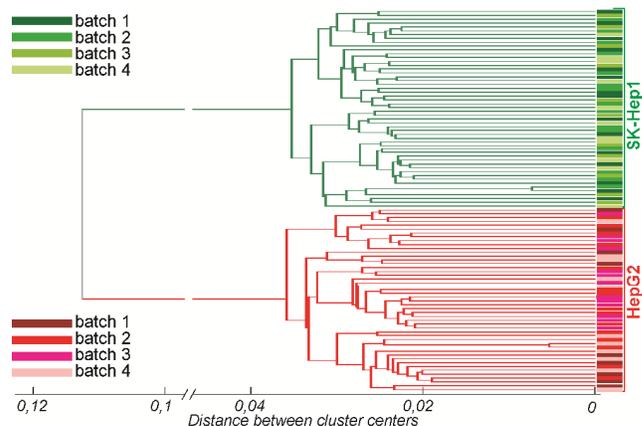


Fig. 3 Dendrogram from HCA of Raman spectra showing the clustering of HepG2 ($n=48$) and SK-Hep1 ($n=52$) cells with assigned sub-clusters. The batches of the cell lines are marked in different colors.

average spectra of the HepG2 and SK Hep1 cells, the spectral variations, shown as a difference spectrum, were obtained (Fig. 2(A-B)c). The difference spectrum exhibits intensities in the range of 2900-2850 cm^{-1} and at 1655, 1440, 1304, 1266, 1060 cm^{-1} .

3.1.1 Hierarchical cluster analysis of inter- and intra-individual cell variabilities. For the detection of the inter-individual variability of the two different cell lines a HCA of all HepG2 ($n=48$) and SK-Hep1 ($n=52$) cells was performed. This way, a clear separation of HepG2 and SK-Hep1 cells was achieved based on the cell-type specific Raman pattern (100% accuracy). The plotted dendrogram in Fig. 3 displays the distinct subdivision according to the cell line.

For the detection of the intra-individual variability each single cell were subdivided by HCA in five clusters corresponding to the different cell compartments, which were confirmed by HE stained microscope images and Raman maps of CH-stretching region (Fig. 4A, B and C). In accordance with the cell nucleus the average Raman spectra showed the typical bands of the nucleic acids at 1575, 1369, 1084 and 782 cm^{-1} , whereas in the average spectra of the cytoplasm spectral pattern of proteins and lipids without bands of nucleic acids were observed (Fig. 4D and E). The data regions of the lipid rich cell components, confirmed by the typical Raman spectra of lipids, demonstrated intensive bands at 3008, 2898, 2854, 1655, 1440, 1304, 1266, 1078, 1059, 968 and 717 cm^{-1} (Fig. 4F). The obtained clusters were used to extract the spectral data of each single cell compartment and to implement them in the SVM-based classification algorithm described below.

3.1.2 Principal component analysis. Using PCA the wide variety of statistical variables of the extensive spectral data set was converted into PCs. The bands at 2900, 2898, 2854, 1660, 1655, 1440, 1304, 1267, 1266, 1060 cm^{-1} were identified in the calculated PCs (Fig. 5). These bands resemble those bands of the difference spectrum (Fig. 2B) and are also similar to the bands of pure lipids spectra (Fig. 4F). Almost the half (48,76%) of the variance in this data set was corresponded to PC1 and can be assigned to lipids (2848, 2854, 1655, 1440, 1304, 1266, and 1060 cm^{-1}).

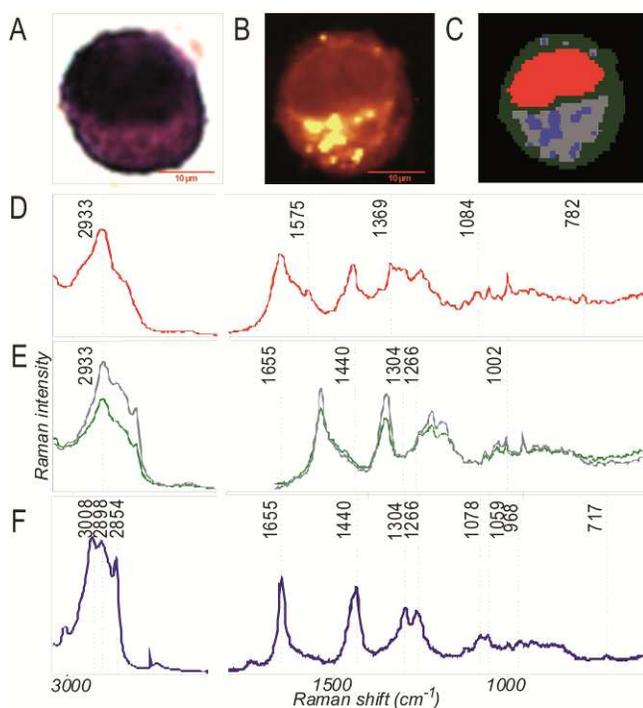


Fig. 4 (A) HE stained microscopy image, (B) Raman image of the CH-stretching region. (C) HCA map with five clusters of the same HepG2 cell, including the cluster for nucleus (red), cytoplasm (grey and green) and lipid droplets (blue). (D-F) Corresponding average spectra to the cluster of (D) nucleus, (E) cytoplasm, and (F) lipid droplets in the spectral region 3100-2800 cm^{-1} (x1) and 1800-600 cm^{-1} (x5).

3.1.3 Support vector machine classification model. A SVM algorithm was trained on the averaged spectra of 60 cells of HepG2 ($n=30$) and SK-Hep1 ($n=30$) to define a Raman pattern that predicts the cell type. For the recognition algorithm the first 20 PCs were used of which the first four PCs captured the main

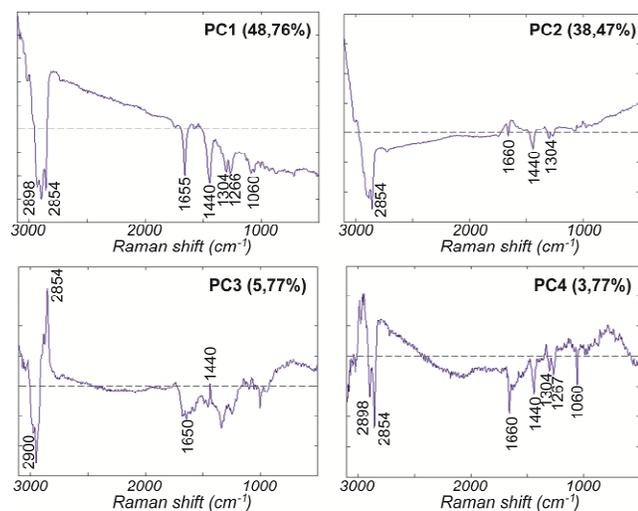


Fig. 5 (A) Loadings for the first fourth principal components of the PCA model of HepG2 and SK-Hep1 cell lines in a spectral region 3100-600 cm^{-1} .

