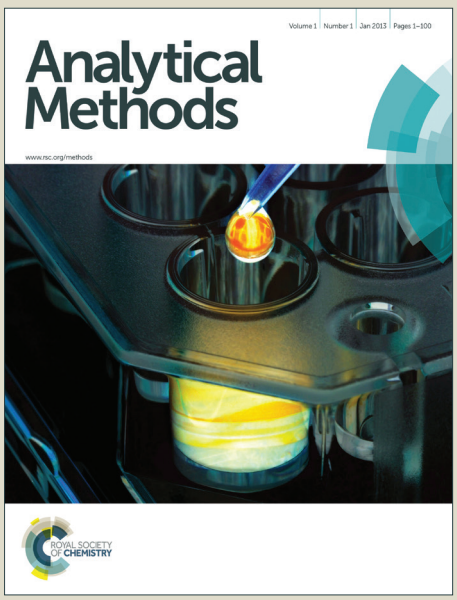
Analytical Methods

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.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Raman spectroscopy and cytopathology of oral exfoliated cells for oral cancer diagnosis

Aditi Sahu[1#], Sneha Tawde[1#], Venkatesh Pai[2], Poonam Gera[3], Pankaj Chaturvedi[4], Sudhir Nair[4], C. Murali Krishna[1]*

(*mchilakapati@actrec.gov.in)

[1]Chilakapati laboratory, ACTREC Kharghar, Navi Mumbai; [2]Mahimkar Lab, ACTREC, Kharghar, Navi Mumbai; [3]Biorepository lab, ACTREC, Kharghar, Navi Mumbai;[4]Head and Neck Surgical Oncology, Tata Memorial Center, Mumbai.

#- Contributed equally

* Address for correspondence

Dr. C. Murali Krishna

Scientific Officer 'F' and Principal Investigator

Chilakapati Laboratory

ACTREC, TMC

Kharghar, Sector '22'

Navi Mumbai – 410210, India.

Telephone: +91-22-2740 5039/5544

Fax: +91-22-2740 5095

Email - mchilakapati@actrec.gov.in, pittu1043@gmail.com

Keywords: Oral cancer, exfoliative cytology, Raman spectroscopy, Pap staining, PCA, PC-LDA

Analytical Methods Accepted Manuscript

Abstract:

For oral cancers, screening and monitoring of high-risk populations can aid in early diagnosis and improve overall outcomes. Of the new methods, approaches based on exfoliative cytology are more practical for mass screening and monitoring of high-risk populations. Raman spectroscopy and exfoliative cytology for cervical cancers has shown promise in differentiating normal and abnormal samples. In this study, feasibility of Raman oral exfoliative cytology along with cytopathology for oral cancer diagnosis was evaluated on 70 specimens. Exfoliated cells were obtained from 15 healthy volunteers (HV), 15 healthy tobacco users (HT), and 20 contralateral or disease control (DC) and 20 tumor (T) sites of oral-cancer patients. Pap staining was carried out post Raman spectral acquisition. Spectral findings demonstrate that with increase in severity of pathology from HV to T, higher DNA and changes in secondary structure of proteins were encountered. Owing to heterogeneity in cellular samples, two different approaches- point-spectra and patient-wise were evaluated for data analysis. PCA and PC-LDA using both approaches indicate that HV and HT are distinct from cancer groups DC and T. Misclassifications were also observed between HT and DC. These findings also correlate with cytopathological findings. Less misclassifications and higher classification efficiency was observed for patient-wise approach. Large-scale validation study needs to be undertaken for evaluating utility of Raman oral exfoliative cytology for screening of oral cancers using patient-wise approach.

Analytical Methods Accepted Manuscript

INTRODUCTION:

The incidence of oral cancer worldwide is around 300000 new cases every year[1]. Tobacco abuse (smoking and smokeless) is the most common etiological factor for oral cancer[2]. The overall 5-year survival rate of around 50% is mainly attributed to delayed diagnosis and recurrence. These rates are less than occult cancers like breast, colorectal, cervix[3]. Early detection of oral cancer and its curable precursors remains the best way to ensure patient survival and improved quality of life[4, 5]. South Asian countries like India are major contributors to the global cancer burden, due to rampant tobacco habits. In fact, it is estimated that over 90% of the global smokeless tobacco burden is in South Asia- around 100 million people use smokeless tobacco in India and Pakistan alone [6, 7]. Due to a known dose-response relationship between tobacco consumption and development of oral cancer, chronic tobacco abusers are at high risk for development of oral cancer [8]. Thus, screening and monitoring of these high-risk populations, along with the general population is crucial.

The current standard screening procedure for oral cancers is visual inspection followed by biopsy and histopathology of suspicious lesions. These currently employed methods have shown limited potential in detecting precancerous or early cancerous lesions[9]. In fact, visual examination was useful as a method of screening for oral cancer only in high risk cases like chronic smokers or alcoholics[10]. The suspicious lesions detected upon visual inspection are biopsied and sent for microscopic examination for confirming abnormality in the sample. The invasiveness of this methodology leads to reduced patient compliance and therefore higher number of defaulter patients. Thus, there is a need for new, less-invasive approaches to complement the existing methodologies to identify dysplasia and early oral cancers in both, symptomatic and

Analytical Methods Accepted Manuscript

asymptomatic patients. Several new diagnostic modalities have been developed in recent times to aid in early detection of oral cancer[3,11]. These include: use of vital staining like Toluidine blue, brush biopsy, salivary biomarkers, and optical methods like Vizilite, Vizilite Plus, Microlux DL, VelScope and Identafi 3000. **These light-based methods rely on the assumption that absorption and reflection of light differs between normal and abnormal (metabolic or structural change) tissues[12]. Vizilite and MicroLux DL are based on the principle of chemiluminescence, wherein a rinse using 1% acetic acid is followed by direct visual examination using a blue-white light source. While Vizilite uses a disposable chemiluminescent light, MicroLux DL employs a reusable, battery-powered source. Oral mucosal abnormalities appear as "aceto-white" against a lightly bluish normal mucosa. VelScope, a portable device for direct visualization of oral cavity using intense blue excitation light, employs the principle of change in tissue autofluorescence during abnormal conditions. Using a selective (narrow-band) filter, dark regions representing abnormal areas with loss of fluorescence interspersed in green fluorescence emitted by normal mucosa are observed. Identafi employs multi-spectral Fluorescence and Reflectance technology to enhance visualization of mucosal abnormalities using 3 different wavelengths. White light is used for conventional examination, violet light to detect changes in tissue autofluorescence while amber light is used to detect morphological abnormalities in vasculature[13, 14].** These new innovative techniques show promising results, however their routine clinical use will depend on the results of methodologically sound clinical trials that demonstrate their impact on patients' survival and effectiveness in early detection.

Exfoliative cytology is a simple, rapid, and less invasive technique[15]: it is thus well accepted by patients and is suitable for routine application in population screening programmes. This

Analytical Methods Accepted Manuscript

methodology has been used as a tool for screening gynecological malignant diseases, after development of Papanicolaou (Pap) staining by George Papanicolaou[16, 17].Pap staining of oral exfoliated cells was first explored to study keratinization patterns and changes during menstrual cycle[18-20], followed by studies on oral cancer diagnosis[21, 22]. At that time, this technique could not achieve the expected sensitivity and specificity[23]. After several modifications in sample collection and staining methods, introduction of computerized image analysis, DNA cytometry and addition of molecular tests like detection of p53 mutation and loss of heterozygosity, this area has received renewed interest in the last few years[24-27]. The clinical utility of these approaches is still under verification.

Raman spectroscopy (RS), a vibrational optical spectroscopic method based on inelastic scattering of light, has been increasingly employed for disease diagnosis. *In vivo* RS has previously shown promise in management of oral cancers, including detection of early changes like cancer field effects (CFE)/ malignancy associated changes (MAC)[28-30]. In vivo applications require on-site instrumentation and stringent experimental conditions. Less invasive samples like exfoliated cells can be collected at multiple screening centers, followed by transportation to and analysis at a centralized facility. RS of exfoliated cells from cervix has shown promise in detecting HPV infection and classifying normal and abnormal specimens[31, 32]. A pilot study has confirmed feasibility of Raman spectral acquisition on oral exfoliated cells[33]. In the present study, pellet of exfoliated cells from buccal mucosa of healthy subjects (with and without tobacco habits) and oral cancer patients were analyzed using RS, followed by Pap staining. Due to inherent heterogeneity of cell samples, multiple spectra were acquired from different areas of the pellet, and both point-spectra and patient-wise approach were evaluated. Findings are presented in the manuscript.

