2012-12

Raman Micro Spectroscopy for the Characterisation of Cervical Cancer.

Nosheen Rashid
Technological University Dublin

Recommended Citation

Rashid, N. (2012). *Raman micro spectroscopy for the characterisation of cervical cancer*. Doctoral Thesis. Technological University Dublin. doi:10.21427/D7DS30

Raman micro spectroscopy for the characterisation of cervical cancer

Submitted by

Nosheen Rashid

M.Phil (Biotechnology)

For the award of Doctor of Philosophy

Radiation and Environmental Science Centre
(RESC)/ School of Physics
Focas Research Institute, Dublin Institute of
Technology

Supervisors:
Prof. Dr. Fiona M. Lyng and Prof. Dr. Hugh J. Byrne

December, 2012

Abstract

The primary purpose of this study is to evaluate the potential of FTIR and Confocal Raman micro spectroscopy (CRM) in elucidating the biochemical changes occurring in different layers of the cervical epithelium including basal, superficial and the underlying connective tissue known as stroma during the progression of Cervical Intraepithelial Neoplasia (CIN) to cancer. Initially the two techniques were compared and Raman was chosen based on its higher spatial and spectral resolution. The sample preparation and spectral measurement procedures were optimised and all samples were formalin fixed paraffin processed, dewaxed using xylene, and measured on calcium fluoride windows. Raman spectra were recorded using a source wavelength of 785nm and a X100 dry objective lens. Raman micro spectroscopy was able to differentiate the normal region of the cervical tissue sample into three layers including stroma, basal/para-basal and superficial layers on the basis of the spectral features of the collagen, DNA bases and glycogen as well as discrimination of the diseased areas from the normal areas. In particular Raman spectroscopy could describe the biochemical changes in the diseased samples in detail. On moving from normal to abnormal regions of the cervical tissue sample, the characteristic Raman features of the basal layer were observed in the superficial layer and in the stroma. Notably, the normal region of a CIN III sample was found to have biochemical information similar to the abnormal region. This has been indicated by the absence of the collagen Raman spectral bands in the stromal layer as well as absence or minute presence of the glycogen bands in the superficial layer. A comparison of the Principal Components Analysis (PCA) loadings of the HPV negative and positive cell lines (*C33A* and *CaSki*), with those of the basal layers of normal and abnormal tissue samples showed no strongly matching Raman signatures which could lead to the identification of signatures of HPV infection in the cervical cancer tissue samples.

However, a feature associated with the amide-III beta sheet (1222 cm^{-1}) was found to be consistently present in the PCA loadings contributed by the basal layers of the intermediate and abnormal samples. This was much reduced in intensity in the most extreme abnormal sample and the carcinoma *in situ* samples. This feature may be considered as an early marker of disease progression, but further investigation is needed to confirm this finding. KMCA indicated the possibility of the migration of basal cells into the superficial and stromal layers and subsequent PCA led to the conclusion that basal cells indicated by high DNA content and lack of glycogen are progressing to the superficial layer due to the progression of the disease. In the case of stroma, the basal like characteristics are actually associated with the biochemical changes in the stromal cells and there is no migration of these cells into the stroma. In addition, during cervical cancer progression, relative to the DNA, collagen has a diminished contribution at some points in the Raman map of the stroma and KMCA recognised the greater similarities with the DNA rich cells of the basal layer. This has been supported by the enhanced expression of p16 protein in the basal and superficial layers rather than in the stroma.

Declaration

I certify that this thesis which I now submit for examination for the doctoral thesis is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for another award in any Institute.

The work reported on in this thesis conforms to the principles and requirements of the institute's guidelines for ethics in research.

Signature __________________ Date ________________

Candidate: Nosheen Rashid

Acknowledgment

All thanks to Allah, the God, the most gracious and merciful who is the creator of everything and who created the opportunity for me to learn the knowledge of the science and excel.

I would like to take this opportunity to mention a number of people who have helped me in many different ways towards the submission of this work.

It’s my pleasure to thank my supervisors Prof. Dr. Fiona M. Lyng and Prof. Dr. Hugh. J. Byrne for their continual advice and encouragement right from the start day of my PhD and throughout the time it has taken me to finish this report.

I would like to thank Dr. Kelvin. W. C. Poon, who provided me valuable assistance many times.

Many thanks are due to Dr. Luke O’ Neil and Anne Shanahan for providing assistance regarding the handling of the Instruments.

Particular thanks are due to Jawad Sarwar, and Syed Mahmood Ali who provided moral support and valuable suggestions and help in a number of ways.

I would like to show my gratitude to my sisters Summera, Dr. Lubna, Dr. Ayesha, my brothers Shah Nawaz, Zeeshan Haidar, all of them are very dear to me, whose prayers are always with me.

This work would not have been possible without the encouragement and support which I have been gathering every time my husband Dr. Haq nawaz and my son Maarij Nawaz, “Thank you” to them for just about everything.

These acknowledgements would not be complete without mentioning my collective source of constant support and encouragement: Abbu Jee, Ammi Jee, Amma and Abba (in laws), their prayers are really a blessing of the God, the gracious.

Abbreviations

°C	degree centigrade
A	adenine (A)
Bend	Bending
Brth.	breathing
C	cytosine (C)
cm^{-1}	per centimeter
CRM	Confocal Raman Microspectroscopy
def	deformation
Deoxyrib	deoxyribose
FTIR	Fourier Transform Infra red
nm	nanometer
O-P-O	Phosphodiester
PC1	1st principal component
PC2	2nd principal component
PCA	Principal Component Analysis
PO_4 group	Phosphate group
Str	stretching
T	thymine (T)
KMCA	K-means cluster analysis
ICA	Independent component analysis
NCLS	Negatively constrained non linear analysis
CIN-I	Cervical intraepithelial neoplasia-I
CIN-II	Cervical intraepithelial neoplasia-II

CIN-III	Cervical intraepithelial neoplasia-III
HPV	Human papilloma virus
Pap	Papanicolaou

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cervical cancer is the third most common cancer in women worldwide, 85% of cases arising in the developing world (Jemal et al., 2011). Mortality associated with cervical cancer can however be significantly reduced if the disease is detected at the early stages of its development or at the pre-malignant state (cervical intraepithelial neoplasia, CIN) (Parkin et al., 2005).

Current methods for identifying neoplastic cells and differentiating them from their normal counterparts are often nonspecific, slow, invasive, or a combination thereof (Chan et al., 2006). The primary screening tool for cervical neoplasia is the Papanicolaou (Pap) smear test, which involves the microscopic examination of exfoliated cells for morphological abnormalities (Papanicolaou and Traut, 1941). The colposcopy is a widely used tool, which usually follows an abnormal Pap smear. The incidences of the disease and associated mortality have been falling in developed countries, where these screening methods are widely practiced. However, these methods have several shortcomings, including high false negative/positive results that could be due to the subjective interpretations of the cytologist/pathologist diagnosing the disease based on morphological abnormalities. There are difficulties in separation of normal, basal cell hyperplasia, immature squamous metaplasia and inflammation associated changes from true koilocytes, which are indicative of low grade squamous metaplasia (McCluggage et al., 1998). There are also difficulties in differentiating low grade squamous intraepithelial lesions from immature metaplastic squamous with atypia, mild abnormality, resulting in inter- and intra- observer difference of opinion (McCluggage et al., 1998).

It is estimated that 90% of deaths can be prevented by early diagnosis, however (Bazant-Hegemark et al., 2008), which emphasizes the need for effective screening methods. Optical methods such as Confocal Microscopy and Optical Coherence Tomography are potential techniques that can be used to visualise the neoplastic changes more sensitively. These tools are emerging as non-destructive, real-time imaging modalities for early diagnosis of precancerous lesions of the cervix (Bazant-Hegemark et al., 2008). Potentially offering higher sensitivity and specificity based on biomolecular analyses, spectroscopic methods such as IR absorption and Raman scattering are attracting increased attention. The suitability of these techniques, particularly Raman spectroscopy, for rapid noninvasive screening of cervical biopsies will be explored in this work.

1.1 Cervix

The cervix is the lower fiberomuscular portion of the uterus. The upper portion adjoining the uterus is known as the endocervix and the lower end adjoining the vagina is known as the exo cervix. The lining of the endocervix is covered by a single layer of tall columnar cells with dark staining nuclei forming the columnar epithelium (**Figure 1.1**). The exocervix, on the other hand, is covered by stratified, glycogen containing squamous epithelium. The point where the two epithelial layers meet is known as the transformation zone (Sellors and Sankaranarayanan, 2003/4). Initially, the epithelium at the transformation zone is comprised of a single layer of mucus secreting columnar epithelium, but the hormonal changes at the onset of puberty make the environment for the exposed epithelium more acidic, causing it to change from a single layer epithelium to a more protective, stratified squamous epithelium.

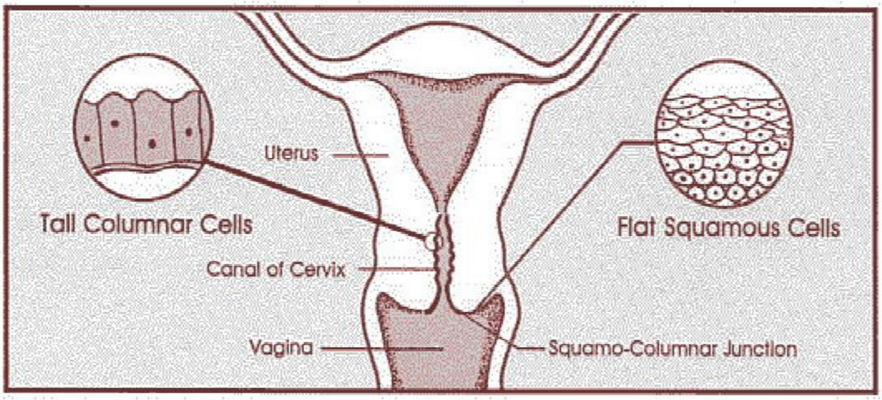


Figure 1.1 Diagrammatical representation of human cervix.
(http://www.gynalternatives.com/structur.htm)

1.2 Cervical Cancer

Cervical cancer is the major cause of gynaecologic-related death worldwide, but is absolutely avoidable with regular cytological screening (Shepherd and Bryson, 2008 Nov). Cervical cancer impacts the lives of women worldwide. Recent reports on cervical cancer indicate that 275,100 women died of cervical cancer in 2008 and 85% of these deaths occurred in developing countries (Jemal et al., 2011). Cervical cancer is only the eleventh most common cancer among women in developed countries, however, due to existing screening practices. In the absence of such established screening procedures, it becomes the second most common cancer (Peto et al., 2004). Most of the precancerous and cancerous lesions arise in the transformation zone, which is the area where metaplasia occurs at the onset of puberty (Cuzick et al., 2008).

1.2.1 Types of cervical cancer

Cervical cancer can be divided into the following types:

1.2.1.1 Squamous cell carcinoma

Squamous cell carcinoma is the most common carcinoma of the cervix (about 90%) and it originates from the squamous epithelium of the exocervix. It is sub divided into three classes:

1. Large cell keratinising tumours
2. Large cell non keratinising tumours
3. Small cell non keratinising tumours

1.2.1.2 Adenocarcinoma

Adenocarcinoma is a type of cancer which arises from the glandular epithelium. Although it is rarely found in the cervix, it develops in the endocervix. It accounts for 10% of cervical cancers and primarily affects women over the age of 60 years (Mera, 1997). Different reports suggest that there has been a recent increase of up to 15% in its occurrence, especially in young women who take oral contraceptives (Quinn, 1998).

Mixed adenosquamous tumours are rare (about 5%). Other malignant tumours such as lymphomas, melanoma, and sarcomas can be diagnosed but are rare (Mera, 1997).

1.3 Risk factors

The risk factors for cervical cancer according to epidemiological studies are discussed in detail below:

1.3.1 Sexual activity

The risk relates to the number of sexual partners for either the woman or her partner. The greater the number of sexual contacts, the greater can be the risk involved (Mera, 1997).

The age of first sexual intercourse is not a risk factor in itself, other than the likelihood that initiation at an earlier age may result in the accumulation of more sexual partners. Most of the other risk factors are associated with this factor.

1.3.2 Infection with Human papilloma virus (HPV)

Infection with the human papilloma virus is the most commonly cited cause of cervical cancer, occurring in 99.7% cases, especially with types 16 and 18 of the virus (Bosch et al., 1995, Walboomers et al., 1999). Over 100 types of human papilloma virus have been identified, of which 40 are known to infect the genital tract. Some of them, such as type 6 and 11, cause only warts and are known as low risk strains, while others such as type 16 and 18 can lead to cervical precancerous states and are thus classed as high risk strains. Type 16 followed by 18 are the most frequently detected types of HPV at the time of diagnosis of squamous cell carcinoma (SCC) (Woodman et al., 2007). Other high risk HPV types include 31,33,35,39,45,51,52,56,58,59,68,73 and 82 and the low risk types include 6,11,40,42,4344,54,61,70,72 and 81 (Villa and Denny, 2006).The integration rate of the HPV 16 genome with that of the infected basal cells increases with severity of cervical neoplasia. However, in some women with invasive disease, only episomal forms, genetic material existing as a free autonomously replicating DNA, are identified (Woodman et al., 2007). On the other hand, HPV 18 is found in an integrated form in most of the cases with high grade cervical intraepithelial neoplasia or invasive disease (Woodman et al., 2007). Infection with HPV 18 is considered to be the cause of adenocarcinoma, which is only 10% of cervical cancer (Woodman et al., 2007).

Pathological examination of cervix cone biopsies reveals that, in many cases, more than one grade of cancer can co-exist in the same sample and this is considered to be due to the infection of the cervix by different strains of HPV and not due to the different stages of evolution in a continuing process that is caused by a single viral type infection (Agorastos et al., 2005).

1.3.3 Irregular cervical cytological screening

Women who do not undergo regular cervical cytological screening may have four times more risk of developing cervical cancer than those who are regularly screened (Mera, 1997).

1.3.4 Cigarette smoking

Cigarette smokers have higher risk of developing cervical cancer than non smokers (Mera, 1997). The metabolites of tobacco smoke have been found in the cervical mucus of smokers. The presence of constituents of tobacco smoke in the cervix may favour infection by the human papilloma virus.(HPV) The DNA of the cells in the cervix forms adducts with some metabolites of tobacco and can act as a tumour promoter after infection with HPV (Collins et al., 2010).

1.3.5 Use of oral contraceptives

The use of oral contraceptives increases the risk of cervical cancer, whereas use of barrier methods of contraception are associated with decreased risk and the reason for these methods being safe is that they do not allow seminal fluid to come in contact with the cervix (Mera, 1997). Prolonged use of oral contraceptives increases the risk of cervical cancer in HPV positive women (Smith et al., 2003).

1.4 Classification system

Cervical cancer arises after progressing through a long phase of pre-invasive disease. This is characterised microscopically as a sequence of events progressing from cellular atypia to various grades of dysplasia or cervical intraepithelial neoplasia (CIN) before progression to invasive cancer (Richart, 1968). It has been observed that some cases of dysplasia regress, while some persist, some progressing to carcinoma *in situ* (CIS). The concept of cervical cancer precursors dates back to the late nineteenth century, when non invasive atypical changes were found in tissue specimens adjoining invasive cancer (Sellors and Sankaranarayanan, 2003/4). The term carcinoma *in situ* was developed in 1932 to represent those lesions which had undifferentiated carcinoma cells in the full thickness of the epithelium without disturbing the basement membrane (Sellors and Sankaranarayanan, 2003/4). A relationship between CIS and invasive cervical cancer was subsequently reported (Reagan et al., 1953). In late 1950, the term Dysplasia was coined, to indicate cervical epithelial atypia, a missing step between normal epithelium and CIS (Reagan et al., 1953). Dysplasia was further classified into three groups, mild, moderate and severe, depending on the levels of atypical cells in the epithelium. As a result, cervical precancerous lesions were described using the categories of dysplasia and CIS for many years. A system that separates the classification of dysplasia from CIS was increasingly accepted on the basis of findings from a number of follow up studies involving women with such lesions (Sellors and Sankaranarayanan, 2003/4). A direct relationship between progression of disease and histological grade of tissue led to the concept of a single continuous disease process (Sellors and Sankaranarayanan, 2003/4). This continuous process leads the normal epithelium through various grades of dysplasia into invasive cancer. On this basis, the term cervical intraepithelial neoplasia (CIN) was developed by Richart in 1968 for cellular atypia confined to the epithelium (Richart, 1968). It classified mild dysplasia as CIN I, moderate as CIN II, with CIN III describing severe dysplasia to carcinoma *in situ*.

HPV was recognised, in 1980, as an infectious agent responsible for pathological changes such as koilocytic atypia (Richart, 1990), which means cells with irregular cytoplasm or a halo; koilocytes are cells with a luminous circle in the cytoplasm (Sellors and Sankaranarayanan, 2003/4).

The hypothesis that cervical cancer may arise from infections with HPV was made by Hausen (Hausen et al., 1974a, Hausen et al., 1974b, Zurhausen, 1976). The isolation of HPV 16 and later HPV 18 and characterisation as infectious agents responsible for cervical cancer (Boshart et al., 1984, Durst et al., 1983) and a detailed study of cervical dysplasia suggesting that its development is not a continuous event, as only 5% of CIN II and less than 12% of CIN III progress to invasive cancer if left untreated (Pinto and Crum, 2000), paved the way for a creation of the Bethesda system of classification. The Bethesda system (TBS) of classification was created in 1988 and revised in 1991 in a workshop organised by the US National Cancer Institute (Jastreboff and Cymet, 2002). This system classifies cervical intraepithelial neoplasia according to the following categories: atypical squamous cells of undetermined significance (ASCUS) refers to squamous atypia, low grade squamous intraepithelial lesion (LSIL) applies to lesions with condylomatous (a skin tumour near the anus or genital organs) atypia, mild dysplasia or CIN I and high grade squamous intraepithelial lesion (HSIL) applies to lesions with moderate to severe or CIN II and CIN III lesions and CIS (Moscicki, 1999, Kaufman and Adam, 1999, Sigurdsson, 1999). TBS was re-evaluated and revised in a 2001 workshop organized by the US National Cancer Institute, at which more than 20 countries were represented (Solomon et al., 2002). The revisons of the 2001 TBS included specific statements about specimen adequacy, general categorization, and interpretation and results. In the adequacy category, “satisfactory” and “unsatisfactory” are retained, but “satisfactory but limited by” is eliminated. The new category of “atypical squamous cells” (ASC) is replaced by the category of “atypical squamous cells of undetermined significance” (ASCUS) while dividing it into qualifiers of (a) ASC of “undetermined significance” (ASC-US) and (b) “cannot exclude high-grade squamous intraepithelial lesion (HSIL),” or (ASC-H). The categories of ASCUS, “favor reactive” and “favor neoplasia” were eliminated. The terminology for low-grade squamous intraepithelial lesions (LSILs) and HSILs remained the same as previous. The category of “atypical glandular cells of undetermined significance” (AGUS) is eliminated to avoid confusion with ASCUS and replaced by the term “atypical glandular cells” (AGC), in an effort to identify whether the origin of the cells was endometrial, endocervical, or unqualified. “Endocervical adenocarcinoma *in situ*” and “AGC, favor neoplastic” were included as separate AGC categories (Apgar et al., 2003).

1.5 Precancerous stages

The normal squamous stratified epithelium in an adult cervical transformation zone has three different types of cell layers, including a single layer of basal cells at the bottom, several layers of round polygonal cells forming an intermediate layer and the flattened epithelium at the surface. A haematoxylin and eosin (H&E) stained tissue section of normal cervical epithelium is shown in **Figure 1.2**.

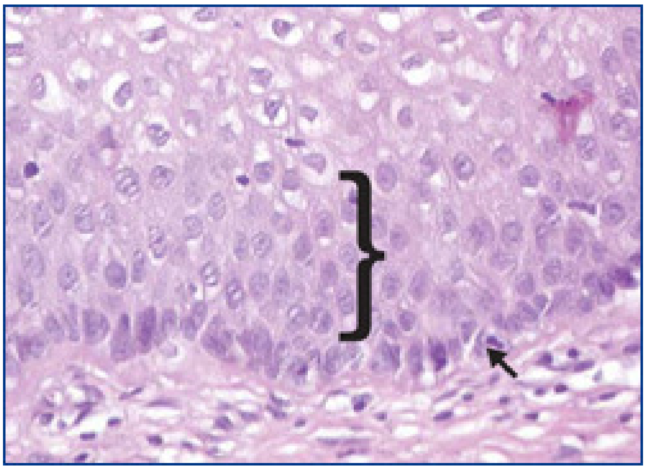


Figure 1.2 Hematoxylin and Eosin (H&E) stained normal stratified squamous epithelium basal and parabasal cell layer. The bracket and arrow show parabasal and basal cell layers respectively.

(http://www.mtmlabs.com/us/index.php?option=com_content&view=article&id=107&Itemid=29)

1.6 Cytological Abnormalities

For the past 50 years, the primary method used for the screening for pre-cancerous lesions in asymptomatic women in the general female population has been the Papanicolaou test (Pap smear). The method was developed by George Papanicolaou, before HPV was implicated in cervical cancer, and boasts a specificity of 95-98% but a sensitivity of lower than 50% (Nanda et al., 2000). In another study where a comparison between the HIV-infected and HIV-uninfected women was made for the diagnostic characteristics of Pap testing, the sensitivity was found to be 86.7% in the HIV-infected group and 94.4% in the HIV-uninfected group ($P= 0.70$). The specificity was 65.8% versus 88.3% ($P= 0.01$) respectively (Anderson et al., 2006). The Pap smear is the microscopic examination of the exfoliated cells of the cervix for morphological abnormalities (Ronco et al., 2006). Abnormal Pap smears are normally followed by colposcopy or histological tests, but the decision still remains in the hands of a pathologist, rendering the interpretation subjective (Papanicolaou and Traut, 1997).

Cellular changes, symptomatic of disease, can be detected in such cytological smears. Cytological smears are classified on the basis of shape and size of the cell with special emphasis on shape and size and staining properties of the cell nucleus. The changes in the structure of nuclei are labelled as dyskaryosis. Dyskaryosis may be classified as mild, moderate and severe, corresponding, as mentioned previously, to CIN I, CIN II and CIN III respectively. Cytology also has the ability to identify invasive cancer and any minute change in the cells termed as borderline changes. Smears do not, however, allow an estimate of the extent of epithelial involvement and the state of the underlying stromal tissue. Overall, cytology tends to underestimate the severity of abnormality (Mera, 1997).

1.7 Histological abnormalities during dysplasia

While progressing towards invasive cancer, a chain of morphological changes can be seen in the epithelium of the cervix (**Figure 1.3**). These changes involve individual cells and the thickness of the abnormal epithelium. The dysplastic changes or grades of CIN are the terms used to define these changes. These changes involve the cell development and maturation and can be seen as atypical maturation of different epithelial cell layers, abnormal keratinisation and nuclear and cytoplasmic changes (Mera, 1997).

An expansion in the layers of epithelium due to the relatively undifferentiated cells in the basal and parabasal cell layers occurs so that they constitute the major portion of the epithelium. The precancerous conditions of the cervix are always confined to the epithelium of the cervix and cause no harm to the underlying stroma (Mera, 1997). The three grade CIN system can be used to describe effects on cell differentiation and the thickness of epithelium affected.

1.7.1 CIN I

In this grade of cervical pre-cancer, the cells in the basal layer of the epithelium have larger nuclei than their normal counter parts. There is an increase in nucleus to cytoplasm ratio and the nuclei are pleomorphic (exhibit a variety of shapes). These types of cells occupy one third of the epithelial layer and the upper two thirds remains as normal stratified squamous epithelium.

1.7.2 CIN II

In the CIN II grade, the abnormal cells occupy the lower half of the epithelium. The upper half still shows normal stratification and maturation. A H&E stained CIN II cervical epithelium is shown in **Figure 1.3**.

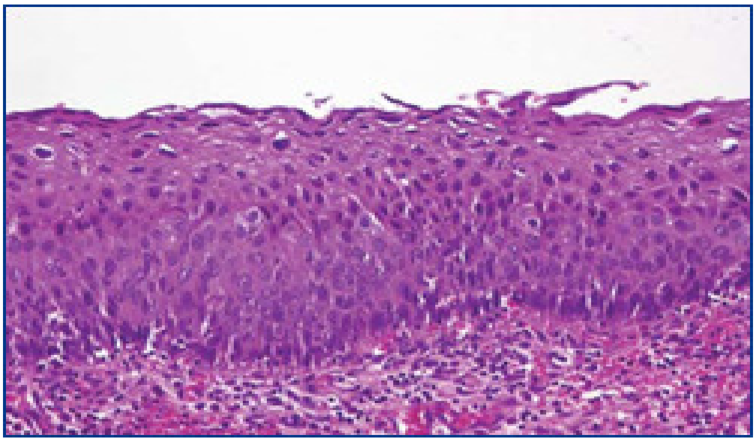


Figure 1.3 Hematoxylin and Eosin (H&E) stained CIN II cervical epithelium.

(http://www.mtmlabs.com/us/index.php?option=com_content&view=article&id=107&Itemid=29)

1.7.3 CIN III

In the CIN III grade, the whole epithelium is covered by abnormal cells, leaving only the upper one or two layers as normal epithelium. In this grade, the cells contain large dark staining nuclei. These are the most severe changes possible before invasive carcinoma. A H&E stained CIN III cervical epithelium is shown in **Figure 1.4**.

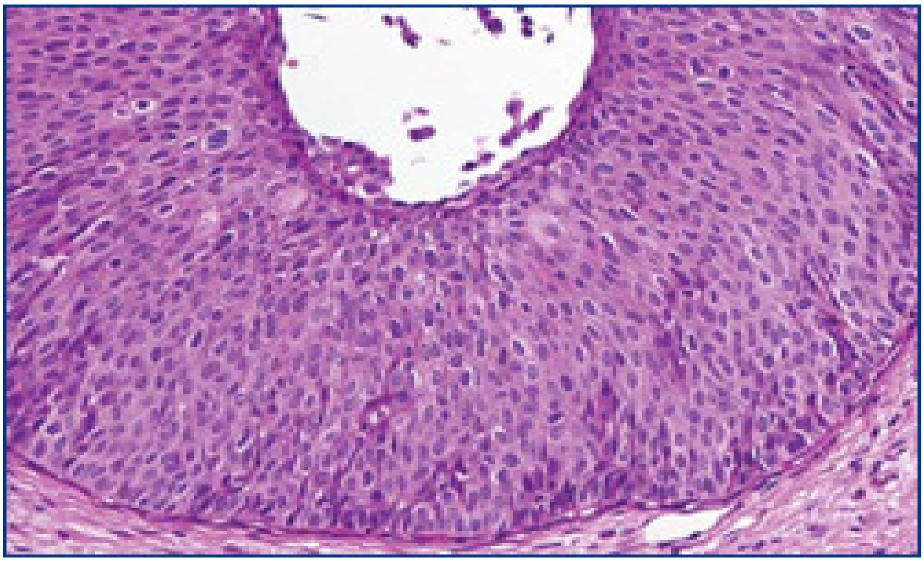


Figure 1.4 Hematoxylin and Eosin (H&E) stained CIN III cervical epithelium. (http://www.mtmlabs.com/us/index.php?option=com_content&view=article&id=107&Itemid=29)

A schematic diagram of these layers of a normal cervical epithelium along with the changes which occur during the progression of cervical intraepithelial neoplasia (CIN) is presented in **Figure 1.5**.

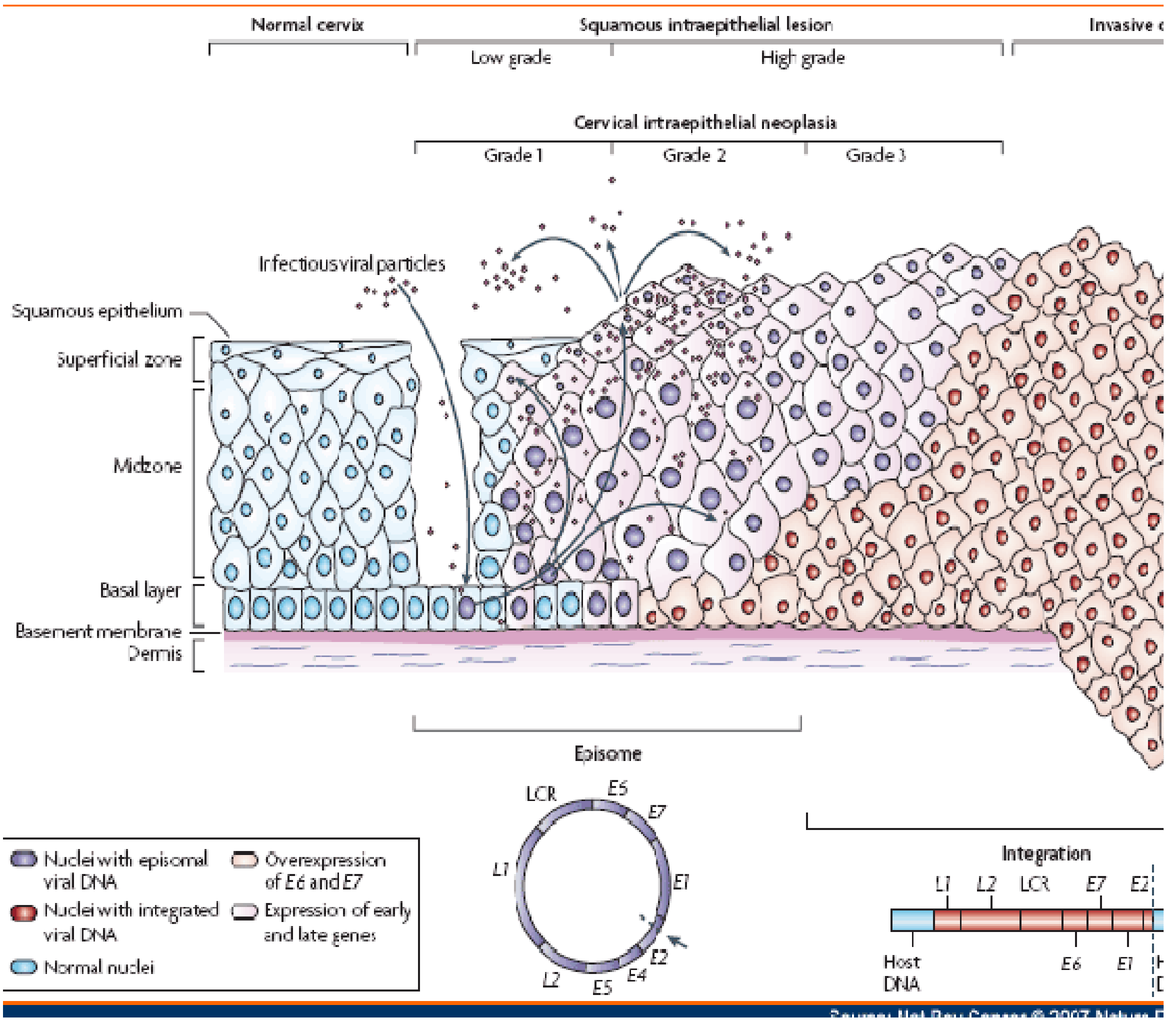


Figure 1.5 Schematic representation of the progression of normal epithelium towards invasive cervical cancer.
(www.medscape.com)

1.8 Progression to invasive cancer

Early dysplastic changes may persist without progression to the next stage and may revert or may progress to invasive cancer. Currently, it is not possible to predict the cases which are going to progress or revert. The degree of regression depends on the severity of the lesion. Most of the CIN I lesions revert to normal, only 1% progressing to invasive carcinoma (Pinto and Crum, 2000) and this is why the treatment of a CIN I lesion is considered unnecessary (Mera, 1997).

Other risk factors can determine progression, including infection with HPV types 16 or 18. Treatments exist to prevent progression, but they are not given until lesions progress to CIN II or CIN III.

1.9 Malignant stages

Cervical cancer is locally invasive, spreading upward in the *corpus uteri,* downward into the vaginal walls, forward in the bladder and backwards to cover the rectum and sacral plexus. Fistulas can be produced by the involvement of the rectum or bladder. Lymphatic spread is usually along the paracervical lymph toward the external iliac lymph nodes. Cervical cancer can metastasise to the liver, lungs and bone but this occurs only at very late stages (Mera, 1997).

On the basis of the degree of invasion and involvement of lymph nodes, malignancy is classified into following stages:

(a) Stage 0 Only few cancerous cells in the epithelium of cervix. They can persist there for several years.

(b) Stage I Cancerous cell spread to the underlying basement membrane, but still limited to the cervix. It is further divided into following sub classes. IA: Cancerous cells invade the connective tissue but there is no involvement of blood and the lymphatic system. Thickness of the tumour is not more than 3mm. IB: Thickness of tumour is more than 3mm and there is involvement of the blood and lymphatic system.

(c) Stage II Tumour spread beyond the cervix into the upper wall of vagina.

(d) Stage III Pelvic wall and lower third of vaginal wall are covered by tumour cells.

(e) Stage IV There is involvement of rectum and bladder.

The survival rate depends on the prognosis of the tumour. Stage 0 tumour is entirely treatable. The 5 year survival rate is 94% for Stage IA but this falls to 30% for Stage III and IV.

1.10 Detection and assessment of cervical lesions

The assessment of cervical lesions is usually done by visual inspection, bimanual pelvic examination, colposcopy and histological examination after the appearance of the symptoms or finding of abnormalities by employing the routinely used Pap smear examination. No useful information is obtained by the visual examination or bimanual inspection if the tumour is in the initial stages, especially if it lies inside the cervical canal. It appears as an inflammable outgrowth of the external opening of the lower part of the cervix in the vagina. As it penetrates the cervix, it becomes a hard, barrel shaped mass. It is important to mention that the cervix is supported by the ligament which connects it to the bony pelvis which means that the cervix is not originally fixed at its position but is caused to be by the development of the tumour (Mera, 1997). The penetration of the tumour into the vaginal fornix, which is an arch shaped depression in the upper vagina, results in the fixation of the cervix at its place. The tumour growth is fragile and bleeds on touching. Cervicitis and infected benign mucous polyps can produce the same symptoms, however (Mera, 1997).

Histological confirmation of the malignancy opens the pathway to several tests to confirm the stage of disease e.g. cystoscopy, proctosigmoidoscopy or barium enema to determine the involvement the bladder or rectum, chest radiography, intravenous pyelogram. The findings of the staging investigation should be confirmed by surgical exploration and histological examination of the excised tissue, unless radiotherapy alone is the treatment of choice (Mera, 1997).

Several different techniques have been used for the diagnosis of cervical cancer. A brief overview of these techniques is presented in this chapter and they are discussed in more detail in **chapter 2**. For the preliminary diagnosis of cervical cancer, the Pap smear test is commonly used. It involves the microscopic examination of the exfoliated cells of the cervix and detects cytological abnormalities (Ronco et al., 2006). The Pap smear test has low sensitivity and, to overcome this, HPV-DNA/RNA based techniques were introduced for the investigation of atypical squamous cells of undetermined significance (ASCUS) and low grade squamous intraepithelial lesions (LSIL) (Cuzick et al., 2008), as infection with Human Papilloma Virus (HPV) is considered an established reason for these pathological changes (Richart, 1990). HPV DNA/RNA testing is usually performed with commercially available tests, eg. Hybrid Capture II (HC II), Amplicor, Cobas4800, PreTectHPV-Proofer, APTIMA and Linear Array test (Ovestad et al., 2011, Clad et al., 2011). It is followed by the evaluation of the cervix with a low power light microscope (colposcope) for changes in the tissue reflectance, a marker for precancerous lesions (Thekkek and Richards-Kortum, 2008). It should be noted that due to the low specificity of such a visual examination, the ultimate diagnosis is recognized to be the histopathological examination of a punch biopsy or excised specimen (Mitchell Md et al., 1998).

In order to use the Pap smear test effectively, a proper infrastructure of laboratory for screening purposes, trained staff and most importantly facilities for recalling a patient for repeated examination are needed, which make this an expensive option in resource poor settings.

On the other hand, although they are more sensitive, interpretation of the results of colposcopy of the cervix and histopathological examination of the cervical tissue sections are in the hands of colposcopist and pathologist, respectively, and can vary according to the expertise of the examiner, giving rise to considerable intra- and inter- observer variability and false positive and negative results, hence making the techniques subjective. Moreover, these techniques are based on the observation of the morphological changes which become evident later than the biochemical changes (Hutchings et al., 2009, Kendall et al., 2009) taking place as a result of the progression of disease and cannot reliably be detected by histopathological techniques (Haggitt, 1994).

In order to provide the opportunity of cervical screening to women in resource poor settings and to increase specificity and sensitivity of diagnosis, new screening techniques are required.

Simple visual approaches such as visual inspection with acetic acid (VIA) have also been evaluated to facilitate cervical cancer screening in developing countries as an alternative to the the Pap smear and colposcopy (Ekalaksananan et al., 2010, Murillo et al., 2010, Aggarwal et al., 2010, Cremer et al., 2010, Ngoma et al., 2010, Nessa et al., 2010, Muwonge et al., 2010, Sankaranarayanan et al., 1998, Sankaranarayanan et al., 2007). More sophisticated optical techniques are being considered as a viable alternative for the early diagnosis of the cervical cancer. These are currently in use for research purposes and a lot of effort is being made to establish them for the detection and diagnosis of cervical cancer. They include confocal microscopy, optical coherence tomography and fluorescence spectroscopy, as discussed in detail in **chapter 2**.

Most techniques, and indeed classification systems, are based on the morphological changes associated with disease. These are a result of and become evident later than the biochemical changes taking place as a result of the progression of disease (Hutchings et al., 2009, Kendall et al., 2009). Ultimately, new techniques should be based on detection of the biochemical changes. Vibrational spectroscopic techniques, including FTIR absorption and

Raman scattering hold such potential and promise disease diagnosis at its earliest stage (Kendall et al., 2009).

Raman spectroscopy is an analytical technique that enables chemical characterisation of molecules within a sample. It is based on the inelastic scattering of photons upon interaction with material vibrations and provides a chemical signature of the material. It has been recognised to hold great potential in tracing the events of biochemical changes involved in the progression of normal tissue toward invasive carcinoma (Chan et al., 2006). Prominent features of malignant cells with respect to normal cells are increased levels of DNA and amino acids and decreased levels of glycogen in epithelial layers (Krishna et al., 2006, Utzinger et al., 2001). Raman spectroscopy has been shown to be capable of discriminating between normal keratinocytes, and cervical cancer cell lines with a sensitivity of 70-100% and specificity of 70-90% and was also able to differentiate between normal and HPV infected keratinocytes (Jess et al., 2007). *In vivo* probe systems that can probe for biochemical changes have been used for Raman spectroscopy (Krafft and Sergo, 2006). The cervical tissue studies to date have been confined to differentiation of normal cervical tissue from high grade or cancerous tissue. Investigation of 150 cervical tissue biopsies by Raman spectroscopy was carried out and normal cervical tissues were separated from malignant tissues with a sensitivity and specificity of 99.5% (Krishna et al., 2006). (Lyng et al., 2007) reported differentiation of 40 histological samples into normal cervical epithelium, invasive carcinoma and cervical intraepithelial neoplasia with nearly 100% sensitivity and specificity. An *in vivo* study on 66 patients showed that Raman spectroscopy has 89% sensitivity and 81% specificity to distinguish high grade dysplasia from benign tissue (Robichaux-Viehoever et al., 2007). The reported results are higher than colposcopy, which has a sensitivity of 87% and specificity of 72%.

1.11 Aim of the study

The aim of the present study is to evaluate the potential of Raman micro spectroscopy for early stage diagnosis of disease of the cervix as well as to elucidate the biochemical origin of the spectral markers associated with the onset and progression of the disease. The initial stages will establish the biochemical signatures of different layers of the cervical epithelium. It is expected that these signatures will be distinct, as the biochemistry of each layer is different. These differences in the biochemistry of each layer in the form of Raman spectral markers may be helpful for the better classification of cervical intraepithelial neoplasia (CIN). This would lead to the classification of cervical cancer on the basis of the biochemical changes occurring during the progression of the disease instead of morphological changes.

Raman micro spectroscopy will then be used to monitor the complete progression of cervical cancer from normal through to the various grades of CIN including CIN I, CIN II, CIN III and invasive carcinoma on the basis of spectral differences obtained from respective tissue samples.

Raman mapping will be carried out to differentiate between different layers of cervical epithelium such as basal, parabasal, intermediate and superficial, in formalin fixed and paraffin preserved (FFPP) normal cervical tissue sections. The next step will be the recording of Raman maps from pathologically diagnosed CIN I, CIN II, CIN III and tumour tissue sections from areas of normal and abnormal morphology.

The Raman spectra of cervical tissues are characterized by strongly overlapping spectral features so it can be difficult to interpret by visual inspection. Hence sophisticated mathematical and statistical data analysis techniques and algorithms are required for the extraction of useful information from these spectra. In recognition of this, multivariate data analysis such as principal component analysis (PCA), K means cluster analysis (KMCA), independent component analysis (ICA) and nonlinear negatively constrained least squares analysis (NNCLSA) will be used as standard protocols for spectroscopic data analyses in an attempt to elucidate the pattern of progression from normal to invasive carcinoma. Complementary techniques such as FTIR imaging and immunohistochemistry will also be carried out in parallel.