A Complete cell					B Nucleus						
Training dataset	H	S	Test dataset	H	S	Training dataset	H	S	Test dataset	H	S
HepG2 (H)	97%	21%	H	97%	11%	HepG2 (H)	93%	21%	H	96%	5%
SK-Hep1 (S)	3%	79%	S	3%	89%	SK-Hep1 (S)	7%	79%	S	4%	85%

C Cytoplasm					D Lipid droplets						
Training dataset	H	S	Test dataset	H	S	Training dataset	H	S	Test dataset	H	S
HepG2 (H)	89%	8%	H	84%	11%	HepG2 (H)	95%	17%	H	93%	0%
SK-Hep1 (S)	11%	92%	S	16%	89%	SK-Hep1 (S)	5%	83%	S	7%	100%

Fig. 6 Confusion tables of the SVM classification models of HepG2 (H) and SK-Hep1 (S) cells. Misclassifications (in %) of (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets in training and test dataset are shown.

variance (accounting for 96.8%). The PCs higher than 20, which mainly contained noise and had less than 1% of the total variance, were not included in the classification algorithm. A performed 10-fold cross-validation, which was calculated by using subsets of the trainings set, estimates the predictive performance of this classification algorithm with 89%. The cross-validation classified not more than one HepG2 cell incorrectly as SK-Hep1 cell and a maximum of seven SK-Hep1 cells were misclassified as HepG2. By the following obtained classification model previously unknown cells of the two cell lines from a validation set (HepG2: $n=18$; SK-Hep1: $n=22$) were identified correctly with a prediction accuracy of 93% (sensitivity of 97%, specificity of 89%).

In addition to the classification model of complete cells, we elaborated also the difference spectra and classification algorithm by using the spectral information of different cell compartments. For this purpose, a discriminant analysis was performed and the data were imported in the SVM-algorithm. Consequently, for

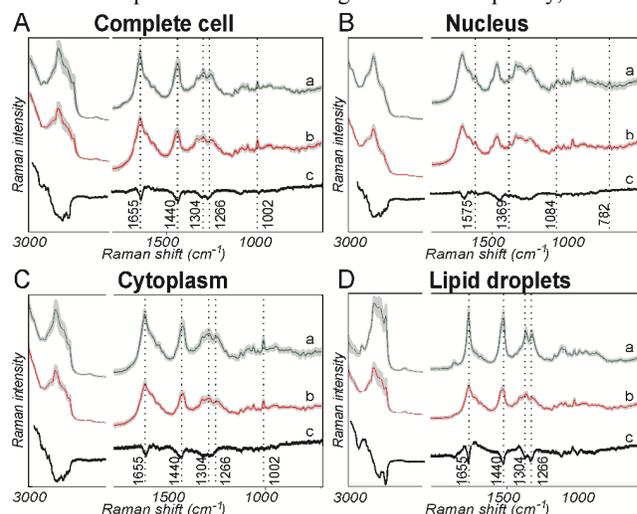


Fig. 7 Average Raman spectra of (a) proliferating and (b) non-proliferating HepG2 cells, and (c) the corresponding difference spectra of (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets. Shaded areas of the spectra a and b represents the standard deviations.

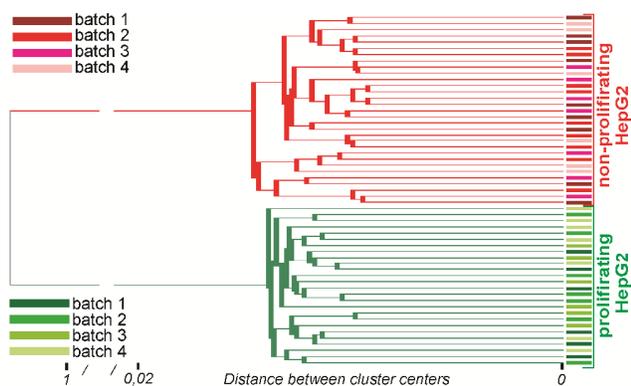


Fig. 8 Hierarchical clustering of proliferating HepG2 cells in the exponential phase ($n=26$) and the non-proliferating HepG2 cells in the plateau phase ($n=30$) demonstrated as a dendrogram with assigned sub-clusters. The batches of the cells are marked in different colors.

each cell compartment, such as the cell nucleus, cytoplasm and the lipid storage organelles, separate classification models were generated. By using both specific data sets as two independent classifiers an accuracy of 91% and 87% for the data of cellnucleus and cytoplasm were obtained, respectively. The spectral information of lipid droplets, confirmed by the typical Raman spectra of lipids, were isolated by HCA and also used to classify HepG2 and SK-Hep1. This classifier showed a very high sensitivity of 93%, specificity of 100% and accuracy of 96%. The details of the prediction for individual classification model are shown in the confusion table of Fig. 6.

3.1.4 Test-retest reliability of the sample measurements. To verify the test-retest reliability of the generated classification algorithm two different data sets of the HepG2 and SK-Hep1 cell lines were acquired under the same conditions, but at different points in time. Data sets were collected from Raman measurements of three independent samples of HepG2 and SK-Hep1 cells from different batches and with a time lag of 3 months. The second data set was used to train the classification model. Therefore, the prediction accuracy of the classification model was first estimated by performing a 10-fold cross-validation of the 100 cells. The cross-validation calculated a sensitivity of 94% and a specificity of 87% for the predictive accuracy. However, the trained classification model based on the first data set applied to the independent second data set resulted then in a sensitivity of 82% and a specificity of 75%.

3.2 Unsupervised analysis and classification of different proliferation performances of HepG2 cells

The difference spectrum of the proliferating (exponential phase) and non-proliferating (plateau phase) HepG2 cells indicated higher intensities within the spectral region between 3020-2800 cm^{-1} and at 1655, 1440, 1304, 1266 and 1002 cm^{-1} of the proliferating cells (Fig. 7A).

3.2.1 Hierarchical cluster analysis. To detect also the inter-individual variability in case of HepG2 cells in the exponential ($n=26$) and plateau ($n=31$) growth stages HCA was performed and plotted as a dendrogram. The tree structure in the

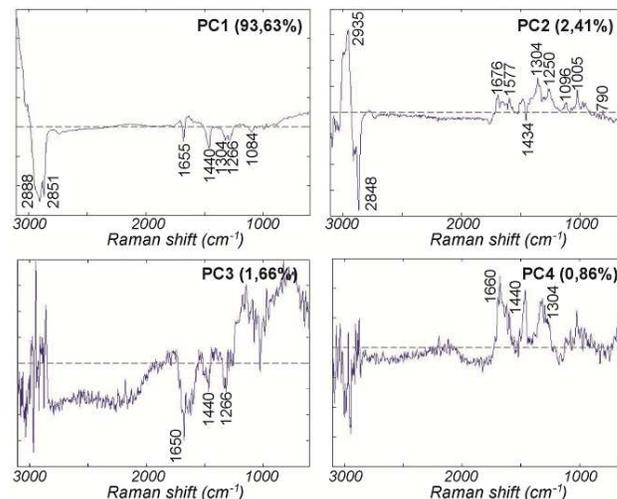


Fig. 9 First four PC from PCA model performed on proliferating (exponential phase) and non-proliferating (plateau phase) HepG2 cells.

dendrogram (Fig. 8) shows a perfect discrimination of the proliferating and non-proliferating HepG2 cells in two separated clusters. Thereby, all HepG2 cells were clustered according to their growth phase and not one cell were misclassified (100% accuracy).