Analytical Methods Accepted Manuscript

MATERIALS AND METHODS

2.1Exfoliative cytology:

Oral exfoliated cells were collected from histopathologically confirmed oral cancer (buccal mucosa) patients with tobacco habits visiting the out-patient department of Head and Neck oncology, Tata Memorial Hospital, Mumbai, India. Exfoliated cells were collected from "site of tumor" (referred to as 'T' henceforth) and respective "Contralateral normal or Disease control sites" (referred to as 'DC' henceforth in the manuscript) from the same patient using Cytobrush (Himedia, India) in normal saline. Exfoliated specimens were also collected from both buccal mucosae of healthy subjects unrelated to cancer patients, subjects with tobacco habits were termed as 'healthy tobacco' (HT) while without tobacco habits were termed as 'healthy volunteers' (HV). No clinical abnormalities were found in mucosa of HT and HV subjects. Demographic details of study group are summarized in Table 1.

Samples were transported in ice at 4° C, and immediately processed upon arrival. Exfoliated cells were harvested after vortexing and centrifugation at 6000 rpm for 1 min. Due to high vascularity of tumors, blood contamination was observed in most cancer samples; RBC lysis protocol was therefore adopted as an integral step during cell processing. For sake of uniformity, all samples were treated with RBC lysis protocol after harvesting cell pellet. The protocol is as follows: 1 ml RBC lysis buffer was added to each tube, incubated for 15 min at room temperature (mixed in intervals of 5 min), and washed in saline two times. Cell counting was performed using Neubauer's chamber after staining with Trypan blue and the final pellet was employed for spectral acquisition. After spectral acquisition, cells were smeared onto glass slides, fixed and subjected to Pap staining.

Analytical Methods Accepted Manuscript

2.2 Raman spectroscopy:

The pellet was placed on a CaF_2 window and spectra recorded using Fiber Optic Raman microprobe (Horiba-Jobin-Yvon, France). The details of the instrument have been described elsewhere[34]. Briefly, this system consists of laser (785 nm, Process Instruments) as an excitation source and HE 785 spectrograph (Horiba-Jobin-Yvon, France) coupled with CCD (Synapse, Horiba-Jobin-Yvon) as dispersion and detection elements, respectively. Optical filtering of unwanted noise, including Rayleigh signals, is accomplished through 'Superhead', the other component of the system. Optical fibers were employed to carry the incident light from the excitation source to the sample and also to collect the Raman scattered light from the sample to the detection system. Raman microprobe was assembled by coupling a 40X microscopic objective (Nikon, Japan) to the superhead. Spectral acquisition parameters were: laser power: 40 mW, acquisition time: 15 s and average: 3. About 6-7 spectra were recorded from several areas to span the pellet. After spectra recording, cells were smeared onto glass slides and subjected to Papanicolaou staining.

2.3 **Papanicolaou staining**

Pap stain employs a combination of 3 stains, namely Haematoxylin, Orange G and Eosin-Azure[35]. Haematoxylin stains cell nucleus a blue color, while the differentiating stains OG and EA provide a subtle range of orange, bluish-green, and pink hues to the cytoplasm according to the amount of keratin and degree of maturation. Post Raman-spectral acquisition, cell pellet was smeared onto the glass slide, fixed in 95% ethanol and Pap stained. The slides were read by a certified pathologist subsequent to mounting. The protocol for Pap staining standardized in the current study is shown in Figure 1.

Analytical Methods Accepted Manuscript

2.5 Spectral pre-processing and data analysis:

The acquired Raman spectra were corrected for CCD response and spectral contaminations from substrate and fiber signals. To remove interference of the slow moving background, first derivatives of spectra (Savitzky-Golay method and window size 3) were computed[36, 37]. Spectra were interpolated in the range 800-1800 cm^{-1} since this region is an important constituent of the finger-print region. Interpolated first derivative and vector normalized spectra were then subjected to multivariate unsupervised Principal Component Analysis (PCA) and supervised Principal Component-Linear Discriminant Analysis (PC-LDA). In brief, Principal Component analysis or PCA is routinely used method for data compression and visualization. It describes data variance by identifying a new set of orthogonal features, called as principal components (PCs) or factors. In LDA, the classification criterion is identified using the scatter measure of within class and between class variance. LDA can be used in conjunction with PCA (PC-LDA) to increase the efficiency of classification. The advantage of doing this is to remove or minimize noise from the data and concentrate on variables important for classification. In our analysis, significant principal components ($p<0.05$) were selected as input for LDA. In order to avoid over-fitting of the data, as a thumb rule, total number of factors selected for analysis were less than half the number of the subjects in the smallest group[38-40]. PC-LDA models were validated by Leave-one-out cross-validation (LOOCV). Leave-one-out cross validation is a type of rotation estimation, a technique used for assessing performance of a predictive model with a hypothetical validation set when an explicit validation set is not available. Leave-one-out involves using a single observation from the original sample as the validation data, and the remaining observations as training data. This is repeated such that each observation in the sample is used

Analytical Methods Accepted Manuscript

once as the validation data and averaged over the rounds[41]. Algorithms for these analyses were implemented in MATLAB (Mathworks Inc., USA) based in-house software[42].

Average spectra were computed from the background-subtracted spectra prior to derivatization for each class and were baseline-corrected by fitting a fifth order polynomial function. These baseline-corrected, smoothed (Savitzky–Golay, 3) and vector-normalized spectra were used for spectral comparisons.

RESULTS AND DISCUSSION:

Early diagnosis of oral cancer may lead to increased disease free-survival rates and better outcomes. Visual inspection is the currently available method for screening which may be useful as a screening tool only in high-risk cases. Exfoliative cytology is a more practical tool for mass screening and also monitoring of high-risk populations: due to ease of collection, safety, cost-effectiveness and possibility of multiple sampling. Like cervix cancers, Pap-based screening programs can help in detecting early cancer-related morphological changes in the oral cavity. However, morphological features alone may prove insufficient in detecting early stages of oral cancer, as very few cytopathologically abnormal cells may be present in the sample. Spectral cytopathology (IR microscopy coupled with multivariate analysis) has been shown to detect disease in even those cells from the oral cavity which did not show any obvious morphological abnormalities[43]. Raman spectroscopy coupled with cytopathology may also have the potential to detect such subtle cellular changes indicative of malignant transformation. Before envisaging screening-related applications to detect early oral cancer, the feasibility to differentiate healthy and tumor exfoliated cells has to be evaluated. In the present study, exfoliated cells were collected from healthy (tobacco and non-tobacco habitués) and cancer (tumor and contralateral

Analytical Methods Accepted Manuscript

sites) subjects. Some studies suggest that despite having normal morphology, cells with MAC and squamous atypia, and even normal appearing cells from the majority of tumor samples, share biochemical features with cells which are morphologically cancerous. Therefore, the high-risk group of chronic tobacco abusers and contralateral site were included to identify biochemical changes induced upon tobacco exposure and early MAC, respectively. Cytological analysis of samples was also carried out using Pap staining for parallel confirmation of cellularity and morphological features of exfoliated cell specimens.

Cytological findings:

Cytological analysis of the same pellet was carried out after Pap staining by a certified pathologist. Apart from the normal cells of the three epithelial layers, namely stratum basale, intermedium and superficial, slides were scored for presence of orangeophilic cells, parakeratotic cells, anucleate squames which mainly indicate keratinization status of the mucosa; inflammatory, dyskeratotic and dysplastic cells, which may indicate tendency of malignant transformation of the mucosa. Representative slides for the 4 groups, i.e. HV, HT, DC and T are shown in Figure 2a-d.

In the healthy volunteers (HV) smears, normal proportions of cells from stratum intermedium, stratum superficial, and few cells from stratum basale were observed. A very small percentage of parakeratotic and orangeophilic cells were detected in few cases but no dysplastic or inflammatory cells were found. Representative cytological smears are shown in Figure 2a. As in HV, healthy tobacco (HT) smears showed no dysplastic or inflammatory cells but higher numbers of orangeophilic, and occasional parakeratotic cells and anucleate squames were

Analytical Methods Accepted Manuscript

observed (Figure 2b). In the oral cancer group- contralateral (DC) and tumor (T) cell smears, higher number of orangeophilic cells, parakeratotic cells and anucleate squames, along with dyskeratotic and dysplastic cells as compared to HT smears were observed. Few instances of dysplasia were observed in DC which were absent in HT. The frequency of dyskeratosis and dysplasia was highest in T (Figure 2c-d).

In the HV group, chronic chemical irritation due to tobacco-related products was absent. The small percentage of orangeophilic and parakeratotic cells could possibly be attributed to exposure to some physical or chemical irritants. In case of HT group, chronic tobacco exposure leads to induction of the protective stimuli of the epithelium wherein the epithelial cells produce keratin. Thus, an increase in the number of orangeophilic cells (indicating abundance in intracellular keratin), parakeratotic (polygonal cells with pyknotic nucleus and abundant intracellular keratin), anucleate squames (orange cells with degenerated nucleus) is observed. The oral cancer subjects recruited in the study were also chronic tobacco abusers. Higher frequency and longer duration of tobacco habits in these subjects is observed in terms of increased degree of keratinization (higher numbers of orangeophilic cells, parakeratotic and anucleate squames in both DC and T groups). As expected, from contralateral to tumor condition, frequency of dyskeratosis (premature, abnormal keratinization) and dysplasia (cellular abnormality associated with malignant tendency) increased. Variable number of dysplastic cells can also be expected in the contralateral group ascribed to two reasons a) cancer-field effects and b) migration of tumor cells by saliva (micro-metastases)[44]. Thus, overall cytological findings indicate very few parakeratotic, orangeophilic cells in HV, higher numbers in HT and DC,

Analytical Methods Accepted Manuscript

presence of few dysplastic cells in DC and highest numbers of parakeratotic, orangeophilic and dysplastic cells in T[45].