1.12 Overview of the thesis

In **chapter 2**, the different techniques which are currently in use for cervical cancer screening and diagnosis are discussed in greater detail. These techniques include the Pap smear test, HPV DNA/RNA based screening methods, visual inspection with acetic acid (VIA), colposcopy, histopathology. Moreover, techniques based on optical methods including confocal microscopy, optical coherence tomography, fluorescence spectroscopy, FTIR and Raman micro spectroscopy are also discussed along with the advantages and disadvantages of each technique with respect to their application for the screening of cervical cancer. As the techniques of choice for the current study are FTIR and Raman micro spectroscopy, these two techniques are compared for the acquired spectral features from the same cervical tissue samples. As Raman micro spectroscopy provides better spectral and spatial resolution, this technique has been selected for the further studies.

In **chapter 3**, for Raman spectral studies, different parameters which need to be optimized for sample preparation and measurement are discussed. In the case of formalin fixed and paraffin processed (FFPP) tissue sections, which are used for all the experiments for Raman spectral measurements, parameters such as the dewaxing solvent, the thickness of the section cut from the block, the substrate for the adhesion of the tissue sections, the laser wavelength and the objective lens to focus the laser and collect the Raman signals are optimised.

Having optimised these parameters for the current study, different cervical tissue samples from CIN II and CIN III patients are analysed with Raman micro spectroscopy. In **chapter 4**, the results of the Raman spectral mapping performed on the cervical tissue samples are presented. The results are analysed in the context of their morphological classification, and trends in biochemical changes are explored. The “Raman based classification” is compared to that of histopathology which suggests Raman micro spectroscopy as an alternative to histopathology.

In order to further elucidate the biochemical basis of the development of the cervical cancer, principal component analysis (PCA) has been performed on the Raman spectral data and presented in the **chapter 5**. The PCA of different regions of the tissue samples, normal and diseased, as identified by KMCA in the previous chapter, may be helpful to describe the biochemical differences between them. The changes occurring in the basal layer can be associated to the infection of the HPV virus and to identify these spectral changes, PCA results of the Raman spectral data of the cervical cell lines with different copy number of HPV will be compared with the PCA of the of the basal layers of the true normal and normal side of the CIN II sample. Further, PCA have been employed to differentiate between the basal layers of the normal, intermediate and abnormal, as mentioned in the chapter 4, to answer the question that was put by the results of the KMCA that either the basal layer proliferates into the stromal and superficial layers or there are the cells of these layers which undergo some biochemical changes during the development and progression of the disease.

To further confirm the information acquired from the PCA, in **chapter 5**, regarding the answer to the question “whether the cells of the basal layer proliferates into the superficial stromal layers or there are the cells of these layers which undergo some biochemical changes during the development and progression of the disease” immunohistochemical staining for the evaluation of the expression of p16 protein is performed and presented in **chapter 6**. The expression of p16 is a biomarker for the presence of HPV and is expected to be enhanced in the basal and superficial cells/layers rather than in the stroma/stromal cells.

Chapter 2

Techniques for cervical cancer screening and diagnosis

2.1 Introduction

The molecular and biochemical changes that ultimately lead to the occurrence of cancer include the disturbance of the basic organisation of nucleic acids, proteins, lipids, and carbohydrates (Ooi et al., 2008) and these changes can be used as diagnostic markers of cancer. Historically, many techniques have targeted the above mentioned biomolecules and sub-cellular structures, nucleus to cytoplasm ratio or cellular morphology for diagnosis as well as for the prognosis of cancer. However, changes at this level are detectable only after they have already caused considerable gross morphological changes in the tissue. Due to technological advances, more opportunities have arisen to investigate the bio-molecular components in tissue samples and to examine malignant abnormalities on this basis. As a result, the mutations/ biochemical changes in bio-molecules appearing before manifestation of disease on a cellular level can be detected earlier, facilitating timely and precise diagnosis, leading to increased patient survival and quality of life (Ooi et al., 2008).

2.2 Techniques Currently Used

Currently, screening for cervical cancer is preliminarily done by the microscopic examination of the exfoliated cells of the cervix for cytological abnormalities (Ronco et al., 2006) and the technique is commonly known as the Papaniculaou test or Pap smear test (Papanicolaou and Traut, 1941). An abnormal Pap smear test is followed by the assessment of the cervix with a low power light microscope (colposcope) for localised variations in the tissue reflectance, a marker for precancerous lesions (Thekkek and Richards-Kortum, 2008). Simple agents such as acetic acid and Lugol's iodine, together with the use of a green illumination filter, can enhance the contrast of suspicious regions. Due to the low specificity of such a visual examination, the ultimate diagnosis is established by the histopathological examination of a punch biopsy or excised specimen (Mitchell et al., 1998). Infection with the Human Papilloma Virus (HPV) is accepted as a root cause of such pathological changes (Richart, 1990). Thus, the identification of abnormalities in exfoliated cells (smears) may be followed by screening using the HPV DNA or RNA detection test (Cuzick et al., 2008). Currently, HPV testing is being introduced as a primary screening tool and an adjunct to cervical cytology for women over 30 years as recommended by American society for Colposcopy and Cervical pathology (Wright et al., 2007). The recommendations have been updated to address the age-appropriate screening strategies, including the use of cytology and HPV testing, follow-up of women after screening which includes management of screen positives and screening interval for screen negatives, age at which to exit screening, future considerations regarding HPV testing alone as a primary screening approach, and screening strategies for women vaccinated against HPV16/18 infections (Saslow et al., 2012).

2.2.1 Pap smear test

In order to determine cytological abnormalities in the cells of the cervix, the Pap smear is the test that is used for the preliminary screening of cervical cancer. This is done by the microscopic examination of the exfoliated cells of the cervix (Ronco et al., 2006).

The collection of cells from the squamocolumnar junction and outer opening of the cervix of the uterus is done by the use of a cyto brush. The cells are then stained by the Papanicolaou technique and are examined under a microscope to look for abnormalities in the nucleus and cytoplasm, such as nuclear to cytoplasm size ratio. Increased ratio is linked with a severe degree of CIN. Symptomatic changes to the nucleus include (a) increase of nuclear size with difference in shape (common feature of all dysplastic cells), (b) intense staining pattern of nucleus, (c) irregular chromatin distribution, (d) uncommon mitotic figures and visible nucleoli.

The proportion of the thickness of the epithelium with mature and differentiated cells is used for grading CIN. More severe degrees of CIN may have greater proportion of the thickness of epithelium (Sankaranarayanan, 2004).

2.2.1.1 Advantages

1. The sensitivity and specificity of Papanicolaou test are reported to be 11 to 99 and 14 to 97% respectively (Boyko, 1996).
2. The technique is relatively non-invasive.

2.2.1.2 Disadvantages

1. Sampling errors may occur and consistent and reliable sampling cannot be guaranteed
2. The technique is subjective and may lead to wrong interpretation.

2.2.2 Human Papillomavirus (HPV) based screening

A large amount of epidemiology and laboratory based research indicates that infection with oncogenic HPV, especially with types 16 and 18, is frequently associated with squamous cell carcinoma in the cervix (Woodman et al., January 2007). This link between infection with high risk HPV in the lower genital tract and cervical cancer has resulted in the development of a HPV DNA or RNA detection system for cervical cancer screening and of vaccines against high risk HPV types. There are currently two vaccines available on the market; Gardasil (Merck and Co.) and Cervarix (GlaxoSmithKline). Gardasil provides immunity against the two high risk HPV types, 16 and 18, which are associated with the development of cervical cancer and two low risk HPV types, 6 and 11, responsible for the majority of genital warts and recurrent respiratory papillomatosis. Cervarix claims to provide immunity against the two high risk HPV types, 16 and 18 (Madrid-Marina et al., 2009).

2.2.2.1 DNA based techniques

Currently, two DNA based tests for HPV detection are approved by the US Food and Drug Administration (FDA); Hybrid Capture II (HC II) (Qiagen Gaithersburg, Inc. MD USA) and the Hologic Cervista high risk HPV DNA test (Hologic Incorporated, Bedford, MA) (Clad et al., 2011). Other DNA based tests available include Roche Amplicor, Roche Linear Array and Cobas 4800 (Ovestad et al., 2011).

Hybrid capture-II has a sensitivity of 93.1% and 95.5% for detecting CIN II and CIN III respectively (Dufresne et al., 2011). The detection of a virus in women with transient stages of infection may increase the burden of screening of the HPV positive woman, as detection of HPV DNA prompts immediate treatment of the lesions, although follow up of these lesions suggest that 45.7% of CIN I and 21% of CIN II lesions regress (Insinga et al., 2009). The technique can be used as a follow up test for women treated for high grade lesions with localized treatment such as ablation or excision therapy to identify women who have been or have not been cured by their treatment (Cuzick et al., 2008). This technique is able to distinguish between high risk and low risk HPV types on the basis of nucleic acid hybridization of the bases of the E-1 gene but does not allow the identification of specific genotype.

Amplicor® is another DNA based HPV test which can target and detect DNA from 13 high risk genotypes. This test allows simultaneous PCR amplification of the target DNA from 13 high risk HPV genotypes (16, 18, 31, 33 35, 39, 45, 51, 52, 56, 58, 59 and 68) and from β-globin DNA as a cellular control. After PCR amplification, the oligonucleotide probes are used to hybridize to the amplified products and they are detected by colorimetric determination (Ovestad et al., 2010).

In the Roche Linear Array test, the biotinylated primers are used to define an approximately 450-bp sequence within the L1 region of the HPV genome. A pool of primers is used to amplify the HPV DNA targets from 37 HPV genotypes (6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 81, 82, 83, 84, 89 and IS39 (subtype of 82)). It should be noted that the group of 16 high-risk genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, 82 and IS39) includes the 13 genotypes targeted by the Amplicor test. Moreover, groups of possible high-risk (26, 53 and 66), unclassified (55, 62, 64, 67, 69, 71, 83, 84 and 89) and low-risk (6, 11, 40, 42, 54, 61, 70, 72 and 81) genotypes can also be detected. The β-globin DNA can be used as a cellular control (Ovestad et al., 2010).

The Cobas 4800 system, another DNA based test for HPV detection, (Roche Molecular Diagnostics, Pleasanton, CA, USA) provides fully automated sample preparation combined with real-time PCR technology which can amplify and simultaneous detect 14 high risk HPV genotypes, identification of HPV 16 and 18 and β-globin used as an internal control for sample validity (Ovestad et al., 2011).

The Hologic Cervista high risk HPV DNA and Cervista HPV 16/18 tests (Third Wave Technologies, Madison, WI; now owned by Hologic, Bedford, MA) were approved by the US FDA in March 12, 2009. Cervista HPV HR (high risk), for use in cervical cancer screening, is a DNA test for 14 carcinogenic HPV genotypes. Cervista HPV 16/18 is approved for the DNA detection of HPV-16 and HPV-18, the 2 most carcinogenic HPV types (Kinney et al., 2010).

2.2.2.2 RNA based techniques

The progression of cervical cancer is dependent on the continuous expression of the E6/E7 genes of HPV, essential for successful transformation and maintenance of the neoplastic phenotypes in cervical carcinoma cells. To detect these genes, on the basis of their mRNA, two techniques are available; PreTect HPV Proofer and APTIMA (Clad et al., 2011).

The PreTect HPV Proofer test is a real time multiplex assay for full length amplification of the mRNA of 5 high risk HPV types and is carried out by extraction of the total RNA from the samples.

The APTIMA HPV messenger RNA assay works on the target capture transcription mediated amplification of E6/E7 mRNA from 14 high risk HPV types and is able to differentiate between these types and identify which HPV type is present as well as which are more likely to persist.

2.2.2.2.1 Advantages

1. HPV DNA detection can be used as a primary screening tool.

2. It is an adjunctive test to cytology; can be used as a classification tool for the management of women with atypical squamous cells abnormalities of undetermined significance.

3. HPV DNA genotyping can be used for the identification of the different types of HPV.

4. Sensitivity and specificity for detection of high-risk HPV are found to be >92% and 99% for the APTIMA HPV Assay and 93% and 82% for the HC2 test (Dockter et al., 2009).

2.2.2.2.2 Disadvantages

1. HPV testing identifies transient infections which are common in women under 30 years and hence no important information is provided below this age.
2. None of these techniques can predict regression of CIN-II and CIN-III.

2.2.3 Colposcopy

Colposcopy is a technique of visualizing abnormalities in the tissue using light microscopy. Contrasting agents such as 3-5% dilute acetic acid, resulting in so-called aceto whitening, are used as contrast agents to help differentiate abnormal areas of the tissue from normal. Aceto whitening is due to the reversible coagulation of intracellular proteins. An increase in intracellular proteins during neoplasia results in a dense white spot (Gaffikin et al., 2003). Recognition of well defined areas near to the squamocolumnar junction indicates a positive test. The colposcopist extracts biopsies from any abnormal area found and sends them to the pathologist for further examination.

2.2.3.1 Advantages

1. This is a relatively non-invasive technique.
2. Detailed and careful examination is possible as the technique provides good magnification.
3. It has sensitivity greater than 90% (Mitchell, 1994).

2.2.3.2 Disadvantages

1. The technique/instrumentation is expensive.
2. Minor signs may be over interpreted and may lead to wrong decisions.
3. It has poor specificity (greater than 50%) (Mitchell, 1994).
4. It is subjective

2.2.4 Visual inspection with acetic acid (VIA)

In addition to more advanced techniques, simple visualisation approaches have also been investigated to facilitate cervical cancer screening in resource poor settings. For example, the use of visual inspection with acetic acid (VIA) is also under study as an alternative to the Pap smear and colposcopy in many developing countries (Ekalaksananan et al., 2010, Murillo et al., 2010, Aggarwal et al., 2010, Cremer et al., 2010, Ngoma et al., 2010, Nessa et al., 2010, Muwonge et al., 2010, Sankaranarayanan et al., 1998, Sankaranarayanan et al., 2007). VIA involves the examination of the cervix with the naked eye with the help of a bright light source after one minute application of contrasting agents such as 3-5% acetic acid. A recent review of the performance of VIA in twenty six studies reported 80% sensitivity and 92% specificity (Sauvaget et al., 2011).

2.2.4.1 Advantages

1. VIA requires minimum support infrastructure.
2. The patient can be referred immediately for testing.

2.2.4.2 Disadvantages

1. It is a subjective technique.

2. There is a need to define consistent criteria for the suspicious lesions and to train the provider.

2.2.5 Histopathology

The examination of the processed tissue samples by a pathologist for diagnosis is called histopathology. The processing of the tissue samples includes a series of steps which preserves the anatomy of the tissue samples as close as possible to their structure *in vivo*. After taking a biopsy, the tissue samples are passed through different processes which include fixation, dehydration, clearing, embedding, section cutting and staining. Tissue processing can be done manually or can be automated. After the final step, staining, the tissue sections on slides are examined by the pathologist. The diagnosis and grading of CIN by a pathologist is done on the basis of histological features including differentiation, maturation and stratification of the cells as well as nuclear abnormalities. The samples with more severe degrees of CIN are expected to have a higher proportion of the thickness of the epithelium composed of undifferentiated cells with only a thin layer of mature, differentiated cells on the surface. Nuclear abnormalities, including enlarged nuclei, increased nucleus to cytoplasm ratio, increased intensity of stained nuclei as well as variation in the size of nuclei are also taken into account (Sellors and Sankaranarayanan, 2003/4).

2.2.5.1 Advantages

1. Exact diagnosis of the severity of the cancer is possible.
2. The margin line of diseased area can be determined.

2.2.5.2 Disadvantages

1. False negative and false positive results may be possible, especially in the case of CIN or atypical cells.

2. The technique is subjective and interpretation of the results is in the hands of the pathologist.

2.3 New techniques based on optical methods

There has been a tremendous increase in the development of more complex optical methods for the detection and diagnosis of cervical cancer and they are now considered as techniques of choice for early detection, because these are fast, can be automated and objective with the help of statistical analysis techniques (Chidananda et al., 2006). The optical techniques currently under development include confocal microscopy, optical coherence tomography, and fluorescence spectroscopy. In the next sections of this chapter, these techniques are discussed in detail with a particular focus on the use of these techniques in studies related to the detection of cervical cancer along with advantages and disadvantages of each technique.

2.3.1 Confocal Microscopy

A confocal microscope works on the principle of optical sectioning and is generally described as a microscope that collects light coming only from the focal plane. Adjusting the focal plane provides different slices for 3D stacks. It allows non-invasive high resolution imaging at the cellular level at varying depths of the cervical epithelium. The light for detection is created by either fluorescence or reflection. Confocal fluorescence microscopy, by probing for fluorophoric metabolites (NADH, FAD), can differentiate fluorescence from the cytoplasm and peripheral membrane of the cell. It can also be employed to differentiate between different layers of the tissue samples, including basal cells from parabasal, intermediate and superficial cells (Pavlova et al., 2003).

Confocal microscopy, while working on the basis of the reflectance, can visualise the indicators of cancer onset and progression, such as morphological changes, nuclear to cytoplasm ratio or irregularity of cell spacing (Carlson et al., 2005, Bazant-Hegemark et al., 2008). The confocal technique has been applied *in vivo* to acquire images of the cervical epithelium with sub-cellular resolution by using a fibre-bundle confocal reflectance microscope (Carlson et al., 2005) in order to differentiate between normal and cervical precancers.

The technique has the following advantages and disadvantages:

2.3.1.1 Advantages:

1. It can provide an axial resolution of ≤1μm.
2. The light for detection is created by either reflection or fluorescence, providing double selectivity of target volume. Simple chemicals such as acetic acid can be used as contrast agents.
3. The technique has high sensitivity (100%) and specificity (91%) (Bazant-Hegemark et al., 2008).

2.3.1.2 Disadvantages:

1. Contrast agents can make the focussing complicated by causing the tissue to swell.
2. The field of inspection is of the order of a few hundred micrometers.
3. The penetration depth of modern probes is ~250 μm but can be impaired by the presence of scatterers in the overlying tissue (Bazant-Hegemark et al., 2008).

2.3.2 Optical coherence tomography (OCT)

Optical coherence tomography (OCT) is an imaging tool which uses the scattering of the incident light by the tissue to obtain quantitative information of the microstructural changes in the tissue sample. The technique is based on low coherence interferometery which is obtained by using typically a near infrared light source and an interferometer and usually its working principal is explained by analogy to ultrasound. The changes in the refractive index determine the contrast for the structural features (Bazant-Hegemark et al., 2008). The higher resolution of OCT compared to ultrasound is due to the use of light, rather than sound or radio waves. The light beam is focussed on the tissue, and a tiny portion of this light which is reflected from sub-surface features is collected. It should be noted that most light is not reflected but, rather, scatters and loses its original direction and does not contribute to forming an image but rather contributes to glare. By using the OCT technique, scattered light is filtered out with the help of an interferometer to completely remove the glare. Even the very tiny proportion of reflected light that is not scattered can then be detected and used to form the image.

The diagnostic potential of the technique has been shown by different researchers, assessing normal and abnormal cancerous tissues (Boppart et al., 1998, Clark et al., 2004, Maitland et al., 2008). Endoscopic OCT was used for the first time for the evaluation of cervical malignancy (Feldchtein et al., 1998, Sergeev et al., 1997). The intactness of the basement membrane has been determined to be a marker for a healthy epithelium, by the evaluation of cervical tissue samples from the genital tract of the female, 21 healthy and 11 CIN III samples, using OCT (Pitris et al., 1999). In another study of samples including CIN I, CIN II/III, it was reported that the thickness of the squamous epithelium varies for healthy samples and a healthy epithelium can be used as a marker for an intact basement membrane (Escobar et al., 2004, Escobar et al., 2005).

The technique has following advantages and disadvantages:

2.3.2.1 Advantages

1. The technique can show diagnostic capability at a tissue depth of 2-3 mm because it uses near IR where tissue is more transparent.

2. The technique is non-invasive, as it makes use of near infrared light sources in the range of 800-1500nm, at powers that are harmless to skin (Fujimoto et al., 2000).
3. The interferometric detection is not disturbed by ambient light.

2.3.2.2 Disadvantages

1. Although it has sensitivity of 95%, it has a specificity of only 46% for the differentiation of precancerous lesions from cancerous lesions of cervical epithelium (Gallwas et al., 2010)
2. The choice of the wavelength of light is limited by the strong absorbers in tissue samples, namely melanin (visible range) and water (infrared range).
3. The light sources to be used need to fulfil the stringent criteria in respect to partial coherence, for the interferometric detection.
4. Low contrast of the technique
5. Poor spatial resolution and sub-cellular structures or even single cells cannot be resolved.

2.3.3 Spectroscopy

Spectroscopy is defined as the interaction of light with matter in terms of wavelength (λ). The three main types of spectroscopy are absorption, emission and scattering spectroscopy. Absorption spectroscopy usually entails ultra violet, visible and/or infrared absorption. It measures the spectrum of light that a substance absorbs to get information about its structure. Emission spectroscopy measures the spectrum of light that a substance emits upon excitation e.g. fluorescence and phosphorescence spectroscopy. Scattering spectroscopy is similar to emission spectroscopy but it detects and measures all the wavelengths that a substance scatters upon excitation e.g. Raman spectroscopy (Latimer, 1967).

2.3.3.1 Fluorescence spectroscopy

This technique is a form of electromagnetic spectroscopy and utilizes light, usually in the UV region of the spectrum, to excite electrons in molecules of specific compounds resulting in the emission of light of a lower energy than the incident light, called fluorescence, which is measured for the analysis of the samples.

Fluorescence spectroscopy has been employed to demonstrate the potential of the technique for detection of pre-cancers in the cervix (Georgakoudi et al., 2002, Ramanujam et al., 1996, Zlatkov, 2009). The fluorescence spectra of normal and cancerous tissue samples have been studied (Zuluaga et al., 1999, Hubmann et al., 1990, Brookner et al., 1999, Grossman et al., 2001, Agrawal et al., 1999, Ramanujam et al., 2001, Ramanujam et al., 1994) in an effort to establish an optical pathology method for the early detection of cancer. By acquiring the laser induced fluorescence spectra from normal and cancerous stage III B cervical tissue samples and employing Principal Component Analysis (PCA), (Chidananda et al., 2006) have concluded that fluorescence spectroscopy can be used as a technique to predict the pathology of samples with high sensitivity and specificity, by performing comparisons with the calibration sets of different stages of cervical cancer.

2.3.3.1.1 Advantages

1. The technique can provide very specific information as specific molecules will give fluorescence at a specific excitation wavelength.
2. The technique has great sensitivity to changes in the structural and dynamic properties of bio molecules.

3. The technique has made it possible to identify cellular components with a high degree of specificity.

2.3.3.1.2 Disadvantages

1. Consumables like the fluorescent tags may be required for the samples to be studied, which are expensive.
2. As it uses UV or visible light, Confocal fluorescence spectroscopy has a penetration depth of only ~0.1mm.

2.3.4 Vibrational spectroscopy

During the vibration of a molecule, movements of the bonded atoms towards and away from each other occur at a characteristic frequency. By taking the example of a diatomic molecule as two spheres of mass m_1 and m_2 connected by a spring (Hollas, 2004), the energy of the system can be related to how much the spring is stretched or compressed. According to a classical mechanics approach, the spring obeys Hooke's law, for which the force, *F,* required to extend the spring is proportional to the displacement from equilibrium Δx,

$$F = -K\Delta X \qquad \text{(Equation 2.1)}$$

The proportionality constant is known as the force constant, K.

The potential energy, *E,* of the system depends on the square of the displacement from equilibrium and is given by:

$$E = \frac{1}{2}K\left(\Delta x\right)^2 \qquad \text{(Equation 2.2)}$$

The characteristic frequency is given by:

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad \text{(Equation 2.3)}$$

Where, the reduced mass, $\mu = m_1.m_2/m_1+m_2$

The frequency of the vibration is proportional to the square root of the ratio of the spring force constant to the reduced mass on the spring. The lighter the mass on the spring or stronger the spring, the higher the vibrational frequency will be. Similarly, vibrational frequencies for stretching bonds in molecules can be related to the strength of the chemical bonds and the masses of the bonded atoms.

It should be noted that the molecules differ from sets of spheres-and-springs in that the vibrational frequencies are quantized. That is, only certain energies for the system are allowed, and only photons with certain energies can excite molecular vibrations. Moreover, the symmetry of the molecule will also decide whether a photon can be absorbed.

2.3.4.1 Infrared spectroscopy

Infrared radiation belongs to that part of electromagnetic spectrum that lies between visible and microwave region. The infrared region is divided into three regions (near, mid and far infrared) according to its distance from the visible region in the electromagnetic spectrum, as illustrated in Table 2.1.

Region	Wavelength range(λ,µm)	Wavenumber range(λ, cm^{-1})	Energy (E) (Kcal/mol)
Near	0.78-3	12820-4000	10-37
Middle	3-30	4000-400	1-10
Far	30-300	400-33	0.1-1

Table 2.1 Regions of the IR spectrum in commonly employed units.

Frequency (ν) is defined as the number of wave cycles that pass through a point in one second. The unit for the frequency is Hz. 1 Hz = 1 cycle/second. The length of one wave cycle is defined as the wavelength (λ) and is measured in microns (µm). Wavelength and frequency are inversely related:

$$\nu = c/\lambda \qquad \text{(Equation 2.4)}$$

Infrared (IR) spectroscopy is based on the absorption of infrared radiation by the sample under study and the fact that molecules absorb specific frequencies of the incident light which are characteristic of their structure.

IR spectroscopy optically probes the molecular vibrations that depend on the composition and structure of the material under study. Infrared light is used to irradiate the sample and changes in the vibration induced are recorded. Disease and other pathological conditions lead to changes in structure and chemical composition. These changes produce alterations in the vibrational spectra that can be used as sensitive phenotypic marker of disease (Krafft et al., 2008b). Near or mid infra red light sources give the best contrast for medical applications.

There may be many vibrations in a molecule, for molecules with N atoms, linear molecules have 3N–5 degrees of vibrational modes, whereas nonlinear molecules have 3N–6 degrees of vibrational modes. A vibration can be "IR active," only if it is associated with changes in the permanent dipole.

FTIR (Fourier Transform Infra red) spectroscopy is a measurement technique for the acquisition of infrared spectra. Frequency differentiation is performed with an interferometer. After passing through the sample, the measured signal is the interferogram. The mathematical Fourier Transform of this signal produces a spectrum similar to that from conventional (dispersive) infrared spectroscopy (Gremlich and Yan, 2001).

The use of FTIR spectroscopy for cervical cancer diagnosis has been demonstrated by studying the exfoliated cells from 156 samples from female patients (Wong et al., 1991) and differentiating them into normal, dysplastic and malignant. The cancerous progression and normal proliferation has been associated with glycogen reduction, increased hydrogen bonding from phosphodiester groups and reduced hydrogen bonding in alcoholic groups of the amino acids.

Analysis of FT-IR spectra of over 2,000 individual cells from 10 normal females, 7 females with dysplasia, and 5 females with squamous cell carcinoma has revealed that the spectra of normal-appearing intermediate and superficial cells of the cervix obtained from women with either dysplasia or cancer are different from those of normal women. It is reported that the structural changes underlying the spectroscopic variations may be more representative of the state of the disease than the morphological changes (Cohenford and Rigas, 1998). This information strongly supports the potential of FTIR spectroscopy for the diagnosis of cervical cancer. As cervical cells mature from basal to parabasal, intermediate, and then superficial cells, an increase of glycogen content is reported (Koss, 1992). This normal progression is disrupted in diseased tissue and hence can be used as a diagnostic marker for cervical cancer by FTIR spectroscopy. Moreover, changes in the levels of biomolecules including RNA, DNA can also be helpful for diagnostic purposes (Mordechai et al., 2004). The infrared spectroscopy has been employed for the analysis of Pap smear cells and an algorithm is developed called the PapMap to process the imaging data which is reported to lead towards the automation of the screening process (Schubert et al., 2010).

2.3.4.1.1 Advantages

1. Larger sample area can be scanned in less time as compared to Raman spectroscopy (see section 2.3.2.3.5).
2. Spatial resolution is ≥ 10 µm.
3. Possibility to measure the secondary structure of protein non- invasively in a native environment.
4. Same samples can be used for further studies.
5. No interference from fluorescence of biological sample during measurements.

2.3.4.1.2 Disadvantages

1. The interference from water at 1640cm^{-1} overlaps IR absorbance of protein. Modern FT-IR instrument and software can overcome this problem to a great extent.
2. There is strong interference from glass. There is need for alternative materials that do not interfere during IR spectroscopy to be used as substrates for biological samples.
3. It needs good data analysis techniques to acquire useful and meaningful information from the spectral data.

2.3.4.2 Raman Spectroscopy

The working principal of Raman spectroscopy is based on a scattering phenomenon. Scattering results from the interaction between the photons and materials, such as molecules. As shown schematically in **Figure 2.1**, most photons colliding with molecules do not change their energy after collision. These collisions are called elastic collisions and the result of these collisions is elastic scattering called **Rayleigh** scattering. Rayleigh scattering is a two photon process as a result of which there may be a change in direction of light but no net change in frequency.

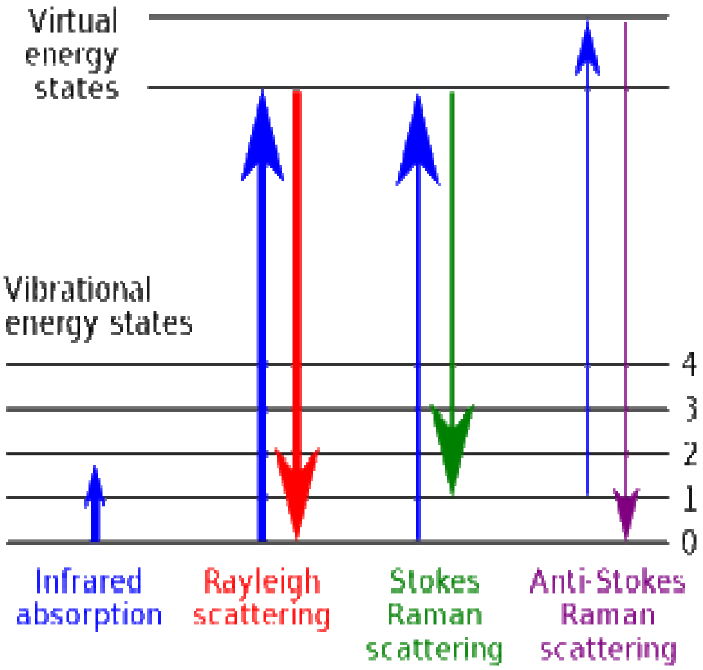


Figure 2.1 Schematic presentation of Rayleigh and Raman scattering of incident light. (http://en.wikipedia.org/wiki/Raman_spectroscopy)

On the other hand, inelastic collisions, after which the incident photons change their energy, lead to inelastic scattering during which exchange of vibrational energy occurs. The process is defined as Raman scattering. Raman scattering is also a two photon process by which a change in the energy of the colliding photon occurs from $h\nu_o$ to $h\nu_o \pm h\nu_s$ where ν_o

is the incident photon frequency, ν_s is the frequency of the scattering vibration and h is Planck's constant.

For a molecular vibration to be Raman active, the net molecular polarizability must be changed. The polarizability (α) represents the ability of an applied electric field, *E,* to induce a dipole moment, μ_0, in an atom or molecule. This process is represented mathematically by the equations below (Ball, 2001):

$$\mu_o = \alpha E \qquad \text{(Equation 2.5)}$$

At the molecule's equilibrium nuclear geometry, the polarizability has a value, α_0. At some displacement, Δr, away from the molecule's equilibrium geometry, the instantaneous polarization α is given by:

$$\alpha = \alpha_o + \left[\frac{\partial \alpha}{\partial r}\right] \Delta r \qquad \text{(Equation 2.6)}$$

where the derivative $\left[\frac{\partial \alpha}{\partial r}\right]$ represents the change in polarizability as a function of displacement. If the molecule is vibrating in a sinusoidal fashion, Δr can be written as some sinusoidal (that is, sine- or cosine-based) function in terms of the frequency of the vibration, ν_s, and the time, *t*:

$$\Delta r = r_{\max} \cos(2\pi v_s t) \qquad \text{(Equation 2.7)}$$

where $r_{\max}$ is the maximum vibrational amplitude. Light that has a particular frequency, ν_o, has an associated electric field, *E,* that also has sinusoidal behavior:

$$E = E_{\max} \cos\left(2\pi v_0 t\right) \qquad \text{(Equation 2.8)}$$

Here, $E_{\max}$ is the maximum electric field frequency. Equations 2.6, 2.7 and 2.8 can be substituted for α, Δr and E, respectively, into equation 2.5 to get:

$$\mu_0 = \alpha_0 E_{\max} \cos(2\pi\upsilon_0 t) + E_{\max} r_{\max} \left[\frac{d\alpha}{dr}\right] \cos(2\pi\upsilon_s t)\cos(2\pi\upsilon_0 t) \qquad \text{(Equation 2.9)}$$

To get a relationship for a product of two cosines, which is what the second term in equation 2.9, is:

$$\cos(a)\times\cos(b) = \frac{1}{2}\left[\cos(a+b)+\cos(a-b)\right] \qquad \text{(Equation 2.10)}$$

In comparing equations 2.9 and 2.10, it can be seen that a is $2\pi\nu_0 t$ and b is $2\pi\nu_s t$. Substituting for the product of the two cosines in equation 2.10, gives the following:

$$\mu_o = \alpha_o E_{\max} \cos(2\pi v_o t) + \frac{E_{\max} r_{\max}}{2} \left[\frac{\partial\alpha}{\partial r}\right] \cos\left(2\pi t\left(v_o + v_s\right)\right) + \cos\left(2\pi t\left(v_o - v_s\right)\right)$$

(Equation 2.11)

The oscillating polarization reradiates light at the same frequency and thus the scattered light has components of the frequencies represented in Equation 2.10. By analyzing the terms in equation 2.10, it can be seen what they imply. The first term contains the variable ν_{in}, which is the frequency of the incoming light. This term relates to the outgoing, scattered photon that has the same frequency as the incoming photon. That is, this term ultimately explains Rayleigh scattering. The second term contains two cosines. One contains the variable $\nu_{in}+\nu$, which relates to an outgoing, scattered photon that has increased in frequency by some amount, ν, which is the frequency of the molecular vibration. The associated increase in energy is the result of the annihilation of a quantum of vibration of energy hν. The last cosine term contains the variable $\nu_{in}-\nu$, which relates to a scattered photon that has decreased its frequency by the same amount, ν. The decrease in energy is the result of the creation of a quantum of vibration of energy $h\nu$. These two terms ultimately explain Raman scattering. The two terms show that incoming photons will shift their frequencies, up *and* down, by amounts equal to the vibrational frequencies of the molecule. This is the basis of Raman spectroscopy.

At ambient temperature, most of the molecules are in their vibrational ground state. According to Boltzman, a much smaller number of molecules are in a vibrationally excited state. Therefore, the chances for a Raman interaction which transfer vibrational energy to a molecule and leave a quantum of lower energy $h\nu_o – h\nu_s$ are much higher than an interaction which leaves a quantum of higher energy $h\nu_o + h\nu_s$. In 1852, while studying fluorescence spectra, Stokes postulated that the wavelength of light produced by phosphorescence or fluorescence is always longer than that of the exciting light. By association, Raman scattering processes are referred to as Stokes, resulting in quanta of lower energy, or anti-Stokes resulting in quanta of higher energy (Ball, 2001, Schrader, 1995). At room temperature, as most of the molecules are generally in the ground vibrational state, Stokes Raman scattering most commonly occurs and is usually measured.

The Raman shift of the photon energy is a measure of the energy of vibration of the molecule. A complex molecule will have many vibrational modes, and the intensity of the scattered radiation versus Raman shift gives a Raman spectrum which is characteristic for each individual substance and for this reason, Raman spectral features can be the identification markers of a substance and can be used to analyse its structure (Sasic, 2008).

Studies ranging from pure biological molecules in isolation to all aspects of the human body have been a matter of strong interest in the field of spectroscopy of bio-molecules. FTIR Spectroscopy was used to study human and animal tissues (Blout and Mellors, 1949) and Woernely, 1952. Due to the lack of well established instrumentation at that time, researchers who attempted to record the Raman spectra of such samples were facing problems of overwhelming fluorescence from biological molecules, long integration times and high power density requirements. Important developments in the field of diode lasers and charged coupled device (CCD) cameras in 1969 (Boyle and Smith, 1970) finally made it possible to record high quality Raman spectra of biological samples in 1970s (Lord and Yu, 1970). Vibrational Spectroscopy has emerged as a major tool for biochemical applications as it is able to provide direct information regarding the molecular constituents of tissue samples at the cellular level.

There has been extensive use of Raman spectroscopy for various applications including studies made on biological tissue. These include bone (Rehman et al., 1995), cornea (Kachi et al., 2000), epithelial tissue including the larynx, tonsil, oesophagus, stomach, bladder and prostate (Stone et al., 2002, Stone et al., 2004, Keller et al., 2008), lung (Notingher et al., 2003, Huang et al., 2003a, Huang et al., 2003b, Kaminaka et al., 2001, Min et al., 2005, Krafft et al., 2008a, Liu et al., 2005, Qian et al., 2008, Short et al., 2008, Bonnier et al., 2010a). The technique has also been applied to study the composition of lymphocytes (Puppels et al., 1993), human red blood cells (Deng et al., 2005) as well as to differentiate between different types of cells including mixed cancer cells (Krishna et al., 2005), human living cells (Kuhnert and Thumser, 2004), individual cells (Chan et al., 2006). In addition, it has been employed to study different biological/macromolecules including DNA (Ruiz-Chica et al., 2004), bodily fluids such as saliva (Farquharson et al., 2005b, Farquharson et al., 2005a), tissue processing (O Faolain et al., 2005), and raft cultures (Viehoever et al., 2003).

The use of Raman spectroscopy to distinguish cervical cancer from normal tissue was first reported in 1992 (Liu et al., 1992). Three significant peaks at 1262 ,1445 and 1657 cm^{-1} were consistently observed in all gynaecologic tissues studied using FTIR and Raman spectroscopy and it was observed that the intensity ratio of 1445/1657 cm^{-1} could be used to differentiate between normal and malignant cervical tissue samples. The intensity ratio of 1445 and 1657 cm^{-1} was found to be higher in normal tissue samples (Liu et al., 1992). Resonance Raman spectroscopy, using a 254 nm laser, was used to differentiate between malignant and normal cells of the cervix and the breast (Yazdi et al., 1999). Differentiation was possible due to changes detected in the peak ratios of different nucleic acid contents in normal and tumour cells. Changes in the vibrational structure of nucleic acids associated with the malignant phenotype of the cell were also detected. The peak ratio 1480/1614 cm^{-1} was found to have a higher value in malignant samples and lower for normal samples (Yazdi et al., 1999).

Spectra obtained from human cervical tissue *in vitro*, using a NIR Raman spectroscopic probe could differentiate between normal and precancerous tissues with a sensitivity of 82% and specificity of 92% (Mahadevan-Jansen et al., 1998). An increase in ratio of intensities at 1454 to 1656 cm^{-1} and decrease in ratio of intensities at 1330 to 1454 cm^{-1} was used to separate the cervical dysplasia from normal tissue types *in vivo* (Utzinger et al., 2001). DNA, phospholipids and collagen were identified as diagnostically significant (Utzinger et al., 2001). The analysis of distinctive peaks and fingerprint region agrees with the portrait provided by infrared spectroscopy i.e. basal cells have characteristic bands of DNA, epithelial cells show high level of glycogen in normal cells (Wong et al., 1991). A statistical method that combines the method of maximum representation and discrimination feature (MRDF) and sparse multinomial logistic regression (SMLR) was developed for multiclass analysis of Raman spectra. Application of this method and incorporation of patient hormonal status increased the sensitivity to 98%, specificity to 96% and accuracy to 94% (Kanter et al., 2009a, Kanter et al., 2009b). Four different cell lines, HPV negative C33A, HPV 18 positive HeLa which has 20-50 integrated copies of HPV per cell, HPV 16 positive SiHa with 1-2 HPV copies per cell and CaSki with 60 -600 integrated copies of HPV per cell, were studied using Raman and FTIR spectroscopy. Principal component analysis (PCA) clearly differentiated the C33A, SiHa group from HeLa and CaSki group (Ostrowska et al., 2010). A recent study of de-paraffinized cervical tissue with Raman mapping and hierarchical cluster analysis (HCA) was able to differentiate between normal squamous epithelium and cervical intraepithelial neoplasia (Tan et al., 2011).

In the next section, FTIR and Raman mapping of tissue sections from cervical tissue biopsies is described.

2.3.4.2.1 Advantages of the technique

1. This is non-destructive technique; the sample is not damaged during the analysis by this technique and can be used again.
2. The technique is fast, about 1 min. per acquisition is required.
3. It is non-invasive method.
4. There is minimal interference from water and CO_2.
5. No labelling is required during the use of this technique.
6. *In vivo* and real time measurement from the biological systems are possible.

2.3.4.2.2 Disadvantages of the technique

1. High laser intensities are required because the Raman scattering is a relatively weak phenomenon and thus the equipment cost can be quite high depending on their applications.
2. The technique, generally, is less sensitive for quantitative analysis as compared to chromatographic techniques like HPLC.
3. It needs good data analysis techniques to acquire useful and meaningful information from the Raman spectra.