Each single HepG2 cell in different proliferating condition was as well clustered in intra-individual components of the cell. The extracted spectral information of the correspondent nucleus, cytoplasm and lipid regions was used for classification (Fig. 7).

3.2.2 Principal component analysis. Analyzing PCs of this data set the bands 2888, 2851, 1655, 1440, 1304, 1266, and 1084 cm^{-1} seen in PC1 includes more than 93% of variance and are equivalent to lipid spectra (Fig. 9A). The positive bands that were detected in PC2 (Fig. 9B) corresponding to DNA (1577, 1096 and 790 cm^{-1}) and proteins (2935, 1676, 1340, 1250 and 1005 cm^{-1}), while the negative bands can be assigned to lipids (2848 and 1434 cm^{-1}).

3.2.3 Support vector machine classification model. In order to classify the HepG2 cells in different cellular growth stage, a training and validation data set of cells in the exponential and plateau phases were calculated by the SVM-based classification model. The performance of the classification model to predict the unknown proliferation phase of HepG2 cells exhibited a sensitivity of 100% and specificity of 98%. As mentioned above for the classification of HepG2 and SK-Hep1, spectra of proliferating and non-proliferating HepG2 cell compartments were classified separately. Over all accuracies of the classification models the classifier based on the Raman spectra of lipid droplets organelles showed the best result. The details of misclassified proliferating and non-proliferating HepG2 cells are demonstrated in the confusion table of Fig. 10.

Discussion

To prove the concept that Raman imaging technique is suitable for the detection of malignant cells in liver lesions the morphological and proliferating features of cancer cells were investigated and classified by this vibrational spectroscopic

A Complete cell						B Nucleus					
Training dataset	P	NP	Test dataset	P	NP	Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	91%	8%	P	100%	2%	Prolifer. (P)	97%	15%	P	90%	3%
Nonprolif. (NP)	9%	92%	NP	0%	98%	Nonprolif. (NP)	3%	85%	NP	10%	87%

C Cytoplasm						D Lipid droplets					
Training dataset	P	NP	Test dataset	P	NP	Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	99%	1%	P	100%	1%	Prolifer. (P)	100%	1%	P	100%	0%
Nonprolif. (NP)	1%	99%	NP	0%	99%	Nonprolif. (NP)	0%	99%	NP	0%	100%

Fig. 10 Confusion tables of the SVM classification models of proliferating (P) and non-proliferating (NP) HepG2 cells. Misclassifications (in %) of (A) complete cell, (B) nucleus, (C) cytoplasm, and (D) lipid droplets in training and test dataset are shown.

method.

In previous studies it has already been demonstrated that it is possible to detect differences in normal and malignant hepatocytes based on single point Raman spectra^{13,38}. Therefore, the focus of this study was set on the differentiation and classification of different tumor cells. In contrast to previous these previous studies, the differentiation was based on spectral information collected from whole cell images, rather than from single point spectra with a sampling volume of about $1 \mu\text{m}^3$. A liver cancer cell line of the hepatocellular carcinoma (HepG2), as the most common primary liver cancer, and an adenocarcinoma (SK-Hep1), that is found as primary liver cancer in the liver, were measured with Raman spectroscopy. The obtained Raman spectra were processed with a multivariate analysis method and implemented into a reliable prediction model. Thus, to the best of our knowledge this is the first study that is using vibrational spectroscopy as a classifying tool for different types of liver cancer cells and their proliferation states.

Like a molecular fingerprint the Raman spectra exhibit spectral differences of cells and to some extent molecular contents of various cellular components, such as nucleic acids, proteins, carbohydrates and lipids. While the specificity of Raman spectroscopy is very high, the signal intensity and thus the level of sensitivity of this technique is quite low, which can cause a reduced detection of molecules with low concentration. Other limitations of this technique can arise from an unwanted background signal originated from the fluorescence of the sample or contamination.

In our investigations the difference spectrum of the observed averaged spectra and the multivariate discriminant analysis of HepG2 and SK-Hep1 data sets were able to extract the subtle spectral differences between these two cell lines with pronounced intensities in the C-H stretching region of $2900\text{--}2850 \text{ cm}^{-1}$ and a characteristic shape in the fingerprint region between $1800\text{--}600 \text{ cm}^{-1}$. The results represent significant spectral differences in the peak intensities at $2900\text{--}2850$, 1655 , 1440 , 1304 , 1266 and 1060 cm^{-1} . This recognition pattern applied in a multivariate -

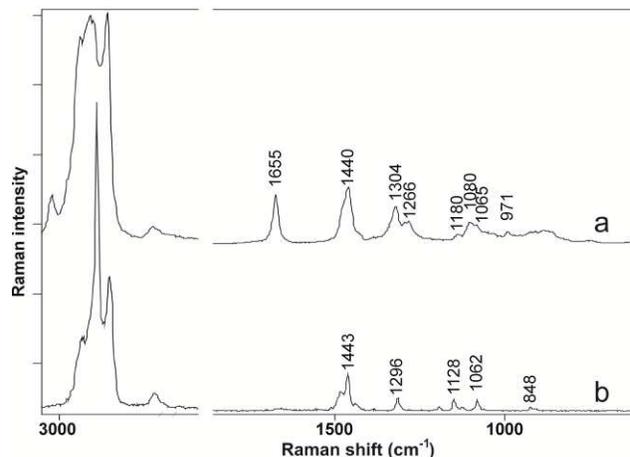


Fig. 11 Reference Raman spectra of (a) unsaturated fatty acid (oleic acid) and (b) saturated fatty acid (stearic acid) acquired with WITec Raman spectrometer.

HCA allows already a precise separation of the HepG2 and SK-Hep1 cell line in two major clusters without any misclassification. Furthermore, by implementing the same spectral pattern into the above-mentioned classification algorithm the previously unknown cell identity can be predicted with an accuracy of 93%. In the discriminant analysis the average spectra of both cell lines differ significantly in the specific spectral characteristics of lipids. Due to this fact, the prediction accuracy of the classification model was improved to 96% by the sole use of the spectral pattern of the cytoplasmatic lipids. Thus, it was possible to show that a majority of spectral information, which characterizes and distinguishes different liver cancer cells, is located in lipids. These detected spectral differences can be assigned exactly to the specific wavenumbers of unsaturated fatty acids and were verified by a contrasting juxtaposition of the averaged spectra according to the lipid clusters to the reference spectra of unsaturated fatty acids (Fig. 11). Hence, the quantity of unsaturated acids is increased in the HepG2 cell line.