Spectral Features:

The average spectra for the four groups are shown in Figure 3. Prominent spectral features include bands at 830 cm^{-1}, 857 cm^{-1}, 1009 cm^{-1}, 1093 cm^{-1}, 1170 cm^{-1}, bands in the region of 1270-1340 cm^{-1}, 1450 cm^{-1} and 1660 cm^{-1} that can be overall assigned to phenylalanine, DNA-phosphate backbone related features, amide III, CH_2 twisting in proteins and lipids, DNA base features, CH_2 bending in protein and lipids and amide I features from proteins[46, 47]. **The band at 1170 cm^{-1} can be tentatively assigned to tyrosine and v (C-C) of skeletal structure in proteins[48-50] and also to tobacco-related amine-containing adducts[51-53]. This peak is present in all groups; the higher presence in cancer groups (both contralateral and tumor) may be attributed to increased tobacco exposure groups.** With the increase in severity of pathology in HV to T, higher DNA (1095 and 1325-30 cm^{-1}), higher CH_2 bending (1450 cm^{-1}), higher amide III was observed. Broadening in the amide I region was also encountered with increase in pathological severity. As compared to the other groups, broad amide I, higher CH_2 bending and amide III regions, and nucleic acid bases features around 1320-1340 cm^{-1} were observed in the average tumor spectrum. Thus, increase in DNA content and changes in protein secondary structures could be hallmarks of severe pathological states.

Multivariate analysis

A total of 386 spectra (89: Healthy volunteer, 100: habit control, 86: disease control, 110: tumor) from 50 cases (15: Healthy volunteer, 15: habit control, 20: disease control and 20: tumor

Analytical Methods Accepted Manuscript

from same oral cancer patient) were acquired. In the first point-spectra wise approach, all spectra from the 50 subjects (total 386 spectra) were used for multivariate analysis. Three hundred and eighty-six spectra were first subjected to PCA to understand trends in data. For PCA, scores of factors 1 and 3 were explored for classification. Factor loadings for factor 1 and 3 are shown in Figure 4a-b. The scatter plot in Figure 4c indicates almost distinct clusters for HV and HT while overlapping clusters are observed between DC and T groups. As PCA is not a classification tool but is used for data compression and visualization to indicate trends in the data, PC-LDA was employed to explore classification between the groups. Six factors (Figure 5a) were used for the analysis which accounted for ~68% classifications. Scores of factor 1 and 2 were employed for obtaining scatter plot, as shown in Figure 5b. As seen in the scatter plot, HV and HT form different clusters while overlap is observed between DC and T. As seen in the PC-LDA confusion matrix (Table 2a), 69/89 HV spectra were correctly classified while 12 misclassified with HT. 68/100 HT spectra were correctly classified while 21 misclassify with HV. Out of 86 DC spectra, 59 got correctly classified while 14 misclassified with T, 7 with HT and 6 with HV. In case of T, 69/116 spectra got correctly classified while 33/116 misclassified as DC, 9/116 with HT and 5/116 with HV. As PC-LDA is a supervised approach, leave-one-out-cross-validation (LOOCV) was carried out to evaluate the results obtained by PC-LDA. As seen in the LOOCV confusion matrix (Table 2b), 69/89 spectra of HV group were correctly classified and 12/89 misclassified with HT. 68/100 HT spectra were correctly classified while 21 spectra misclassified with HV. 58/86 DC spectra were correctly classified while misclassification of 14 spectra with T was observed. 68/116 spectra of T got classified correctly, while 34 misclassified as DC, 9 as HT and 5 as HV. Thus, misclassifications of HV were seen mainly with HT (14%) while for HT, major misclassifications were with HV (20%), and some with DC (8%). In case of

Analytical Methods Accepted Manuscript

DC and T, major misclassifications were between these groups: 17% DC misclassified with T and 29% T misclassified with DC. Some DC and T also misclassified with HT (~7%) and HV (6% DC and 4% T).

A second approach, termed as patient-wise approach, wherein average spectrum of each sample was used for data analysis was also explored. It is known that the laser spot size is 4-5 micron and the penetration is ~40 micron. The laser probing volume accommodates several stacked cells and their intracellular components. Studies suggest that intracellular and cell-to-cell variation can be detected by Raman spectroscopy[48, 54, 55]. Thus, average spectrum was calculated to yield a true representative of the sample and also to circumvent the intra-sample heterogeneity. These average spectra were then subjected to multivariate analysis PCA and PC-LDA. PCA was carried out using 10 factors, scores of factor 2 and factor 4 were used to obtain scatter plot. Factor loadings are shown in Figure 6a-b. As seen in Figure 6c, slightly distinct clusters for healthy, and tumor groups while overlap of contralateral with both healthy tobacco and tumor groups was observed. As compared to the first approach, better classification trends were observed for the patient-wise approach. PC-LDA was then carried out to build standard models, using 4 factors (Figure 7a). The scatter plot shown in Figure 7b indicates almost exclusive clusters for healthy and tumor groups. The PC-LDA confusion matrix is shown in Table 3, 13/15 HV, 12/15 HT, 11/16 DC and 12/19 T were correctly classified. LOOCV was carried out to validate the results which show that 13/15 HV, 11/15 HT, 9/16 DC and 12/19 T were correctly classified. Thus, 86% HV were correctly classified, 6% misclassifications were with HT and DC, none with T. Similarly for HT, 13% misclassifications were found with HV and DC, none with T. For DC, major misclassifications were with HT (19%), and 12% with HV and T. 21% T misclassified

Analytical Methods Accepted Manuscript

with DC, only 10% and 5% with HT and HV, respectively. Thus, better classification efficiency and lower misclassification rate was obtained for all groups using the patient-wise approach. As averaging leads to true representation of the cell pellet, less-cross talk between samples of different groups was observed.

In case of healthy tobacco users, it is known that chronic tobacco exposure leads to biochemical changes in oral mucosa leading to cytological changes in the cells. However, the tobacco exposure in the buccal mucosa area and the response to tobacco exposure may not be uniform for all cells. Thus, presence of normal cells in the HC group smears is expected- also observed during cytological analysis. Minor misclassifications between these two groups can thus be explained. Some misclassifications were also observed between the HT and DC groups. The contralateral mucosa of the oral cancer patient has also been chronically exposed to tobacco. This persistent and chronic irritation first leads to biochemical changes in the cells (thus similarity with HT) eventually followed by malignant transformation. In case of cancer group, major misclassifications were observed between DC and T. These misclassifications could be attributed to tumor heterogeneity, and collection of other cells (from contralateral site) is also a possibility. Further, as previously stated, contralateral mucosa may have presence of genetically altered cells due to field cancerization or micro-metastases[44, 56]. This is evident even in the cytopathological findings, where parakeratotic, orangeophilic and dysplastic cells are present in both groups, only higher numbers are evident in the T group[57, 58.]

It has been observed that cervical cancer screening using Pap-based cytological test dramatically reduced cervical cancer incidence and mortality in developed countries[59]. It is known that such diagnoses based on human interpretations of morphological changes in cells/tissues are

Analytical Methods Accepted Manuscript

associated with low sensitivity and specificity. Early pathological conditions usually comprise of fewer cells with definite abnormal structural changes interspersed between vast majority of normal cells, thus diagnosis is further complicated. The criteria of only morphological changes may therefore be insufficient for early cancer detection. Optical spectroscopies like FTIR and Raman yield a global spectroscopic signature by probing inherent cellular biochemistry. Further, these are objective, rapid approaches independent of human errors like fatigue, inexperience and inter-observer variability. Thus, they may prove to be potential tools for early cancer diagnosis. Several cell-based Raman spectroscopic studies using either cell pellets, cell suspensions and thin-cell layers have been carried out. Recent Raman spectroscopic studies have demonstrated difference between normal and CIN cytology using ThinPrep approach **(an improvised exfoliated cell sample collection and processing method)**, while another study using monolayer of cells has discriminated normal, dysplastic and SCC cell lines[60, 61]. Thus, Raman mapping of oral cell smears using automated scanning may be an ideal approach for early oral cancer diagnosis. This approach requires recording of spectra from individual cells, and thus longer spectral acquisition times are required. **Coherent anti-Stokes Raman spectroscopy (CARS) and stimulated Raman spectroscopy (SRS) are non-linear, high-speed biomolecular imaging approaches, but they may have limited clinical utility[62, 63].**