2.4 Conclusion

Different techniques currently in use for research or clinical purposes for the screening of cervical cancer have been introduced in this chapter. These techniques, Pap smear, colposcopy, HPV DNA/RNA based techniques, VIA, histological examination, confocal microscopy, optical coherence tomography and vibrational spectroscopy are discussed along with their advantages and disadvantages, with particular focus on vibrational spectroscopy.

In the next chapter, the use of the FTIR and Raman micro spectroscopy is initially demonstrated and the spectral data acquired by using these techniques is compared to choose the best technique for further analysis. Moreover, the optimisation of different parameters, for the chosen technique to perform the spectral mapping, including dewaxing agent, laser line selection, substrate, objective lens and thickness of the cervical tissue sections is presented.

Chapter 3

Optimisation of vibrational spectroscopy for cervical tissue Imaging

As described in **Chapter 2**, one of the several advantages of FTIR and Raman spectroscopy is the ability of the techniques to provide information about the biochemical composition of biological specimens with little or no sample preparation. In the current study, both techniques have been demonstrated for the imaging of cervical tissue section and compared to choose the best one for further analysis.

3.1 Comparison of FTIR and Raman spectral mapping of cervical tissue sections

For this study, two parallel tissue sections from a FFPP cervical tissue block were cut, one used for H&E staining and other for FTIR and Raman spectral mapping.

For H&E staining, first of all the tissue section was dewaxed using xylene and then stained in Harris' haematoxylin for five minutes, and then washed in running tap water for three minutes. After washing, the sample was differentiated in 1% acid alcohol for 1 or 2 seconds. The cellular nuclei should be blue and the background should be clear when observed under the light microscope. This was followed by staining in 1% Eosin for one minute and then rinsing well in tap water. The sample was dehydrated by dipping three times in methylated spirit followed two baths of absolute alcohol for three and two minutes respectively and then two baths of histoclear for four and five minutes respectively. Finally, the stained tissue section was mounted with glass cover slip using DPX mounting medium.

The stained tissue section was used as a morphological standard to identify abnormal and normal epithelium against the unstained tissue sections.

3.1.1 FTIR Data acquisition

FTIR images were recorded using a Perkin Elmer Spotlight 400N FTIR imaging system, incorporating a liquid nitrogen cooled mercury cadmium telluride 16x1, 6.25µm pixel array detector, and were acquired by the Spectral Image software . The FTIR image from dewaxed cervical tissue section mounted on a CaF_2 slide was recorded over the range 4000-800 cm^{-1} in transmittance mode with a resolution of 4 cm^{-1} and interferometer speed of $1.0cm^{-1}$/second at continuously varying magnification. The scans per pixel for back ground were 120 and, for images, 16 per pixel respectively.

3.1.2 Raman Data acquisition

Raman maps were recorded using a HORIBA Jobin Yvon HR 800 Raman microscope (LabSpec V5.58) with a 785 nm laser as source. The X100 dry objective was used to focus on the sample and collect the Raman scattered light in a backscattering geometry. Raman scattering was collected through a 100 µm confocal hole onto a Synapse air-cooled CCD detector for the range of 400-1800 cm-^{1} using a 300 Gr/mm diffraction grating. The instrument was calibrated using the 520.7 cm^{-1} peak of silicon. Raman spectral mapping was done using 15 seconds x2 acquisition with a step size of 18µm.

3.1.3 Data pre-processing of FTIR and Raman maps

For the FTIR spectra, the correction to atmospheric CO_2 absorption bands in the spectra was done by a built in function of Perkin Elmer Spotlight software. All data processing

(FTIR and Raman spectral data) was done directly on the spectral image using MatLab 7.2 and the protocols which have been established in the Dublin Institute of Technology (DIT) based on the internationally recognized methods (Kniefe, 2010). Data pre-processing included smoothing, baseline correction and normalization. All spectra, including calibration and substrate backgrounds, were vector normalized and smoothed using a Savitzky Golay smoothing method. A rubber band correction for baseline removal for all the spectra was carried out and the substrate spectra were subtracted from each spectrum. This has been demonstrated by presenting the raw Raman spectral data and the processed data in **Figure 3.1** and **3.2** respectively.

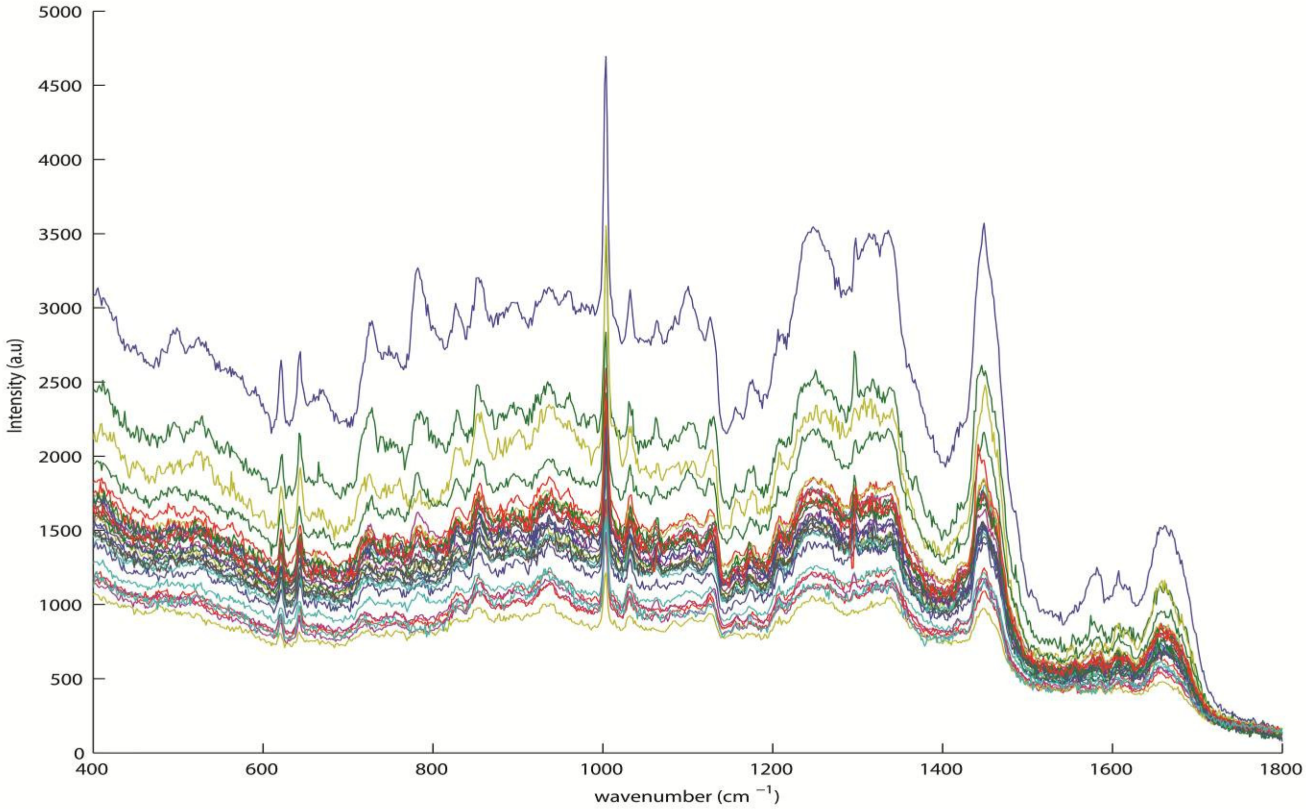


Figure 3.1 Raw Raman spectral data of basal-true normal cervical tissue sample.

All the data sets have been processed by using the same protocols and there is no manipulation of the data done which could potentially contribute to the changes in the Raman spectra and into the results of the PCA loading and KMCA. For example, the vector normalisation of the spectra, in fact removes the intensity variation of different peaks (if it exists) and all the data would reflect only the changes associated with the changes occurring as a result of the disease progression.

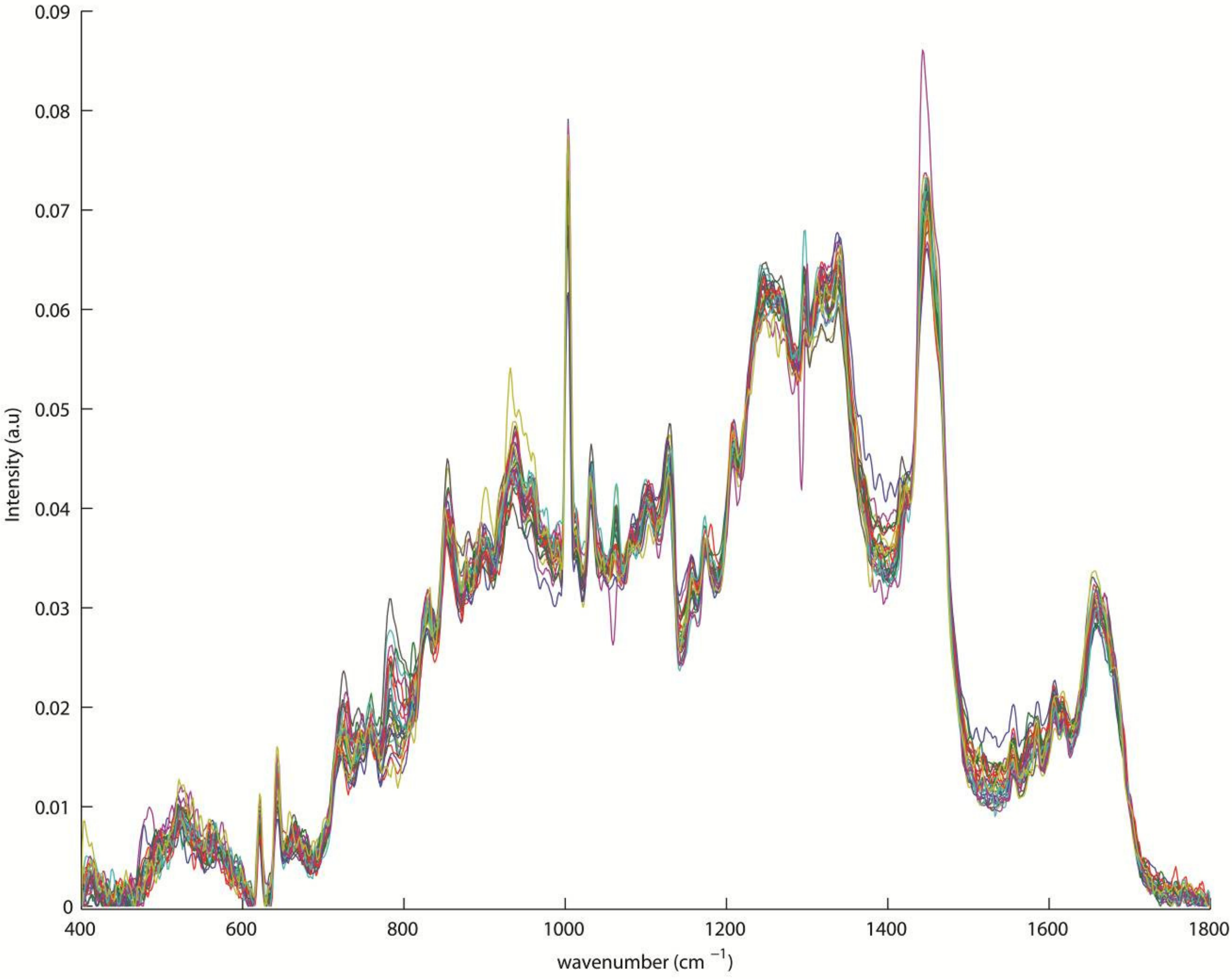


Figure 3.2 Processed Raman spectral data of basal-true normal cervical tissue sample.

3.1.4 Data analysis

K-means cluster analysis (KMCA) and Principal Component analysis (PCA) were used to analyse the spectral data sets obtained from the FTIR and Raman maps. The results of the

KMCA for the initial study are presented in the current chapter and some detailed analysis in **chapter 4** and those of the PCA in the **chapter 5**.

KMCA is an unsupervised non hierarchical method of clustering the cases randomly into user predefined clusters and cycles until a local minimum is found by using the Euclidean sum of squares as a descriptor (Wang and Mizaikoff, 2008). The false colour map generated then shows the clusters with similar spectral and hence biological properties.

Principal component analysis (PCA) consists of a mathematical procedure involving the transformation of possibly correlated variables into a smaller number of uncorrelated variables known as principal components (PC). The first principal component accounts for as much variability in the data as possible, and each succeeding principal component accounts for as much of the remaining variability as possible. The loadings of the PC can describe the biochemical differences which are used by the PCA to determine the variability in the Raman spectral data and hence separation of different groups of spectra.

PCA was developed in 1901 by Karl Pearson (Pearson, 1901). Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA involves the calculation of the eigenvalue decomposition of a data matrix, usually after mean centering the data for each attribute. The results of a PCA are usually discussed in terms of component scores and loadings (Martens and Naes, 1994). The PCA results are presented as loadings and the scatter plot. The first principal component accounts for as much variability in the data as possible, and each succeeding principal component accounts for as much of the remaining variability as possible. A PCA scatter plot is used to group and differentiate datasets according to their similarities and differences, based on the identified principle components. The weightings, or “loadings”, of the principle components represent the spectral profile which is responsible for the grouping and differentiation of the datasets.

The loadings of the PC can therefore describe the biochemical differences which are used by the PCA to determine the variability in the Raman spectral data and hence separation of different groups of spectra.

3.1.5 Results of FTIR and Raman spectral mapping of cervical tissue sections

Figure 3.3 a, shows a H&E stained cervical tissue section, the red box indicating the area from where Raman and FTIR maps are acquired. The tissue region has been diagnosed as normal by the histopathology department of the Coombe Women and Children's Hospital, Dublin. **Figure 3.3 b and 3.3 d** show the optical images of the parallel unstained tissue section as viewed by the FTIR and Raman micro spectrometer. KMCA of the FTIR and Raman maps of the same sample are shown in **Figure 3.3 c** and **Figure 3.3 e,** respectively. KMCA of the FTIR image divides the map into four parts, indicated by blue and brown and orange clusters representing stroma, basal and substrate respectively while green and mid blue, both associated to the superficial layer.

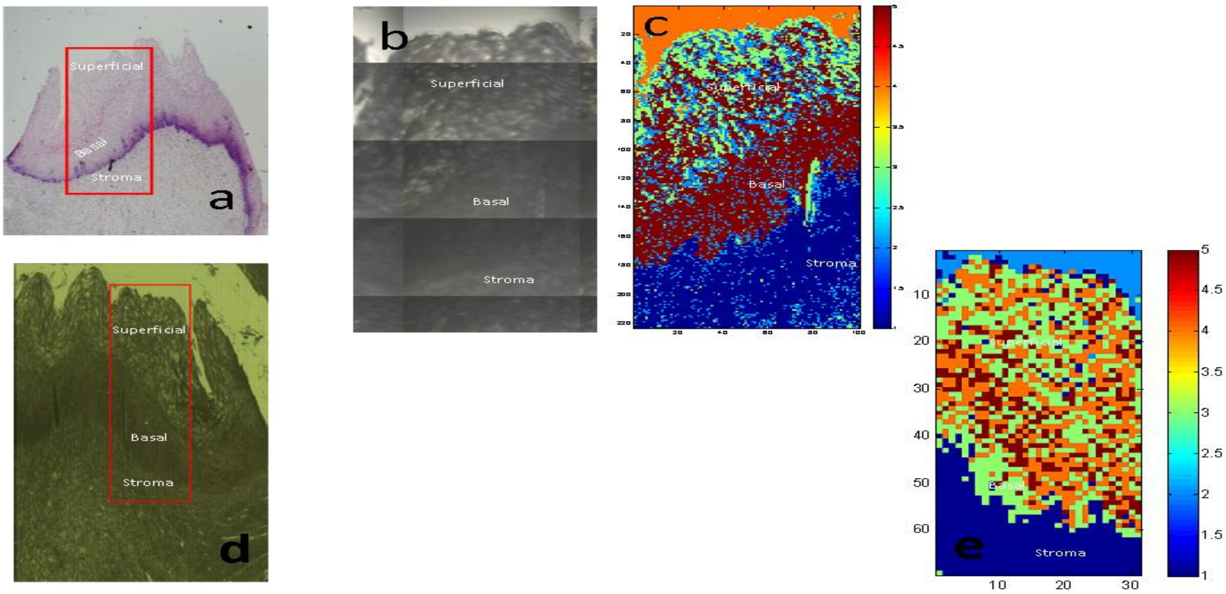


Figure 3.3 H&E stained cervical epithelium (a) unstained tissue under FTIR spectroscope (b) Five cluster KMCA map generated from the FTIR map (c) unstained tissue under Raman spectroscope (d) Five cluster KMCA map generated from the Raman map (e).

Figure 3.4 shows the mean spectra for the five K-means clusters in the FTIR map. The specific vibrational modes are labelled with solid lines. The bands at 1028, 1080 and 1152 cm^{-1} are assigned to glycogen and are only observed in the superficial layer (green and mid blue). These bands are absent in the stroma (blue). This is understandable, as no glycogen is expected in this layer (Chowdhury and Chowdhury, 1981). The FTIR feature at 1236 cm^{-1} can be assigned to the asymmetric stretching of the O-P-O of DNA/RNA in the basal layer (brown) and for collagen proteins in the stroma (blue) (Andrus and Strickland, 1998). This FTIR band is present in both these layers. The stromal layer is mainly composed of fibroblasts that are rich in collagen and the basal layer contains more dividing cells which leads to a greater concentration of DNA and hence this band is more prominent in these layers.

The bands at 1402 and 1454 cm^{-1} can be assigned to the CH deformation modes of carbohydrates (glycogen) (Dovbeshko et al., 2000, Wood et al., 1998). These FTIR bands are dominant in the squamous epithelium (green and mid blue spectra) as expected. The difference between green and mid blue is only in the intensity/concentration of glycogen, but otherwise they give the same biochemical information.

The features of the FTIR spectra at 1654 cm^{-1} (amide-I) and 1546 cm^{-1} (amide-II), both assigned to proteins, are present in all three layers, but their intensity seems to decrease from epithelial, to stroma and to superficial layers respectively. In the epithelial layer, these bands are probably due to the proteins which are also present inside the nucleus where required for folding of the DNA and are almost always seen in any spectra taken from the cells and tissues. The band at 2300 cm^{-1} cannot be assigned to any biological feature and is probably due to the incomplete atmospheric correction, therefore assumed to be spectral artefacts. The sharp FTIR bands at 2848, 2916 and 2956 cm^{-1} are due to the lipids (Fabian et al., 1995, Huleihel et al., 2002).

The KMCA of the FTIR map shows the separation of cervical epithelium into three distinct layers, but the analysis of the mean representative spectrum of each cluster does not elucidate the biological differences underlying this separation. As the FTIR marker bands for DNA (basal layer) and collagen (stromal layer) at ~ 1236 cm^{-1} are not resolved, so there is no band to differentiate between these two layers.

KMCA of the Raman map is shown in **Figure 3.3 e**. The mid blue cluster corresponds to the substrate. The KMCA Raman map differentiates the tissue section into stroma (blue), with characteristic Raman bands of collagen including 849 cm^{-1}, 921 cm^{-1}, 938 cm^{-1} and 1245 cm^{-1}, basal (green) with Raman bands for DNA bases, thymine (755 cm^{-1}), adenine (722cm^{-1}) and cytosine (782 cm^{-1}) and superficial layers (orange and brown) with Raman bands of glycogen at 480, 849 and 938 cm^{-1}, as shown in **Figure 3.5**.

The layer differentiating bands for the mean spectra of FTIR map were limited to those at 1028, 1080 and 1152 cm^{-1}, related to glycogen, and 1236 cm^{-1} related to asymmetric stretching of the O-P-O of DNA/RNA and amide III mainly due to collagen. However, in the case of Raman spectroscopy, there are distinctly resolved bands for collagen, DNA bases and glycogen. It should be noted that these bands related to DNA were not observed in the FTIR spectra of this sample. Thus, a comparison of **Figure 3.4** and **Figure 3.5** indicates the rationale for preference of Raman micro spectroscopy over FTIR for the current studies. Moreover, the spatial resolution of Raman spectroscopy is less than 1µm while FTIR is~5-10 µm, offering the prospect of investigations at sub cellular level.

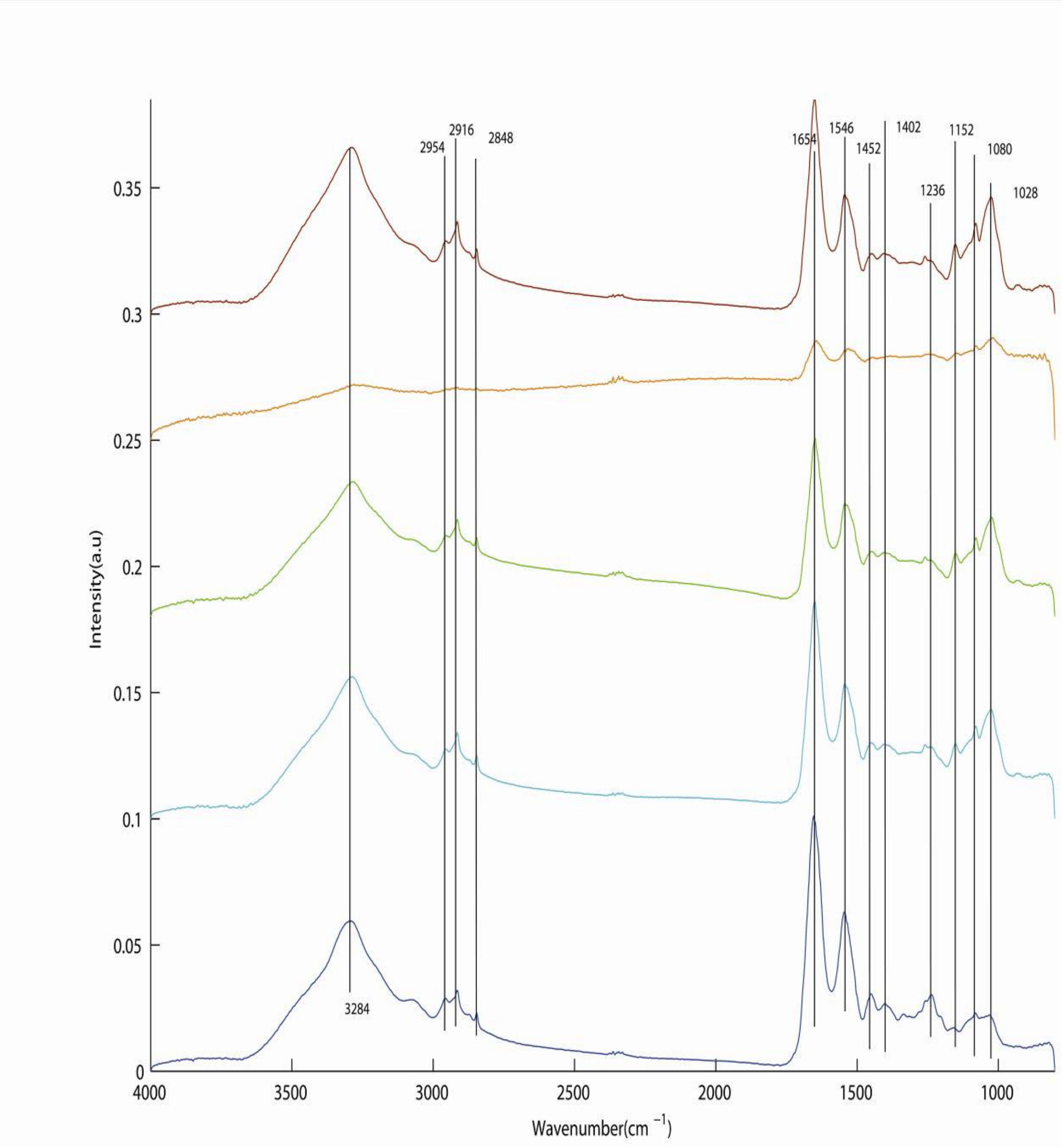


Figure 3.4 K-means cluster spectra for the FTIR map, blue cluster corresponding to stroma, brown to basal layer, green and mid blue colours to superficial epithelium and orange representing substrate of region in figure 3.3 c.

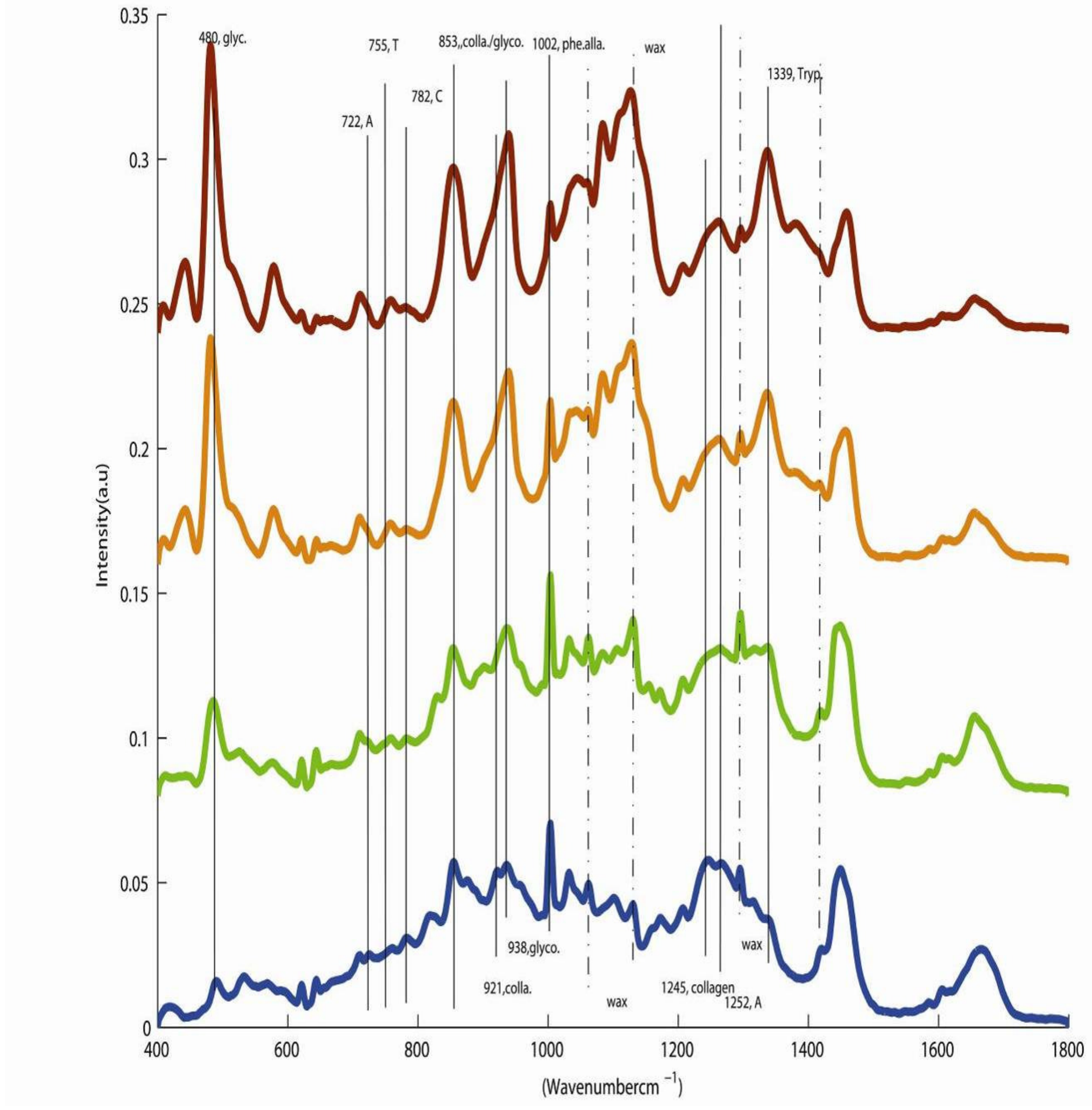


Figure 3.5 Mean Raman spectra from KMCA of Raman map, blue cluster corresponding to stroma, green to basal layer, brown and orange colours to superficial epithelium from region in figure 3.3 e.

Regarding the sample preparation and measurement conditions, there are certain parameters that need to be optimized, however. In the case of FFPP tissue sections, these include the dewaxing protocol, the thickness of the section cut from the block, the substrate for the adhesion of the tissue sections, the laser wavelength and the objective lens used to focus the laser and collect the Raman signals.

3.2 Optimisation of the parameters for Raman spectral Imaging

3.2.1 Introduction

Vibrational spectroscopy is a very sensitive technique and so precise sample preparation is a pre-requisite in order to acquire reproducible results and for standardisation between laboratories. The major advantage of this technique is that of minimal sample preparation. However, depending on the experimental set-up and the origin of the tissue, sample processing is often required. For the Raman analysis of *ex-vivo* specimens, the samples can be either snap frozen, transferred to an optimal cutting temperature (OCT) medium or formalin fixed and embedded in paraffin, in order to maintain the biochemical state of the tissue samples (Salzer and Siesler, 2009). Formalin helps to preserve the sample and paraffin serves as a support medium to cut thin tissue sections using a microtome. For the optical pre evaluation and selection of the sampling area, a parallel section can be prepared on a standard glass slide, H&E stained and analysed. Samples must be fixed and dehydrated, before embedding in paraffin, followed by cutting the sections. After sectioning, the paraffin must be removed by the use of an organic solvent such as xylene, hexane or cyclohexane for H&E staining, and for spectroscopic investigation, the paraffin itself can give strong spectroscopic signatures (Faolain et al., 2005, Salzer and Siesler, 2009).

The above mentioned sample processing steps may affect the biochemical composition of the tissue samples, e.g. lipids of the sample may dissolve in the organic solvents used for dewaxing, or protein can be denatured due to the high temperatures of 45ºC and 60ºC during section cutting or embedding. However, variations in the effects, if any, of the processing to the Raman spectral profiles can be minimised by keeping the heating temperature and time constant, and adopting a standardised processing protocol throughout the study (Salzer and Siesler, 2009).

Although, it has already been demonstrated that dewaxing in hexane for 18 hours is the most efficient procedure for removal of the paraffin from tissue sections (O Faolain et al., 2004), it was required during the present study to compare this protocol with the standard histopathology laboratory procedure for dewaxing in xylene. It was considered important for this study to be consistent with procedures used by the pathologist in the hospital.

3.2.2 Raman Spectral acquisition

The details of the Raman spectral acquisition have already been provided in (**section 3.1.2**). It should be noted that, for fresh tissue sections, Raman spectral acquisition was done by using an X50 long working distance objective lens. Spectral quality was monitored by observing the amide I peak and comparing it with spectral features of pure paraffin wax and solvents used to remove the wax.

3.2.3 Raman spectra of the wax and marking ink

Although, wax is expected to be removed completely after dewaxing the tissue samples with the suitable dewaxing agent, there may remain some residual wax which can contribute to the tissue spectra. The ink which is used by the pathologist to mark the tissue samples can also contribute to the Raman spectra of the sample of interest. The Raman spectra of pure wax and marking ink are presented in **Figure 3.6** and **Figure 3.7**, respectively, for reference.

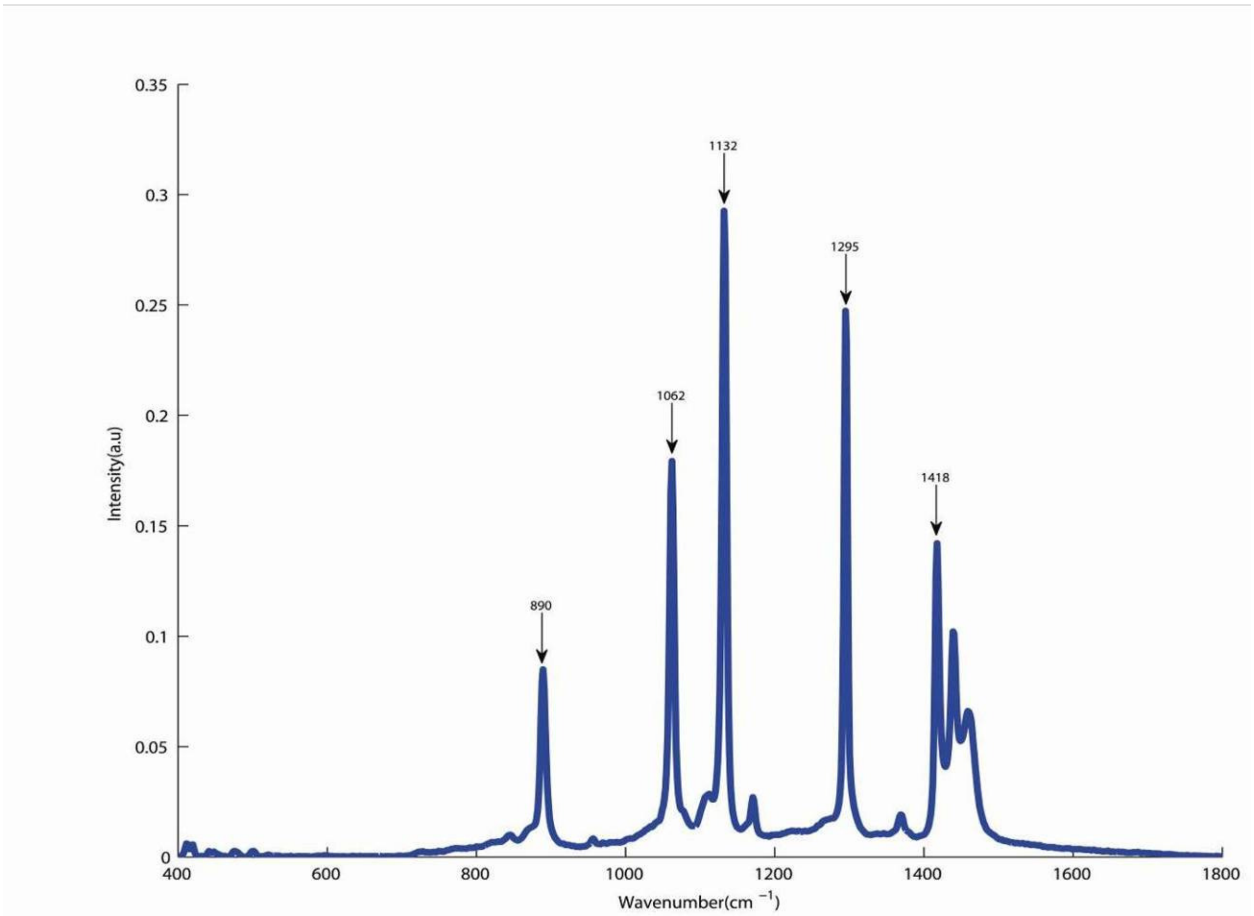


Figure 3.6 Raman spectrum of pure wax.

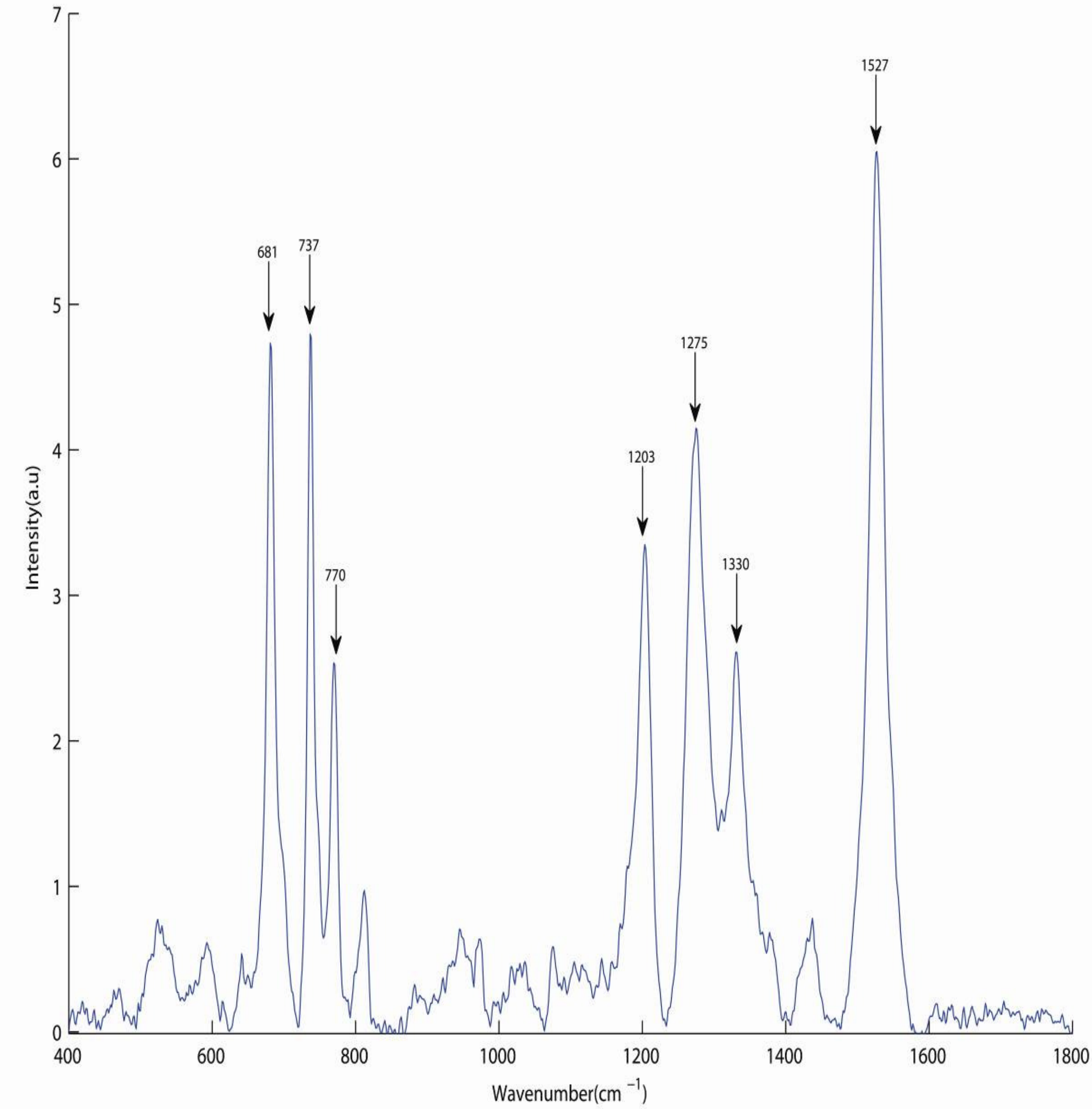


Figure 3.7 Raman spectrum of pure marking ink.

3.2.4 Selection of suitable tissue for optimisation study

Initially, Raman spectra were acquired from fresh liver, oesophagus and skin tissues of rat, to determine the tissue section that would be most suitable for the optimisation studies i.e. producing Raman spectra with low fluorescence / scattering or background noise and matching with the spectra of cervix. Raman spectra from the fresh tissue of rat kidney, liver oesophagus and skin are compared with a Raman spectrum from dewaxed cervical tissue in **Figure 3.8.** Skin was found to be the optimal tissue as it produced good quality spectra in terms of low background and similarity to the spectrum of cervical tissue. The dewaxing protocol was therefore examined using rat skin as the experimental specimen. The selection of the dewaxing agent can be done on the basis of the general rule which is "like dissolves like" (Burke, 1984). As wax is non-polar, so the solvent required for its removal from the tissue samples should be non-polar. Xylene and histolene are commonly employed in hospitals and for consistency with hospital protocols, for the current study, these two agents were included in addition to hexane, cyclohexane, ether and dishwashing liquid. With the exception of the dishwashing liquid, all these compounds are non polar and therefore should show very good solubility for paraffin wax, although they vary in performance. Their ability to dissolve the wax can be related to the Hildbrand solubility parameter value of the solvent (Barton, 1983). This value increases in the sequence hexane, ether, cyclohexane, xylene to histolene. It should be noted that hexane has already been proven to be very useful for such application (O Faolain et al., 2004) and xylene is in use in hospitals. On considering the solvency behaviour of these two, the other compounds including ether and cyclohexane lie between the above mentioned already in use dewaxing agents and the rationale for their selection for the current study is that they could also potentially behave as dewaxing agents as their Hildebrand solubility parameter values lie between the hexane and xylene. Moreover, dishwashing liquid was selected because of its degreasing ability.