By inspecting liver cancer cells with Raman microscopic imaging, the results confirm that a lot of molecular differences are hidden in lipids. Nowadays, it is believed that aberrant lipid biosynthesis is involved in the hepatocarcinogenesis. Already in 2005, the mRNA expression of lipogenic enzymes, which are involved in the production of fatty acids, was investigated in 10 human HCC tissue samples and were compared with the surrounding non-cancerous liver tissue. An elevated expression of mRNA for fatty acid synthase (FASN), acetyl-CoA carboxylase (ACAC) and ATP citrate lyase (ACLY) in the HCC tissue was described³⁹. In a recent large-scale study the expressions of enzymes that regulate lipogenesis were evaluated in HCC tissues and corresponding surrounding non-tumor liver tissues of 68 patients as well as in eight normal (disease-free) human liver tissue of healthy individuals. It was shown by this study that all relevant lipogenic enzymes, which are involved in the fatty acid synthesis (FASN, ACAC, ACLY, ME, SCD1), the cholesterol biosynthesis (SREBP2, HMGCR, MVK, SQS) and their upstream inductors (chREBP, SREBP1, LXR- β) were progressively upregulated from non-tumorous liver tissue toward the HCC. Analogously to the upregulation of the lipogenic enzymes and

their upstream inductors also the quantity of the chemical end products of the respective lipid synthesis (fatty acids, triglycerides, cholesterol) increased. Furthermore, the study demonstrated that a higher expression of these enzymes correlate with an activation of a lipogenic signaling pathway (AKT–mTORC1–RPS6 pathway) and the clinical aggressiveness, characterized by shorter (<3 years) and longer (>3 years) survival of the patients following partial liver resection.⁴⁰ In another study of three HCC cell lines, it has been observed that the FASN protein expression and basal activity levels were higher in HepG2 cells than in the other two HCC cell lines. The inhibition of the FASN activity with C75 induced a cell cycle arrest in all three HCC cell lines.⁴¹

As set out in the study of Gao et al.⁴¹ in our comparative analysis also a higher content of fatty acids in the HepG2 cells were detected by the Raman imaging technology. Additionally, the amount of fatty acids in proliferating HepG2 cells obtained by Raman imaging in the exponential phase was higher than in the plateau phase, in which the cells are usually not proliferating. These findings are also consistent with data of Gao et al.,⁴¹ which indicate that an inhibition of the fatty acid synthesis causes a cell cycle arrest. By performing a classification using the spectral pattern of HepG2 cells in the exponential and plateau phase the state of cell proliferation can be predicted with almost 100% accuracy (sensitivity of 100%, specificity of 98%).

In a previous study it was already possible to detect differences in normal and malignant hepatocytes based on single point Raman spectra.^{13,42} Furthermore it was demonstrated that Raman spectroscopy is able to distinguish a metastatic from a non-metastatic cell line and to associate the detected spectral differences, which were assigned to a higher unsaturated fatty acid content, to the metastatic ability of the two isogenic cancer cell lines.¹⁷

Conclusions

In summary, this study has clearly demonstrated that Raman imaging spectroscopy in combination with multivariate data analysis is able to detect the cell-type specific molecular variations and the variation in cell proliferation of liver cancer cells. Furthermore, by using a support vector machine-based classification algorithm it was even possible to predict with a high accuracy the unknown identity and proliferation behaviour of liver cancer cells and these results have proven to be reproducible. The majority of the detected spectral differences were attributed to a larger quantity of unsaturated fatty acids in the HCC cells and during the proliferation phase. The cell type and the proliferation behaviour are essential distinguishing features of malignant tumors and Raman spectroscopy is a particularly suitable label-free method to identify easily this de novo lipogenesis in liver cancer.

Acknowledgements

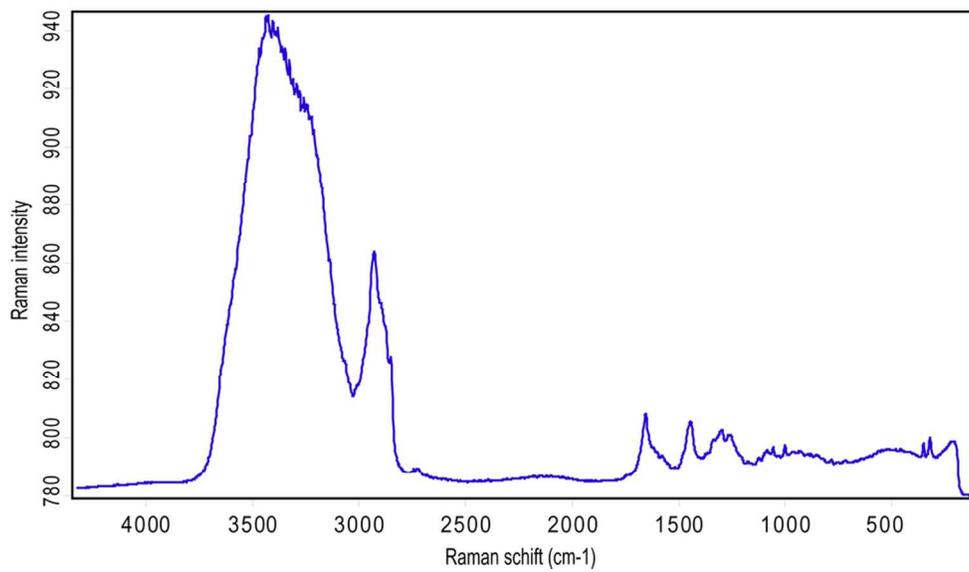
We are grateful to co-workers from gastroenterology laboratory for support, Thomas Bocklitz for critical discussions and Marita Vetterlein for the excellent technical assistance. Financial support from the European Union via the “Europäischer Fonds für Regionale Entwicklung (EFRE)” and the “Thüringer Ministerium

für Bildung, Wissenschaft und Kultur (TMBWK)” is highly acknowledged.