Studies on cell-pellets have differentiated HPV+ and HPV- cell lines; normal and abnormal cervical exfoliated samples and identified HPV infection in cervical cancer patients. Even a single cell population could be identified in mixed cancer cell populations using this approach[55]. One of the major restraints for this approach is cell sample heterogeneity, as Raman probing involves accumulating information from several stacks of cells. Multiple spectra have to be recorded from cell pellets and then tested against robust standard models. If even a single

Analytical Methods Accepted Manuscript

spectrum is classified as abnormal, as per standard histopathological guidelines, the sample is treated as abnormal. For this diagnosis, pure standard models are required for normal and abnormal groups. Because of heterogeneity in malignant and pre-malignant lesions, it is difficult to obtain pure samples for spectral acquisition. An alternative to this point-spectra approach is the patient-wise approach, where all spectra from a sample are averaged to yield a representative spectrum which is subsequently employed for data analysis. It is known that different sub-cellular regions of a cell, different cells within a single population and different cell types can be clearly distinguished due to the inherent sensitivity of the Raman spectroscopy. Thus, by calculating average spectra, contributions from all cellular features of the pellet are included and intra-sample variation is thereby minimized. Several previous studies have also employed a similar approach: single cell studies have compared the averaged Raman spectra of many single cells from each sample, while some pellet based studies compared spatially averaged spectra, obtained from many cells at once from a pellet[64-67]. In the present study, Raman measurements were made at several areas of the pellet and then used to compute a representative average of each sample, which was then employed for data analysis. This approach demonstrated higher true classification for all groups, as compared to the point-spectra approach. Thus, the patient-wise approach may be a more practical and faster approach in low-resource settings as well as an alternative to Raman mapping mode which requires very long spectral acquisition times and sophisticated instrumentation. After preliminary screening using this approach, suspicious samples can be followed up and tested for malignancy-related abnormalities using confirmatory procedures.

Conclusions:

Analytical Methods Accepted Manuscript

Exfoliative cytology is actively being pursued as an adjunct tool for oral cancer screening. However, restricting disease diagnosis only to morphological alterations may not be adequate. In the present study, Raman spectroscopy and Pap staining was carried after oral exfoliative cytology. Spectra were acquired from different areas of cell pellet and subjected to Pap staining after smearing on glass slide. Major spectral features indicate higher DNA, CH_2 bending and amide III and amide I broadening with increase in pathological severity. Cytopathology indicates higher parakeratotic, orangeophilic and dysplastic cells with increasing severity of pathological condition. For data analysis, both spectra-wise and patient-wise approaches was explored in this study. A patient-wise approach leads to better representation of cell pellets. Consequently, higher true classification and lower misclassifications were observed due to decreased cross-talk between samples. This study on 70 specimens serves as a proof of concept for use of Raman exfoliative cytology for oral cancer management, especially screening of high-risk populations. Extensive validation studies using benign, inflammatory and premalignant conditions may throw light on actual potential of this methodology for prospective use in clinics.

ACKNOWLEDGEMENT:

The Raman spectrometer employed in the study was procured from DBT project BT/ PRI11282/MED/32/83/2008 (“Development of in vivo laser Raman spectroscopy methods for diagnosis of oral precancerous and cancerous conditions”), Department of Biotechnology, Government of India. The authors would also like to acknowledge the help of senior residents, especially Dr. Akshat Malik of the Head and Neck ‘C’ department, TMH, Mumbai for their help in collection of exfoliated cells from oral cancer patients.

Analytical Methods Accepted Manuscript

REFERENCES:

1. D. M. Parkin, F. Bray, J. Ferlay and P. Pisani, *CA: a cancer journal for clinicians*, 2005, **55**, 74-108.
2. S. Khandekar, P. Bagdey and R. Tiwari, *Indian J Community Med*, 2006, **31**, 157-159.
3. D. V. Messadi, *International journal of oral science*, 2013, **5**, 59-65.
4. S. S. Napier and P. M. Speight, *Journal of oral pathology & medicine*, 2008, **37**, 1-10.
5. I. van der Waal, *Oral oncology*, 2010, **46**, 423-425.
6. R. John, H. Sung and W. Max, *Tobacco control*, 2009, **18**, 138-143.
7. S. Z. Imam, H. Nawaz, Y. J. Sepah, A. H. Pabaney, M. Ilyas and S. Ghaffar, *BMC Public Health*, 2007, **7**, 231.
8. H. Kuper, P. Boffetta and H. O. Adami, *Journal of internal medicine*, 2002, **252**, 206-224.
9. O. Kujan, A. M. Glenny, R. Oliver, N. Thakker and P. Sloan, *Australian Dental Journal*, 2009, **54**, 170-172.
10. R. Sankaranarayanan, K. Ramadas, G. Thomas, R. Muwonge, S. Thara, B. Mathew, B. Rajan and T. O. C. S. S. Group, *The Lancet*, 2005, **365**, 1927-1933.
11. R. Mehrotra and D. K. Gupta, *Head & neck oncology*, 2011, **3**, 1-9.
12. E. Omar, *Head & face medicine*, 2015, **11**, 6.
13. M. W. Lingen, J. R. Kalmar, T. Karrison and P. M. Speight, *Oral oncology*, 2008, **44**, 10-22.
14. A. F. Zuluaga, N. Vigneswaran, R. K. Bradley, A. M. Gillenwater, C. M. Nichols and C. Poh, 2010.

Analytical Methods Accepted Manuscript

15. M. Diniz-Freitas, A. García-García, A. Crespo-Abelleira, J. Martins-Carneiro and J. Gándara-Rey, *Medicina oral: órgano oficial de la Sociedad Española de Medicina Oral y de la Academia Iberoamericana de Patología y Medicina Bucal*, 2004, **9**, 355.

16. G. N. PAPANICOLAOU, *Annals of internal medicine*, 1949, **31**, 661-674.

17. G. N. Papanicolaou and H. F. Traut, *New York*, 1943, **46**.

18. J. Weinmann, *Journal of Dental Research*, 1940, **19**, 57-71.

19. P. K. D.E. Ziskin, I. Kitley, *J Dent Res.* , 1941, **20**, 386-387.

20. G. N. Papanicolaou, *Ex [oliative Cytology*, 1954, 13-21.

21. P. W. Montgomery and E. Von Haam, *Journal of dental research*, 1951, **30**, 308-313.

22. H. Sandler and S. S. Stahl, *Journal of oral surgery*, 1958, **16**, 414.

23. R. Mehrotra, *Oral Cytology: A Concise Guide*, Springer Science & Business Media, 2012.

24. R. Mehrotra, M. Hullmann, R. Smeets, T. Reichert and O. Driemel, *Journal of Oral Pathology & Medicine*, 2009, **38**, 161-166.

25. J. J. Sciubba and U. C. O. S. Group, *The Journal of the American Dental Association*, 1999, **130**, 1445-1457.

26. N. K. Proia, G. M. Paszkiewicz, M. A. S. Nasca, G. E. Franke and J. L. Pauly, *Cancer Epidemiology Biomarkers & Prevention*, 2006, **15**, 1061-1077.

27. R. Mehrotra, *Diagnostic cytopathology*, 2012, **40**, 73-83.

28. S. Singh, A. Deshmukh, P. Chaturvedi and C. M. Krishna, 2012.

29. S. Singh, A. Deshmukh, P. Chaturvedi and C. M. Krishna, *Journal of biomedical optics*, 2012, **17**, 1050021-1050029.

Analytical Methods Accepted Manuscript

30. S. Singh, A. Sahu, A. Deshmukh, P. Chaturvedi and C. M. Krishna, *The Analyst*, 2013, **138**, 4175-4182.

31. E. Vargis, Y.-W. Tang, D. Khabele and A. Mahadevan-Jansen, *Translational oncology*, 2012, **5**, 172-179.

32. S. Rubina, M. Amita, R. Bharat and C. M. Krishna, *Vibrational Spectroscopy*, 2013, **68**, 115-121.

33. A. Sahu, N. Shah, M. Mahimkar, M. Garud, S. Pagare and C. M. Krishna, 2014.

34. A. Sahu, K. Dalal, S. Naglot, P. Aggarwal and C. M. Krishna, 2013.

35. M. E. Boon and A. J. Suurmeijer, *The pap smear*, CRC Press, 1996.

36. A. Nijssen, K. Maquelin, L. F. Santos, P. J. Caspers, T. C. B. Schut, J. C. den Hollander, M. H. Neumann and G. J. Puppels, *Journal of biomedical optics*, 2007, **12**, 034004-034004-034007.

37. S. Koljenović, T. C. B. Schut, J. M. Kros, H. J. van den Berge and G. J. Puppels, *Laboratory Investigation*, 2002, **82**, 1265-1277.

38. P. Crow, B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad and N. Stone, *British journal of cancer*, 2005, **92**, 2166-2170.

39. T. C. Bakker Schut, R. Wolthuis, P. J. Caspers and G. J. Puppels, *Journal of Raman Spectroscopy*, 2002, **33**, 580-585.