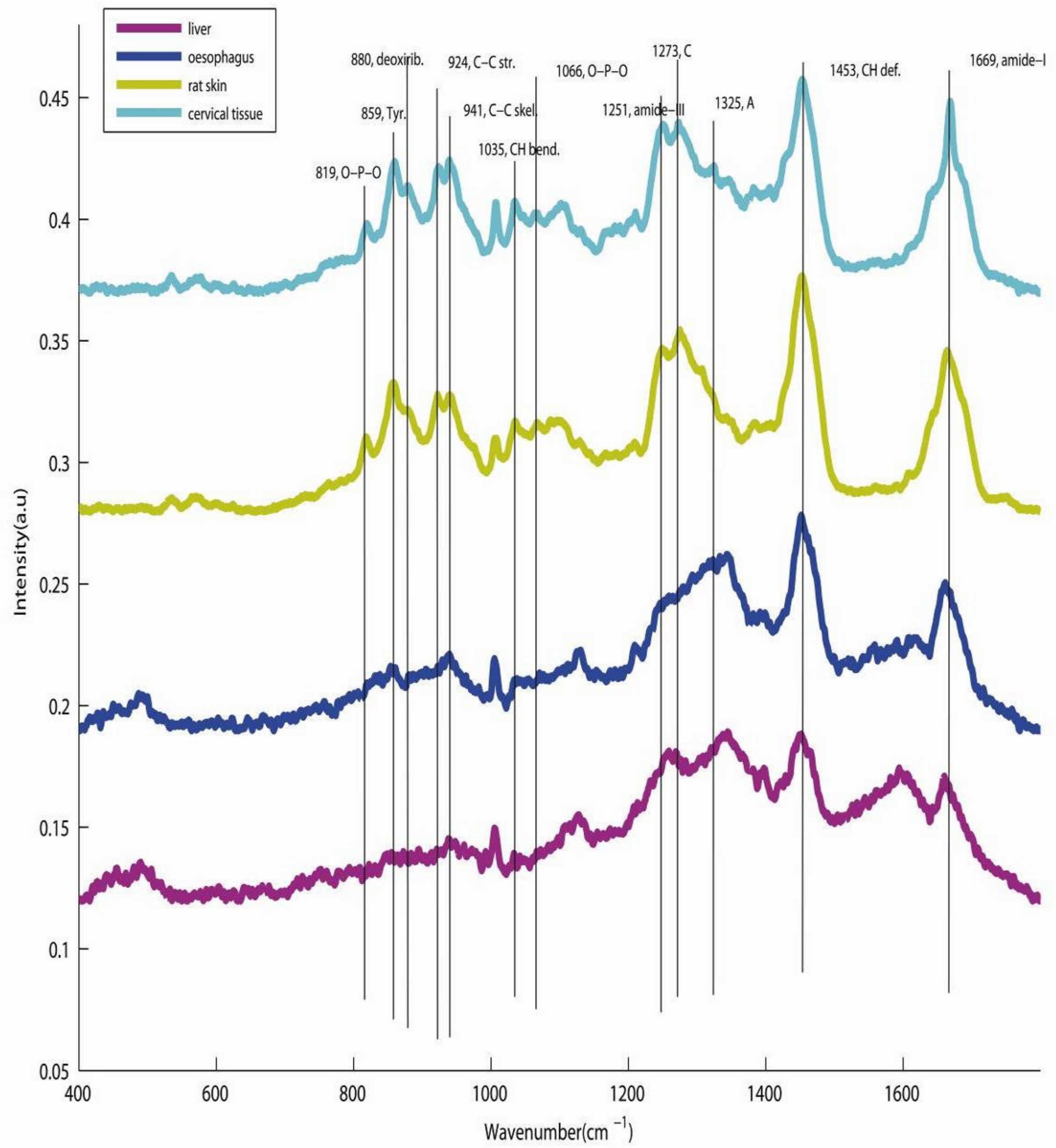


Figure 3.8 The Raman spectra of different organs of the rat compared with one from the cervical tissue section.

3.2.5 Sample preparation-fixation, waxing and section cutting,

The rat skin tissue was fixed in formalin, dehydrated through various grades of alcohol and embedded in paraffin using in house facilities. A microtome (Leica RM 2135) was used to cut the sections and these were mounted onto quartz slides. These sections were dewaxed using (a) ether; (b) cyclohexane; (c) xylene; (d) histolene; (e) hexane; (f) 1.7% solution of dishwashing liquid. The tissue sections were immersed in two baths of dewaxing solvent for 5 and 4 minutes followed by two baths of ethyl alcohol for 3 and 2 minutes and one bath of methylated sprit. Dewaxing solvents were used to immerse the tissue sections for a further 18 hours to remove any remaining wax. In the case of the dishwashing liquid solution, tissues were washed in two baths of soap solution at 90ºC for one minute each and then twice washed in distilled water at 90ºC and once at 45ºC for 30 seconds. After soaking in distilled water at room temperature, the tissue sections were oven dried for 5 minutes at 60ºC. Raman spectra were acquired from formalin fixed and paraffin preserved (FFPP) fresh rat skin tissue samples dewaxed in the above mentioned solvents.

3.2.6 Results and Discussion

Figure 3.9 shows the spectra from a fresh tissue section and from a piece of bulk tissue from the rat skin. The most prominent Raman bands are labelled and include phenylalanine (1005 cm^{-1}), lipids (1451 cm^{-1}) and amide-I (1670 cm^{-1}), amide-III (1273 cm^{-1}). The Raman spectra from the fresh piece of skin are noisy due to the scattering of the laser from the uneven surface of the tissue (Bonnier et al., 2011). The Raman spectrum from the fresh tissue section has more defined Amide I, phenylalanine and lipid bands. Due to this reason, the fresh rat skin "section" was selected for the comparison to dewaxed skin tissue sections in this chapter.

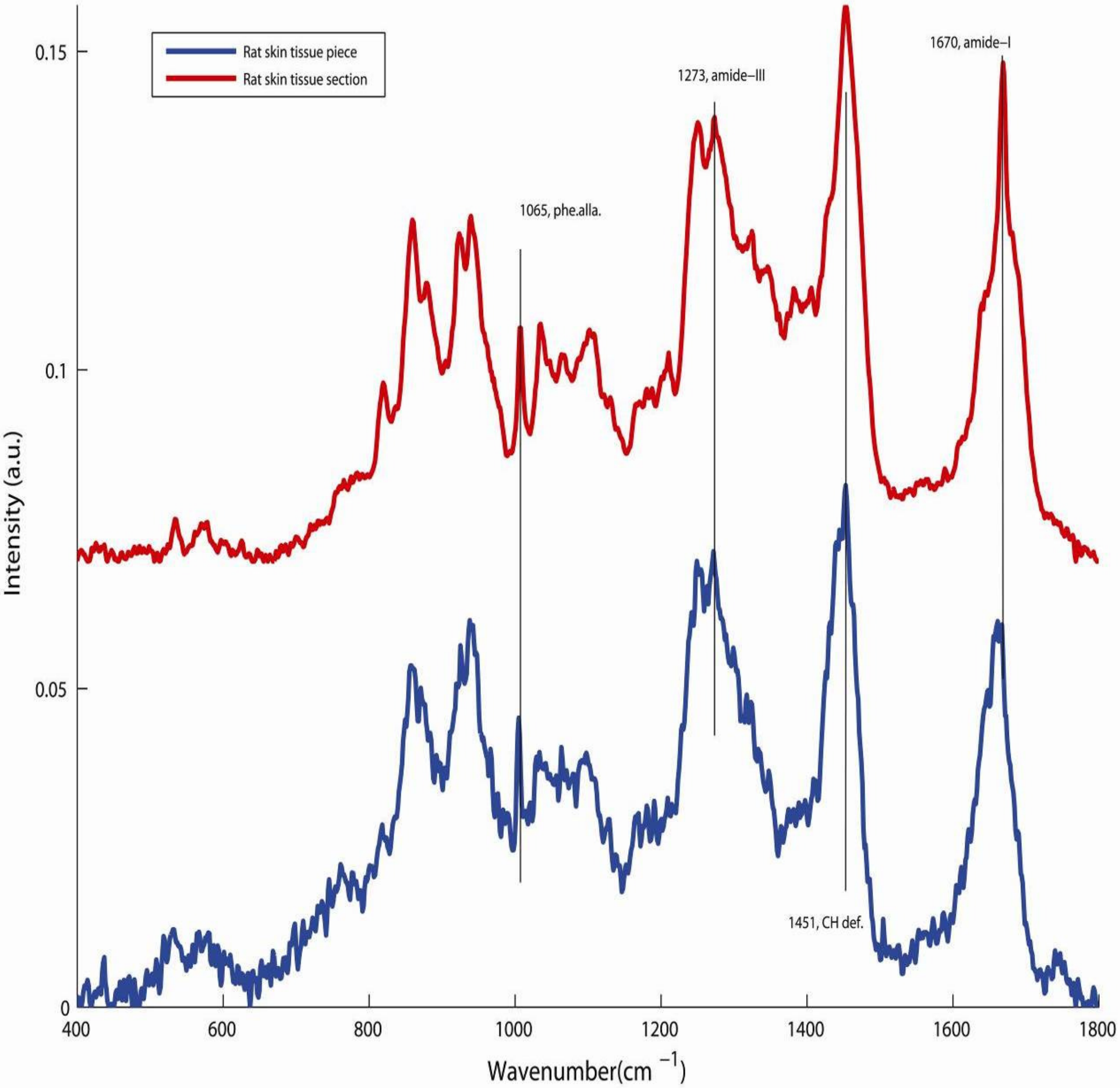


Figure 3.9 Raman spectra of rat skin tissue piece and tissue section.

A comparison of the spectra of the rat skin tissue section with the spectra obtained from rat skin after formalin fixation, paraffin preservation and dewaxing in different solvents is presented in **Figure 3.10**. The dishwashing liquid results in the inefficient removal of wax and Raman spectra have paraffin bands at 1061, 1131, 1293 cm^{-1} and 1439 cm^{-1}. In the case of cyclohexane, the bands present at 787 and 802 cm^{-1} in the Raman spectra can be assigned to cyclohexane itself and thus overall spectral quality was compromised due to these bands. The use of ether did not appear to be suitable as a dewaxing agent as it interfered with the proteins in the tissue section and resulted in the deformation of amide I band. The Raman bands of the amide I and phenylalanine were not as sharp when compared with the spectra obtained after dewaxing in hexane or xylene. The xylene and hexane were determined to be good dewaxing solvents. Histolene did not show any better dewaxing than xylene, most probably because Histolene is 80% xylene in composition.

From these observations, it can be concluded that xylene can be adopted as solvent of choice to keep the protocol consistent with the one used in the hospital histopathology laboratory.

It should be noted that in the previously reported protocol (O Faolain et al., 2005), which included 18 hour soaking in the dewaxing solvent in addition to the normal protocol, was followed for the initial studies (all experiments presented in this chapter) to select a solvent of choice. For the final protocol to be adopted for the further experiments, **chapter 4**, xylene is selected as solvent of choice after the initial experiments reported above. The soaking for 18 hours has not been used, for consistency with the protocol used in the hospitals.

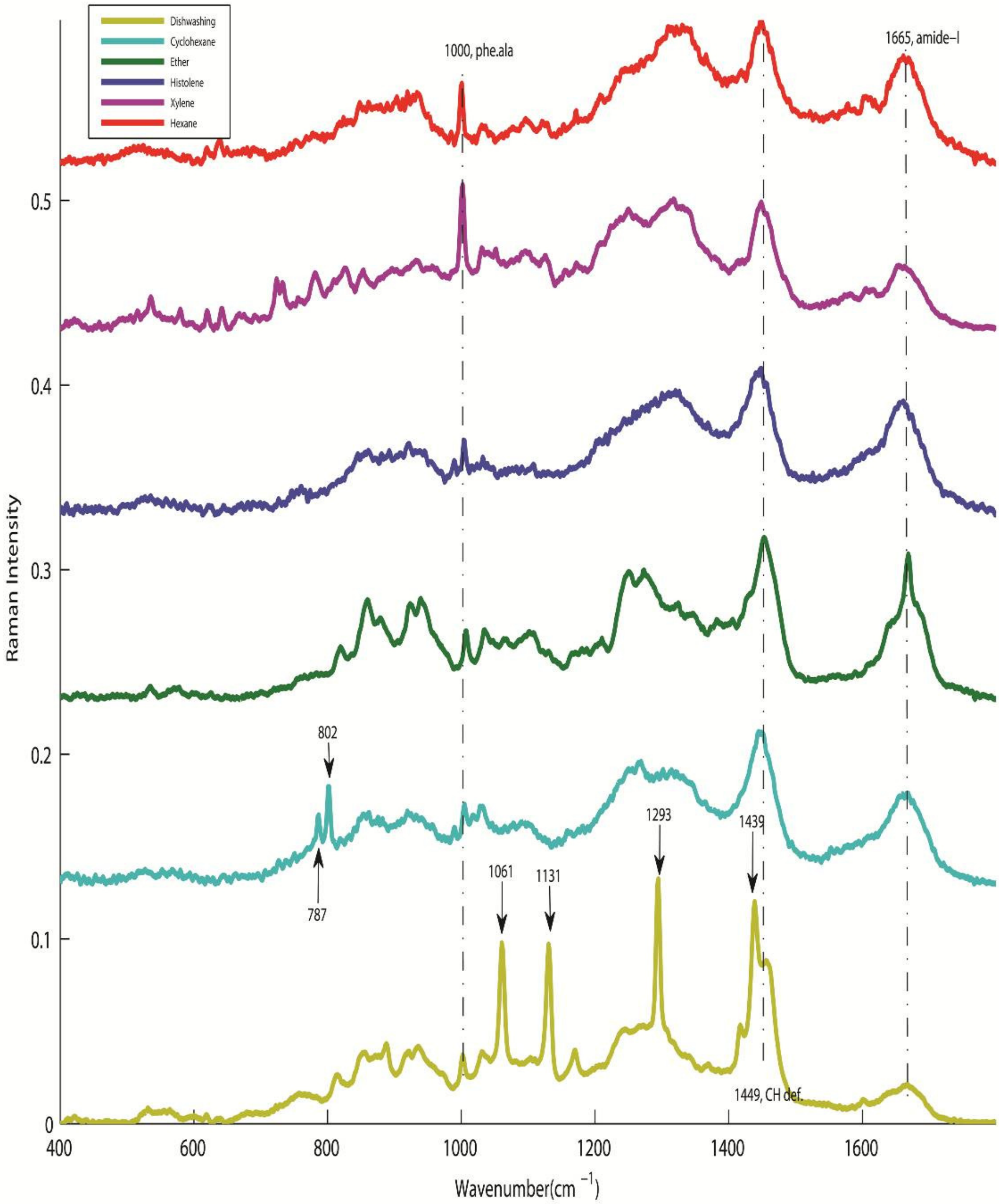


Figure 3.10 Raman spectra of the fresh rat skin after dewaxing in different solvents.

3.3 Laser line selection

3.3.1 Introduction

The choice of laser wavelength is of critical importance for the observation of the spectra of biomaterials without the related fluorescence emission. Raman and Rayleigh scattering depend on the inverse of wavelength of incident light to the power of λ_{in}^{-4}, suggesting that the choice of source should be that of shortest wavelength. However, if the sample of interest is resonant with the incident wavelength, fluorescent emission may arise, which can dominate the Raman signal on the Stokes side. Similarly, photo degradation can occur at shorter wavelengths when there is absorption (Byrne et al., 2011). Notably, in biological tissue, in the absence of melanin, there is significantly less fluorescence at wavelengths higher than ~600 nm (Kollias et al., 1998, Na et al., 2001). The large background observed at wavelengths higher than this has been proposed to arise from stray light due to scattering (Bonnier et al., 2010a). For a given sample, the applied laser power and time allowed for signal collection are the main factors responsible for high signal to noise ratio (S/N) of a Raman spectrum. An increase in laser power will improve the S/N ratio in a given interval of time but may also increase the risk of sample photodegradation. Therefore, it is essential to explore the stability of the sample in relation to the intensity of the laser source and to assess the possibility of sample damage (Gremlich and Yan, 2001). Different groups have investigated biochemical changes caused by different cancers using Raman spectroscopy with different laser lines. Different studies include, analysis of the normal and diseased human breast tissue samples (Frank et al., 1995), differentiation between the normal and transformed (Jurkat and Raji) human lymphocytes using a 633nm laser line (Chan et al., 2006), lung tissue of normal and infant patient using a 785 laser line (Krafft et al., 2008a) differentiation between the cultured normal and malignant cervical and breast cells (Yazdi et al., 1999) and diagnosis of cancer in human cervical tissue using 514nm (Lyng et al., 2007), and 785 nm (Kanter et al., 2009a, Krishna et al., 2006).

For the present study, longer wavelengths such as 532nm, 660nm and 785 nm were investigated to allow mapping to be carried out without photodegradation of the tissue.

3.3.2 Methods

Cervical tissue sections of 5µm thickness were mounted on Mirr-IR slides and dewaxed in hexane for 18 hours. A HORIBA Jobin Yvon HR 800 UV confocal Raman microscope equipped with lasers of wavelength 532, 660 and 785nm was used to acquire the spectra. Other parameters include a grating of 300gr/mm, a 100µm confocal hole, 100% filter and X50 objective lens with numerical aperture of 0.75. The acquisition time was varied from 5, 10, and 30 seconds with 2 acquisitions per spectrum for 532, 660 and 785nm laser lines respectively. The laser power measured at source was found to be ~80 mW for 532 nm, ~90 mW for 660 nm and ~300 mW for 785 nm. The acquisition time varied from one laser to another. Changing from stronger laser power to weaker did not allow the same acquisition time for the three laser lines as prolonged irradiation with stronger laser power resulted in photodegradation of the tissue section.

3.3.3 Results and Discussion

Figure 3.11 represents the Raman spectra obtained from a dewaxed cervical tissue section using three different laser lines. The background at 532 nm and 660 nm laser lines were high but low background was observed using 785 nm. This could be indicative of fluorescence. The background has been removed in Figure 3.6, but the spectra are offset for clarity. The spectrum acquired using the 532 nm laser line (red) has high signal to noise ratio as compared to the 660 nm laser line spectrum (green). This spectrum has normal biological peaks at 620, 642 cm^{-1} (C-C twisting of proteins), 763 cm^{-1} (O-P-O DNA), 850, 940 cm^{-1} (collagen) 1002 cm^{-1} (phenylalanine), 1318 cm^{-1} (thymine), 1340 cm^{-1} (adenine), 1451 cm^{-1} (CH deformation of lipids or proteins) and 1665 cm^{-1} (amide I).

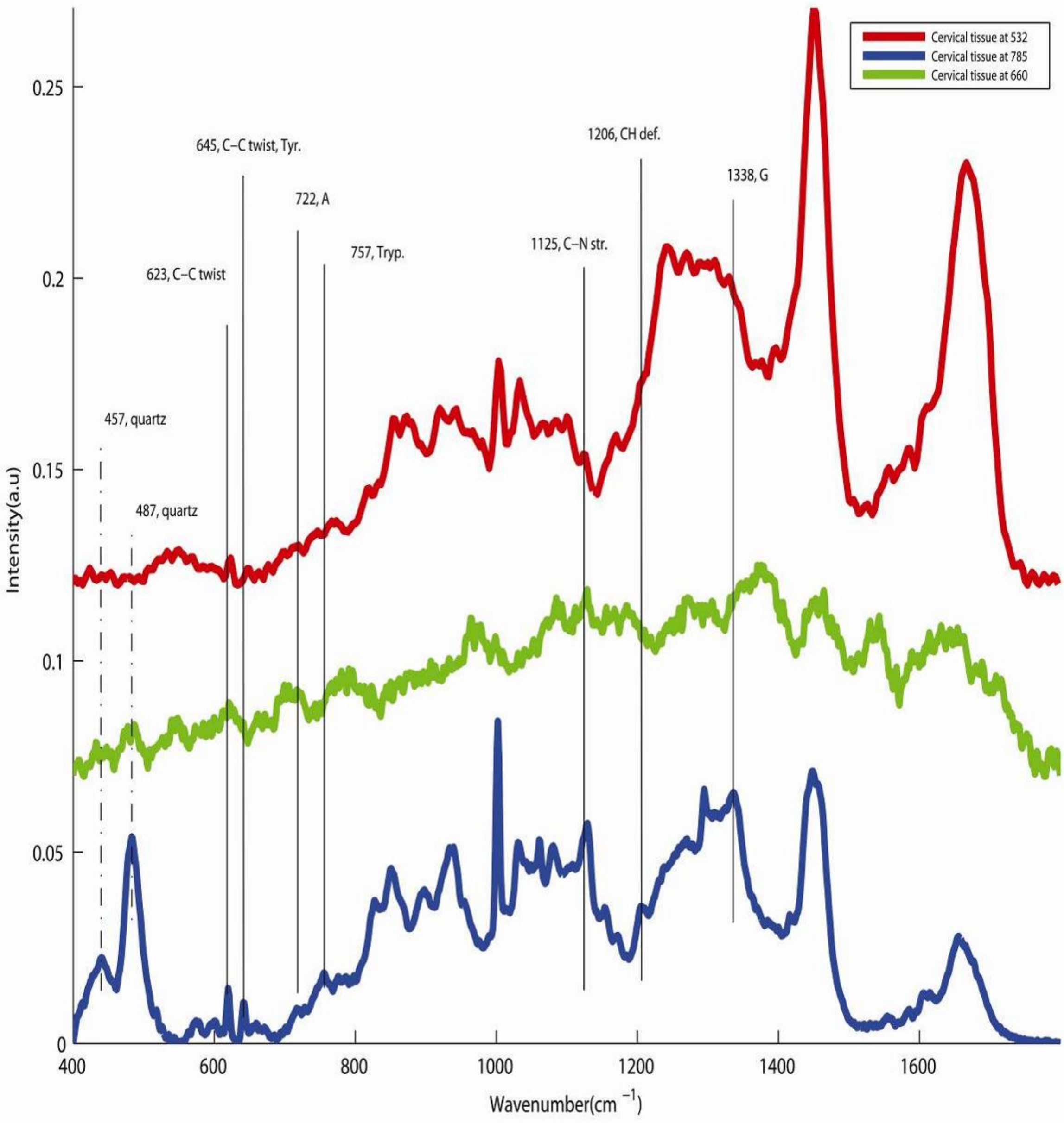


Figure 3.11 Raman spectra of the cervical tissue section acquired with different laser lines.

In **Figure 3.11**, the Raman spectrum acquired from the cervical tissue samples by the use of 785 nm laser line (blue), has more prominent bands as compared with the spectrum obtained with the 532 nm laser line, including 620, 642 cm^{-1} (C-C twisting of proteins) 849 cm^{-1}, 938 cm^{-1} (collagen), 1002 cm^{-1} (phenylalanine) amide III and amide I. The shift of the excitation source from shorter wavelength towards longer wavelength resulted in the minimization of background from tissues as demonstrated in literature and experimentally (Carter and Edward, 2001). It is important to mention that there is no contribution of the two bands of quartz, 457 cm^{-1} and 487 cm^{-1} (dotted lines), in the biological spectra acquired with the 532 laser line as compared to those recorded using 785nm.

The Raman spectrum acquired with the 660nm laser line appears to have a low signal to noise ratio. This may be due to the fact that the Raman signal depends on $1/\lambda^4$ of the laser used and the intensity of the laser. This means the Raman signal will decrease significantly with increasing wavelength i.e. moving from 532, 660, to 785nm. However, in the experimental set up used in these studies, the intensity increases with increasing wavelength. As a result of these two effects, a good Raman signal is expected from 532 nm (~80mW) and 785 nm (~130mW) laser lines and poor signal from 660 nm (~90mW) laser line. Thus with the 660 nm laser line, no meaningful information could be obtained even after the normalization and baseline correction of the spectra.

On comparing the results of the three laser lines 532 nm, 660 nm and 785 nm, the Raman bands are better defined for the 785 nm laser line as compared to the 532nm laser line. All these Raman bands are not clear for 660 nm laser line indicating that it is the least suitable of the three for the current study. Although, the 532 nm laser line also produced reasonably good Raman spectra of the tissue sample, the major concern in the use of this laser line was photodamage of the tissue section during mapping and was observed experimentally during the point spectra acquisition using the 660 nm and 532 nm laser lines. It can be concluded that the 785 nm laser line is most appropriate for Raman mapping of cervical tissue samples.

3.4 Substrate Optimisation

3.4.1 Introduction

The next issue was the selection of an appropriate substrate for mounting the tissue section in order to carry out Raman spectroscopy. **Figure 3.12** represents the Raman spectra of the cervical tissue on CaF_2 and quartz along with the spectra of the substrates at the 785 nm laser line. The Mirr IR substrate is also shown for comparison.

The broad background from the Mirr IR slide under 785nm laser line dominates the Raman spectrum and no spectral features from the tissue sample could be discerned and thus it is not shown in **Figure 3.12**. On the other hand, the spectra of tissue on the quartz substrate contain the peaks at 426 and 790 cm^{-1} and the Raman spectra of the tissue contains a broad band at about $1358cm^{-1}$ with the 785 nm laser line.

The Mirr-IR substrate gives rise to a large background under 785nm laser illumination and as such was not considered suitable. The quartz and calcium fluoride (CaF_2) substrates do not produce a significant background with the 785nm laser line. Furthermore, both of these can also be used with FTIR and Raman spectroscopy. The CaF_2 contribution in the biological spectrum is negligible as compared to a large band from quartz. Thus, CaF_2 was selected as substrate for further experiments with Raman micro spectroscopy.

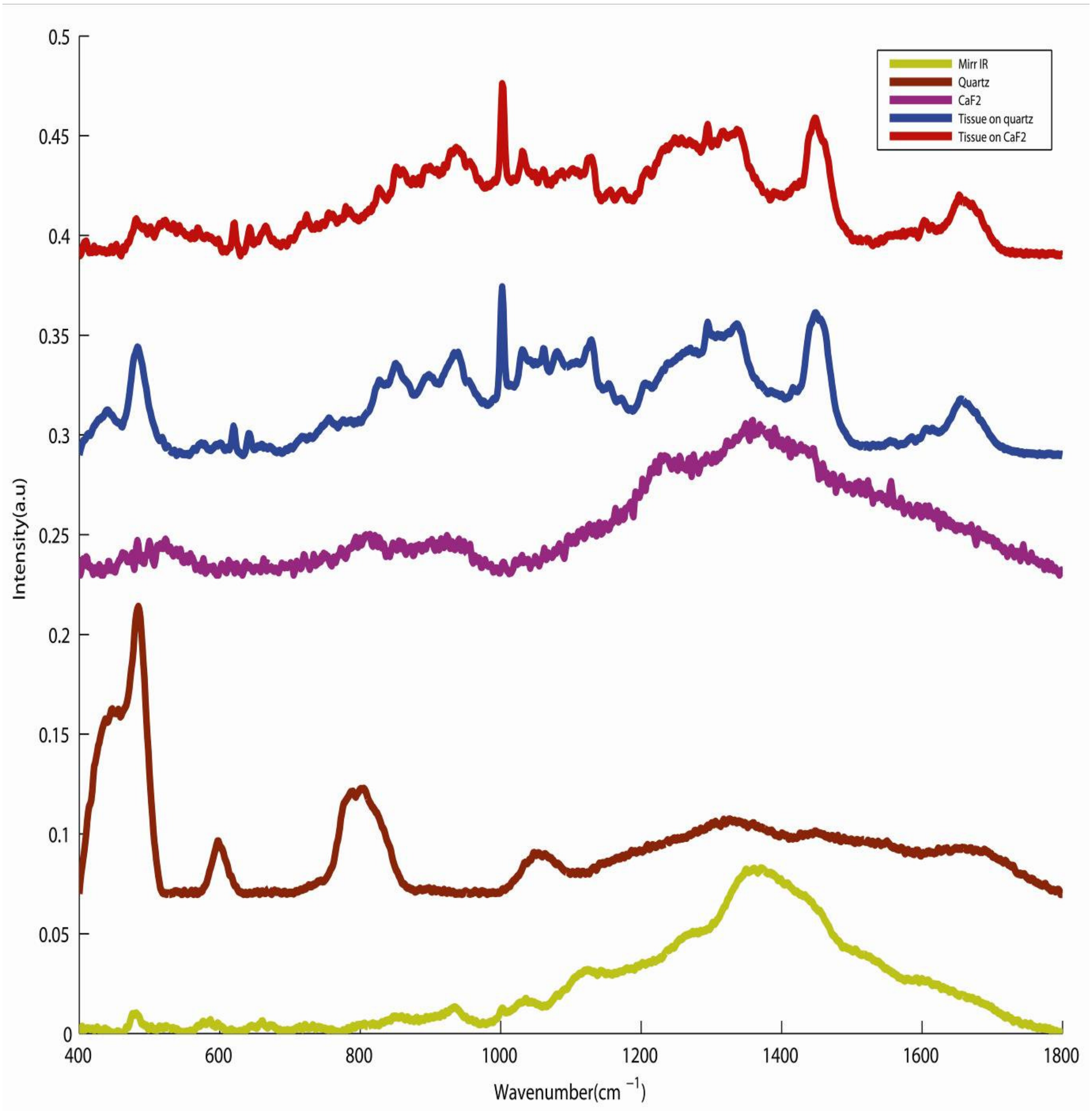


Figure 3.12 Spectra from cervical tissue on quartz and CaF_2 slides and substrates using 785 nm laser line.

3.5 Objective Optimisation

3.5.1 Introduction

Once the tissue processing protocol, laser wavelength and substrate for the mounting of the tissue section was optimized, the next target was to decide on the objective lens used to focus on the tissue section. The option of objective lenses of various magnification and numerical aperture was available i.e. X50, X100 dry, X60 and X100 water immersion objectives with numerical aperture 0.75, 0.90, 1.0 and 1.0 respectively. They are used to focus the laser on the sample to be analysed and to collect the back scattered light. The aim of the study was to evaluate the potential of Raman mapping using objectives of different magnification to differentiate between the layers of normal cervical epithelium on the basis of spectral differences obtained from the respective layers.

3.5.2 Methods

Three parallel tissue sections of 10 µm thicknesses were cut using a microtome, one for H&E staining and two unstained for Raman spectroscopy. The latter two samples were mounted on CaF_2 slides. The tissue sections were dewaxed using the protocol already mentioned in section **3.2.5**.

The point spectra and maps were acquired from the normal region of CIN II cervical tissue sections using lenses of different magnification. The tissue sections mounted on CaF_2 slides were fixed to a normal glass slide with blue-tack for Raman spectral acquisition using the X50 and X100 dry objective and to a plastic Petri dish for spectral acquisition with the X100 water immersion objective. Deionised water was used as an immersing medium for the water immersion objective. Raman spectra from the CaF_2 slide lying above the Petri dish and glass slide were also recorded.

3.5.3 Results and Discussion

3.5.3.1 Point spectral acquisition

Figure 3.13 shows the Raman spectra acquired from cervical tissue sections using different objectives. The decrease in spectral background noise is evident from the observation of the bands labelled with solid lines assigned to DNA bands in the spectra, from dry objective lens to water immersion objective lens and lens of lower magnification to higher magnification because there is a shift in numerical aperture from 0.75, 0.90 for dry or air objectives to 1.00 for immersion objectives. The higher the numerical aperture, the more signal will be collected from the area of interest.

$$\text{NA (numerical aperture)} = n\sin\theta \qquad \text{(Equation 3.1)}$$

As indicated by the **equation 3.1**, this depends on refractive index and intake angle of the lens, so the greater the NA the more the signal collected from the sample.

Immersion of the tissue section in water makes the tissue surface less scattering, resulting in better defined peaks, as shown by (Bonnier et al., 2010b). The comparison of spectral bands at 1665 cm^{-1} for amide I and other structural proteins such as for collagen 856 and 938 cm^{-1} shows an improvement in spectral quality from X50 dry (green spectra) to X100 dry (sky blue spectra) and to X60 and X100 immersion objective (purple and red spectra).

On the basis of these findings, both the X100 dry and X100 immersion objectives were found to be most suitable for the tissue samples for point spectral acquisition. Moreover, for X100 immersion, the Raman spectra are clearer as compared to X100 dry, even without the baseline correction and normalization, as shown in **Figure 3.14**.

It should be noted that from these two objective lenses, the X100 dry objective lens was selected for further experiments, Raman mapping, in the current study. This was due to the reason that the tissue samples, if needed, could be used again and again for mapping and the dry lens would avoid the problem of immersing the tissue sections in the water repeatedly. Furthermore, it was found that the sample positioning was less stable in the immersed state, and thus it was less suitable for prolonged mapping times.

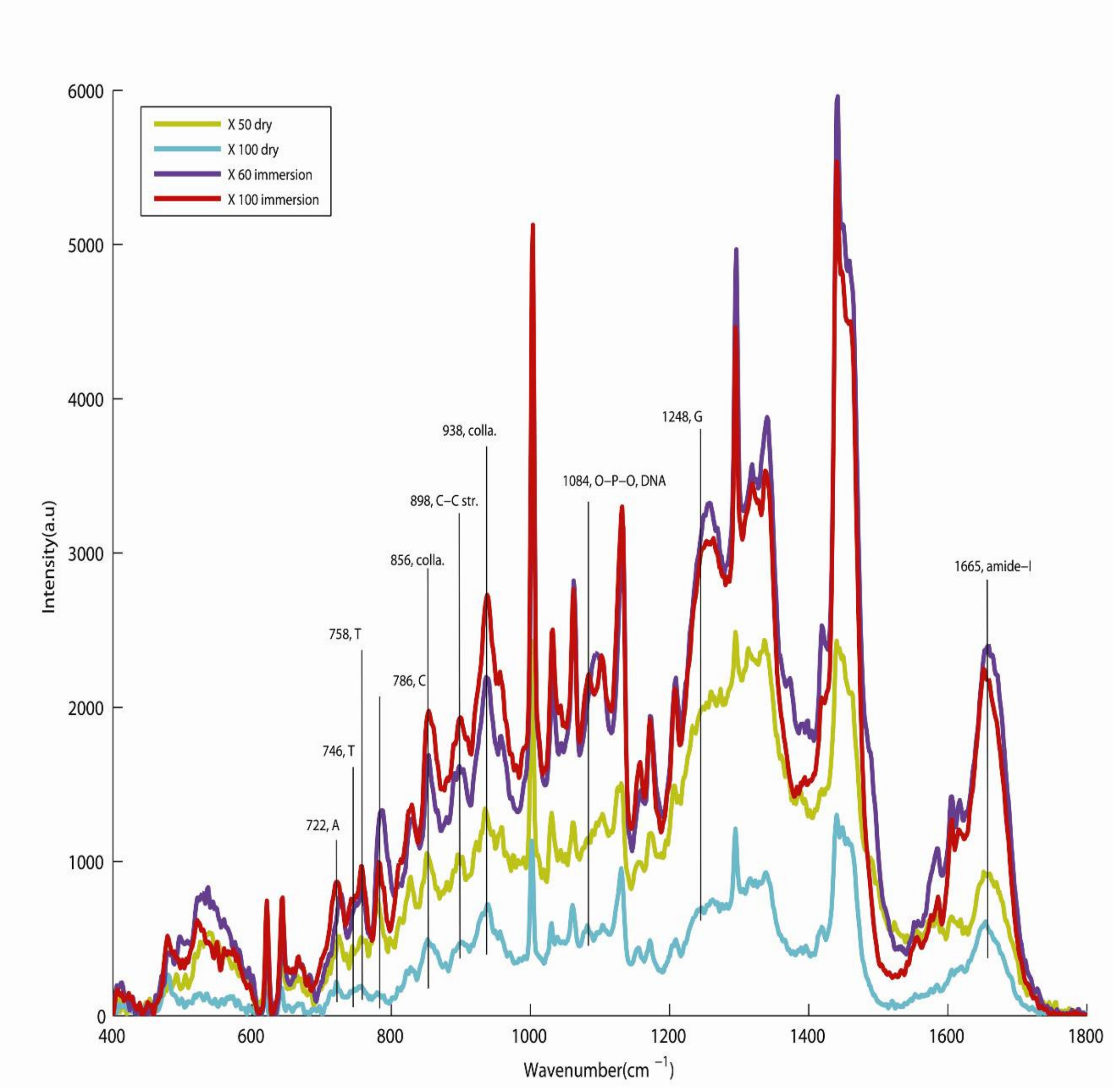


Figure 3.13 Raman Spectra from cervical epithelium using different objective lenses green (X50 NA 0.75) mid blue (X100 NA 0.90) magenta (X60 immersion, NA 1.00) red (X100 immersion, NA 1.00.

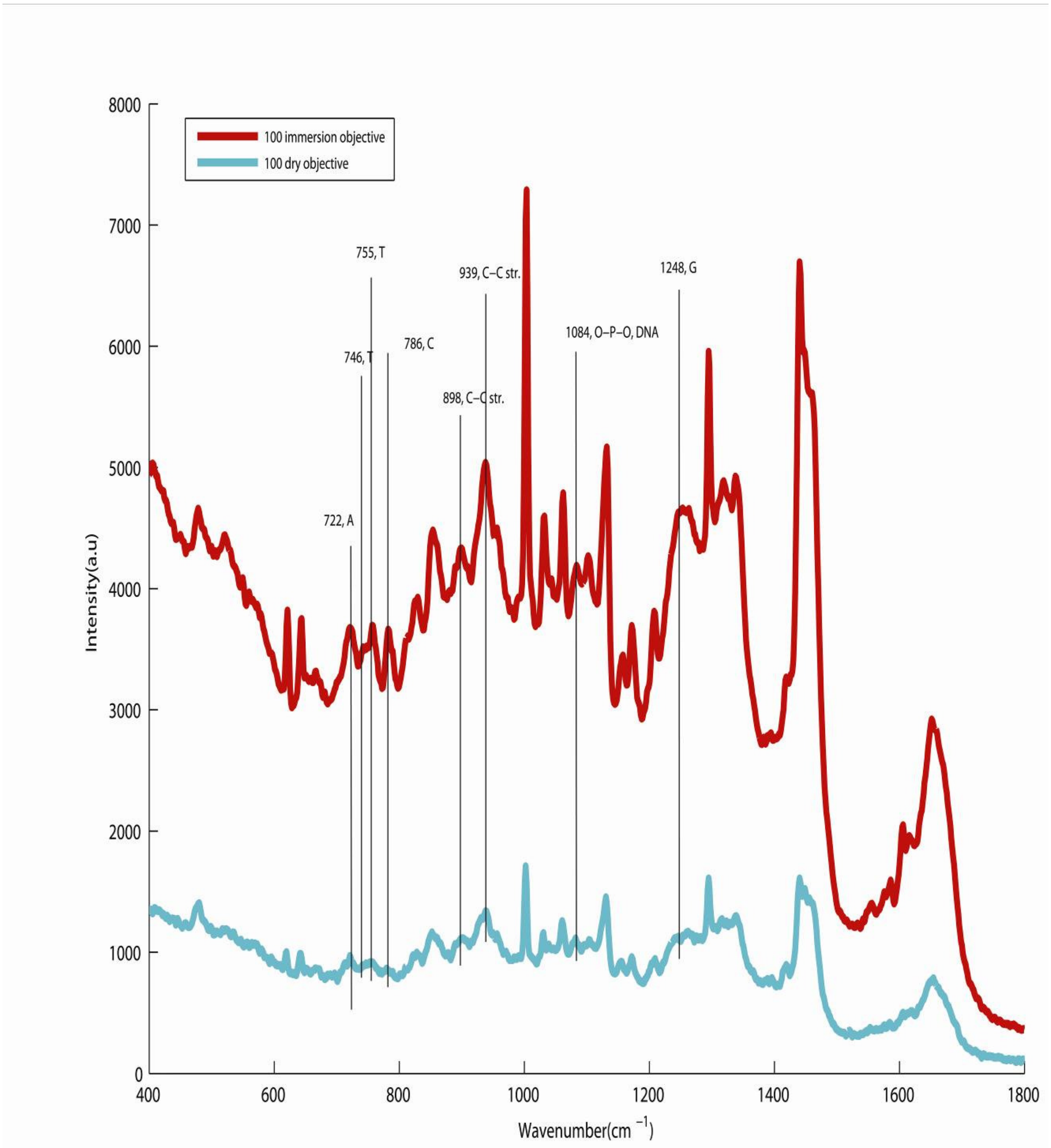


Figure 3.14 Raman Spectra from cervical epithelium tissue section using X100 dry and X100 immersion objectives.

3.5.3.2 Raman spectral mapping

The same procedure was tried for mapping of the cervical epithelium from superficial cells to stroma. The aim of acquiring these maps was to get an initial understanding of the layer separation of cervical epithelium on the basis of collagen, deoxyribonucleic acid (DNA) and glycogen. Also, in order to choose the objective lens that could best reveal the layer separation.

The normal cervical epithelium along with the stroma from an area distant to the CIN II area, as marked by the pathologist, was analysed by Raman micro spectroscopy. **Figure 3.15** presents the results of a preliminary map with a step size of 5µm and acquisition time of 15 seconds x2 per spectrum using X50 magnification objective. **Figure 3.15 A** and **Figure 3.15 B** show images of the H&E stained tissue section taken with a X10 objective and an extended video image of the unstained tissue section with X50 objective. The red box corresponds to the area scanned by the Raman microscope. KMCA was used to classify the spectral data set obtained from the Raman map.

A Raman map of the tissue layers was generated by the use of KMCA, as shown in **Figure 3.15 C**. The Raman map is clustered into five clusters with a different colour being assigned to each cluster. Average Raman spectra representing the respective colour cluster/group are shown in the **Figure 3.16**. The comparison between the KMCA of the Raman map and average spectra obtained as a result of this KMCA shows that the average plot of the mid blue cluster corresponds to the spectrum of the substrate obtained experimentally. Based on literature, (Jess et al., 2007, Notingher, 2007, Notingher et al., 2003, Lyng et al., 2007) the Raman bands in the green plot can be assigned to thymine (480cm^{-1}), adenine (722cm^{-1}), collagen (850, 940cm^{-1}) and phenylalanine (1002cm^{-1}). This indicates that this cluster might be from connective tissue or the basal and parabasal layer. The predominant bands (480, 849 and 938 cm^{-1}) in the blue, orange and brown average Raman spectra of the respective clusters are assigned to glycogen, suggesting different layers of epithelium. The peaks from all spectra at 1061 cm^{-1}, 1133 cm^{-1}, 1294 cm^{-1}, 1339 cm^{-1} correspond to peaks of the Raman spectra of paraffin wax obtained experimentally. This Raman map was able to differentiate between basal, parabasal and squamous epithelium with varying concentration of glycogen.