Notes and references

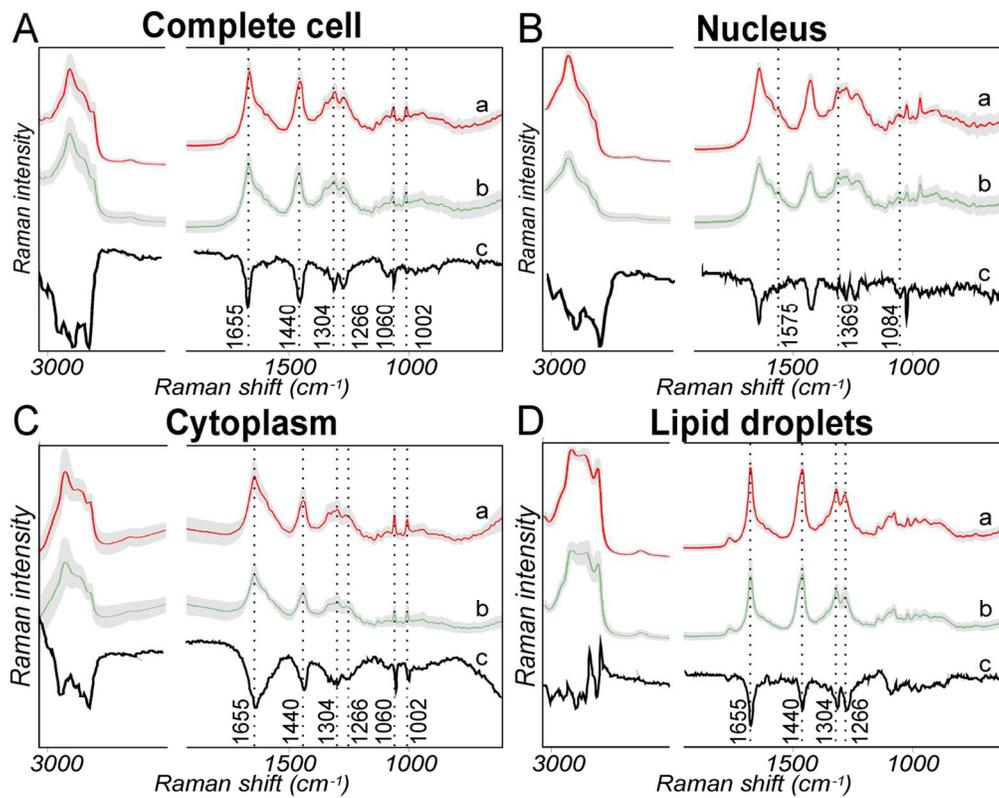
- ^a Department of Internal Medicine IV, Division of Gastroenterology, Hepatology and Infectious Diseases, Jena University Hospital, Erlanger Allee 101, 07747 Jena, Germany. Fax: (0049) 03641/9324222; Tel: (0049) 03641/9324221; E-mail: tatiana.tolstik@med.uni-jena.de
- ^b Institute of Photonic Technology, Albert-Einstein-Straße 9, 07745 Jena, Germany. Fax: (0049) 03641/206099; Tel: (0049) 03641/2060; E-mail: juergen.popp@uni-jena.de
- ^c Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller University Jena, Helmholtzweg 4, 07743 Jena, Germany. Fax: (0049) 03641/948302; Tel: (0049) 03641/948320; E-mail: juergen.popp@uni-jena.de
- [#] Both authors contributed equally
1. A. T. Harris, M. Garg, X. B. Yang, S. E. Fisher, J. Kirkham, D. A. Smith, D. P. Martin-Hirsch, and A. S. High, *Head Neck Oncol*, 2009, **1**, 38.
 2. C. Krafft, B. Dietzek, and J. Popp, *Analyst*, 2009, **134**, 1046–1057.
 3. T. Upile, W. Jerjes, H. J. C. M. Sterenborg, A. K. El-Naggar, A. Sandison, M. J. H. Witjes, M. A. Biel, I. Bigio, B. J. F. Wong, A. Gillenwater, A. J. MacRobert, D. J. Robinson, C. S. Betz, H. Stepp, L. Bolotine, G. McKenzie, C. A. Mosse, H. Barr, Z. Chen, K. Berg, A. K. D’Cruz, N. Stone, C. Kendall, S. Fisher, A. Leunig, M. Olivo, R. Richards-Kortum, K. C. Soo, V. Bagnato, L.-P. Choo-Smith, K. Svanberg, I. B. Tan, B. C. Wilson, H. Wolfßen, A. G. Yodh, and C. Hopper, *Head Neck Oncol*, 2009, **1**, 25.
 4. C. Krafft, B. Dietzek, M. Schmitt, and J. Popp, *J Biomed Opt*, 2012, **17**, 040801.
 5. P. Matousek and N. Stone, *J Biophotonics*, 2013, **6**, 7–19.
 6. D. I. Ellis, D. P. Cowcher, L. Ashton, S. O’Hagan, and R. Goodacre, *Analyst*, 2013, **138**, 3871–3884.
 7. P. Chen, A. Shen, X. Zhou, and J. Hu, *Anal. Methods*, 2011, **3**, 1257–1269.
 8. A. Shen, B. Zhang, J. Ping, W. Xie, P. Donfack, S.-J. Baek, X. Zhou, H. Wang, A. Materny, and J. Hu, *Journal of Raman Spectroscopy*, 2009, **40**, 550–555.
 9. M. S. Bergholt, W. Zheng, K. Lin, K. Y. Ho, M. Teh, K. G. Yeoh, J. B. Y. So, and Z. Huang, *Analyst*, 2010, **135**, 3162–3168.
 10. M. Harz, M. Kiehnopf, S. Stöckel, P. Rösch, T. Deufel, and J. Popp, *Analyst*, 2008, **133**, 1416–1423.
 11. T. Bocklitz, M. Putsche, C. Stüber, J. Käs, A. Niendorf, P. Rösch, and J. Popp, *Journal of Raman Spectroscopy*, 2009, **40**, 1759–1765.
 12. U. Neugebauer, J. H. Clement, T. Bocklitz, C. Krafft, and J. Popp, *Journal of Biophotonics*, 2010, **3**, 579–587.
 13. S. R. Hawi, W. B. Campbell, A. Kajdacsy-Balla, R. Murphy, F. Adar, and K. Nithipatikom, *Cancer Letters*, 1996, **110**, 35–40.
 14. C. Yu, E. Gestl, K. Eckert, D. Allara, and J. Irudayaraj, *Cancer Detection and Prevention*, 2006, **30**, 515–522.
 15. P. R. T. Jess, D. D. W. Smith, M. Mazilu, K. Dholakia, A. C. Riches, and C. S. Herrington, *Int. J. Cancer*, 2007, **121**, 2723–2728.
 16. Y. Oshima, H. Shinzawa, T. Takenaka, C. Furihata, and H. Sato, *J Biomed Opt*, 2010, **15**, 017009.
 17. M. Hedegaard, C. Krafft, H. J. Ditzel, L. E. Johansen, S. Hassing, and J. Popp, *Anal. Chem.*, 2010, **82**, 2797–2802.