40. T. Harvey, E. Gazi, A. Henderson, R. Snook, N. Clarke, M. Brown and P. Gardner, *The Analyst*, 2009, **134**, 1083-1091.

41. K. Varmuza and P. Filzmoser, *Introduction to multivariate statistical analysis in chemometrics*, CRC press, 2009.

Analytical Methods Accepted Manuscript

42. A. Ghanate, S. Kothiwale, S. Singh, D. Bertrand and C. M. Krishna, *Journal of Biomedical Optics*, 2011, **16**, 025003-025003-025009.

43. K. Papamarkakis, B. Bird, J. M. Schubert, M. Miljković, R. Wein, K. Bedrossian, N. Laver and M. Diem, *Laboratory Investigation*, 2010, **90**, 589-598.

44. M. G. van Oijen and P. J. Slootweg, *Cancer Epidemiology Biomarkers & Prevention*, 2000, **9**, 249-256.

45. R. S. Hoda and S. A. Hoda, *Fundamentals of Pap test cytology*, Springer Science & Business Media, 2007.

46. F. S. Parker, *Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry*, Springer Science & Business Media, 1983.

47. Z. Movasaghi, S. Rehman and I. U. Rehman, *Applied Spectroscopy Reviews*, 2007, **42**, 493-541.

48. N. Stone, C. Kendall, N. Shepherd, P. Crow and H. Barr, *Journal of Raman spectroscopy*, 2002, **33**, 564-573.

49. A. R. Offenbacher, J. Chen and B. A. Barry, *Journal of the American Chemical Society*, 2011, **133**, 6978-6988.

50. R. Kengne-Momo, P. Daniel, F. Lagarde, Y. Jeyachandran, J. Pilard, M. Durand-Thouand and G. Thouand, *International Journal of Spectroscopy*, 2012, **2012**.

51. Q. Hu and H. Hou, *Tobacco Smoke Exposure Biomarkers*, CRC Press, 2015.

52. D. Lin-Vien, N. B. Colthup, W. G. Fateley and J. G. Grasselli, *The handbook of infrared and Raman characteristic frequencies of organic molecules*, Elsevier, 1991.

53. S. Gunasekaran, E. Sailatha, S. Seshadri and S. Kumaresan, *Indian J Pure Appl Phys*, 2009, **47**, 12-18.

Analytical Methods Accepted Manuscript

54. Q. Matthews, A. Jirasek, J. Lum, X. Duan and A. G. Brolo, *Applied spectroscopy*, 2010, **64**, 871-887.

55. C. M. Krishna, G. D. Sockalingum, G. Kegelaer, S. Rubin, V. B. Kartha and M. Manfait, *Vibrational Spectroscopy*, 2005, **38**, 95-100.

56. D. P. Slaughter, H. W. Southwick and W. Smejkal, *Cancer*, 1953, **6**, 963-968.

57. S. Khandelwal and M. C. Solomon, *International journal of oral science*, 2010, **2**, 45.

58. M. Babshet, K. Nandimath, S. Pervatikar and V. Naikmasur, *Journal of cytology/Indian Academy of Cytologists*, 2011, **28**, 165.

59. D. M. Eddy, *Annals of internal medicine*, 1990, **113**, 214-226.

60. F. Bonnier, D. Traynor, P. Kearney, C. Clarke, P. Knief, C. Martin, J. J. O'Leary, H. J. Byrne and F. Lyng, *Analytical Methods*, 2014, **6**, 7831-7841.

61. L. F. C. Carvalho, F. Bonnier, K. O'Callaghan, J. O'Sullivan, S. Flint, H. J. Byrne and F. M. Lyng, *Experimental and molecular pathology*, 2015, **98**, 502-509.

62. D. Fu, J. Zhou, W. S. Zhu, P. W. Manley, Y. K. Wang, T. Hood, A. Wylie and X. S. Xie, *Nature chemistry*, 2014, **6**, 614-622.

63. M. Bonn, M. Muller, H. A. Rinia and K. N. Burger, *Journal of Raman Spectroscopy*, 2009, **40**, 763-769.

64. S. Verrier, I. Notingher, J. Polak and L. Hench, *Biopolymers*, 2004, **74**, 157-162.

65. I. Notingher, S. Verrier, S. Haque, J. Polak and L. Hench, *Biopolymers*, 2003, **72**, 230-240.

66. N. Kunapareddy, J. P. Freyer and J. R. Mourant, *Journal of biomedical optics*, 2008, **13**, 054002-054002-054009.

Analytical Methods Accepted Manuscript

67. K. W. Short, S. Carpenter, J. P. Freyer and J. R. Mourant, *Biophysical journal*, 2005, **88**, 4274-4288.

Analytical Methods Accepted Manuscript

Table 1. Demographic and subject accrual data

Sr. no.	Category	No. of subjects	Age range/median age	Tobacco habits	Total no. of spectra acquired
1.	No-habit control	15	29-55/ 44 years	No	89
2.	Habit control	15	20-63/ 34 years	Yes	100
3.	Tumor	20	27-70/ 53 years	Yes	116
4.	Contralateral	20	27-70/ 53 years	Yes	86

Analytical Methods Accepted Manuscript

Table 2. PC-LDA for exfoliated cells from healthy volunteers, habit controls, disease control and tumor using point-spectra approach, a) Confusion matrix of PC-LDA, b) Confusion matrix of LOOCV

	HEALTHY VOLUNTEER	HABIT CONTROL	DISEASE CONTROL	TUMOR	TOTAL
HEALTHY VOLUNTEER	69	12	4	4	89
HABIT CONTROL	21	68	8	3	100
DISEASE CONTROL	7	6	59	14	86
TUMOR	9	5	33	69	116

(a)

	HEALTHY VOLUNTEER	HABIT CONTROL	DISEASE CONTROL	TUMOR	TOTAL
HEALTHY VOLUNTEER	69	12	4	4	89
HABIT CONTROL	21	68	8	3	100
DISEASE CONTROL	7	6	58	15	86
TUMOR	9	5	34	68	116

(b)

Analytical Methods Accepted Manuscript

Table 3. PC-LDA for exfoliated cells from healthy volunteers, habit controls, disease control and tumor using patient-wise approach, a) Confusion matrix of PC-LDA, b) Confusion matrix of LOOCV

	HEALTHY VOLUNTEER	HABIT CONTROL	DISEASE CONTROL	TUMOR	TOTAL
HEALTHY VOLUNTEER	13	1	1	0	16
HABIT CONTROL	2	12	1	0	15
DISEASE CONTROL	2	1	11	2	15
TUMOR	1	2	4	12	19

(a)

	HEALTHY VOLUNTEER	HABIT CONTROL	DISEASE CONTROL	TUMOR	TOTAL
HEALTHY VOLUNTEER	13	1	1	0	16
HABIT CONTROL	2	11	2	0	15
DISEASE CONTROL	2	3	9	2	15
TUMOR	1	2	4	12	19

(b)

Analytical Methods Accepted Manuscript

Figure Legends:

Figure 1. Protocol for Pap staining

Figure 2. Pap stained cytological smears (200X) from- a) Healthy volunteers showing cells from superficial and intermediate layers with occasional presence of orangeophilic cells, b) Habit-controls showing increased number of orangeophilic cells and occassional parakeratotic cells and anucleate squames , c) Disease control or contralateral sites from cancer patients showing increased number of orangeophilic cells with few parakeratotic cells and anucleate squames, and occasional cells with minimally increased nuclear:cytoplasm ratio (indicative of mild dysplasia), d) Tumor showing dysplastic cells with high nuclear :cytoplasm ratio and prominent nucleolus, multinucleation with dirty background (indicative of necrosis) and higher frequency of anucleate squames

Figure 3. Mean Raman spectra of all groups
a) Healthy volunteers, b) Habit-controls, c) disease control, d) Tumor

Figure 4. PCA for healthy volunteers, habit controls, disease control and tumor using point-spectra approach, a) Loadings of factor 1, b) Loadings of factor 3, c) Scatter plot.