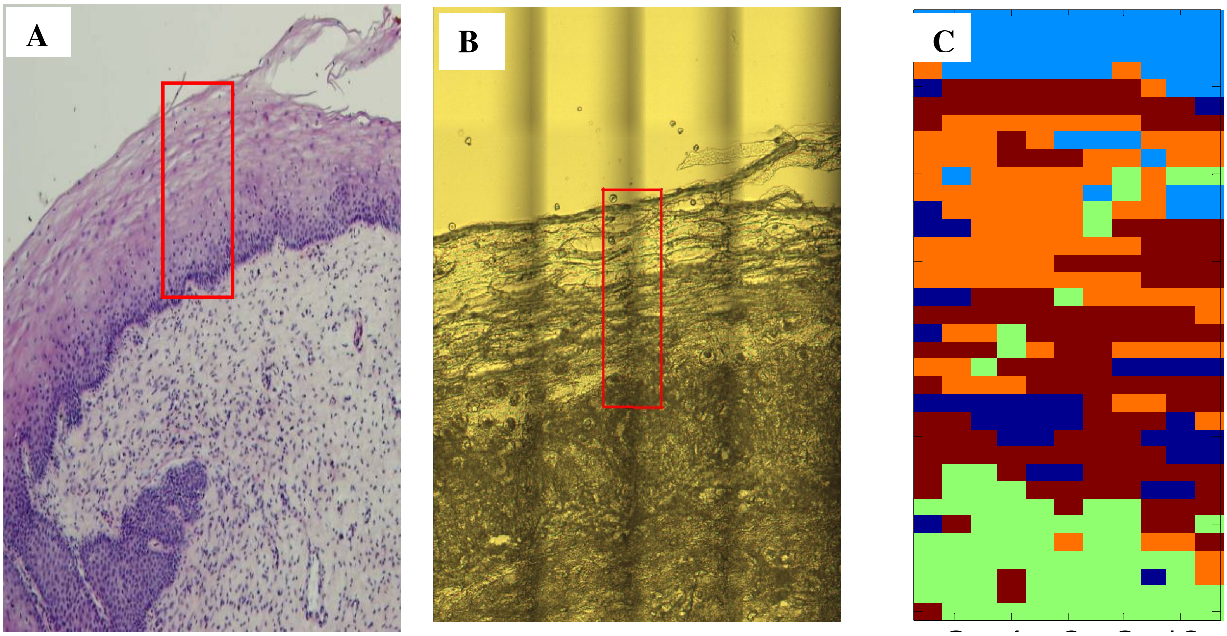


Figure 3.15 H&E stained de-waxed cervical epithelium (A) Unstained cervical epithelium under Raman micro spectroscope with selection showing the area for mapping with X50 dry objective (B). Five cluster K-means analysis of Raman map (C).

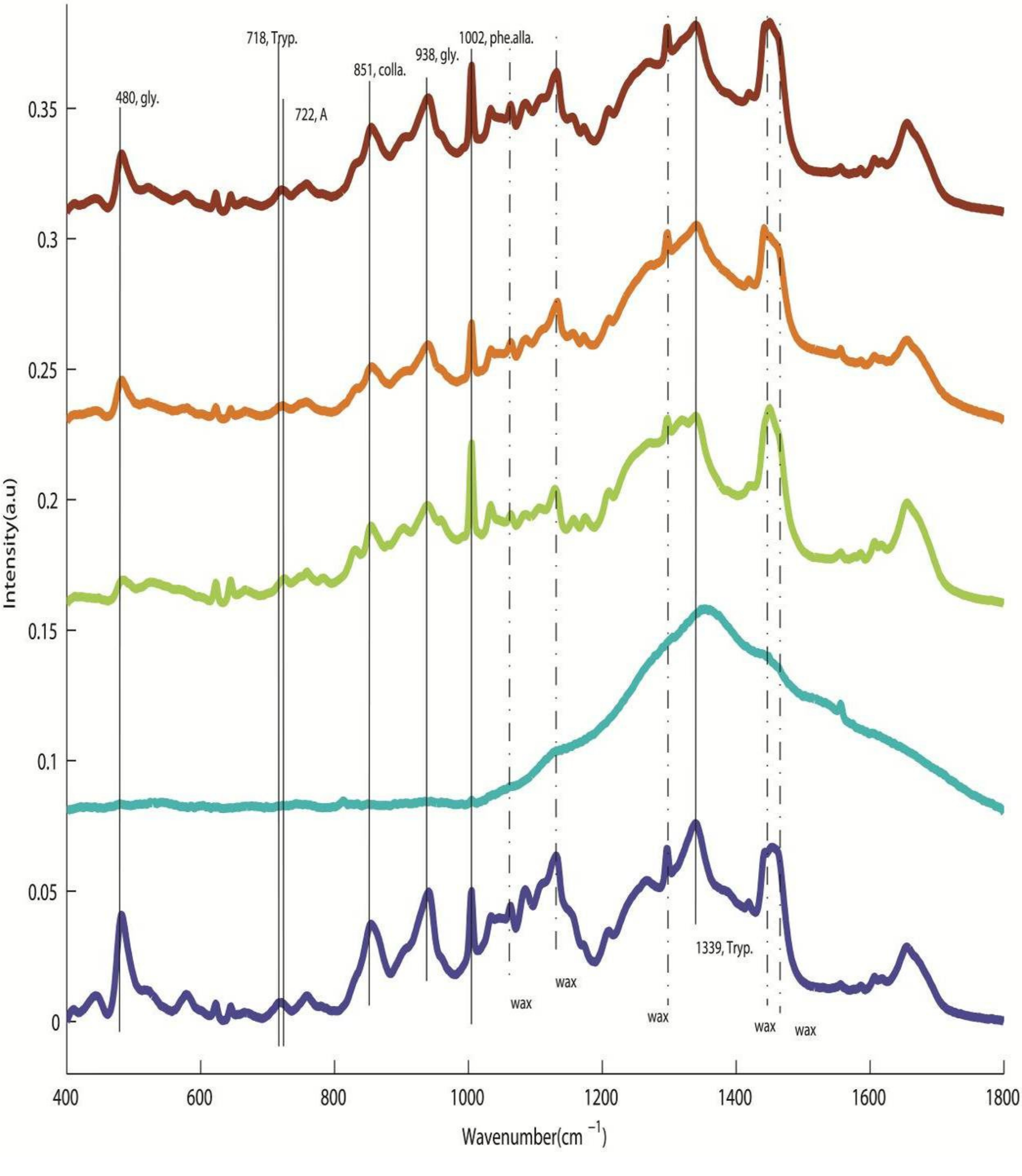


Figure 3.16 Mean spectra of K-means cluster analysis, green representing stroma, orange to basal, brown & blue to superficial layer and mid blue to substrate of the Raman map with X50 dry objective.

Figures 3.17 A & B show the extended video image of the unstained tissue section of the same sample, as in **Figure 3.15 B,** with X100 dry objective, the red box corresponding to the area scanned and the Raman map of tissue layers obtained as a result of the KMCA. The Raman map is clustered into five clusters with a different colour being assigned to each cluster.

Mean Raman spectra representing the respective colour cluster/group are shown in **Figure 3.18**. The KMCA of the Raman map generated five well separated clusters. There exists a very good correspondence between the KMCA of the Raman map and the unstained section presented in the **Figure 3.17 B**. The predominant Raman bands in the sky blue plot (850, 940 cm^{-1}), the green plot (729 cm^{-1}), the orange and brown plots (480, 849 and 938 cm^{-1}) can be assigned to collagen, DNA base adenine, and glycogen respectively. The Raman plot of the blue cluster corresponds to the spectrum of the substrate. This Raman map divides the cervical tissue into connective, basal or parabasal and epithelial layers of different maturity. A contribution of the paraffin wax at 1061, 1133, 1294, 1339 cm^{-1} is also seen in the Raman spectra.

The experiments of objective optimisation were conducted by employing X50 dry, X100 dry and X60, X100 water immersion objectives for the Raman spectral acquisition. Raman spectral features become clearer in the spectra acquired with X100 dry and X100 immersion objectives as compared to X50 dry objective as shown in **Figure 3.13**. Moreover, similar spectral features were obtained with both X100 dry and X100 immersion objectives but the spectra acquired by the use of the X100 immersion had lower background as compared to the X100 dry. This is due to the reason that immersion of the tissue section in the water makes the tissue surface less scattering and results in better defined peaks. It should be noted that the immersion lens did not allow to repeat the experiment if needed for further investigations, however. Use of the dry lens would avoid the problem of immersing the tissue sections in the water repeatedly. So, from these two objective lenses, the X100 dry objective lens was selected for further experiments of Raman mapping, in the current study.

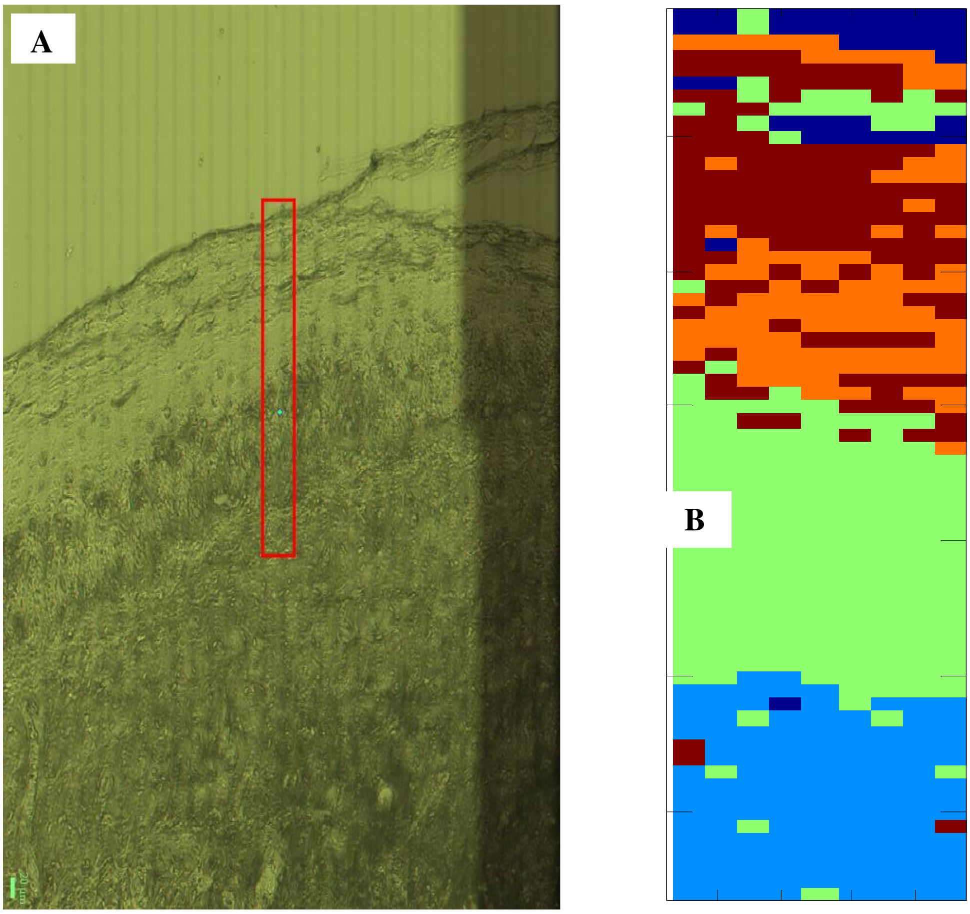


Figure 3.17 Unstained cervical epithelium under Raman micro spectroscope with selection showing the area for mapping with X100 dry objective (A). Five cluster K-means analysis of Raman map, mid blue representing stroma, green to basal, brown & orange to superficial layer and blue to substrate (B).

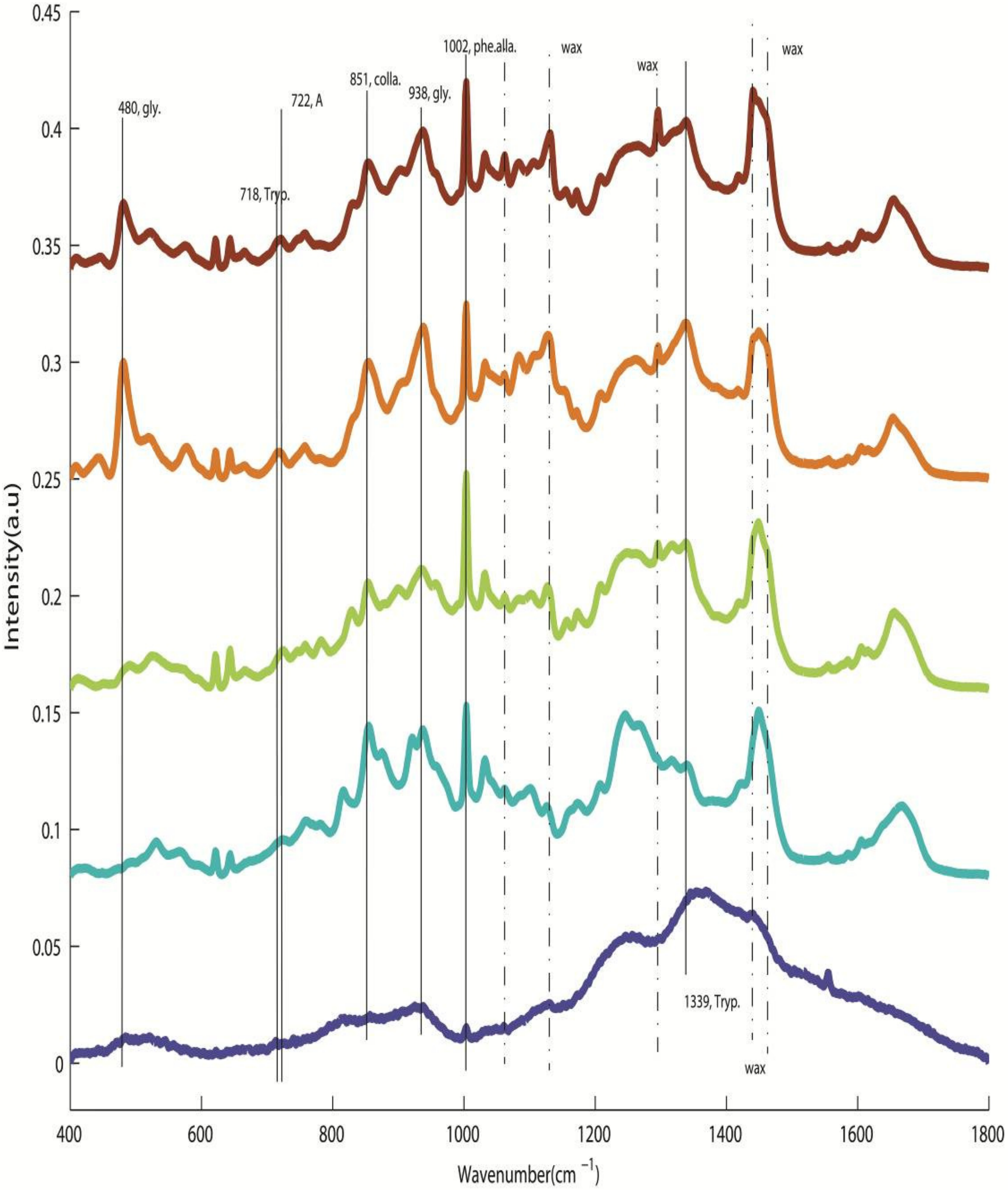


Figure 3.18 Mean spectra of K-means cluster analysis of the Raman map, mid blue representing stroma, green to basal, brown & orange to superficial layer and blue to substrate, acquired with X100 dry objective.

3.5 Conclusion

FTIR and Raman micro spectroscopy have been employed in the initial experiments to acquire spectral maps from the cervical tissue sections. This is done to compare the results of both the techniques and choose one for the further studies based on the biochemical information attained.

The results presented and discussed so far for FTIR and Raman micro spectroscopy and their comparison points towards the greater suitability of the latter technique for the current study. In each sample studied, parallel analysis of the spectral data revealed more biochemical changes in the Raman spectra. Particularly, more DNA related bands are detected in the case of the Raman spectral data. The higher content Raman spectral information achieved is due to the higher spatial and spectral resolution of Raman micro spectroscopy compared to FTIR spectroscopy.

Optimisation of the sample preparation procedures (dewaxing and section thickness) and the recording parameters (laser line, substrate, and objective lens) for Raman micro spectroscopy was carried out.

Different dewaxing agents including hexane, cyclohexane, xylene, histolene, ether and dish washing liquid were utilised to determine their ability to remove the wax from the rat skin tissue samples. On the basis of the spectral quality obtained with dewaxed rat skin, xylene and hexane were found to be equally good and thus these two agents were considered to be almost equally good for the removal of the wax. The dishwashing liquid results in the inefficient removal of wax and Raman spectra exhibited the paraffin bands even after the application of the agent. Cyclohexane itself contributed to the Raman spectra as indicated by the bands present at 787 and 802 cm^{-1}. The use of ether did not appear to be suitable as a dewaxing agent due to its interference with the proteins in the tissue section as indicated by the observation of deformation of the amide I band. Xylene and hexane were found to be the best dewaxing solvents followed by histolene as compared to the other dewaxing agents employed in current study. Xylene was adopted as solvent of choice to keep the protocol consistent with the one used in the hospital histopathology laboratory.

The selection of the most suitable laser line was made from three laser lines including 532 nm, 660 nm and 785 nm. The Raman bands were better defined for 785nm laser line as compared to the 532nm laser line but not observed at all for 660 nm laser line, which showed that it is the least suitable of the three for the current study. The 785nm laser line was preferred over 532nm laser line due to the major concern of photo degradation of the tissue section.

For the selection of the most suitable substrate, three different substrates including Mirr-IR, quartz and calcium fluoride were used and results were compared. The Mirr-IR substrate gives rise to broad background under 785nm laser illumination and as such was not considered suitable. The quartz and CaF_2 substrates do not cause fluorescence with the 785nm laser line. The CaF_2 contribution in the biological spectrum is negligible as compared to a prominent band from quartz. This led to the conclusion that the CaF_2 substrate should be used for the further experiments with Raman micro spectroscopy.

The optimization of the objective lens was done by employing X50 dry, X100 dry and X60, X100 water immersion objectives for the Raman spectral acquisition. Similar spectral features were obtained with X50 dry, X100 dry and X100 immersion objectives but the spectra acquired by the use of the X100 immersion were cleaner as compared to the X50 and X100 dry and thus less preprocessing was required for this objective. This is due to the reason that immersion of the tissue section in the water makes the tissue surface less scattering and results in better defined peaks. But tissue immersion does not readily allow for the repeat of the experiment if needed for further investigations. Use of the dry objective would avoid the problem of immersing the tissue sections in the water repeatedly. So, from these two objective lenses, the X100 dry objective lens was selected for further experiments of Raman mapping, in the current study.

Chapter 4

Raman micro spectroscopy of cervical tissue samples

4.1 Introduction

This chapter will describe the use of Raman micro spectroscopy, initially for the differentiation between different layers of normal cervical epithelium on the basis of the Raman spectral signatures of these layers, and then to explore the changes to these spectral signatures, representative of the biochemical content, associated with pathologically defined diseased states.

The samples used in this study were obtained from the Histopathology Department, Coombe Women and Infants University Hospital. Tissue sections from cervical tissue biopsies clinically characterised as normal, low grade squamous intraepithelial lesion (LSIL) or CIN I, high grade squamous intraepithelial lesion (HSIL) or CIN II, CIN III and carcinoma, were analysed. Raman spectral maps from cervical tissue samples of seventeen different patients, including, two true normal, two CIN I, five CIN II, five CIN III and three carcinoma samples were acquired. Raman spectral mapping of regions identified by the pathologist as normal and diseased was carried out. For each sample, two parallel tissue sections of 10µm thickness were cut from the FFPP cervical epithelial tissue block using a microtome. One section was mounted on a calcium fluoride slide and used for the Raman spectroscopy analysis and the other section was mounted on a glass slide for H&E staining. The tissue sections were dewaxed and stained according to the protocol described in **chapter 3, section 3.2.5**.

Tissue sections from cervical tissue biopsies were mapped with Raman spectroscopy from the regions that were identified as normal and as high grade squamous lesions by the pathologist to elucidate the layer differentiation within the normal tissue and how this differentiation is affected by the cervical intraepithelial neoplasia, based on the spectroscopic signatures of the biochemical composition. The analysis of the Raman maps from the cervical epithelium of different tissue sections revealed a serial change associated with the progression of disease.

4.2 Raman spectral data acquisition

The details of the Raman spectral acquisition were as described in chapter 3, **section 3.1.2**. However, it should be noted that all the results presented in this chapter were acquired by using X100 dry objective and the 785 nm laser.

4.3 Data pre-processing

The details of the data pre-processing are same as described in chapter 3, **section 3.1.3**.

4.4 Results

The main goal of the current study is to demonstrate the capability of Raman micro spectroscopy to classify different stages of progression of normal cervical epithelium to carcinoma *in situ* on the basis of biochemical changes observed. To achieve this classification solely on the basis of Raman spectroscopy, results are presented in the following sections in a sequence which is based on the Raman analysis, independent of the clinical classification of the samples.

4.4.1 Normal Stages

4.4.1.1 Normal sample

Figure 4.1 a, shows the H&E stained tissue section and **Figure 4.1b** shows the extended video image of the unstained tissue section with a X100 objective acquired using the Raman micro spectrometer. The sample was diagnosed as normal as there was no evidence of the metaplasia, and was mapped and analysed, as indicated by the red box. KMCA of the Raman map is shown in **Figure 4.1 c**. The analysis employs 5 clusters which divide the cervical epithelium into three distinct layers. The comparison of the unstained tissue section with the result of the K- means cluster analysis for this map suggests that the blue cluster represents the stroma, the red cluster the basal and parabasal layers, the green cluster the superficial layer and the black cluster the substrate, while the yellow cluster represents the presence of residual wax in the tissue section.

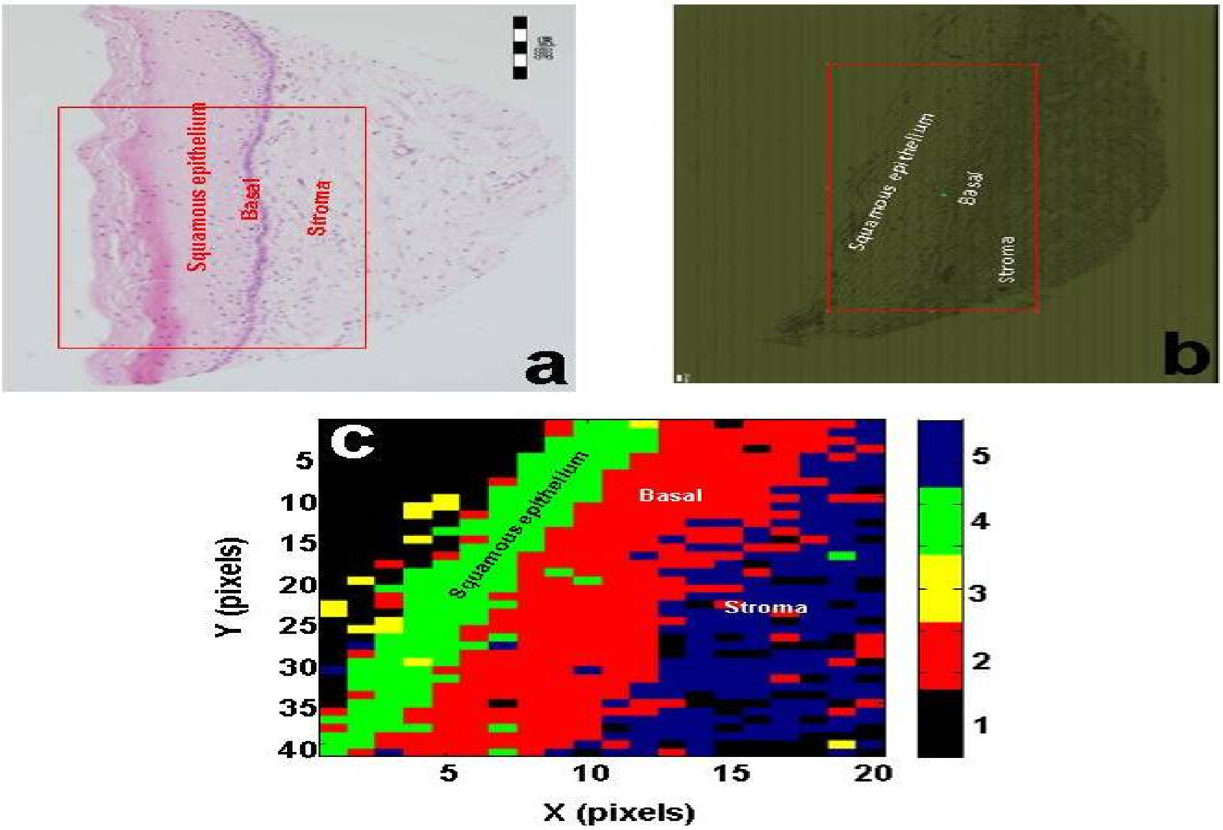


Figure 4.1 H&E stained normal diagnosed tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 50µm (b) Five cluster K means cluster map generated from the Raman map (c).

Mean Raman spectra representing the respective colour cluster/group are shown in **Figure 4.2,** except the clusters describing the substrate and residual wax.

It should be noted that two important Raman markers of collagen and glycogen can be used to differentiate the Raman spectra of the layers of squamous epithelium and stroma. Based on the analyses of (Jess et al., 2007, Notingher, 2007, Notingher et al., 2003, Lyng et al., 2007), the Raman bands at 853, 921, 938 and 1245 cm^{-1} are indicative of collagen and those at 480, 849 and 938 cm^{-1} can be assigned to glycogen. It should be noted that the two bands at 853 cm^{-1} (collagen) and 849 cm^{-1} (glycogen) overlap and are not well differentiated by Raman spectroscopy. By considering the pattern of the collagen and glycogen bands in the current study, it is observed that 921 cm^{-1} and 938 cm^{-1} are normally co-incident in spectra of collagen and the 480 cm^{-1} and 853 cm^{-1} bands are normally present in the spectra of glycogen. The Raman spectra of the pure collagen and glycogen used for reference are described by (Lyng et al., 2007). These Raman bands are therefore good differentiating bands between collagen and glycogen. Another differentiating marker can be that at 1245 cm^{-1}, which is observed as a very sharp band if associated with the presence of collagen, while otherwise it is very much diminished and broad.

The most prominent Raman bands in the blue spectrum of **Figure 4.2**, indicative of the blue cluster of **Figure 4.1c**, can be assigned to collagen (853, 921, 938 and 1245 cm^{-1}) and phenylalanine (1002cm^{-1}). Collagen is a structural protein and is a major component of the connective tissue layer.

The predominant bands in the red spectrum, corresponding to the red cluster, are of DNA bases, thymine (755 cm^{-1}), adenine (722 cm^{-1}) and cytosine (782 cm^{-1}). On the basis of this information, it can be concluded that the red cluster represents the basal and parabasal layers. The bands present at 480, 849 and 938 cm^{-1} in the Raman spectrum suggest that accumulation of glycogen starts in the basal and parabasal layers, which are also rich in

DNA. The major Raman features observed in the mean representative spectrum of the green cluster can be assigned to glycogen (480, 849 and 938 cm^{-1}), and are Raman

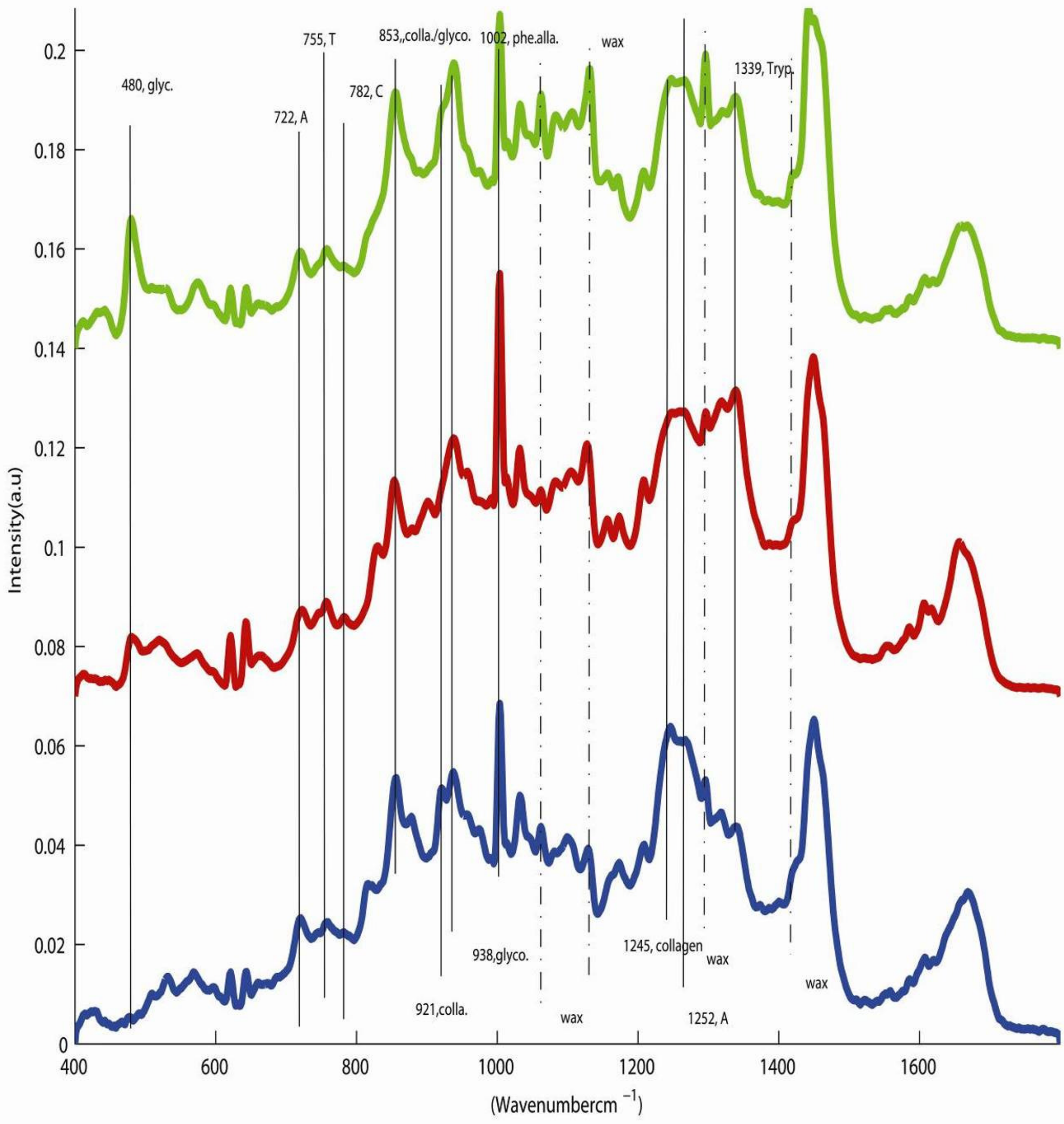


Figure 4.2 K-means cluster spectra, blue representing stroma (collagen), red representing basal layer (DNA) and green representing superficial layer (glycogen), for the Raman map of the marked region in figure 4.1 c.

signatures of the squamous epithelium of the cervix, as cervical epithelium cells accumulate glycogen with maturity (Lyng et al., 2007, Salzer and Siesler, 2009).

It has been observed that the basal layer shares some features with the stromal layer, such as that at 1318 cm^{-1}, assigned to guanine, and the superficial layer, such as that at 1084 cm^{-1}, assigned to the O-P-O stretching of DNA, but these Raman bands are not differentiating bands between these layers. The Raman band at 1339 cm^{-1} is assigned to tryptophan and is present in the Raman spectra of all the layers.

The comparison between the H&E stained tissue section, the unstained tissue section and the KMCA of the Raman map suggests that Raman spectroscopy is able to differentiate normal cervical epithelium into three layers including connective, basal, and squamous epithelium on the basis of the Raman signatures of collagen in the connective tissue, DNA bases in the basal and parabasal layer and glycogen in the superficial layer. The prominence of glycogen bands in the mean spectrum of the basal layer may indicate that cells are dividing and gradually maturing to accumulate glycogen.

4.4.1.2 Normal region of CIN II sample

For the purpose of the study, the normal side of another sample which was CIN II diagnosed, as indicated by the red box in **Figure 4.3a**, was mapped and analysed. KMCA of the Raman map is shown in **Figure 4.3 c**. **Figure 4.3 a,** shows the H&E stained tissue section and **Figure 4.3b** shows the extended video image of the unstained tissue section with a X100 objective acquired using the Raman micro spectrometer. The analysis employs 5 clusters which divide the cervical epithelium into three distinct layers. The comparison of the unstained tissue section with the result of the K- means cluster analysis for this map suggests that the blue cluster represents the stroma, the red cluster the basal and parabasal layers, the green cluster the superficial layer and the black cluster the substrate, while the yellow cluster represents the presence of residual wax in the tissue section.

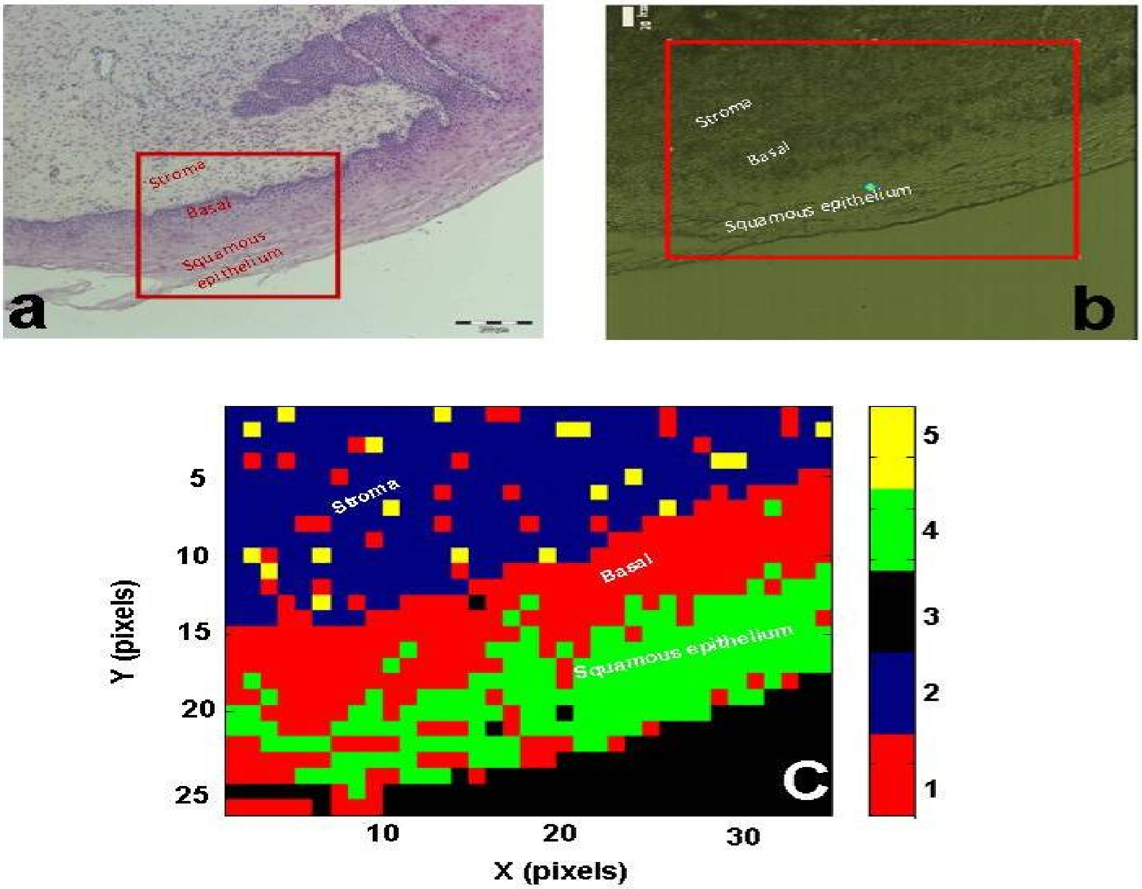


Figure 4.3 H&E stained normal region of CIN II diagnosed tissue with scale bar of 200μm (a) unstained tissue under Raman spectroscope with scale bar of 50μm (b) Five cluster K means cluster map generated from the Raman map (c).

The representative Raman spectra are presented in **Figure 4.4**. As discussed above, in case of the true normal cervical tissue sample, all the information associated with biological spectra of the corresponding layers were the same. The yellow spectrum is presented here, to highlight the peaks related to the residual wax. The peaks present in the yellow spectrum at 1061 cm^{-1}, 1133 cm^{-1}, 1294 cm^{-1}, 1339 cm^{-1} are identical to those of the Raman spectrum of paraffin wax, **chapter 3**, **Figure 3.6**, indicating that the dewaxing procedure was not completely effective in removing the wax from the cervical tissue sections under study.

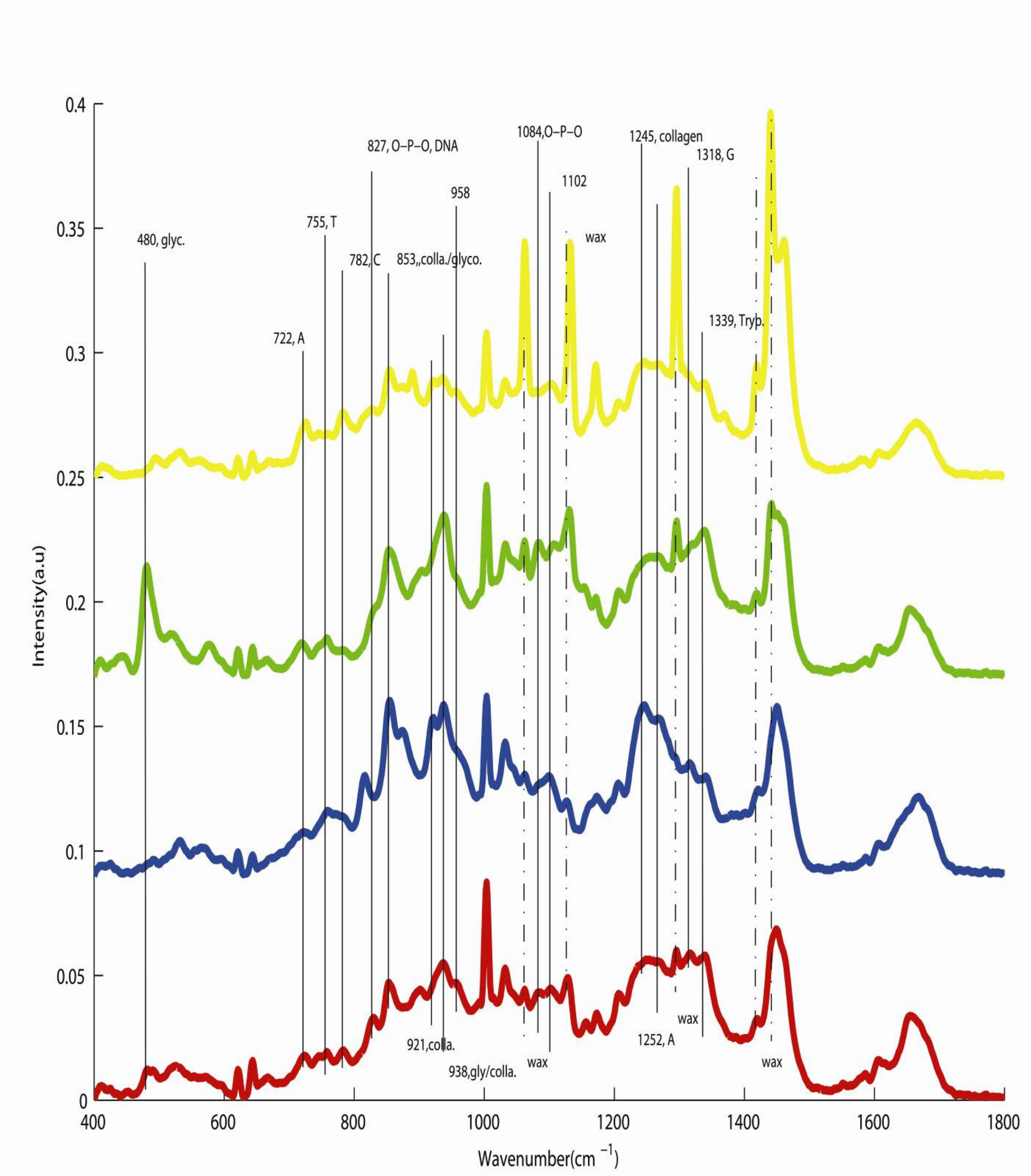


Figure 4.4 K-means cluster spectra, blue representing stroma (collagen), red representing basal (DNA), green representing superficial (glycogen) and yellow representing residual wax, for the Raman map of marked region in figure 4.1 c.