 18. J. W. Chan, D. K. Lieu, T. Huser, and R. A. Li, *Anal. Chem.*, 2009, **81**, 1324–1331.
 19. A. Ghita, F. C. Pascut, M. Mather, V. Sottile, and I. Notingher, *Anal. Chem.*, 2012, **84**, 3155–3162.
 20. K. W. Short, S. Carpenter, J. P. Freyer, and J. R. Mourant, *Biophys. J.*, 2005, **88**, 4274–4288.
 21. S. Verrier, I. Notingher, J. M. Polak, and L. L. Hench, *Biopolymers*, 2004, **74**, 157–162.
 22. C. Krafft, T. Knetschke, R. H. W. Funk, and R. Salzer, *Anal. Chem.*, 2006, **78**, 4424–4429.
 23. M. R. Guarracino, P. Xanthopoulos, G. Pyrgiotakis, V. Tomaino, B. M. Moudgil, and P. M. Pardalos, *Artif Intell Med*, 2011, **53**, 119–125.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
24. N. Bergner, C. Krafft, K. D. Geiger, M. Kirsch, G. Schackert, and J. Popp, *Anal Bioanal Chem*, 2012, **403**, 719–725.
25. N. Bergner, T. Bocklitz, B. F. M. Romeike, R. Reichart, R. Kalff, C. Krafft, and J. Popp, *Chemometrics and Intelligent Laboratory Systems*, 2012, **117**, 224–232.
26. C. Krafft, M. A. Diderhoshan, P. Recknagel, M. Miljkovic, M. Bauer, and J. Popp, *Vibrational Spectroscopy*, 2011, **55**, 90–100.
27. C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, and J. Popp, *Anal. Chim. Acta*, 2013, **760**, 25–33.
28. U. Neugebauer, T. Bocklitz, J. H. Clement, C. Krafft, and J. Popp, *Analyst*, 2010, **135**, 3178–3182.
29. S. Dochow, C. Krafft, U. Neugebauer, T. Bocklitz, T. Henkel, G. Mayer, J. Albert, and J. Popp, *Lab Chip*, 2011, **11**, 1484–1490.
30. C. Krafft, B. Belay, N. Bergner, B. F. M. Romeike, R. Reichart, R. Kalff, and J. Popp, *Analyst*, 2012, **137**, 5533–5537.
31. J. W. Chan, *Journal of Biophotonics*, 2013, **6**, 36–48.
32. J. W. Chan and D. K. Lieu, *Journal of biophotonics*, 2009, **2**, 656–668.
33. D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, *CA Cancer J Clin*, 2005, **55**, 74–108.
34. P. A. Farazi and R. A. DePinho, *Nat. Rev. Cancer*, 2006, **6**, 674–687.
35. M. Roncalli, Y. N. Park, and L. Di Tommaso, *Dig Liver Dis*, 2010, **42 Suppl 3**, S228–234.
36. P. Lasch, *Chemometrics and Intelligent Laboratory Systems*, 2012, **117**, 100–114.
37. A. I. Mazur, J. L. Monahan, M. Miljković, N. Laver, M. Diem, and B. Bird, *J Biophotonics*, 2013, **6**, 101–109.
38. C. Bielecki, T. W. Bocklitz, M. Schmitt, C. Krafft, C. Marquardt, A. Gharbi, T. Knösel, A. Stallmach, and J. Popp, *J Biomed Opt*, 2012, **17**, 076030.
39. N. Yahagi, H. Shimano, K. Hasegawa, K. Ohashi, T. Matsuzaka, Y. Najima, M. Sekiya, S. Tomita, H. Okazaki, Y. Tamura, Y. Iizuka, K. Ohashi, R. Nagai, S. Ishibashi, T. Kadowaki, M. Makuuchi, S. Ohnishi, J. Osuga, and N. Yamada, *Eur. J. Cancer*, 2005, **41**, 1316–1322.
40. D. F. Calvisi, C. Wang, C. Ho, S. Ladu, S. A. Lee, S. Mattu, G. Destefanis, S. Delogu, A. Zimmermann, J. Ericsson, S. Brozzetti, T. Staniscia, X. Chen, F. Dombrowski, and M. Evert, *Gastroenterology*, 2011, **140**, 1071–1083.
41. Y. Gao, L.-P. Lin, C.-H. Zhu, Y. Chen, Y.-T. Hou, and J. Ding, *Cancer Biol. Ther.*, 2006, **5**, 978–985.
42. J. Guo, B. Du, M. Qian, W. Cai, Z. Wang, and Z. Sun, *Chin. Opt. Lett.*, 2009, **7**, 60–63.