Figure 5. PC-LDA for healthy volunteers, habit controls, disease control and tumor using point-spectra approach, Scree plot b) Scatter plot

Figure 6. PCA for healthy volunteers, habit controls, disease control and tumor using patient-wise approach, a) Loadings of factor 2 b) Loadings of factor 4 c) PCA scatter plot

Figure 7. PC-LDA for healthy volunteers, habit controls, disease control and tumor using patient-wise approach, Scree plot b) Scatter plot

Analytical Methods Accepted Manuscript

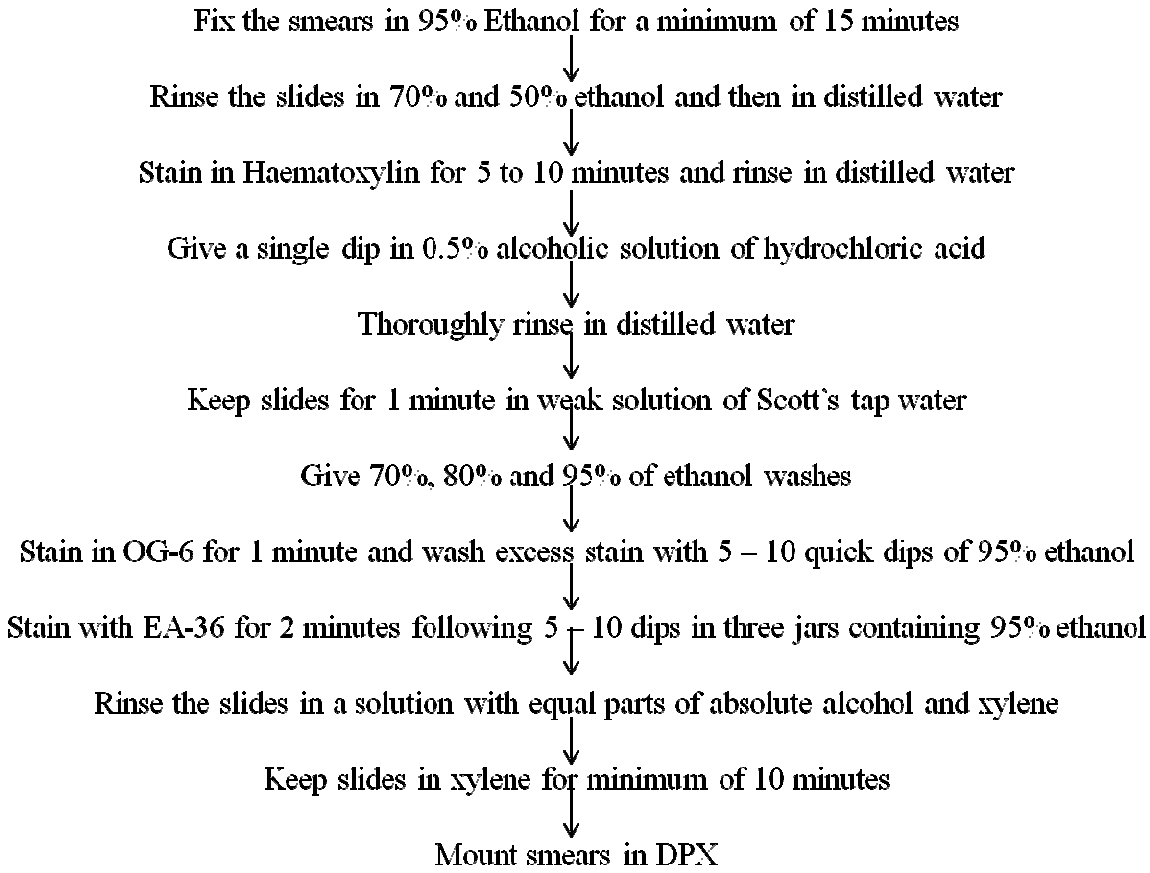


Figure 1

Analytical Methods Accepted Manuscript

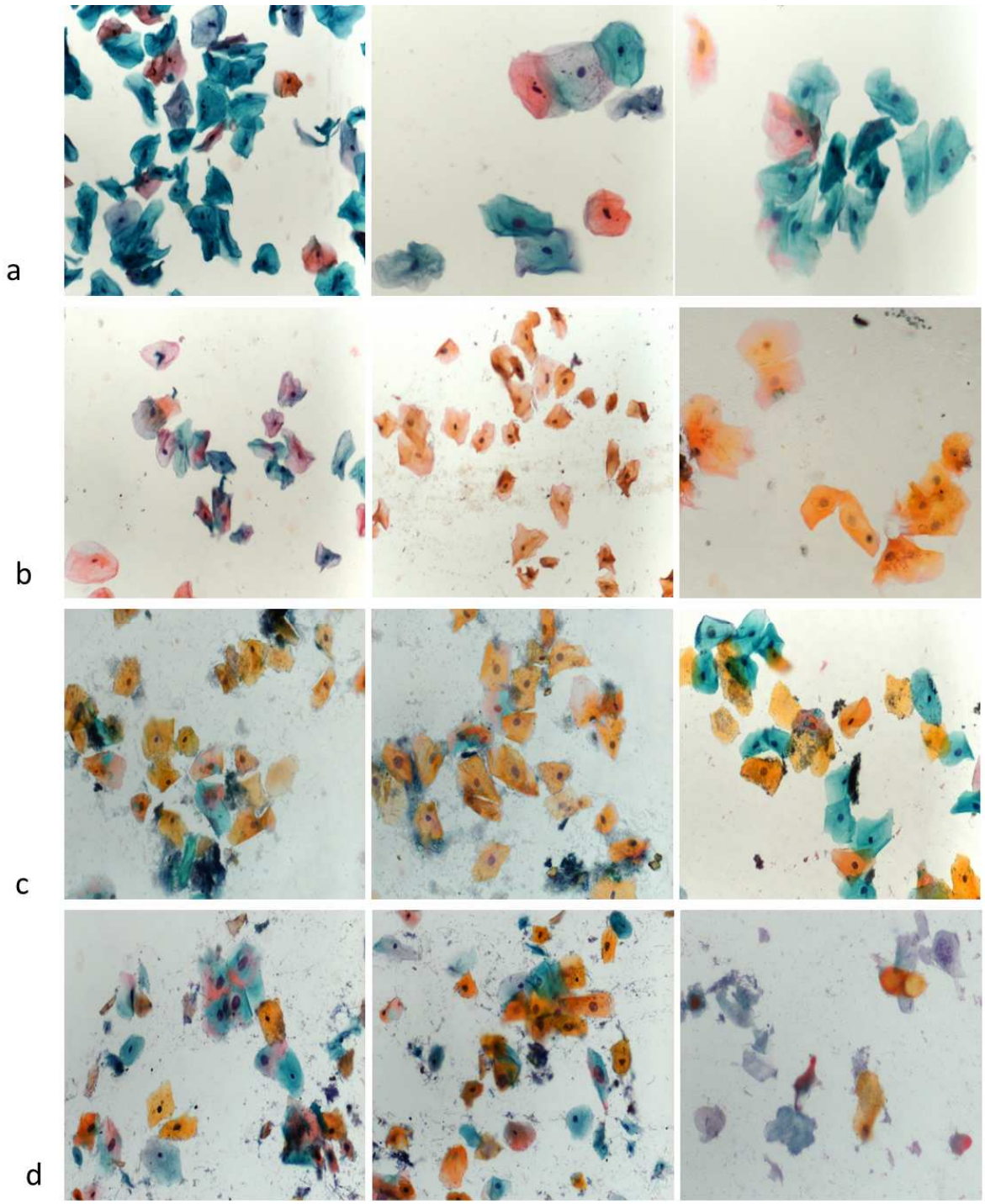


Figure 2

Analytical Methods Accepted Manuscript

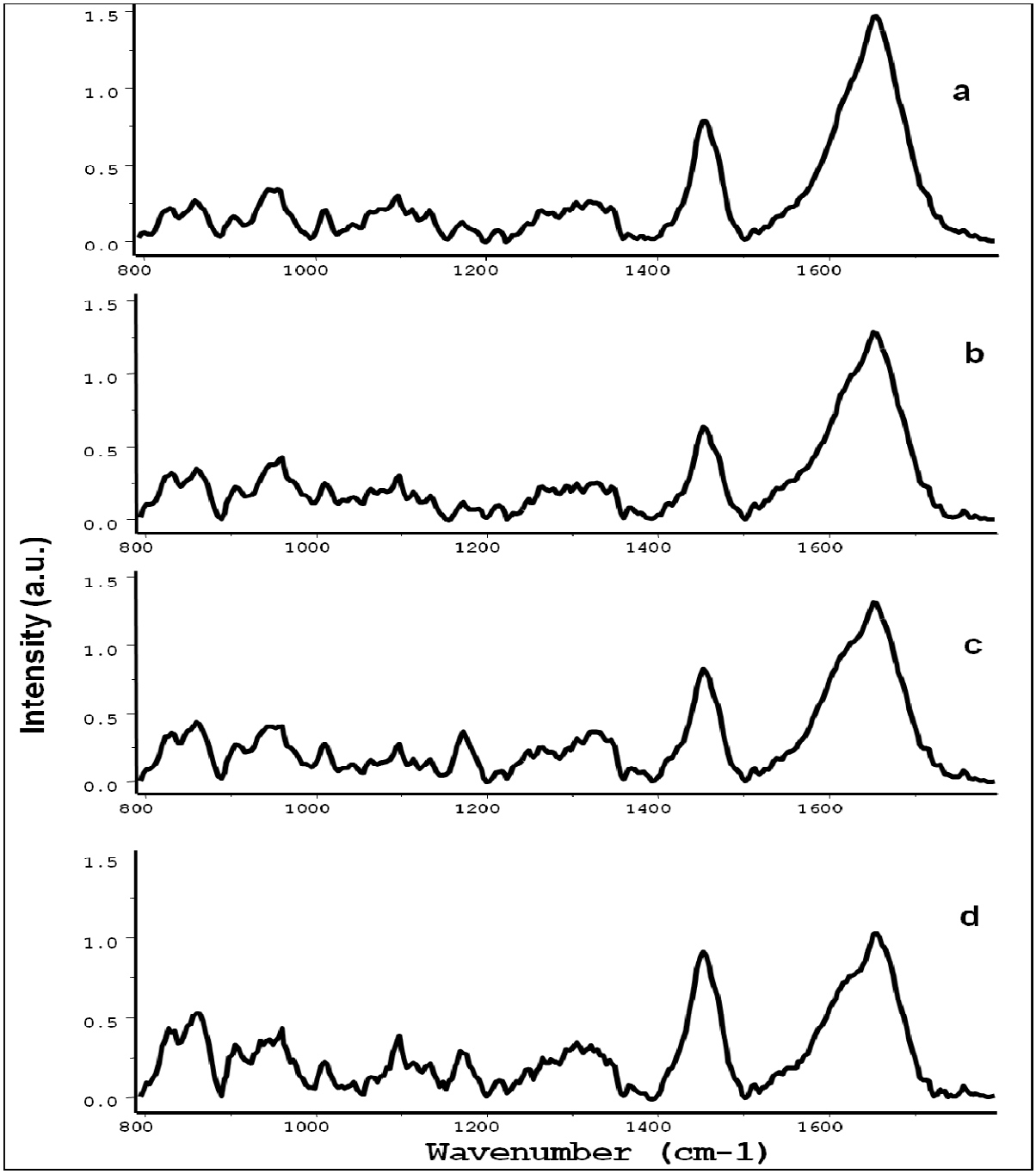


Figure 3

Analytical Methods Accepted Manuscript

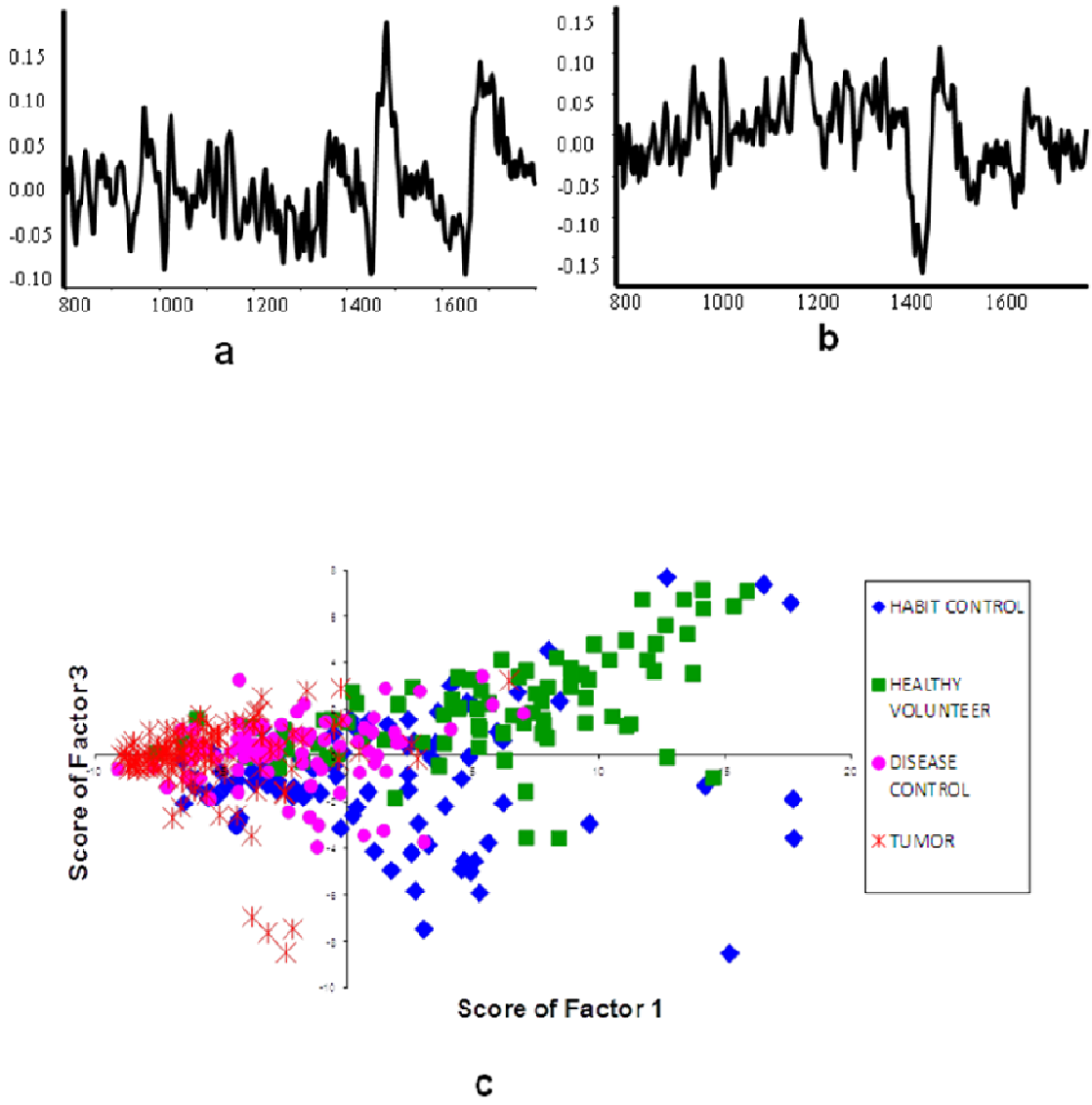


Figure 4