4.4.2 Abnormal Stages

4.4.2.1 Abnormal sample I

Figure 4.5a, b and c shows the H&E stained tissue section, unstained tissue section and KMCA Raman maps of a CIN III diagnosed cervical tissue section, respectively. It can be seen (**Figure 4.5 a**) that one side is more intensely stained compared to the other, indicative of a higher degree of abnormality, and this region, indicated by the red box, was analysed by Raman spectroscopy.

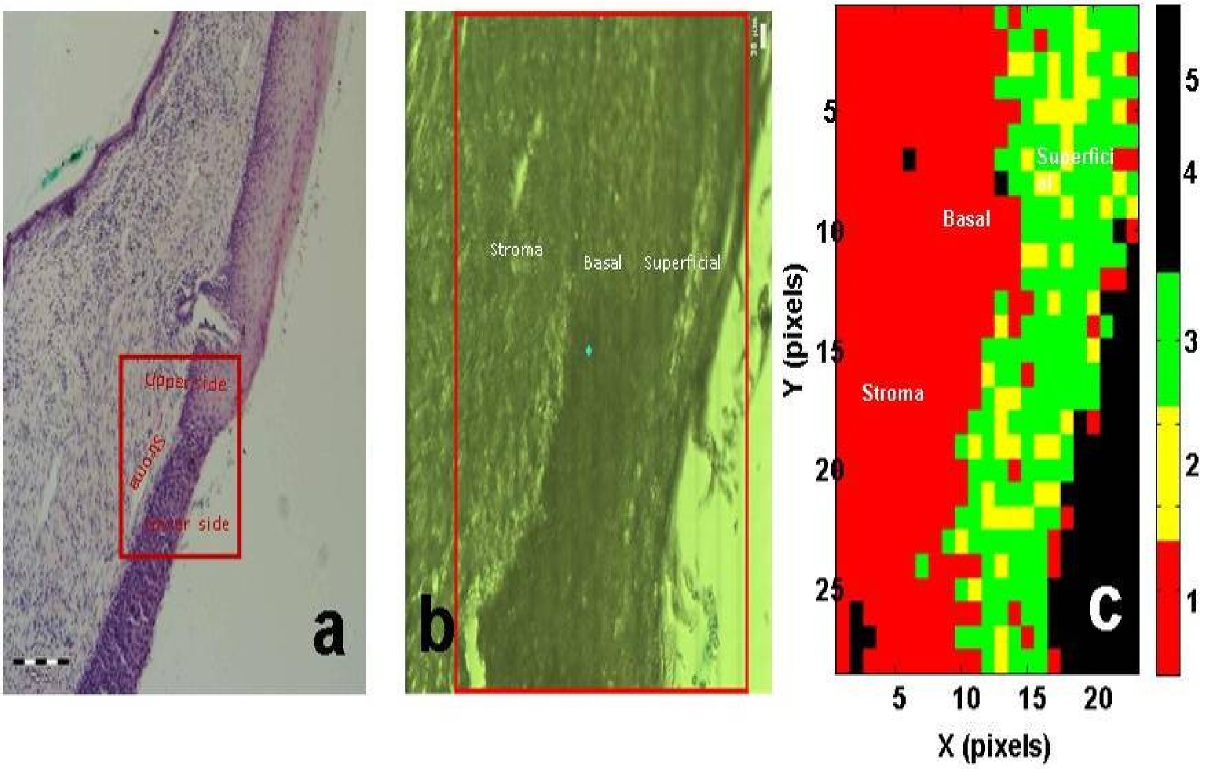


Figure 4.5 H&E stained diseased side of CIN III marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 20µm (b) Five cluster KMCA map generated from the Raman map (c).

The mean Raman spectra of the KMCA are presented in **Figure 4.6.** The KMCA of the Raman map shows a two layer separation, as indicated by two prominent clusters. The red cluster corresponds to the basal layer as indicated by the characteristic Raman markers in the mean spectrum (the Raman bands of the DNA bases including Raman spectral features of thymine (755 cm^{-1}), adenine (722 cm^{-1}) and cytosine (782 cm^{-1}), characteristic of the basal layer). The green cluster corresponds to cervical epithelium. The mean Raman spectrum of the yellow cluster has prominent Raman features of glycogen.

Comparison with the KMCA map of the normal region, shown in **Figure 4.1c**, shows that the squamous epithelial layer seems to have some features similar to the basal layer as observed and discussed in **section 4.4.1**. Notably, there is no distinct representation of the stromal layer in the KMCA of this tissue section, the clustering resulting in only two layers, one with distinct Raman features of DNA and the other with those of glycogen. The absence of the Raman spectral features of collagen may indicate some biochemical changes in the stroma due to the progression of disease and need to be further evaluated to understand the basis of these changes.

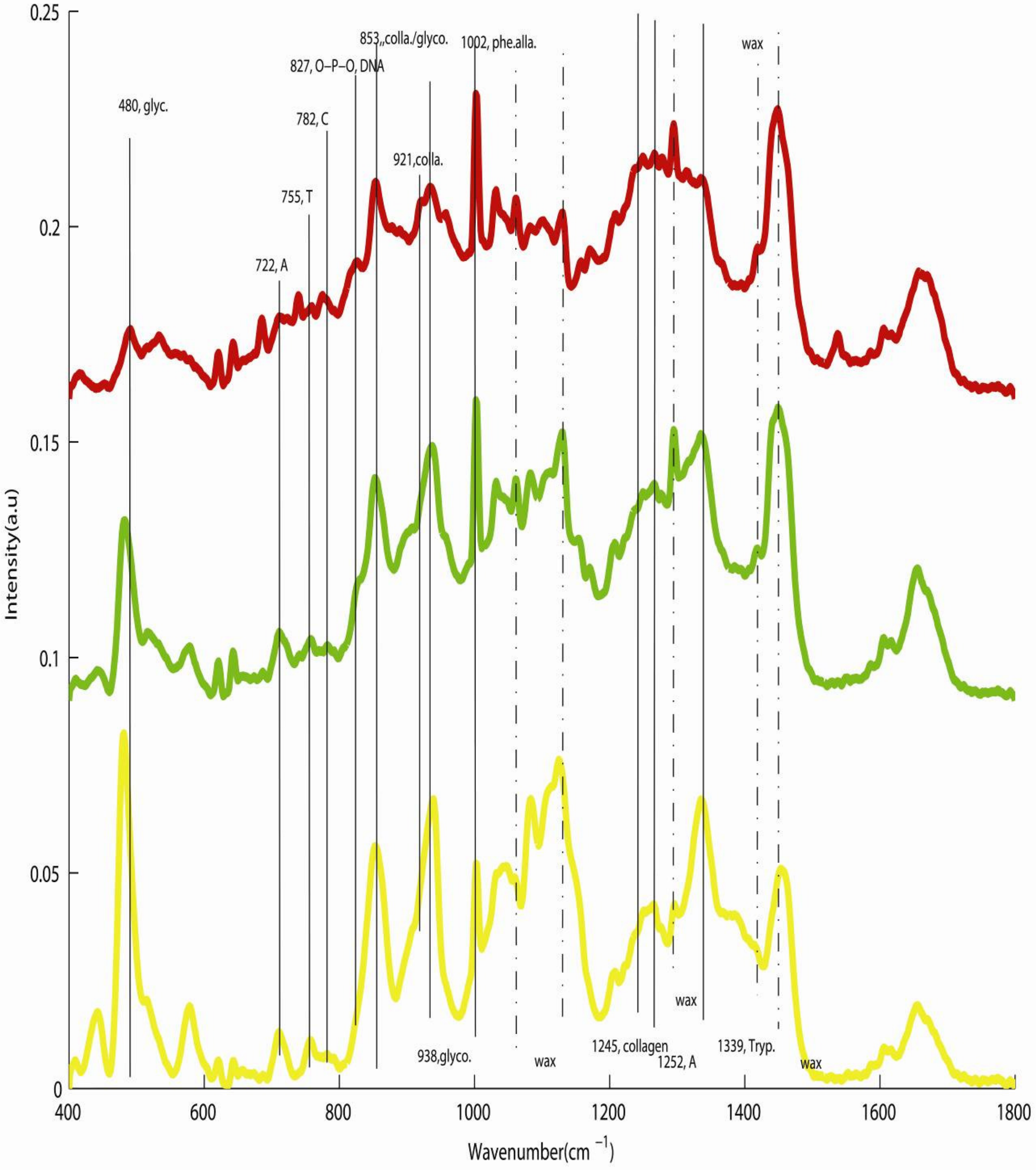


Figure 4.6 Mean Raman spectra of, basal (red), and squamous epithelial (green & yellow) layers from KMCA of Raman map in Figure 4.5 c.

4.4.2.2 Abnormal sample II

Another example of a Raman classified abnormal region is presented in **Figure 4.7** and **Figure 4.8**. **Figure 4.7** shows the KMCA of tissue section from a CIN III region which gives three prominent clusters.

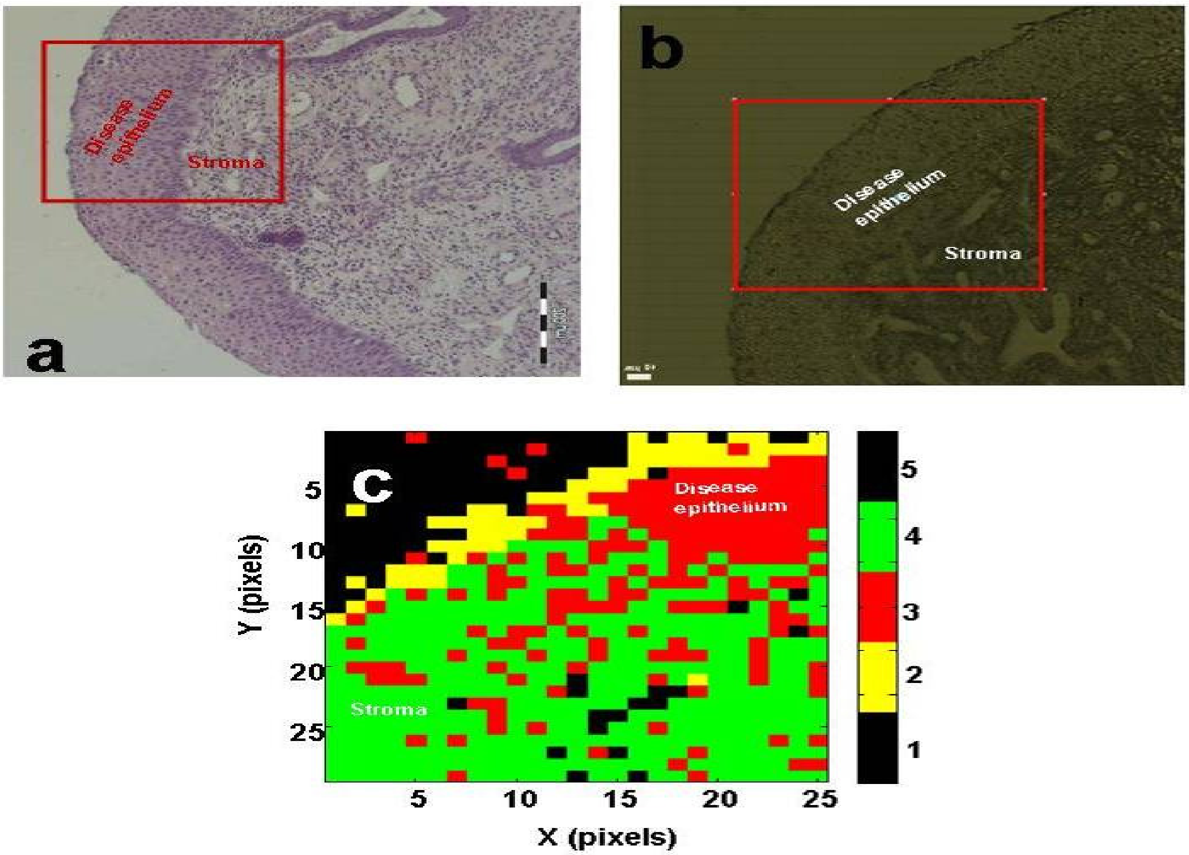


Figure 4.7 H&E stained disease side of CIN III marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster KMCA map generated from the Raman map (c).

Analysis of the mean spectrum of the yellow cluster (**Figure 4.8**) correlates it with the marking ink already presented in **chapter 3**, **Figure 3.7**. In the KMCA map, the green cluster corresponds to stroma as well as some portion of the epithelial layer, while the red cluster corresponds to the basal layer only. On the other hand, the mean Raman spectrum of the red and green clusters (**Figure 4.8**) can both be correlated with the basal layer, but the differentiating factor between the two clusters is the absence of the bands for glycogen/collagen in the red spectrum and the presence of these bands in the green spectrum. It should be noted that the substrate is divided into two types of clusters in this map, represented by the black cluster which is due to substrate only while the yellow cluster is due to a combination of marking ink and substrate.

The results of this Raman map, once again, show the presence of DNA related Raman bands and the absence of collagen Raman spectral bands. This, again, may have some biochemical relation to the progression of disease before the morphological changes take place. The absence of glycogen bands in the green spectrum may indicate the abundance of rapidly growing abnormal cells which are unable to accumulate glycogen due to the rapid cell division and lack of maturity.

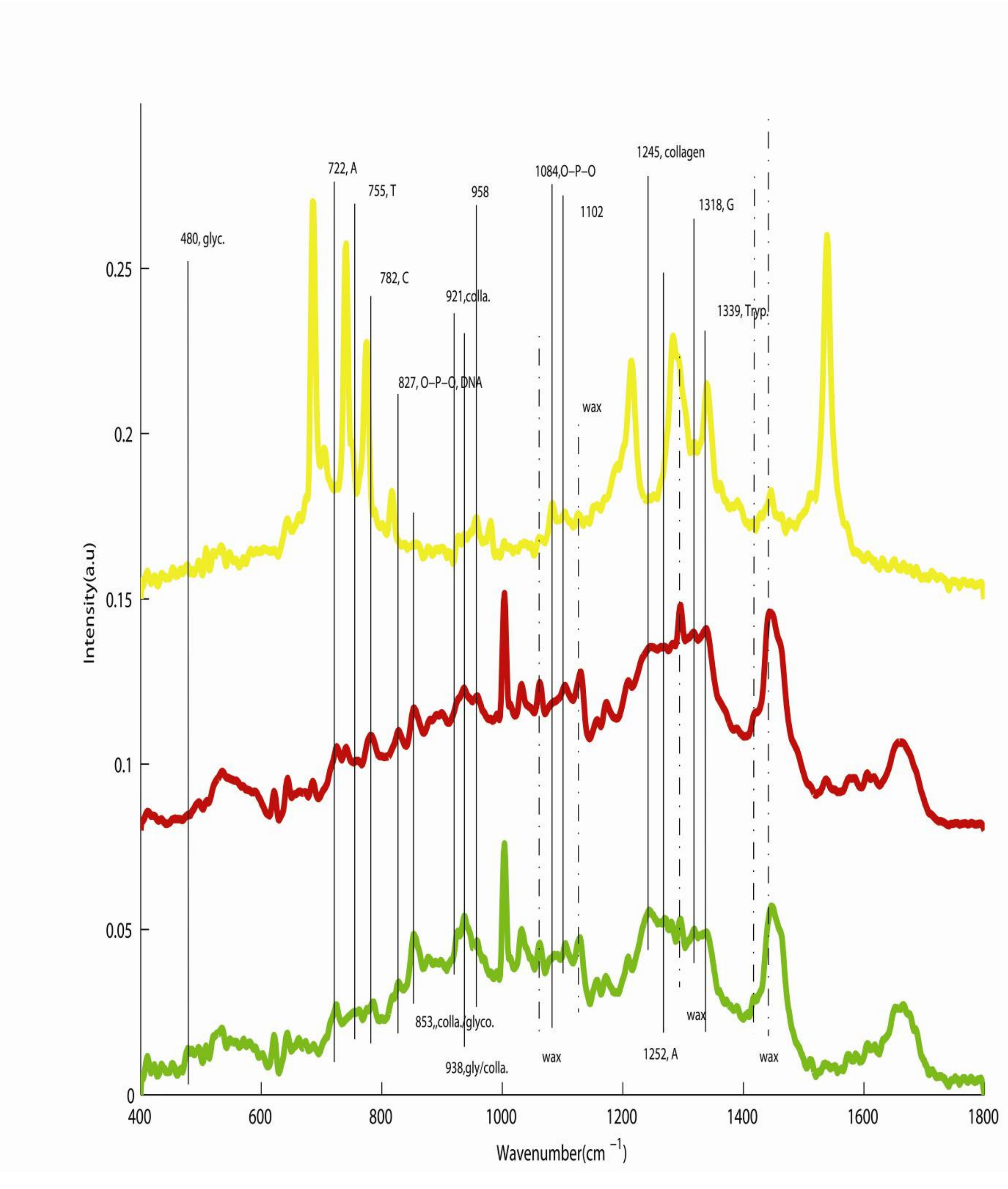


Figure 4.8 Mean representative Raman spectra, red and green spectra correspond to basal (DNA) and yellow spectrum corresponds to marking ink, from KMCA of Raman map in Figure 4.7 c.

4.4.2.3 Abnormal sample III

Another example of a Raman classified abnormal region is presented in **Figure 4.9** and **Figure 4.10**. **Figure 4.9** shows the KMCA of a carcinoma *in situ* tissue section which does not give any separation of the prominent clusters. All the three representative Raman spectra, red, green and yellow contain the features related to DNA, although the yellow spectra also show some features of the residual wax.

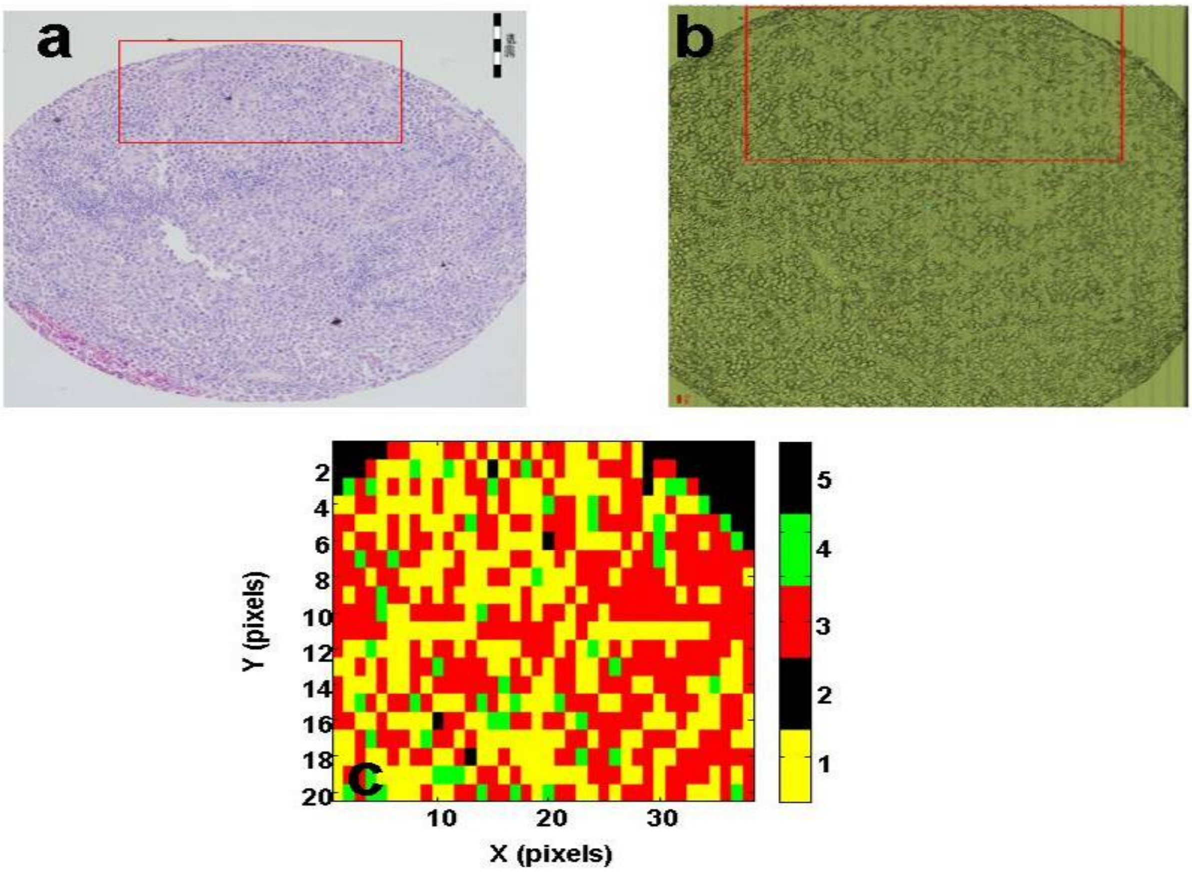


Figure 4.9 H&E stained carcinoma *in situ* marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster KMCA map generated from the Raman map (c).

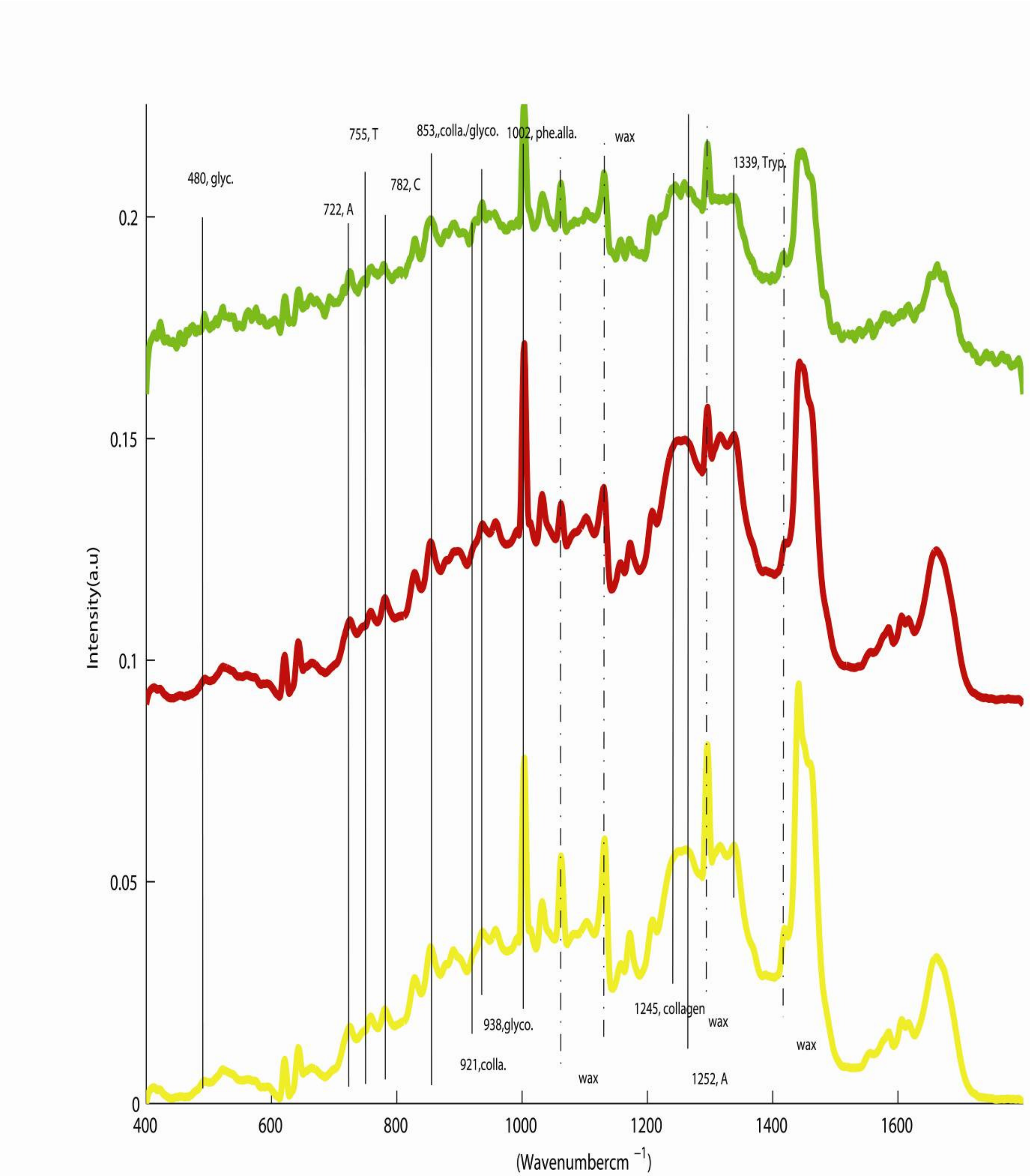


Figure 4.10 Mean representative Raman spectra, red and green spectra show features of the basal layer (DNA) and yellow spectrum shows features of DNA and marking ink, from KMCA of Raman map in Figure 4.9 c.

4.4.3 Intermediate Stages

In this section, Raman spectral results of four different cervical tissue samples are presented to show an increase in the degree of the abnormality in the form of biochemical changes observed by Raman micro spectroscopy.

4.4.3.1 Intermediate sample I

Figure 4.11 a shows the H&E stained normal region of a CIN III diagnosed tissue section taken with a X10 objective which exhibits the normal pattern of the stroma, basal, parabasal, intermediate and superficial epithelium.

Figure 4.11 b shows the extended video image of the same unstained tissue section with a X100 objective, acquired using the Raman micro spectrometer and the red box indicates the area analysed. KMCA of the Raman map is shown in **Figure 4.11 c**.

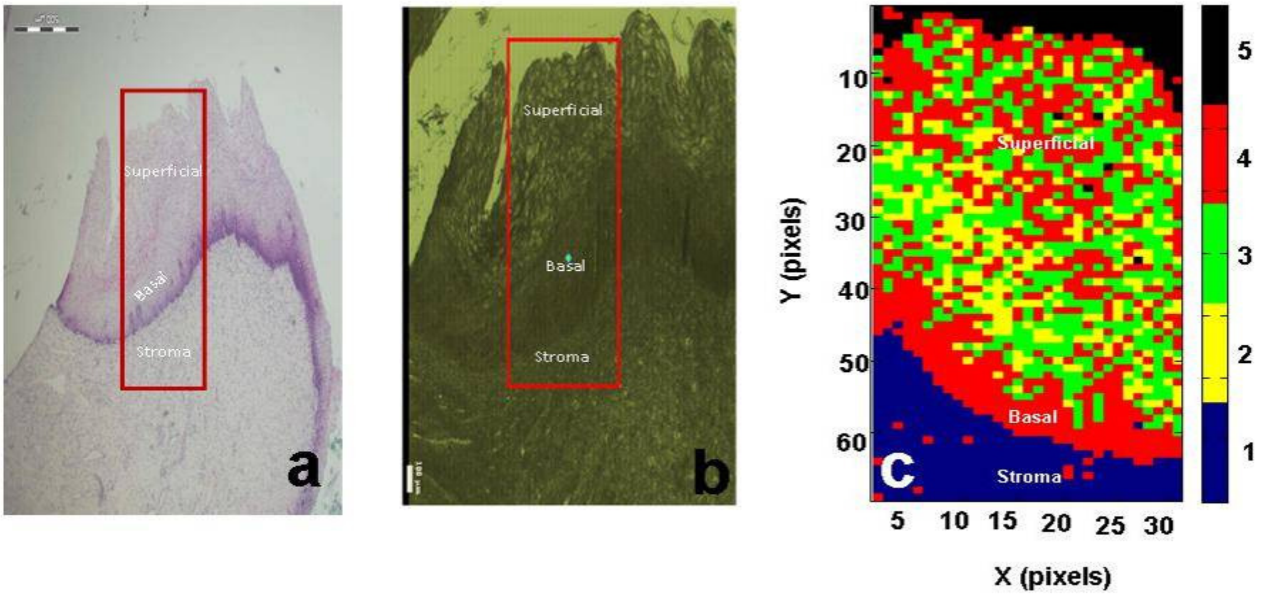


Figure 4.11 H&E stained normal region of CIN III diagnosed tissue with scale bar of 500µm (a) unstained tissue under Raman spectroscope, with scale bar of 100µm (b) Five cluster KMCA map generated from the Raman map (c).

It shows intact stroma, separated from the basal layer, as indicated by the Raman bands of the collagen, including 849 cm^{-1}, 921 cm^{-1}, 938 cm^{-1} and 1245 cm^{-1}, of the blue mean spectrum in **Figure 4.12**. The separation of the stroma and basal layer is very clear, but at the same time, the squamous epithelium has three types of clusters, two with strong Raman spectral features of glycogen (green and yellow clusters) and one with the characteristic Raman spectral features of the basal layer (red cluster), indicating characteristics of basal cells in this layer. This information indicates that the squamous layer shows evidence of disease, based on strong DNA features throughout this layer, indicating rapidly proliferating cells.

4.4.3.2 Intermediate sample II

The other interesting observation concerned with the normal region of the CIN III marked tissue sections. **Figure 4.13 a**, **4.13 b** represent H&E stained and parallel unstained tissue section of a normal region of a CIN III marked tissue section. The KMCA pseudo colour map and mean Raman spectra are presented in **Figure 4.13c** and **Figure 4.14**, respectively.

The mean Raman spectra are presented in **Figure 4.14,** the green spectrum represents the superficial layer (glycogen) while the red spectrum represents the basal and stromal layers as indicated by the presence of characteristic Raman features of both DNA and collagen in the mean Raman spectrum of the red cluster. It is observed that, although the histological appearance of these tissue samples is normal, as indicated by the pathologist, the KMCA divides the Raman spectral data of these “normal regions” into just two clusters.

The Raman spectral features of these two clusters suggest that there are two types of layers, one prominent, the basal layer with strong DNA bands, and the other diminished, the superficial layer as characterised by strong glycogen bands. No features of the stromal layer are evident in these clusters. This trend of the absence of collagen related Raman bands and presence of DNA related Raman bands in the stromal layer, where collagen related Raman bands are normally prominent, has been observed in the normal region of the four CIN III marked tissue sections examined. These observations lead to the conclusion that Raman spectroscopy can detect the biochemical changes occurring before the morphological manifestation of the disease.

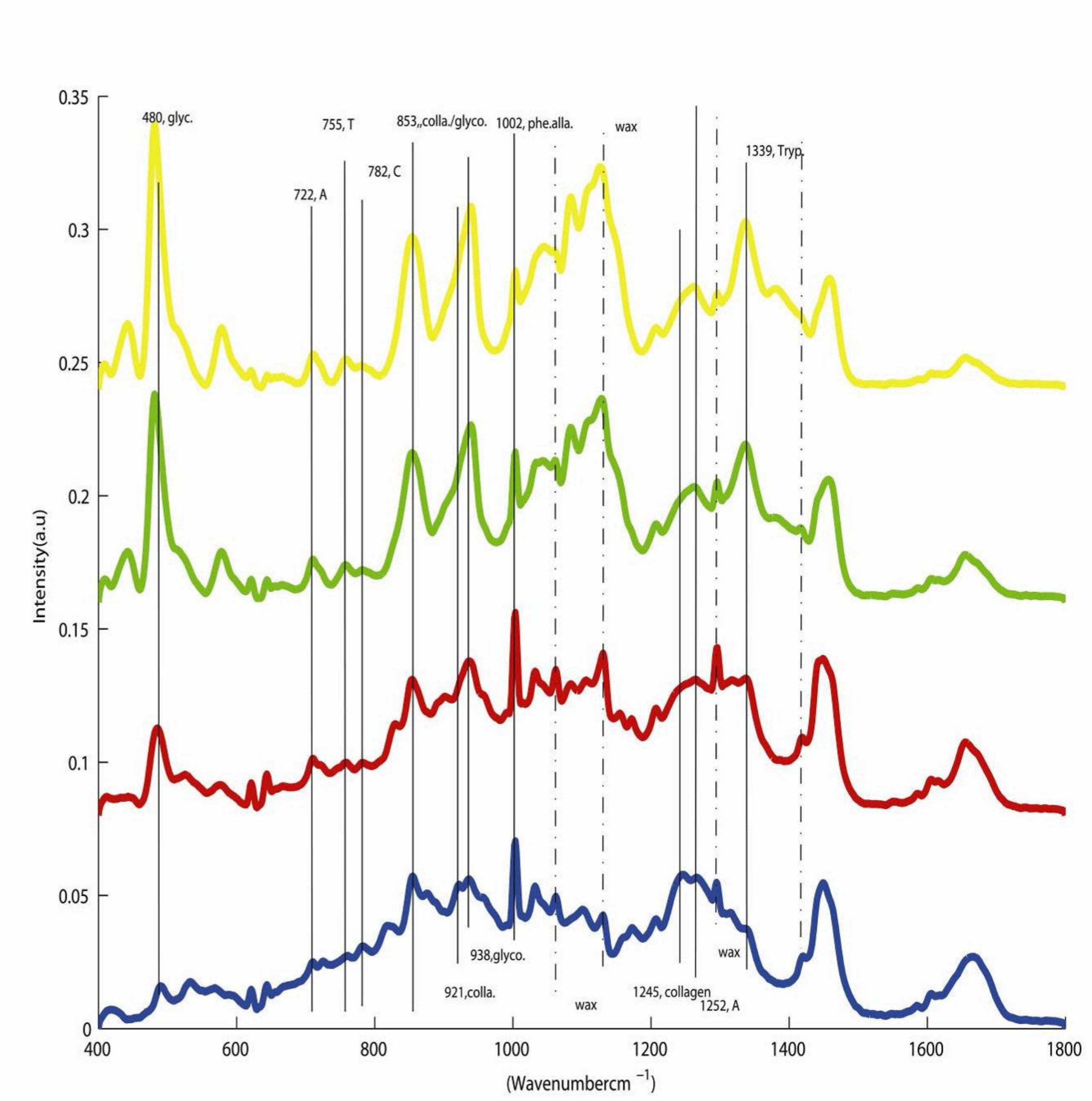


Figure 4.12 K-means cluster spectra, representing blue to stroma (collagen), red to basal (DNA), green & yellow to superficial (glycogen) layers, for the Raman map of marked region in figure 4.11 c.

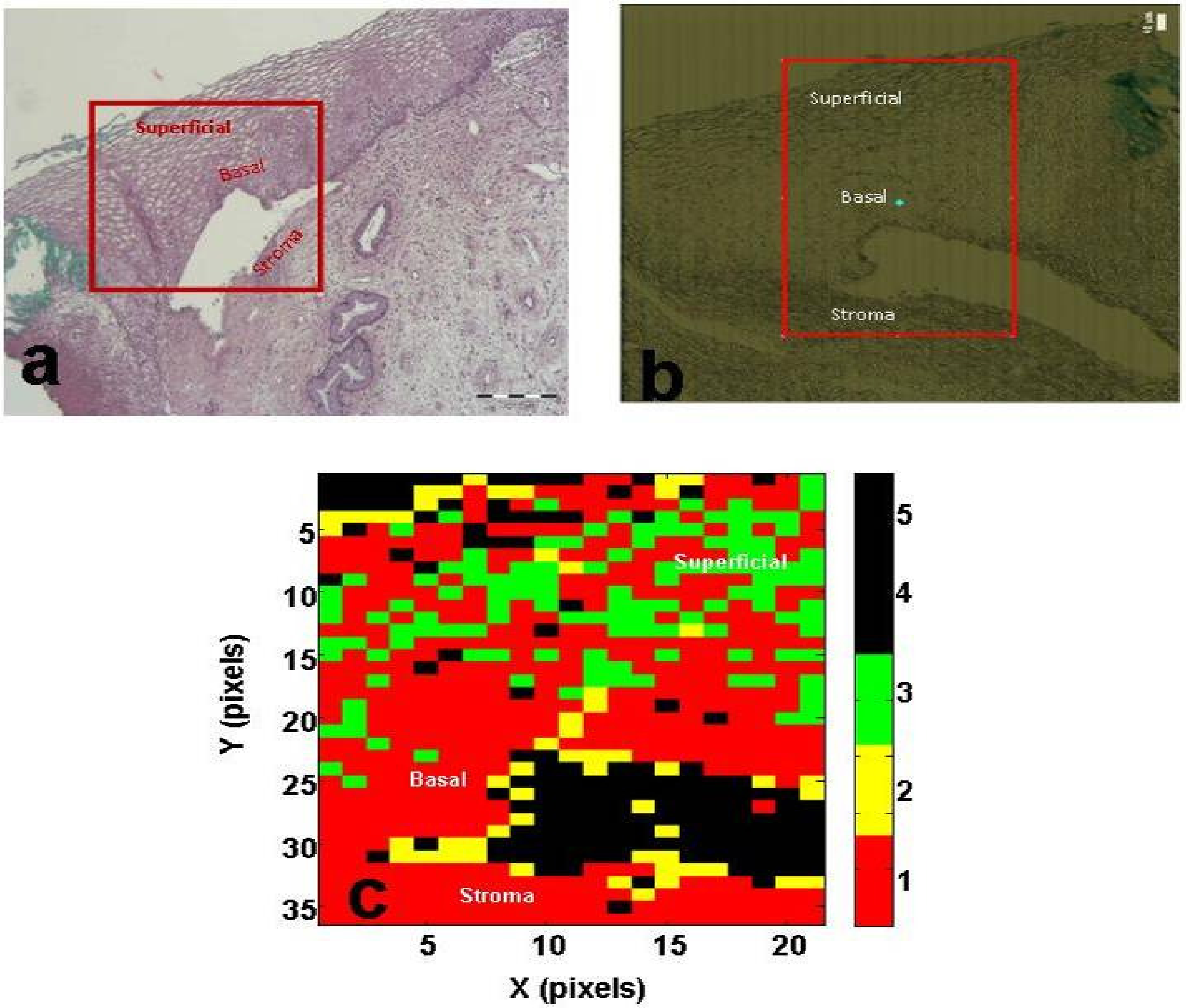


Figure 4.13 H&E stained normal side of CIN III marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster K mean map generated from the Raman map (c).

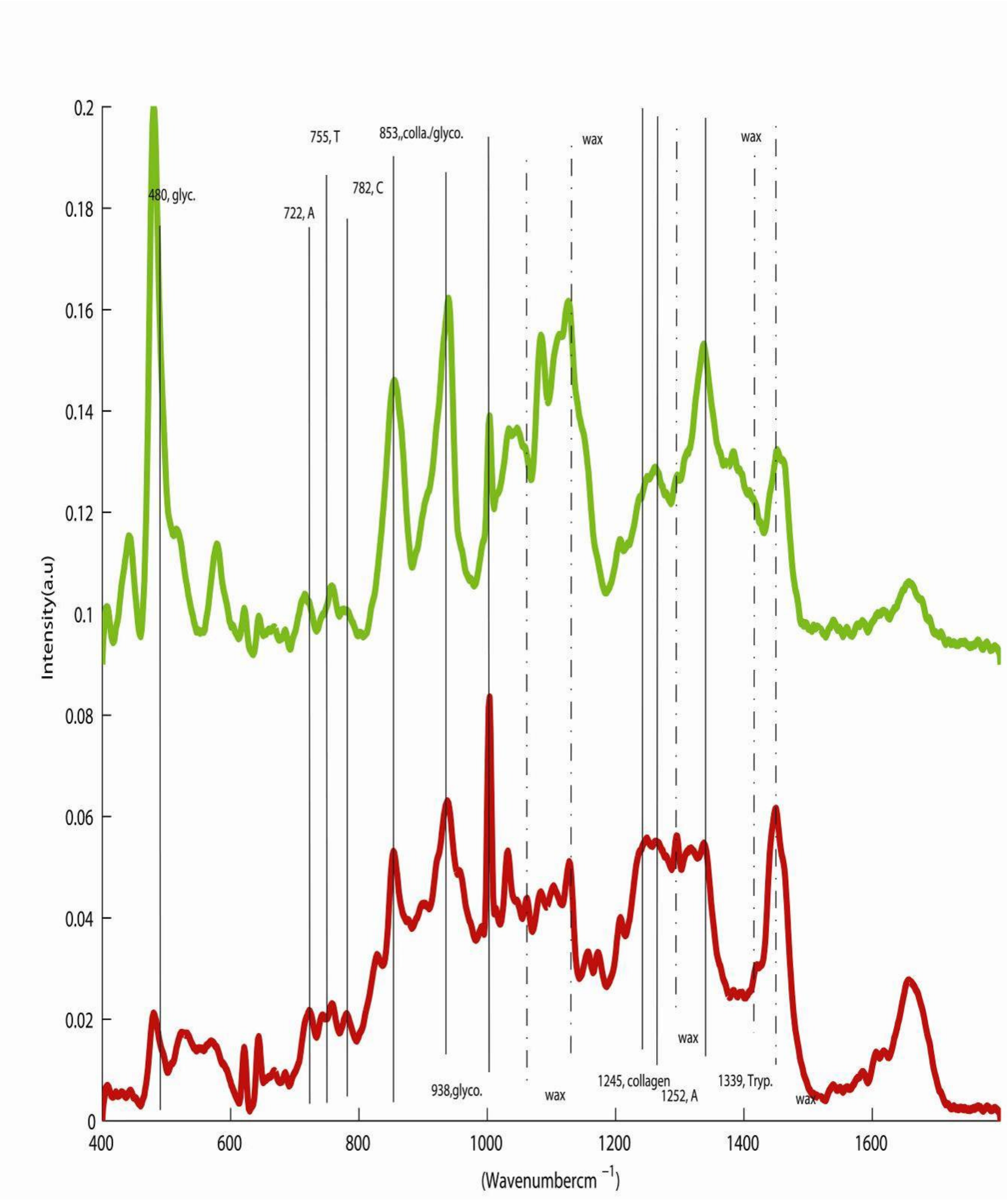


Figure 4.14 Mean representative spectrum, red spectrum representing basal layer (DNA bases) and green spectrum indicating superficial layer (glycogen) from KMCA of Raman map of Figure 4.13 c.