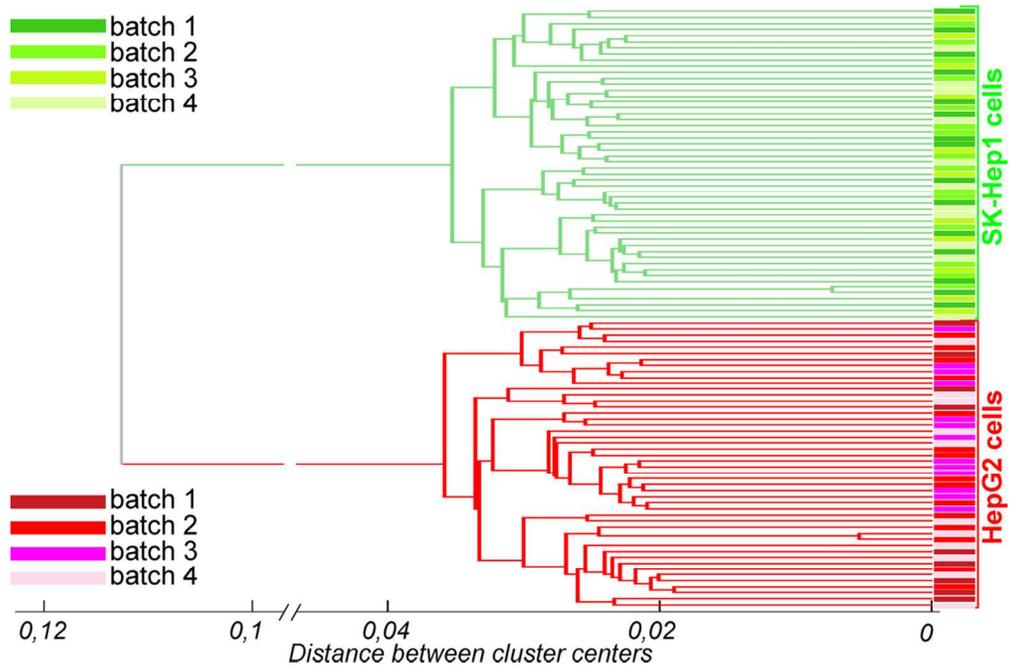


90x52mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

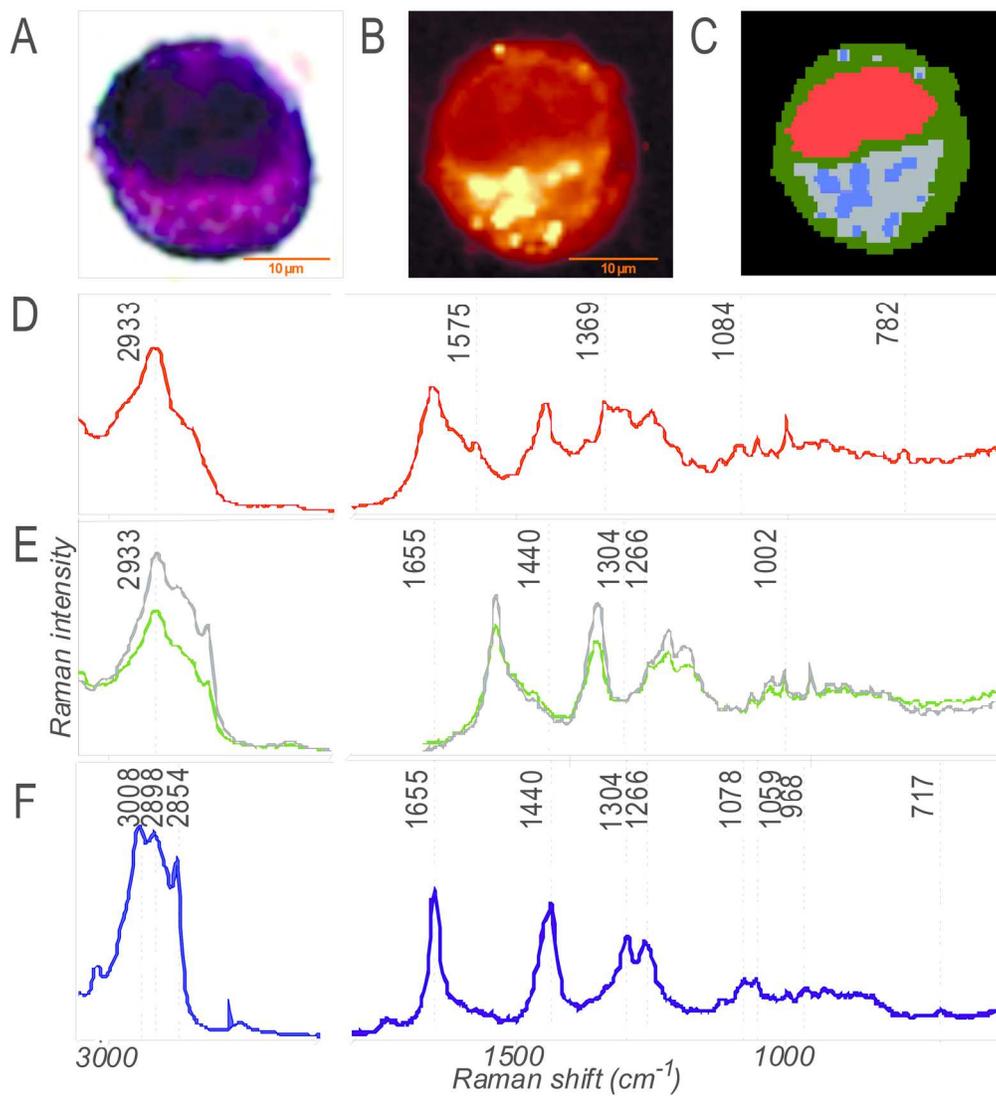


123x97mm (300 x 300 DPI)

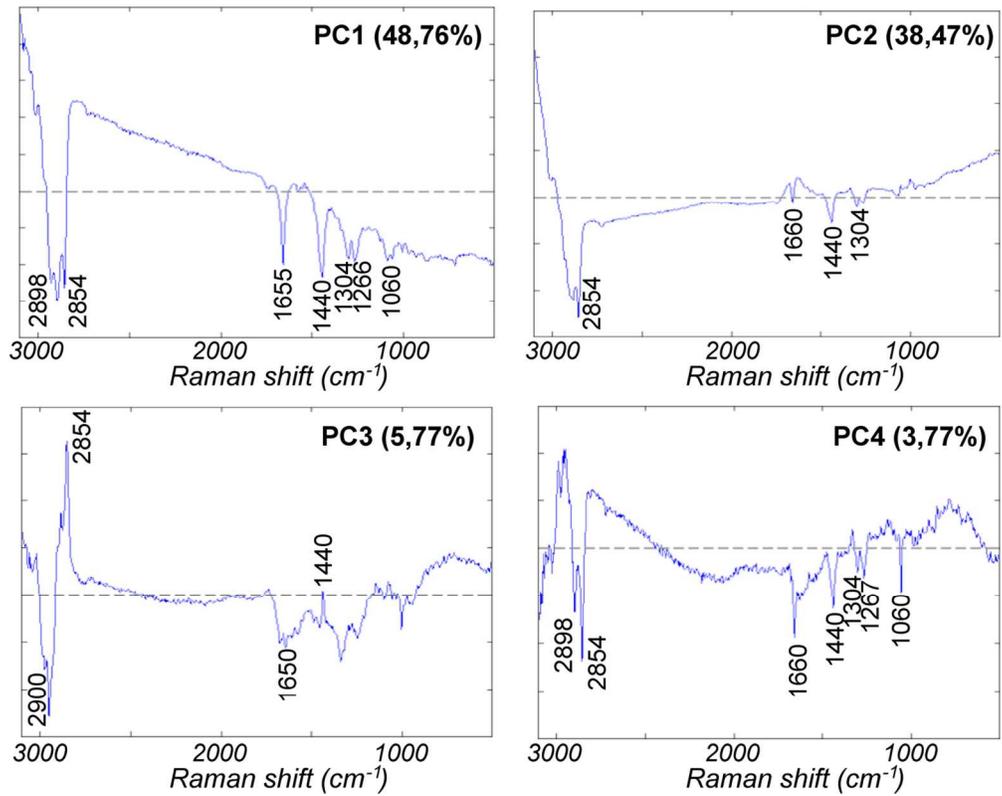


103x68mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



164x180mm (300 x 300 DPI)



124x98mm (300 x 300 DPI)

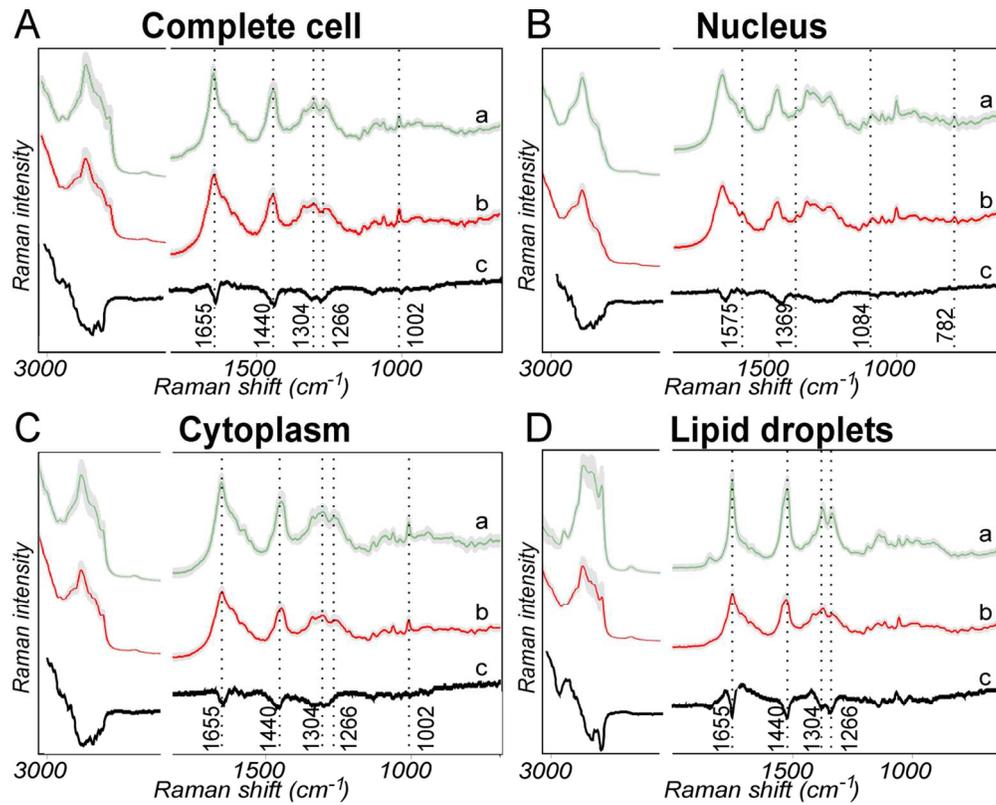
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A Complete cell						B Nucleus					
Training dataset	H	S	Test dataset	H	S	Training dataset	H	S	Test dataset	H	S
HepG2 (H)	97%	21%	H	97%	11%	HepG2 (H)	93%	21%	H	96%	5%
SK-Hep1 (S)	3%	79%	S	3%	89%	SK-Hep1 (S)	7%	79%	S	4%	85%