Analytical Methods Accepted Manuscript

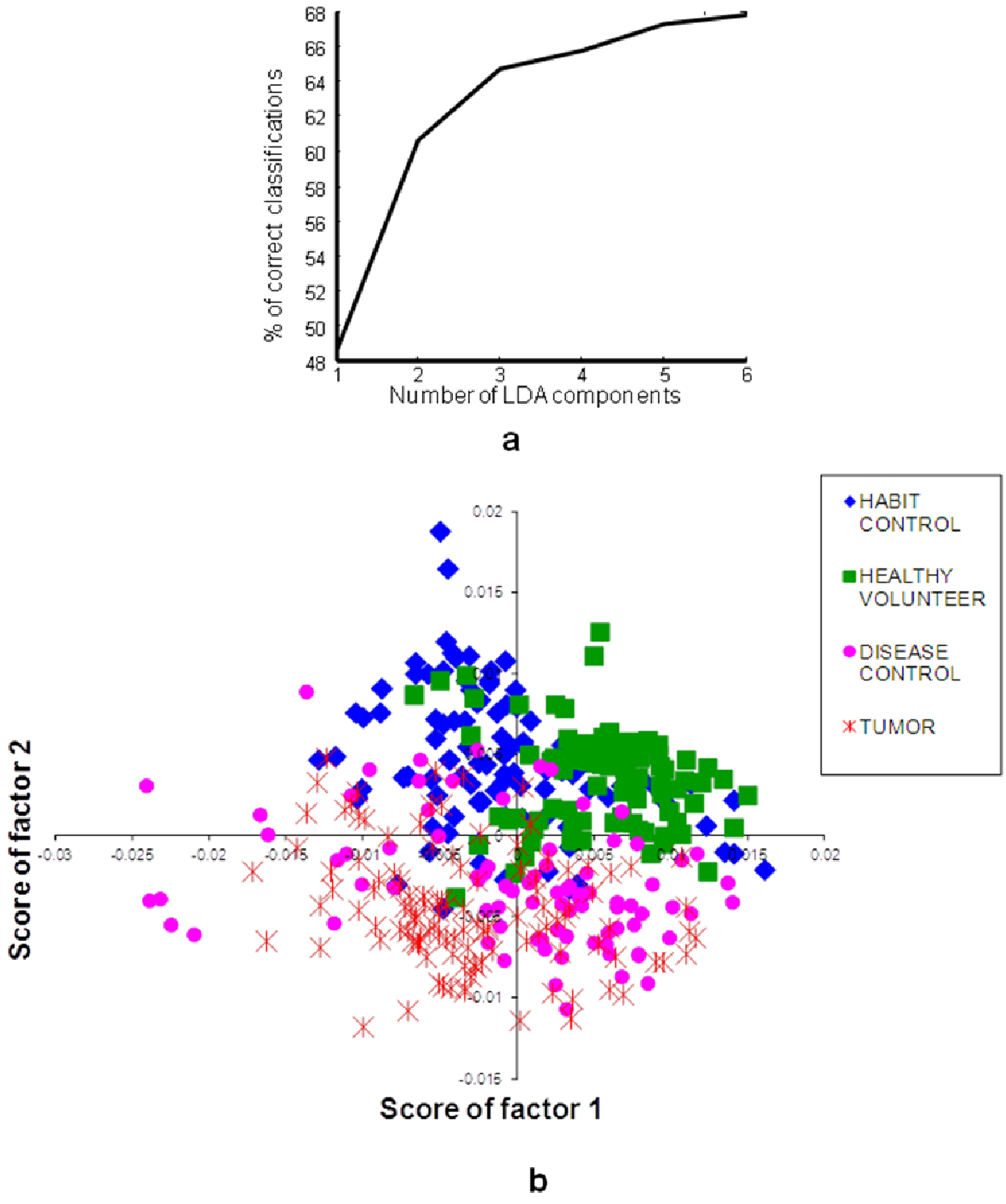
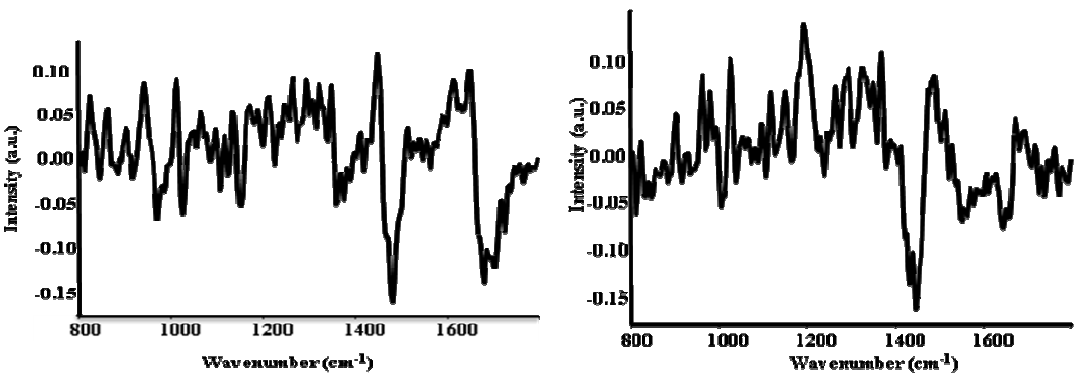
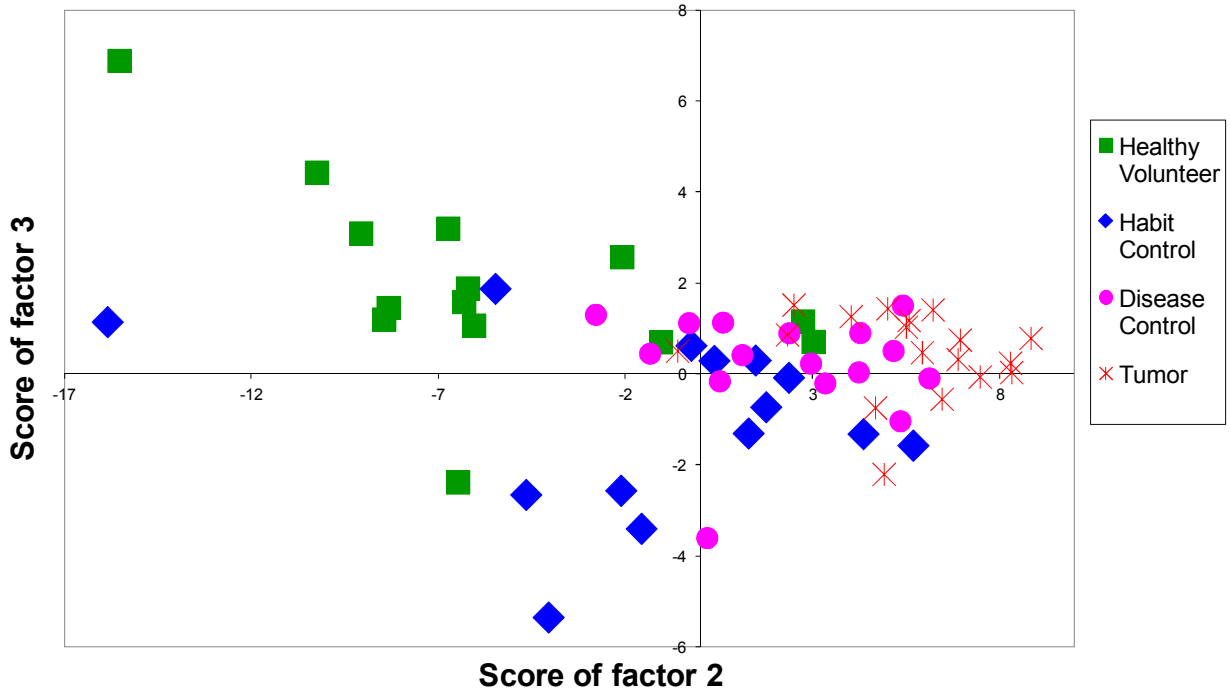


Figure 5

Analytical Methods Accepted Manuscript



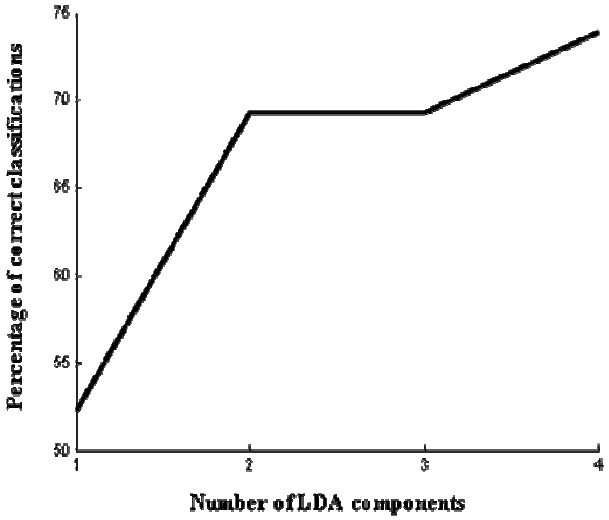
(a) (b)



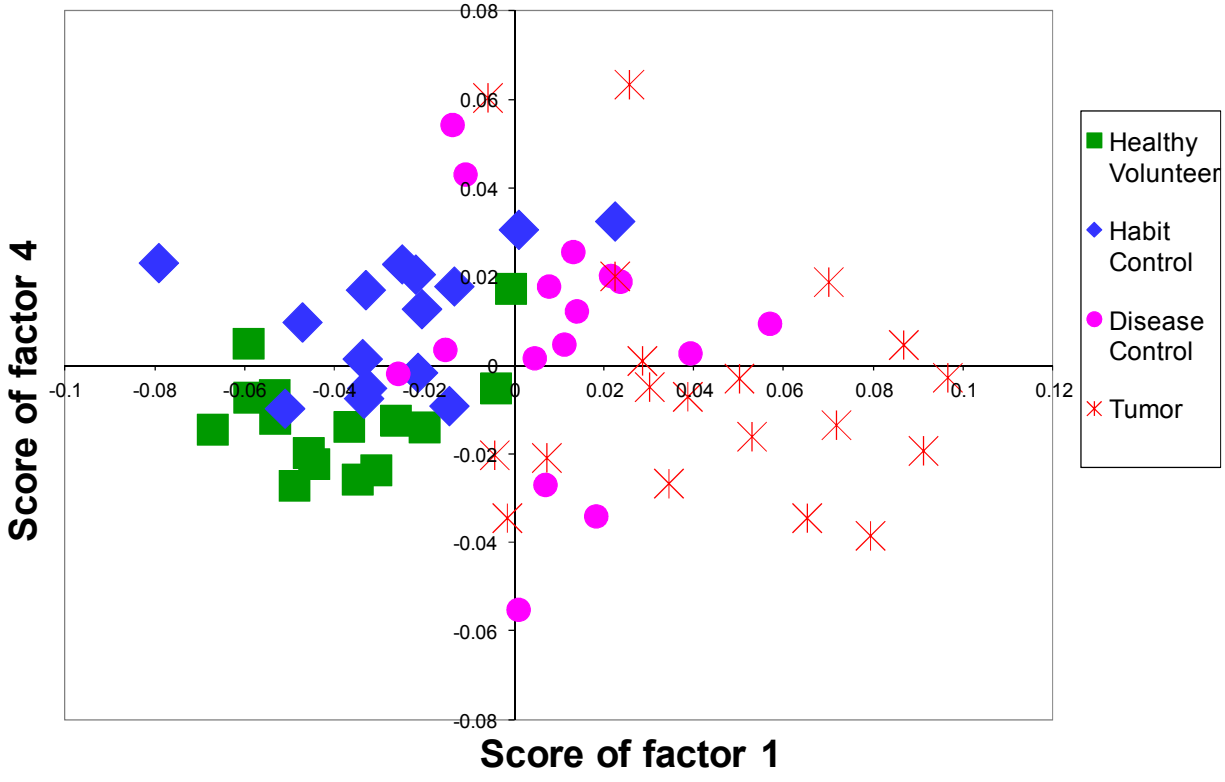
(c)

Figure 6

Analytical Methods Accepted Manuscript



(a)



(b)

Figure 7

Analytical Methods Accepted Manuscript