4.4.3.3 Intermediate sample III

The CIN II marked cervical tissue sample, classified by Raman spectroscopy as intermediate between normal and abnormal is presented here. **Figure 4.15 a** shows the H&E stained tissue section of a formalin fixed paraffin preserved tissue section which has only abnormal regions. **Figure 4.15 b** shows the optical images of the parallel unstained section as viewed by the Raman micro spectroscope. **Figure 4.15 c** shows the five cluster KMCA of the Raman map.

KMCA of Raman micro spectral mapping of this sample divides the tissue section into three layers, stroma (blue cluster), basal (red cluster) and superficial layer (green cluster). The yellow cluster shows residual wax in the tissue and the black cluster corresponds to substrate. Careful analysis of the KMCA pseudo image reveals that the superficial layer exhibits regions with spectral characteristics of the basal layer. In the pseudo image of this sample, it appears as though the basal layer invades into the superficial layer in such a way that the typical characteristics of the latter are hugely reduced and are represented only by a few pixels. The stromal layer (blue cluster) also shows characteristic spectral features of the basal cell layer.

Figure 4.16, elucidates the discriminating Raman marker bands between the layers. Raman bands assigned to the stromal layer are from collagen (853, 921, 938 and 1245 cm^{-1}), those assigned to the basal layer are DNA related bands including thymine (755 cm^{-1}), adenine (722cm^{-1}) and cytosine (782 cm^{-1}) and the bands at 480, 849 and 938 cm^{-1} suggest accumulation of glycogen in the superficial layer.

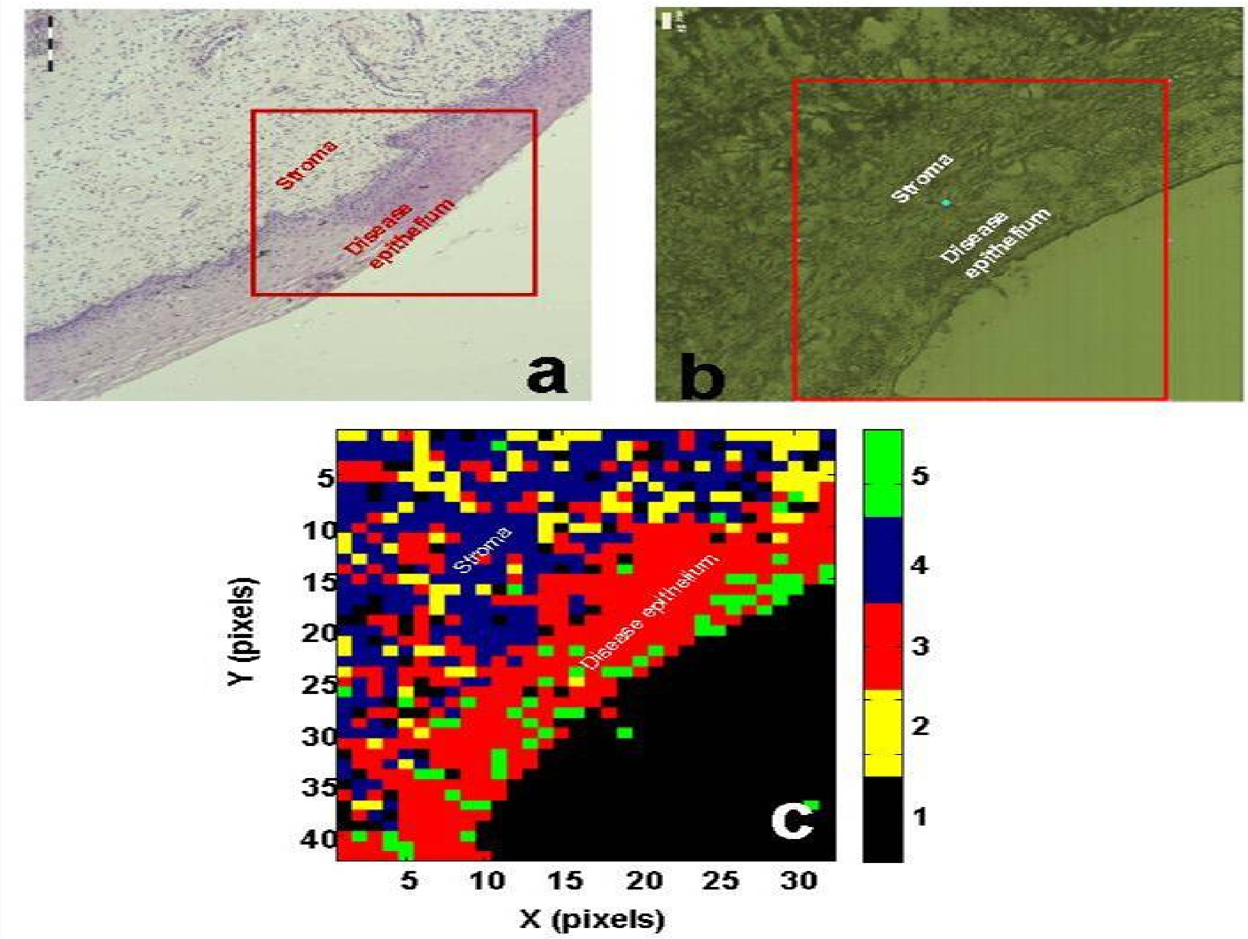


Figure 4.15 H&E stained disease side of CIN II marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope, 40µm (b) Five cluster KMCA map generated from the FTIR map (c).

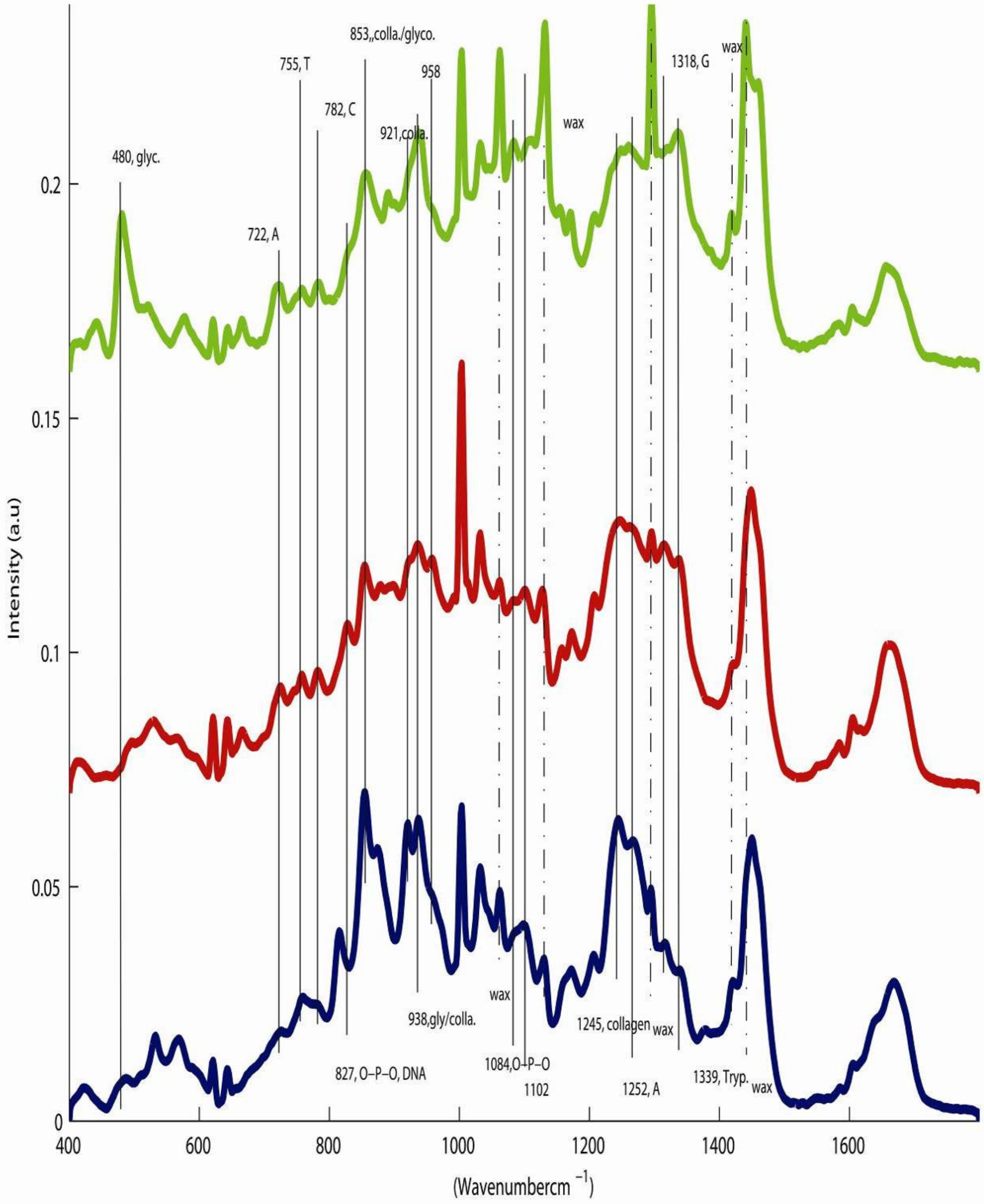


Figure 4.16 Representative mean Raman spectra from KMCA map in Figure 4.15 c where blue spectrum representing stromal layer (collagen), red spectrum representing basal layer (DNA) and green representing superficial layer (glycogen).

4.4.3.4 Intermediate sample IV

Raman spectral results of a CIN II marked cervical tissue sample which was classified as intermediate between normal and abnormal by Raman micro spectroscopy, are presented in **Figure 4.17** and **Figure 4.18**. **Figure 4.17 a** shows a H&E stained tissue section which has diseased and normal tissue areas adjacent to each other. These regions are ideal to investigate how the layer differentiation is affected by the proliferation of disease. **Figure 4.17 b** shows the optical image of the parallel unstained tissue section as viewed by the Raman micro spectrometer. KMCA of the Raman map of the same sample is shown in **Figure 4.17 c**.

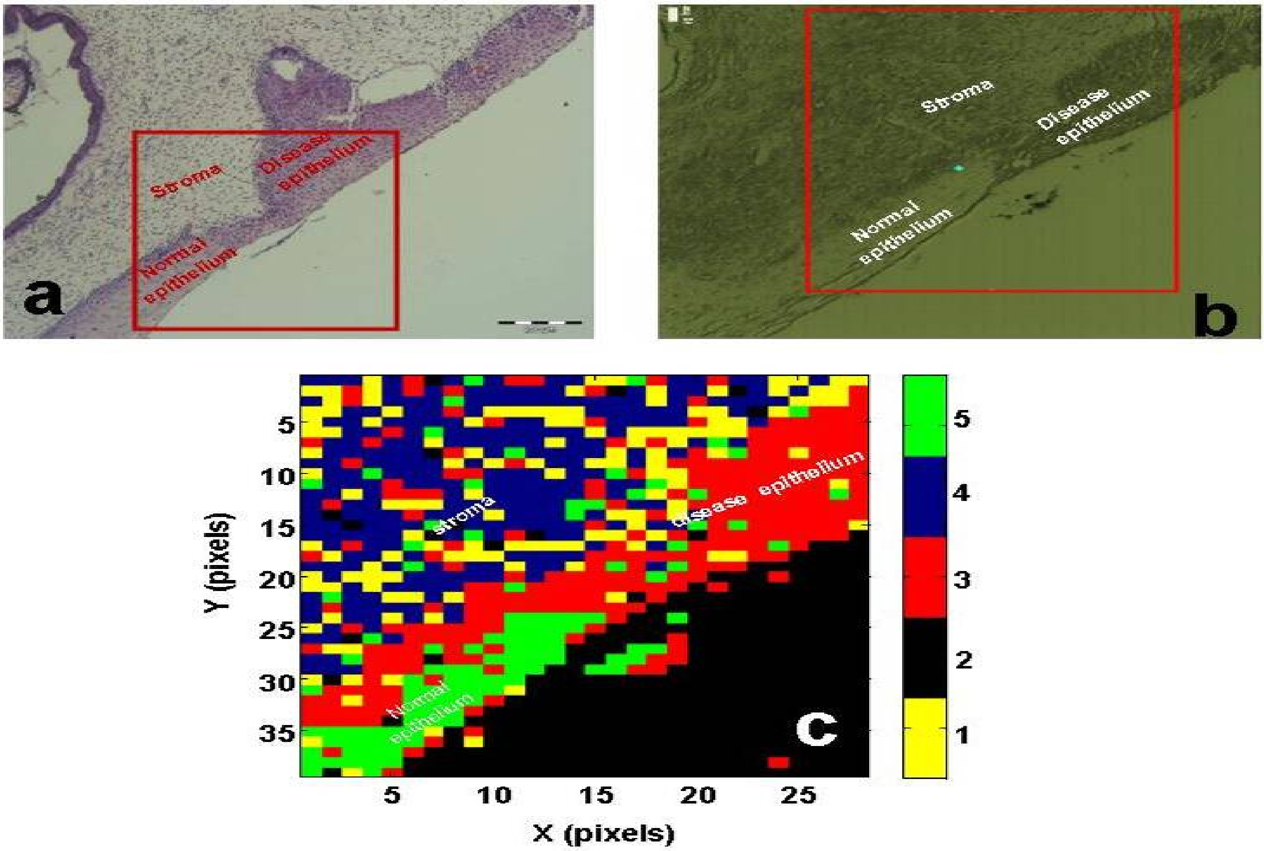


Figure 4.17 H&E stained normal & disease region of CIN II diagnosed tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster KMCA map generated from the Raman map (c).

The KMCA of the Raman map (**Figure 4.17c)** clearly shows the separation of the normal and abnormal regions. The normal region is separated into three layers, including stroma (indicated by the yellow and blue colour clusters with a mean spectrum having characteristic bands of collagen, **Figure 4.18**), basal layer (indicated by the red colour cluster with mean spectrum having characteristic bands of DNA bases) and squamous epithelium (indicated by the green colour cluster, with mean spectrum having characteristic bands of glycogen). The reason for the two clusters (yellow and blue) in stromal layer is the difference in the intensity of the collagen related bands. The abnormal region is separated into only two layers representing stromal and basal layers. The KMCA of the Raman map of this region indicates the presence in the stromal or connective tissue layer of the characteristic features of the basal layer, as characterised by the Raman spectral features of thymine (755 cm^{-1}), adenine (722cm^{-1}) and cytosine (782 cm^{-1}). Notably, the existing system of classification (CIN system) does not consider the involvement of the stromal layer when categorising the degree of abnormality or grading of the cervical intraepithelial neoplasia (CIN) (Sellors and Sankaranarayanan, 2003/4). It is clear, however, that cervical intraepithelial neoplasia turns the three layer separation into two layers.

This increase in DNA content and decrease in the collagen content may raise the question of whether this increase in DNA content is associated with the interaction of the basal cells with stromal cells or is due to some biochemical changes in the stroma cells themselves. Notably HPV infection also results in higher DNA content but is normally understood to primarily affect the epithelial layer and not the stromal layer. Further elucidation of the origin and nature of the biochemical phenomena indicated by spectral analysis requires the PCA analysis and the immunohistochemistry of proteins involved in cervical cancer such as p16, details of which are presented in Chapter 5 and Chapter 6 respectively.

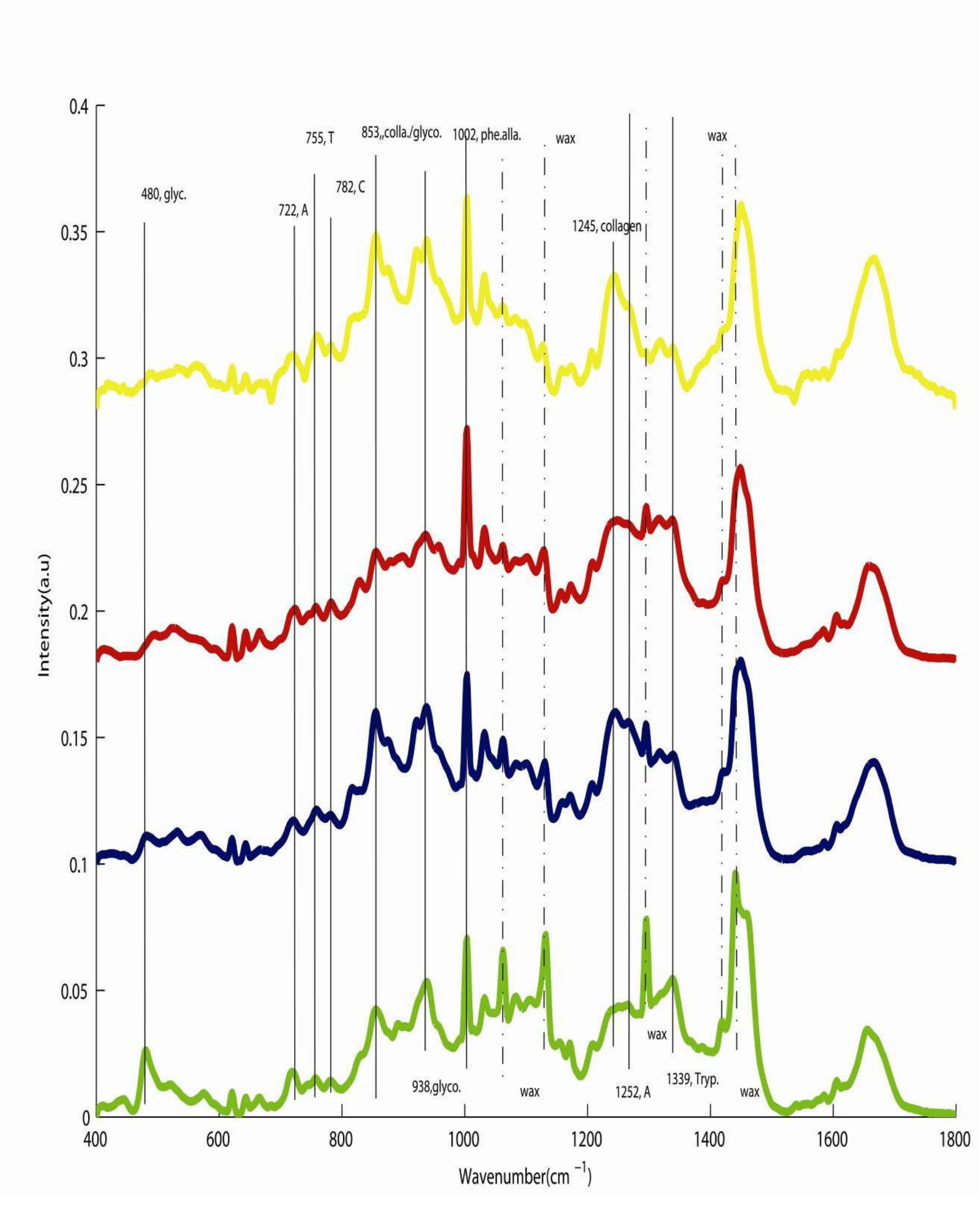


Figure 4.18 Representative mean Raman spectra from KMCA of Raman map, yellow and blue clusters corresponding to stroma, red corresponding to basal layer, and green corresponding to squamous epithelium from the marked region in figure 4.17 c.

4.5 Conclusion

The major purpose of this study is to evaluate the potential of the technique in elucidating the biochemical changes occurring during the progression of Cervical Intraepithelial Neoplasia (CIN) to cancer. The pathologist, in general, divides the cervical epithelium into basal, parabasal, intermediate and superficial layers. The cervical epithelium is supported by the underlying connective tissue known as stroma. Raman micro spectroscopy was employed to visualize this layer separation and also how this separation is affected due to the biochemical changes caused by the progression of the disease.

Raman micro spectroscopy was able to differentiate the normal region of the cervical tissue sample into three layers including stroma, basal/parabasal and superficial layers on the basis of the spectral features of collagen, DNA bases and glycogen. Raman micro spectroscopy reveals more detailed biochemical changes related to the DNA as compared to FTIR spectroscopy as already discussed in **chapter 2**.

The results of the Raman micro spectroscopy presented in this chapter have proven that the technique is very useful in the discrimination of the diseased areas from the normal areas and particularly to describe the biochemical changes in the diseased samples in detail. The Raman spectra contain the signatures of the DNA bases and can be used as a discriminating factor for the differentiation between the layers and between normal and abnormal regions.
On moving from normal to abnormal regions of the cervical tissue sample, the characteristic Raman features of the basal layer are first observed in the superficial layer

and then in the stromal layer. Notably, the region of the CIN III sample which was pathologically classified as normal has been recognised as having biochemical information similar to the abnormal region. This has been indicated by the absence of the collagen Raman spectral bands in the stromal layer as well as absence or minute presence of the glycogen bands in the superficial layer.

All the results, based on KMCA of Raman spectroscopic maps of different cervical tissue samples, presented in this chapter lead to the conclusion that Raman micro spectral mapping of the cervical tissue sections has been able to identify the biochemical changes, occurring during disease progression. The information and conclusion made on the basis of the biochemical changes would not be possible only on the basis of the morphological observations/changes occurring in the tissue.

The presence of the Raman signatures associated with an increase in the DNA content and decrease in the collagen content, due to the progression of disease demand further study to establish the reasons for these changes and understand whether these are due to an interaction of the basal cells with stromal cells or are due to some biochemical changes in the stromal cells themselves. This requires further study on normal and tumour samples as well as an immunohistochemical study. The PCA of these cervical tissue samples are presented in **Chapter 5** and an immunohistochemical study for the evaluation of the expression of the p16 protein, a well known marker protein for cervical cancer is presented in **Chapter 6**.

Chapter 5

Principal Component Analysis of the Raman spectral data from basal, superficial and stromal layers

5.1 Introduction

In the previous chapter, Raman spectral data acquired from cervical tissue sections, analysed by KMCA, was presented. The process differentiates diseased areas from normal areas and particularly helps to elucidate the biochemical changes in the diseased samples. Raman signatures of different representative bio-molecules, including DNA bases, collagen, and glycogen were identified and used as discriminating factors between the basal, stromal and superficial layers of the cells in the cervical tissue samples classified with normal and abnormal areas.

KMCA of the Raman spectral data from the normal and abnormal regions of the cervical tissue samples can lead to the conclusion that, as the disease progresses, the characteristic Raman features of the basal layer initially progress to the superficial layer and then to the stromal layer, based on an increased contribution of DNA features and reduction of the characteristic collagen Raman spectral bands in the stromal layer as well as the glycogen bands in the superficial layer. This information is helpful in the elucidation of the biochemistry of the progression of cervical cancer from normal to abnormal stages and indicates that the decisions made about the stage of the cervical cancer, only on the basis of the morphological changes occurring in the tissue, is not the most appropriate and for this purpose the biochemical changes can be more accurate.

However, the results of the KMCA of the Raman maps pose many further questions. (i) Are the normal regions of a diseased sample biochemically similar to the true normal samples? (ii) Can signatures of HPV infection be observed in early stage diseased samples? (ii) Why does KMCA classify regions of the superficial layer as basal in character? (iii) Why does KMCA classify regions of the stroma as basal in character? While the proliferation of basal-like cells into the superficial layer may be expected in diseased samples, the presence of the Raman signatures associated with an increase in the DNA content and decrease in the collagen content, due to the progression of disease, demands further study to establish the reasons for these changes and to understand whether these are due to an interaction of the basal cells with stromal cells or are due to some biochemical changes in the stromal cells themselves.

KMCA provides a separation of the different layers of the cervical epithelium on the basis of the similarities of the Raman spectral profiles and the clustered data is represented by an average of all the spectra of one layer. In order to further elucidate the biochemical basis of this, and to attempt to address the questions above, principal components analysis (PCA) will be performed on the Raman spectral data and interpreted and presented in this chapter. To demonstrate the process, the Raman spectral maps of true normal tissue will initially be analysed. A pair wise PCA of different regions of tissue, in normal and diseased samples, as identified by KMCA, can highlight the specific biochemical differences (Bonnier and

Byrne, 2012). In terms of the changes occurring in the basal layer, an obvious candidate is infection by the HPV virus. In an attempt to identify spectral signatures of HPV infection, cervical cell lines with different degrees of HPV infection will be compared (Ostrowska et al., 2010).

5.2 Data preprocessing

In order to undertake a detailed analysis of the biochemical differences underlying disease progression, data were preprocessed to remove any contributions of substrate, paraffin or marking ink. All spectra, including calibration and substrate backgrounds, were vector normalized and smoothed using a Savitzky Golay smoothing algorithm of order 5. A rubber band correction for baseline removal for all the spectra was carried out and the substrate spectrum was subtracted from each spectrum (Knief, 2010a). The contribution of the wax was modelled and removed by using Classical Least Square analysis (CLS) (Zhou et al., 2011). All spectral processing and analysis was performed in the Matlab 7.2 (The Mathworks Inc.) environment and the algorithms were developed in-house (Knief, 2010b).

5.3 Data Analysis

Regions of the original Raman data maps clustered by KMCA, and presented in **chapter 4**, were separated by picking the related Raman spectra from that specified area, as indicated by the KMCA in the pseudo colour map. Where identified regions were subdivided, the details are provided in the relevant sections.

5.3.1 Principal component analysis

PCA has been explained in detail in **chapter 3** and in the following sections, it is firstly demonstrated to confirm the differentiation of superficial/basal and basal/stromal Raman spectra from normal tissue, as indicated by KMCA in the previous chapter, but also to confirm the biochemical origin of that separation in a multivariate approach rather than that of identifying individual peaks.

Having established the validity of the PCA, it is then used to explore any biochemical differences between samples classified similarly by KMCA in the previous chapters, and indeed within regions which are clustered. For example, biochemical differences between the basal layers of the normal, intermediate and abnormal samples, as identified in **chapter 4**, are compared to address the question that was raised by the KMCA; whether the basal layer proliferates into the stromal and superficial layers or whether the cells of these layers undergo some biochemical changes during the disease development and progression. It is hoped that this analysis can shed further light on the progression of cervical cancer and may help identify markers of the disease progression associated with each stage of the disease.

5.4 Results: Normal tissue

The KMCA pseudo colour map of the sample of the normal diagnosed tissue section, as well as the visual images of the stained and unstained sections are presented in **Figure 5.1**. This figure is reproduced here to use as a reference for the ability of Raman micro spectroscopy to differentiate between the three layers of the cervical tissue sample, namely the basal, superficial and stromal layers. This layer differentiation was achieved by using KMCA of the Raman spectral data, presented in **chapter 4 (Figure 4.2)**.

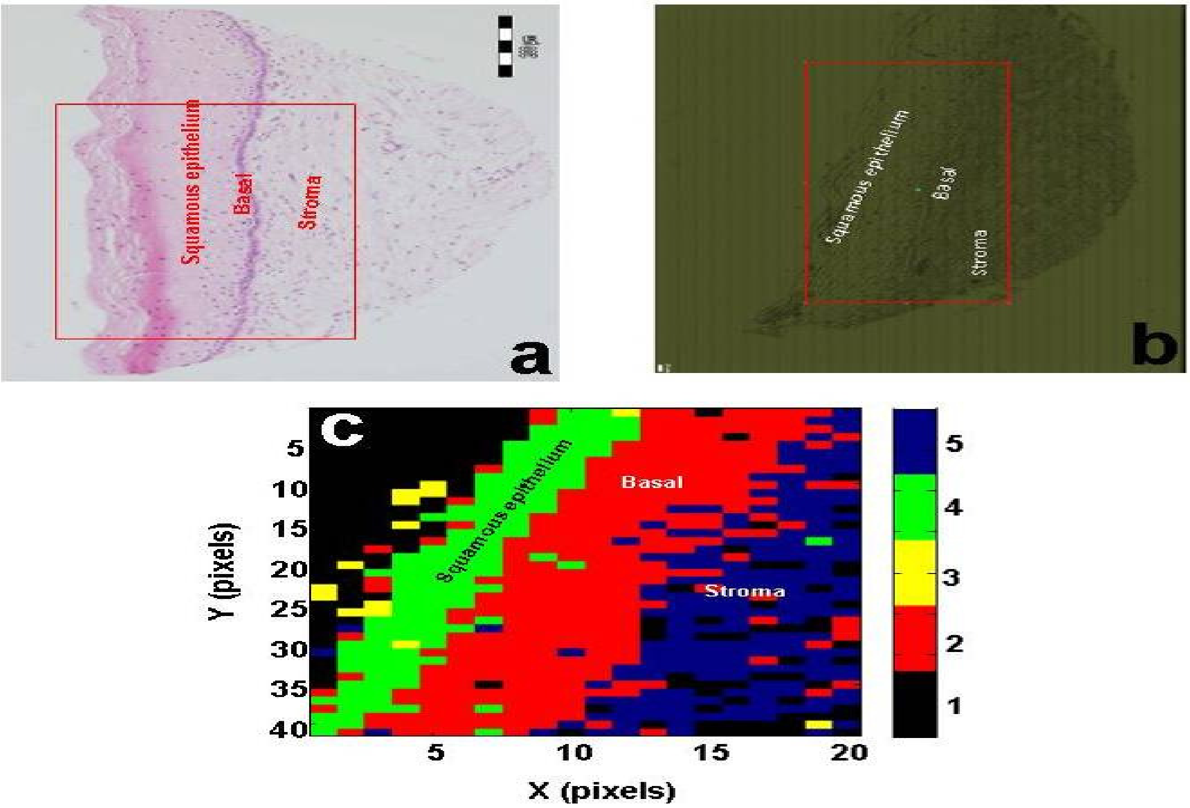


Figure 5.1 H&E stained normal diagnosed tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 50µm (b) Five cluster K means cluster map generated from the Raman map (c).

In order to elucidate the biochemical features responsible for this layer differentiation in addition to those which were identified by the KMCA, PCA was performed on the Raman spectral data associated with each cluster, by comparing the basal layer with the stroma and superficial layers, separately.

5.4.1 PCA of basal-normal vs stroma-normal

Figure 5.2 represents the scatter plot of the PCA for basal vs stroma. These two layers are well differentiated by PC1, which accounts for 67.4 % of the variance. The loading of PC1 is presented in **Figure 5.3**.

Notably, in the PCA scatter plot, the cluster of the Raman spectra of the basal layer is distributed negatively with respect to PC1 and the Raman spectra of the stroma are clustered on the positive side. This is translated to the loadings of PC1 (**Figure 5.3**), as all the negative loadings are associated to the characteristic Raman features of DNA bases and are clearly indicative of the characteristics of the basal layer as it is richer in DNA as compared to the stromal layer. The features associated with DNA, contributed by the basal layer, include those at 728 cm^{-1} (A), 827 cm^{-1} (O-P-O of DNA), 898 cm^{-1} (deoxiribose ring breathing) ,1276 cm^{-1}and 1340 cm^{-1} (G). Only a few features related to proteins are observed, including those at 509 cm^{-1} (S-S stretching), 642 cm^{-1} (C-C stretching), 1129 cm^{-1} (C-N stretching) and 1449 cm^{-1} (CH deformation). On the other hand, the positive loadings can be assigned to the characteristic Raman features of the stromal layer. These loadings are assigned to collagen and differentiate this layer from the basal layer. The features contributed by the stroma include 853 cm^{-1}, 921 cm^{-1}, 972 cm^{-1} and 1242 cm^{-1}, assigned to the collagen of the stromal layer and 1165 cm^{-1}, 1195 cm^{-1} (C-N stretching) and 1655 cm^{-1} (amide-I).

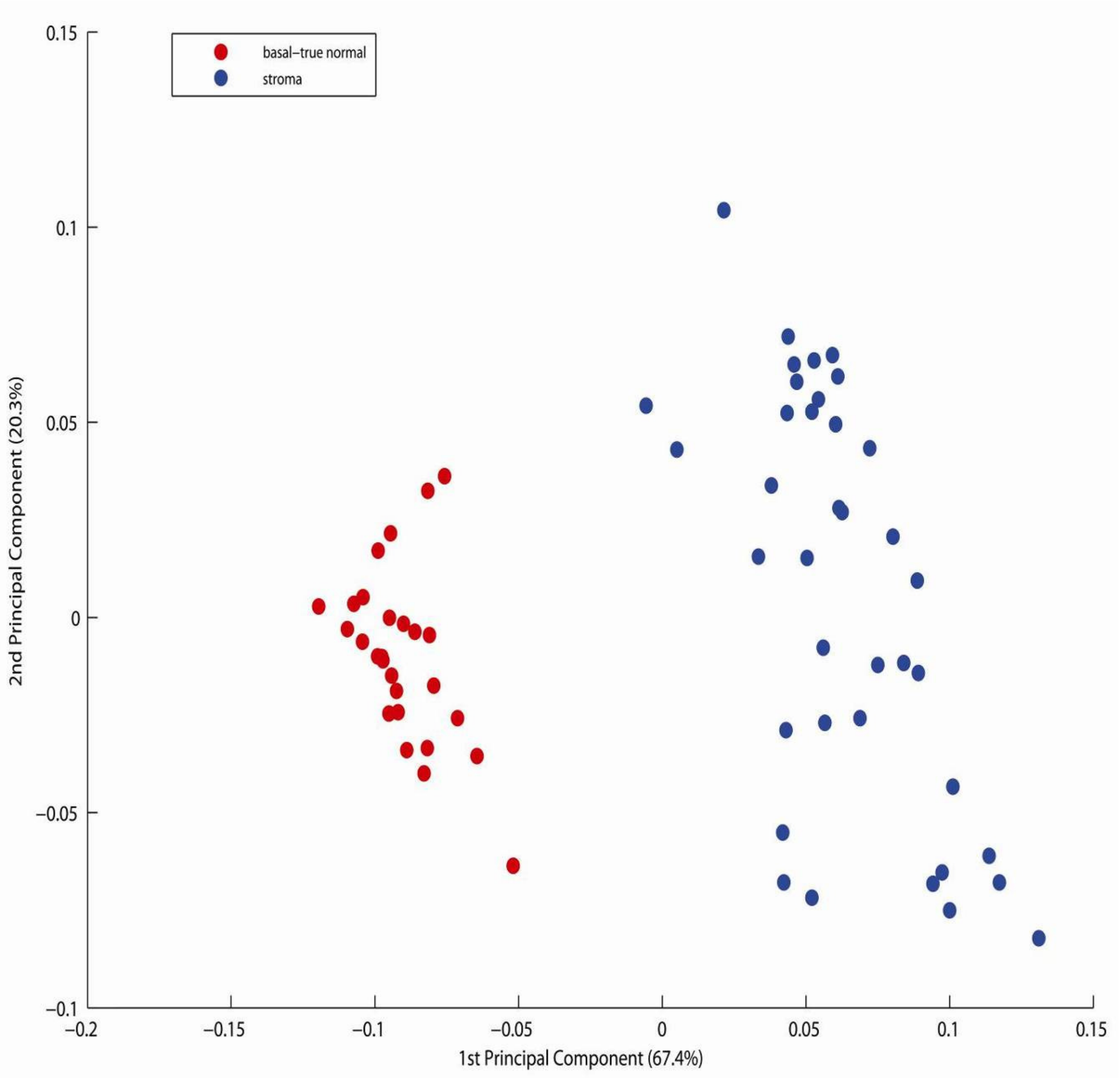


Figure 5.2 PCA scatter plot for basal-normal vs stroma-normal.

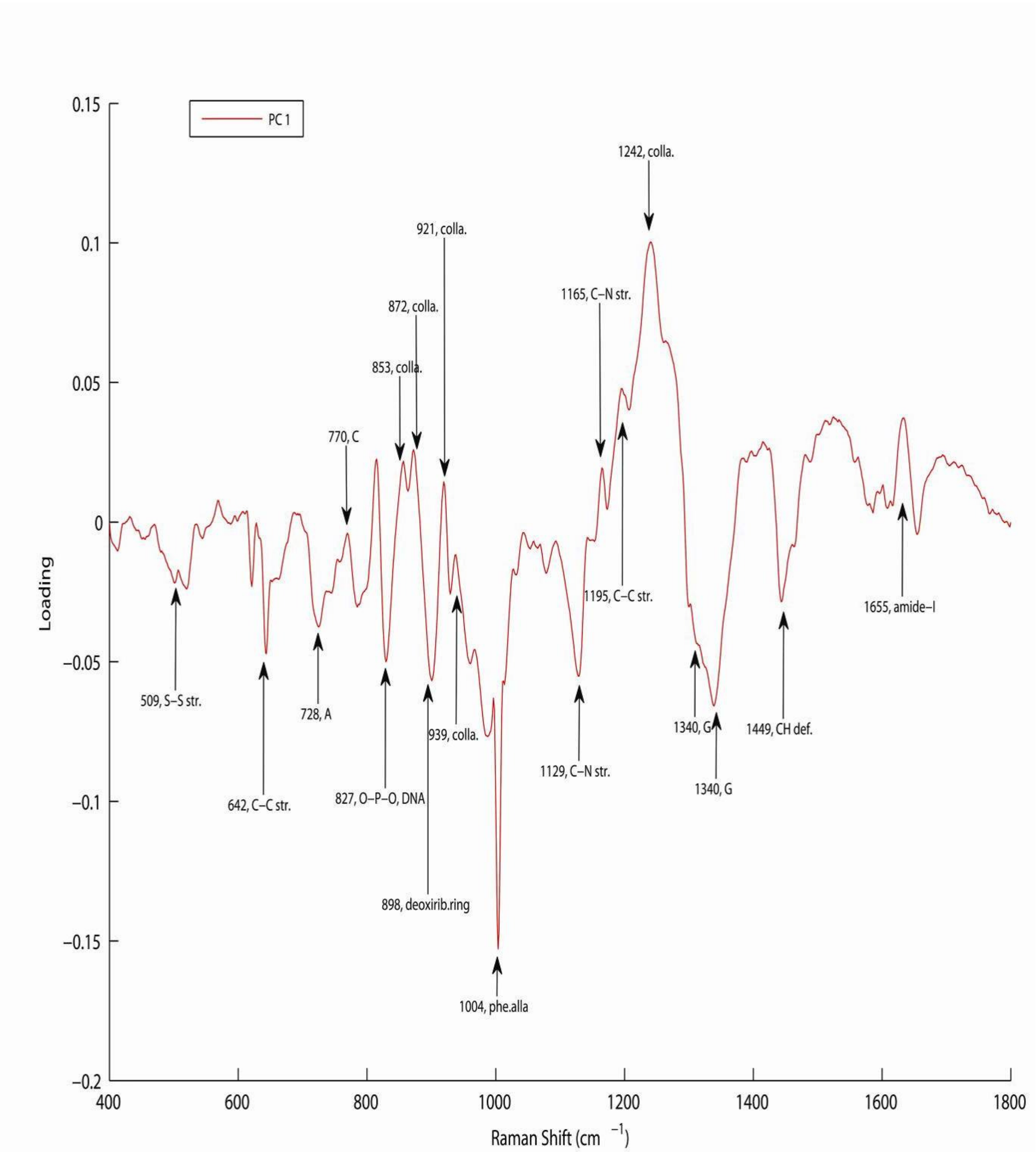


Figure 5.3 PCA loading (PC-1) for basal-normal vs stroma-normal.

5.4.2 PCA of basal-normal vs superficial-normal

Figure 5.4 and **Figure 5.5** represent the PCA scatter plot and loadings of the PCA for the basal-normal versus superficial-normal, respectively. A good separation of the two layers is observed as two well distinguished clusters of the spectra in the PCA scatter plot in the **Figure 5.4**. In the PCA scatter plot, the cluster of the Raman spectra of the basal-normal layer is observed in the negative axis and Raman spectra of the superficial layer are clustered in the positive axis which is consistent with the loadings of the PCA (**Figure 5.5**), as the negative loadings are related to the characteristic Raman features of DNA bases which are characteristic of the basal layer. The clustering of the Raman spectra of the superficial layer is observed in the positive axis, which is consistent with the loadings of the PCA, because the positive loadings are assigned to the characteristic Raman features of glycogen, a specific marker of this layer.

The loadings (PC1) show the Raman features which can be related to the basal layer include the DNA related Raman bands and glycogen related Raman bands which can be attributed to the superficial layer. Specifically, the loadings of PC1 show features including 755 cm^{-1} (T), 728 cm^{-1}, 1338 cm^{-1} (A), 775 cm^{-1} (C), 1182 cm^{-1}, 1313 cm^{-1} (G) and 1080 cm^{-1} (O-P-O of DNA). Other features observed including 480 cm^{-1}, 853 cm^{-1} and 938 cm^{-1} are assigned to glycogen of the superficial layer.

The PCA of the basal layer of the two normal samples designated “true normal-1” and “true normal-2” is presented in the Appendix, A1.1 and A1.2, as scatter plot and loadings respectively. The scatter plot shows some differentiation of the two clusters of the Raman spectra which are attributed to some degree of inter-patient variability, largely based on intensity variations of peaks associated with proteins and lipids.