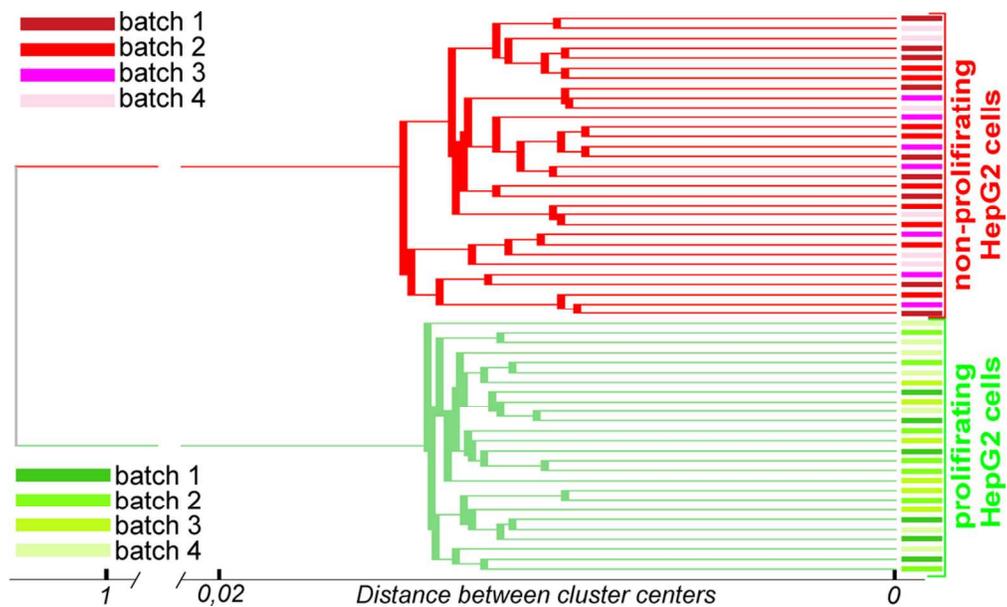
C Cytoplasm						D Lipid droplets					
Training dataset	H	S	Test dataset	H	S	Training dataset	H	S	Test dataset	H	S
HepG2 (H)	89%	8%	H	84%	11%	HepG2 (H)	95%	17%	H	93%	0%
SK-Hep1 (S)	11%	92%	S	16%	89%	SK-Hep1 (S)	5%	83%	S	7%	100%

124x77mm (300 x 300 DPI)

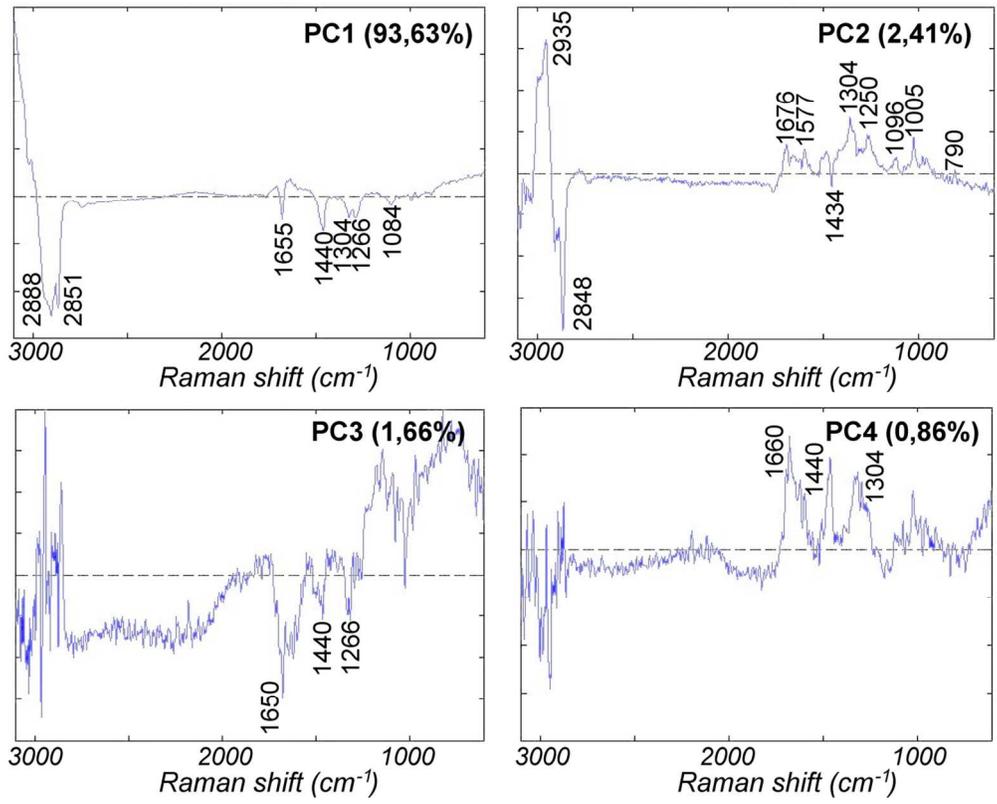
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



124x99mm (300 x 300 DPI)



94x56mm (300 x 300 DPI)



124x99mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A Complete cell						B Nucleus					
Training dataset	P	NP	Test dataset	P	NP	Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	91%	8%	P	100%	2%	Prolifer. (P)	97%	15%	P	90%	3%
Nonprolif. (NP)	9%	92%	NP	0%	98%	Nonprolif. (NP)	3%	85%	NP	10%	87%

C Cytoplasm						D Lipid droplets					
Training dataset	P	NP	Test dataset	P	NP	Training dataset	P	NP	Test dataset	P	NP
Prolifer. (P)	99%	1%	P	100%	1%	Prolifer. (P)	100%	1%	P	100%	0%
Nonprolif. (NP)	1%	99%	NP	0%	99%	Nonprolif. (NP)	0%	99%	NP	0%	100%

124x77mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60