5.4.3 Summary

These results of the PCA of the normal basal layer versus the normal stromal and superficial layers demonstrate the ability of the PCA to highlight the spectral features which lead to the differentiation of these layers on the basis of the Raman spectral features characteristic for each layer. A similar analysis will now be employed to investigate whether the “normal” regions of diseased samples are biochemically similar to samples classified as “true normal”.

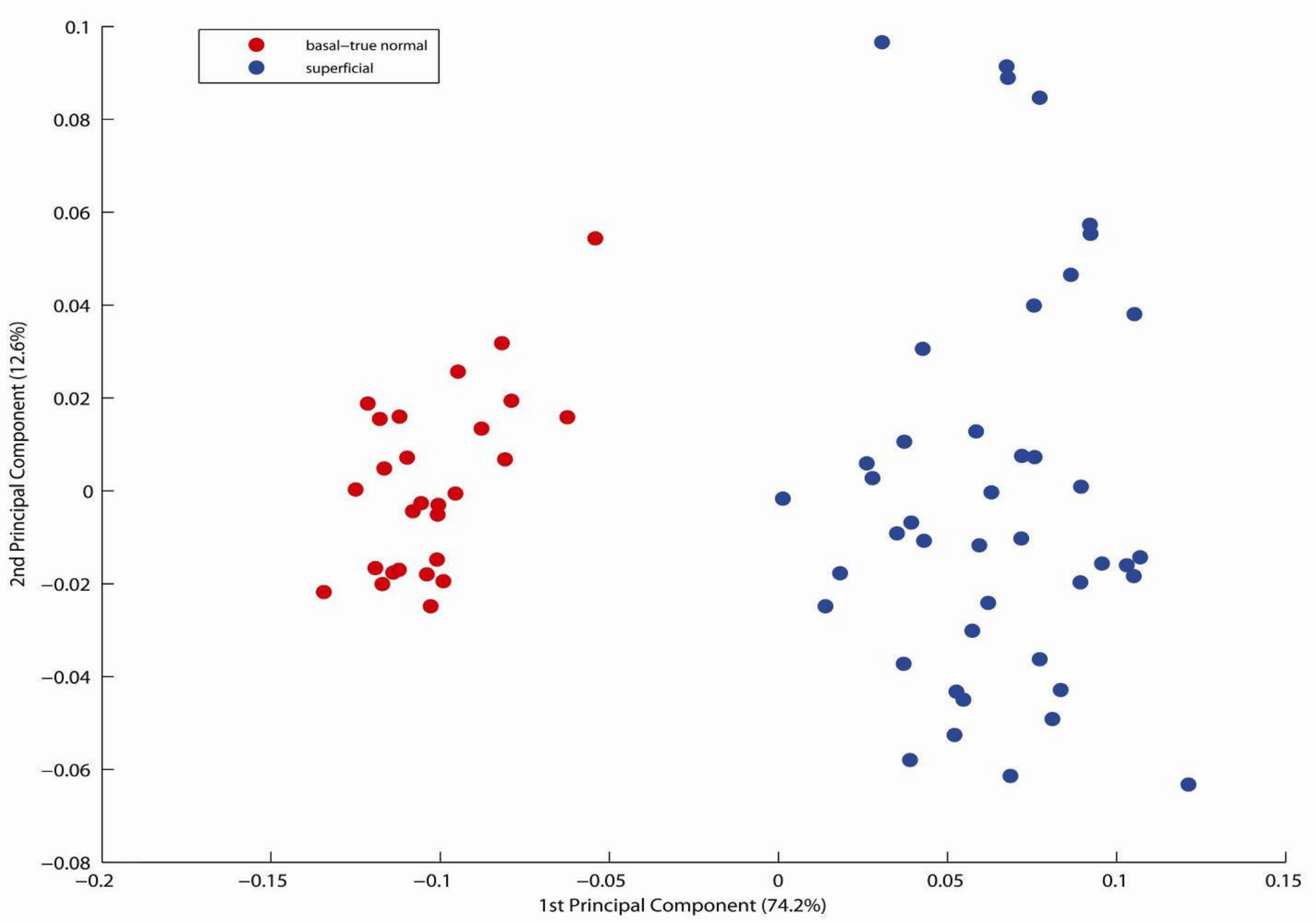


Figure 5.4 PCA scatter plot for basal-normal vs superficial-normal.

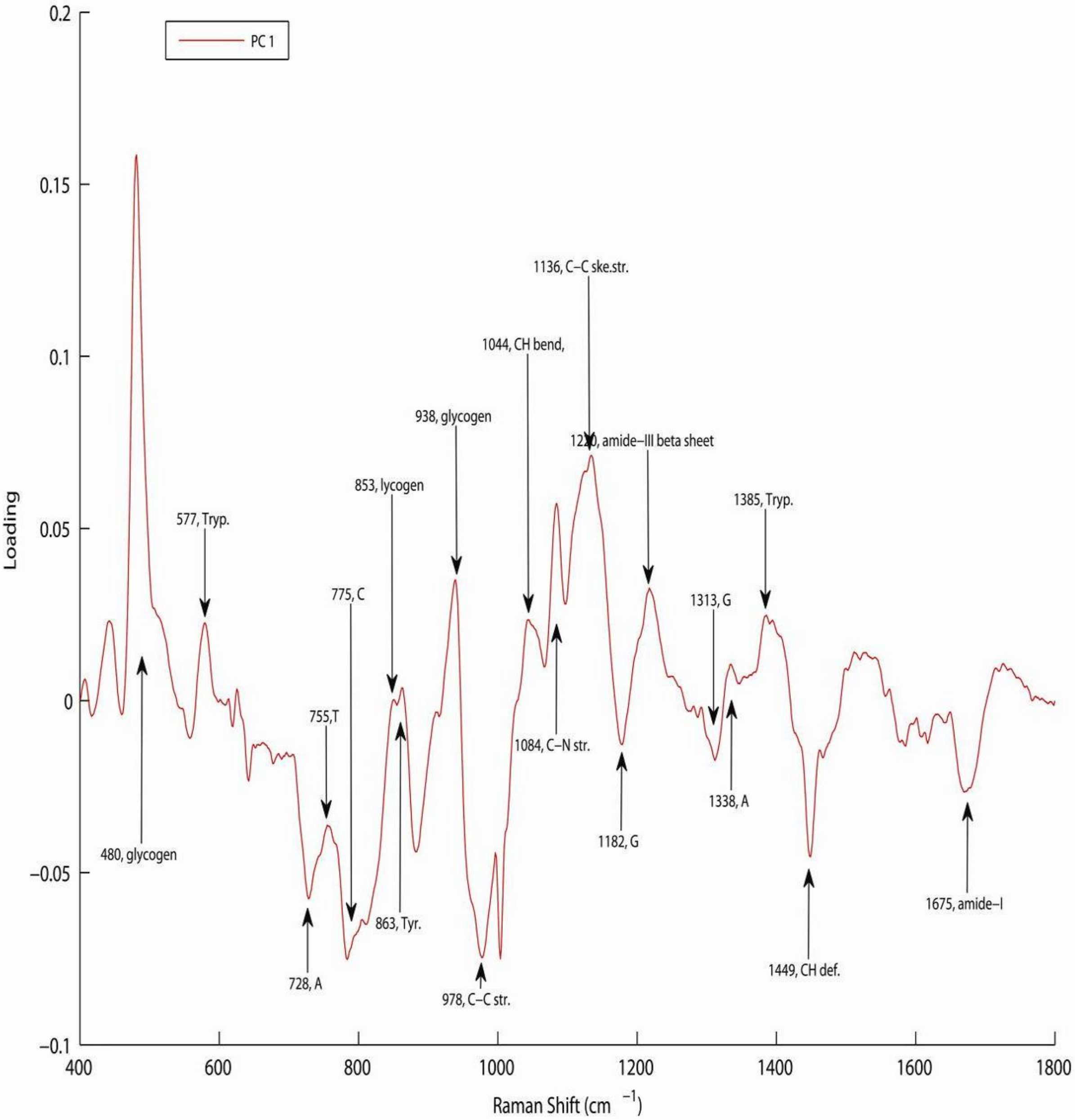


Figure 5.5 PCA loading (PC-1) for basal-normal versus superficial-normal.

5.5 Results: True normal versus diseased samples

5.5.1 PCA for Raman spectra of the basal-true normal vs CIN II normal

The basal layer of the "true-normal" of **Figure 4.1** in chapter 4 and here in the **section 5.4.1** (**Figure 5.1**) was compared to a region of a CIN II sample of **Figure 4.3** in chapter 4 (**Figure 5.6**), characterized by the pathologist as "normal" in order to establish whether Raman spectral signatures associated with disease are identifiable outside of the diseased region. The results are presented as PCA scatter plot and the loadings in **Figure 5.7** and **Figure 5.8** respectively. The two layers are well differentiated by PC1, which accounts for 74.5 % of the variance. Notably, in the PCA scatter plot, the cluster of the Raman spectra of the basal layer from the true normal sample (basal-true normal) is distributed positively with respect to PC1 and the Raman spectra of the basal layer of the CIN II normal sample (basal-1) are clustered on the negative side. This is translated to the loadings of PC1 (**Figure 5.8**) as all the positive loadings are associated with the characteristic Raman features of DNA bases and are clearly indicative of the characteristics of the basal layer. The loadings contributed by the true normal basal layer include 728 cm^{-1} (A), 795 cm^{-1} (C), 828 cm^{-1} and 1088 cm^{-1} (O-P-O, DNA), 881 cm^{-1} (deoxiribose ring breathing), 849 cm^{-1} (glycogen) and 1449 cm^{-1} (CH deformation).

On the other hand, in the negative loadings, contributed by the basal-1 (CIN II normal basal layer), the major difference as compared to the true normal basal is the one assigned to amide-III beta sheets of the proteins at 1222 cm^{-1}. Notably, this feature is not observed in the loadings of the PCA of the basal layers of the two normal samples including "true normal-1" and "true normal-2" (Appendix, **A1.1** and **A1.2**). Moreover, PCA of the basal-1 (CIN II normal basal layer) with the second true normal sample "true normal-2" was also done and presented in the appendix (**A1.3** and **A1.4**) as scatter plot and loadings respectively. Notably, the loadings are the same and 1222 cm^{-1} feature is also observed there which further confirms the above findings.

Other loadings can be assigned to the characteristic Raman features of the DNA, including 1095 cm^{-1} (O-P-O, DNA), 1318 cm^{-1}, 1585 cm^{-1} (G). One band related to glycogen at 482 cm^{-1} is also observed. These results of the PCA show that the normal side of the diseased sample (basal-1) is not the same as the "true normal".

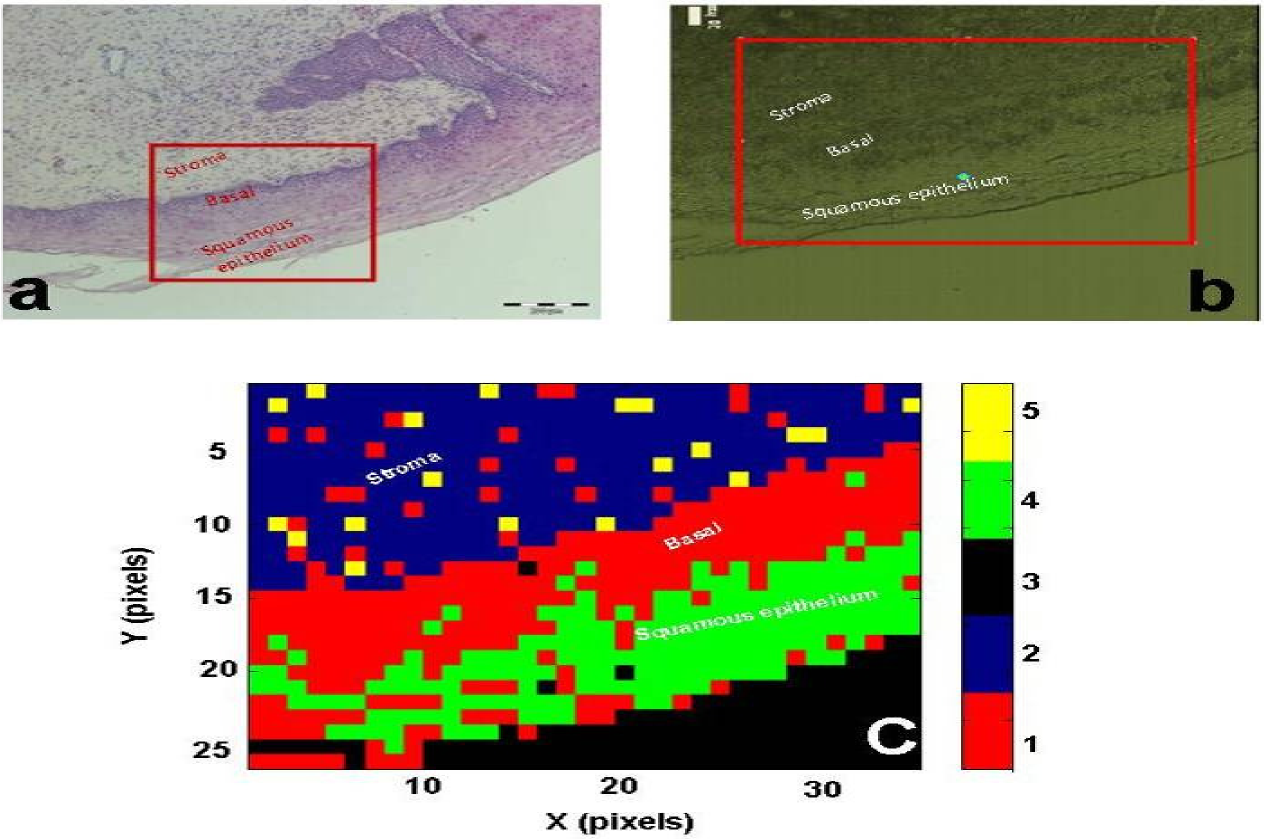


Figure 5.6 H&E stained normal region of the CIN II diagnosed tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 50µm (b) Five cluster K means cluster map generated from the Raman map (c).

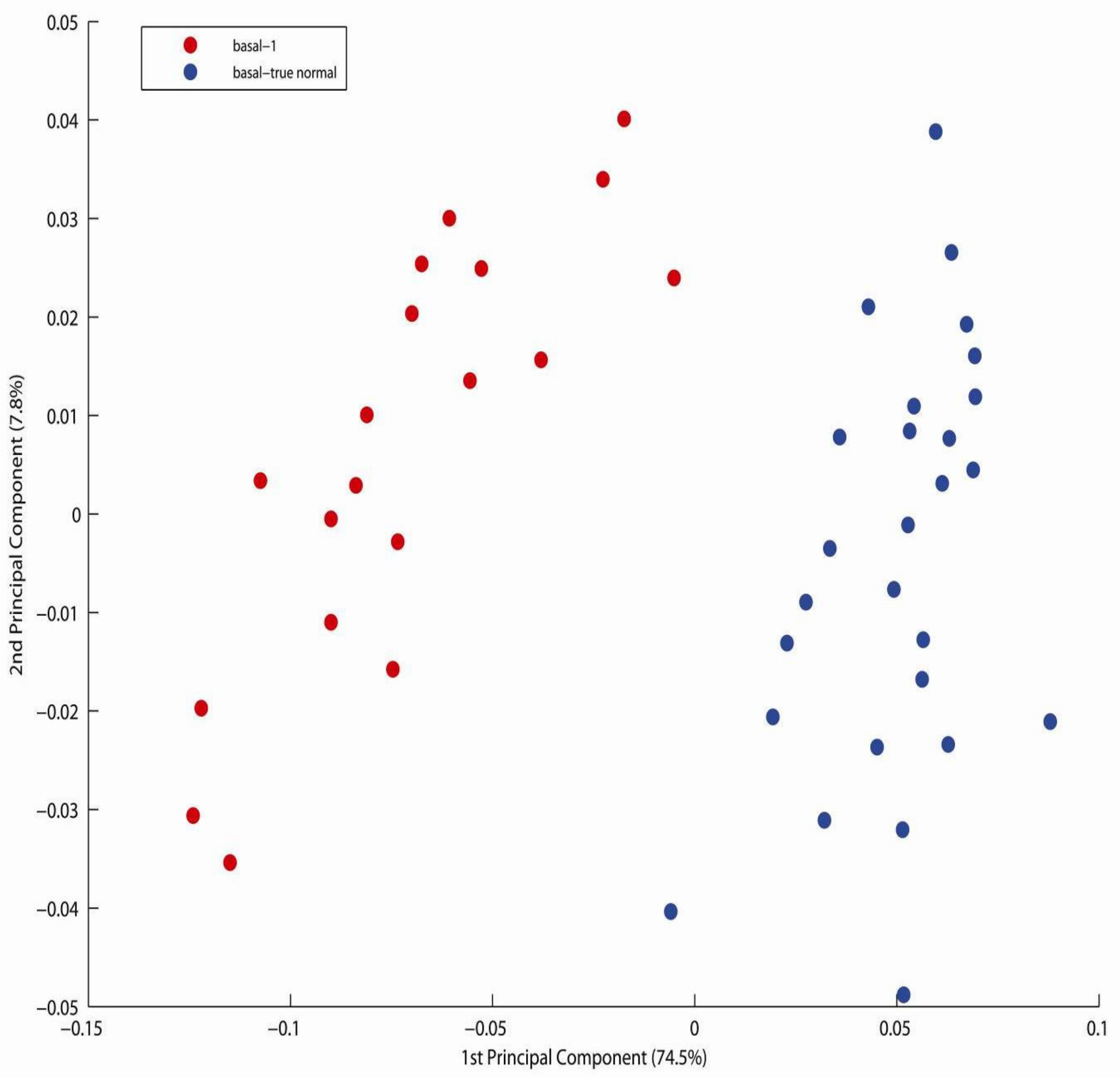


Figure 5.7 PCA scatter plot of basal true normal vs CIN II normal (basal-1).

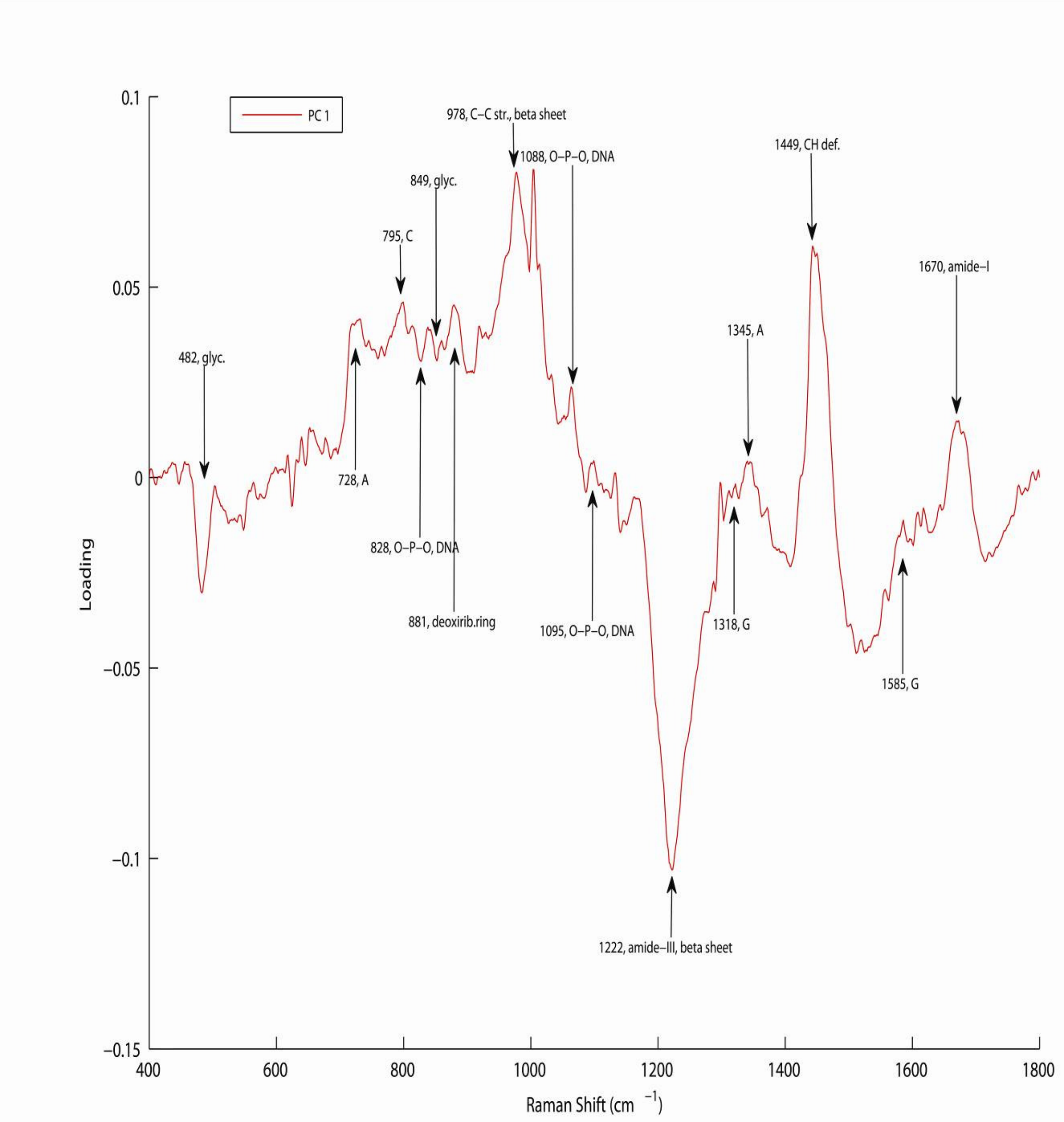


Figure 5.8 PCA loadings of basal true normal vs CIN II normal (basal-1).

5.5.2 PCA of Normal versus Intermediate

It has been observed in **chapter 4** that in the CIN II diagnosed tissue sample which has been identified as Intermediate by KMCA, the relative strengths of the DNA bands (including those at 728cm^{-1} and 856cm^{-1}) increase in the superficial layer as the disease moves from the basal layer to the superficial layer. In addition, some bands of glycogen were also observed in the superficial layer. KMCA showed the presence of the DNA related bands and clustered regions of the superficial layer with the basal layer further supporting this. PCA of the normal tissue region gives a more complete picture of the spectral differences between the two layers, however, and thus may be more precise in elucidating the biochemical differences associated with this early stage of disease progression. Moreover, it is of interest to establish whether the characteristics of the basal layer itself remain the same or not from one stage of the disease to another.

The KMCA classified basal layer (basal-2, **Figure 4.13)** of the diseased region of the CIN II diagnosed tissue, presented here as **Figure 5.9,** was divided into two parts; (i) basal 2a (ii) basal 2b, which are histologically less diseased and more diseased respectively and thus PCA was applied to determine any biochemical differences between the two parts of the basal layer. The basal 2b region was further subdivided into two sub-regions (basal 2b-i and basal 2b-ii, according to the line of separation of basal-2a and the superficial layer, as indicated in **Figure 5.9**, and PCA was performed on this data to explore whether these two groups are similar or different in their biochemical composition, affected by the progression of the disease. Basal-2b-i is expected to be closer in biochemical composition to basal-2a and indeed that of the normal region characterised in the previous section (basal-1) and basal-2b-ii may be expected to resemble the normal superficial layer. These results are presented in the following sections.

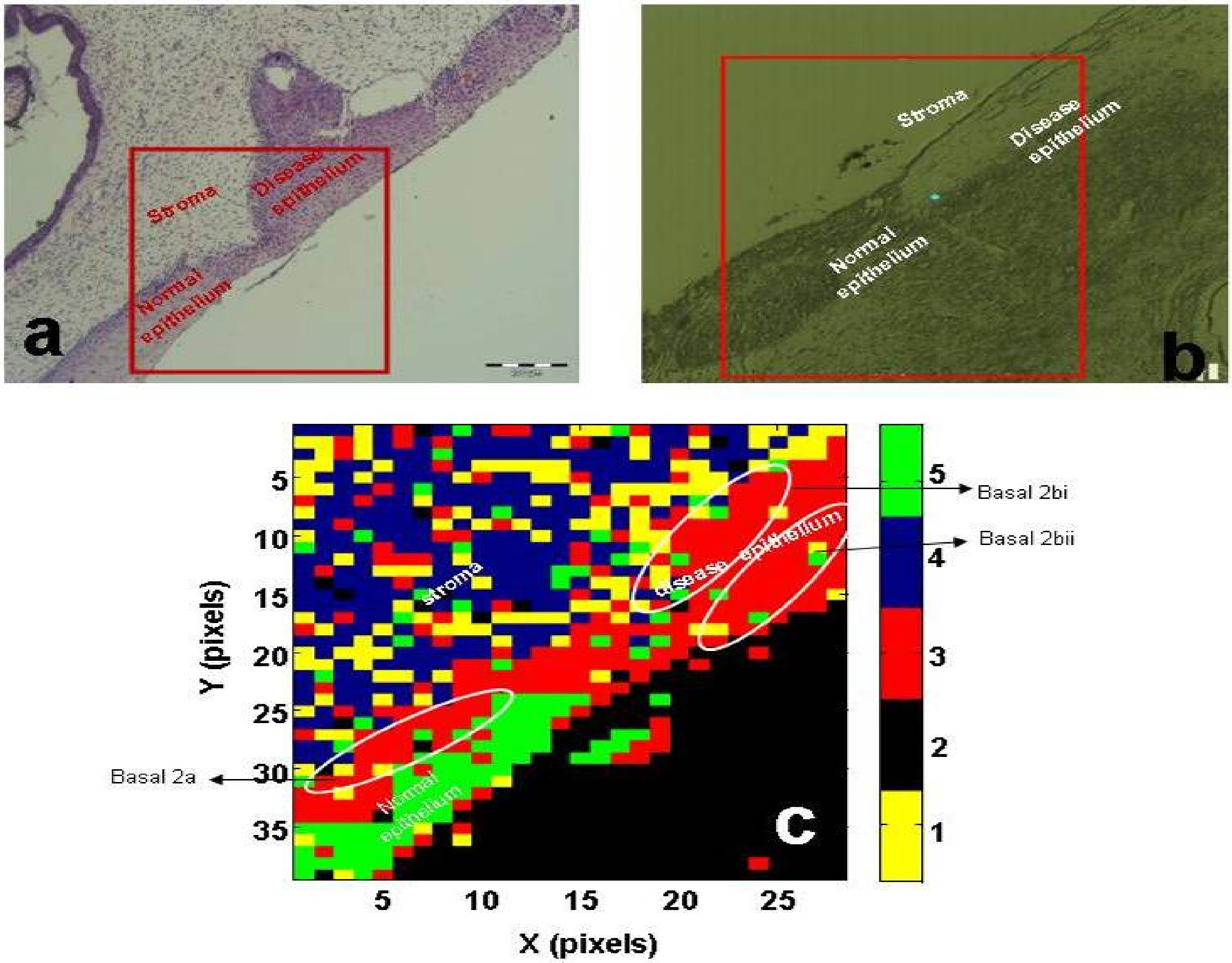


Figure 5.9 H&E stained normal & disease region of CIN II diagnosed tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster KMCA map generated from the Raman map (c).

5.5.3 PCA results of basal true normal vs basal 2a

PCA of the basal layer of the true normal vs the normal side of the basal layer of an intermediate sample (basal-2a) is presented in **Figure 5.10** and **Figure 5.11** as scatter plot and loadings of the PCA, respectively. It was expected that the basal-true normal would be differentiated from the basal-2a in a similar way as basal 1 of the normal-CIN II section. Moreover, these differentiating Raman spectral features may be taken as the markers of the early stage of the disease or a pre-disease stage.

As expected, the PCA scatter plot differentiates the two regions. The Raman spectra of the basal-true normal are clustered in the negative axis and spectra of the basal-2a in the positive axis of the scatter plot. This indicates that the loadings of the PCA which are negative, including those at 480 cm^{-1}, 849 cm^{-1} and 939 cm^{-1} (glycogen), and some loadings related to DNA, including the feature at 718 cm^{-1} (Adenine), are contributed by the basal-true normal layer. The positive loadings associated with the DNA, including 1088 cm^{-1} (O-P-O of DNA) and 1346 cm^{-1} (Guanine) and 1202 cm^{-1}, 1222 cm^{-1} (amide-III beta sheet) and 1675 (amide-I) are contributed by the basal-2a. This can lead us to conclude that PCA helps to differentiate the basal-2a from the basal-true normal and confirms that basal-2a is similar to the basal-1 in biochemical composition, as both layers of the cells are closely related to normal but not identical. This is further supported by the observation of the similar loadings contributed by the basal-2a region (**Figure 5.11**) as contributed by the basal-1 region, when compared to the basal-true normal by using PCA (**Figure 5.8**).

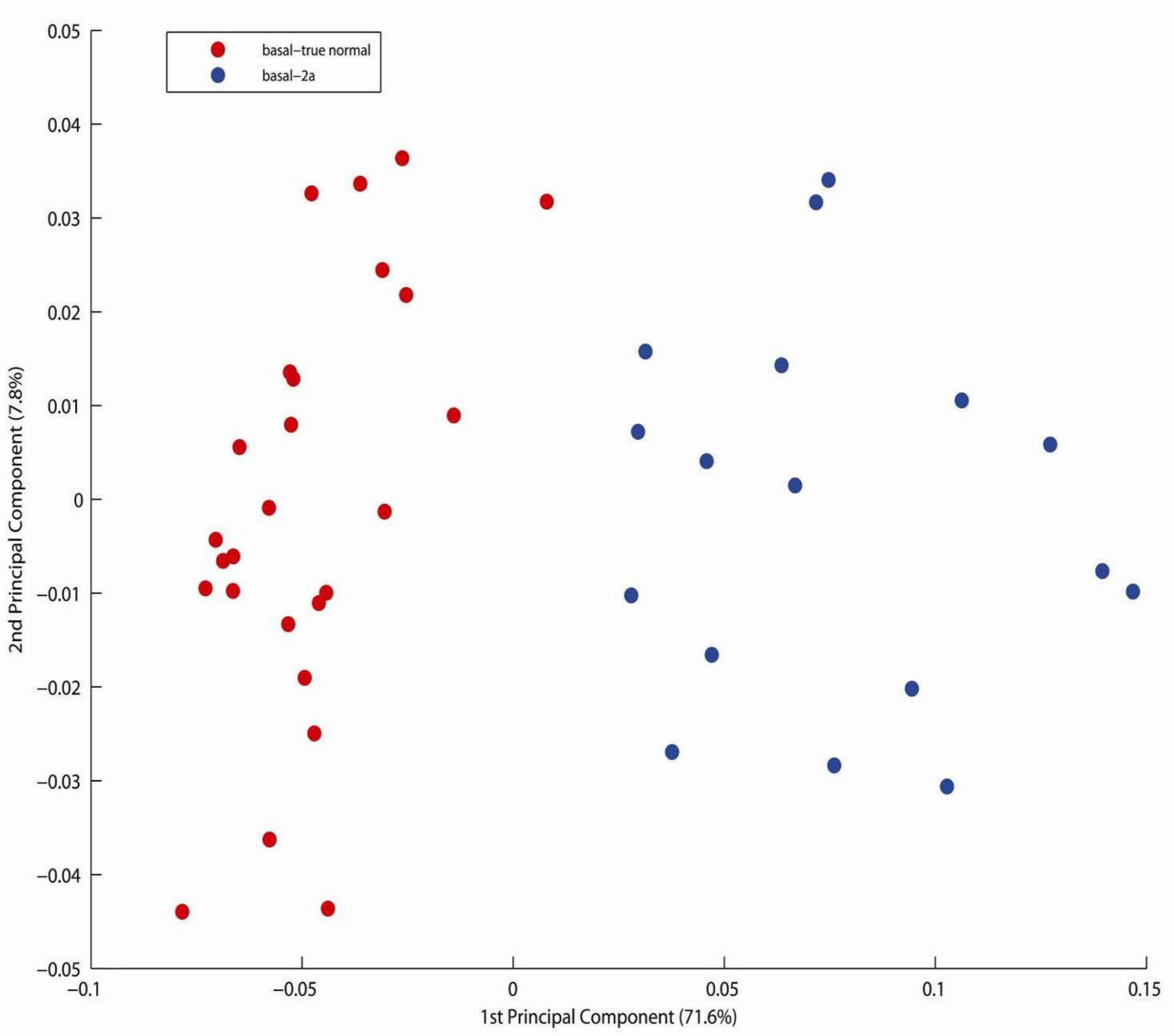


Figure 5.10 PCA Scatter plot for basal-true normal vs basal-2a.

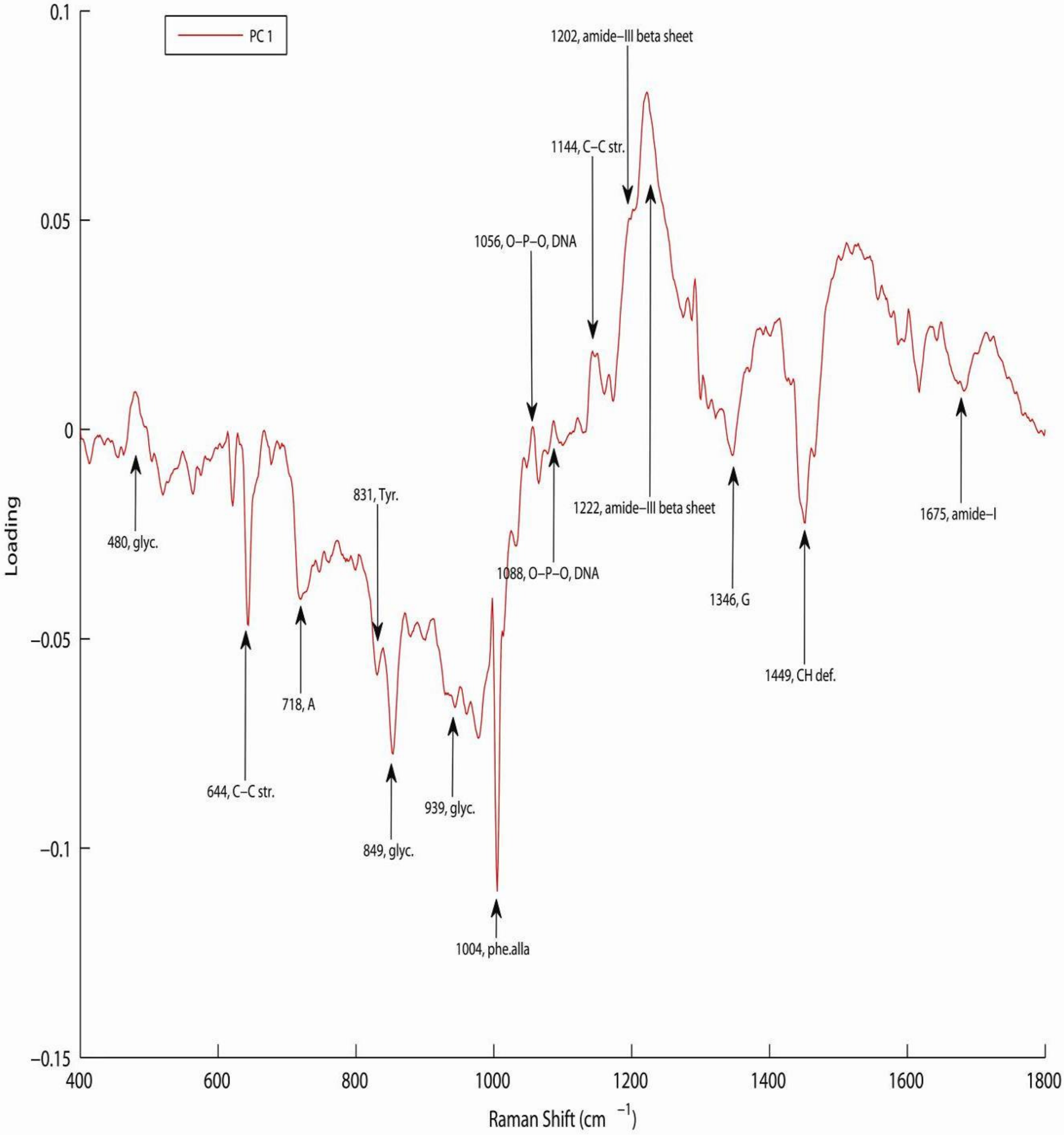


Figure 5.11 PCA loading (PC-1) for basal-true normal vs basal-2a.

5.5.4 PCA results of CIN II normal (basal-1) vs basal 2a

To further confirm the PCA results presented in the previous section, PCA of the basal-2a and basal-1 was carried out. As both the basal-1 and basal-2a are the normal sides of an intermediate cervical tissue sample, it was expected that there will be no proper differentiation between their Raman spectral data.

PCA of these Raman spectral data sets show that PC1, which accounts for 54% of the observed variance, does not differentiate them as indicated by the PCA results of the basal-1 vs the basal-2a (normal side of the basal layer of an intermediate sample) presented in **Figure 5.12** and **Figure 5.13** as scatter plot and loadings of the PCA, respectively. Although, PC2 does differentiates the, as indicated by the PCA scatter plot, it explains only 26 % of the variance in the Raman spectra. Notably, in the PC2 loadings, the major features are associated with glycogen, including those at 480, 849, 939cm^{-1}, which may indicate that there is difference of the glycogen level in basal-1 compared to basal-2a. As the Raman spectra related to basal-1 are clustered in the positive axis and those of the basal-2a in the negative (**Figure 5.13**), the loadings associated with the glycogen in the positive, are contributed by the basal-1 region.

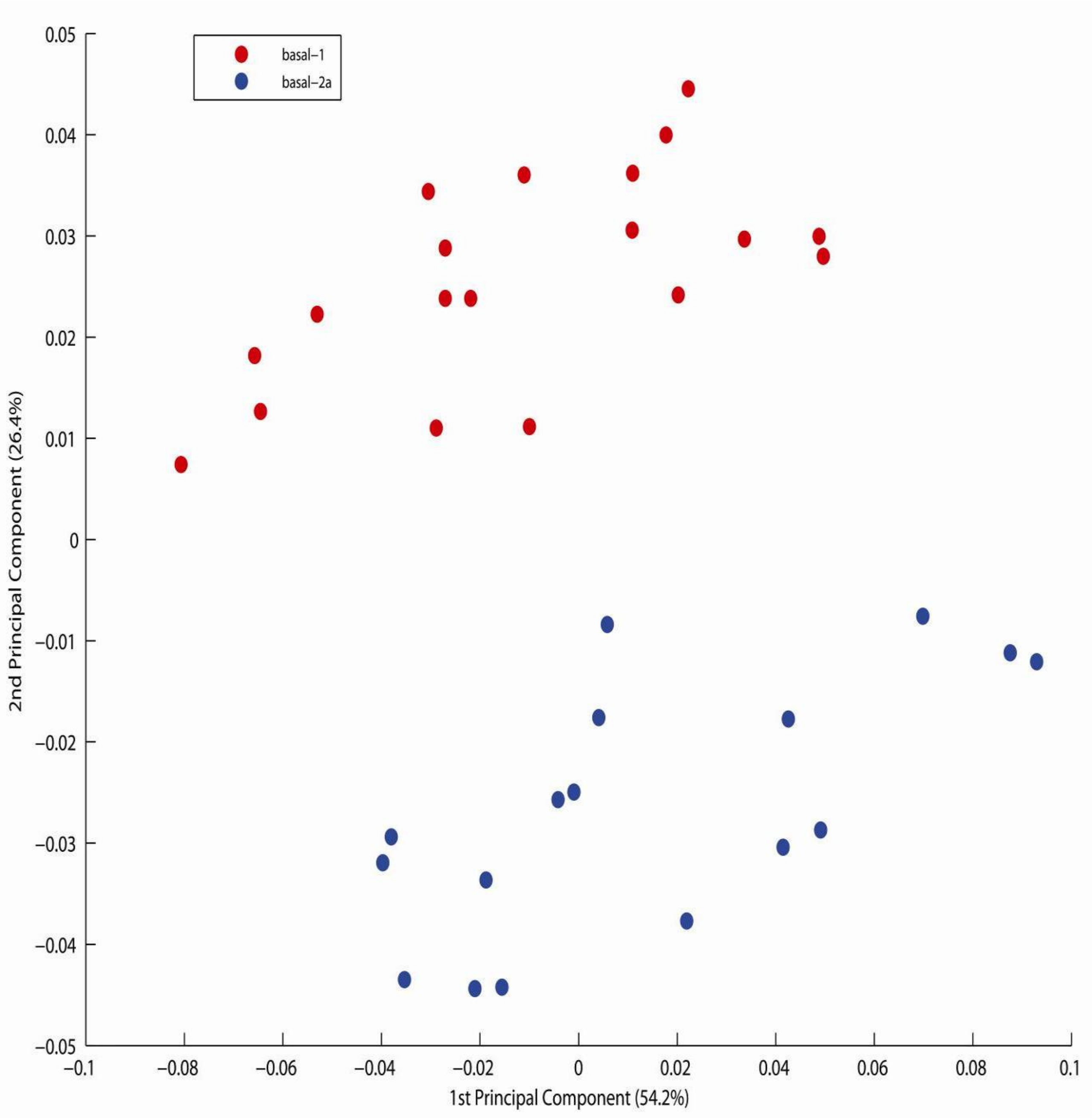


Figure 5.12 PCA scatter plot for CIN II normal (basal-1) vs basal-2a.

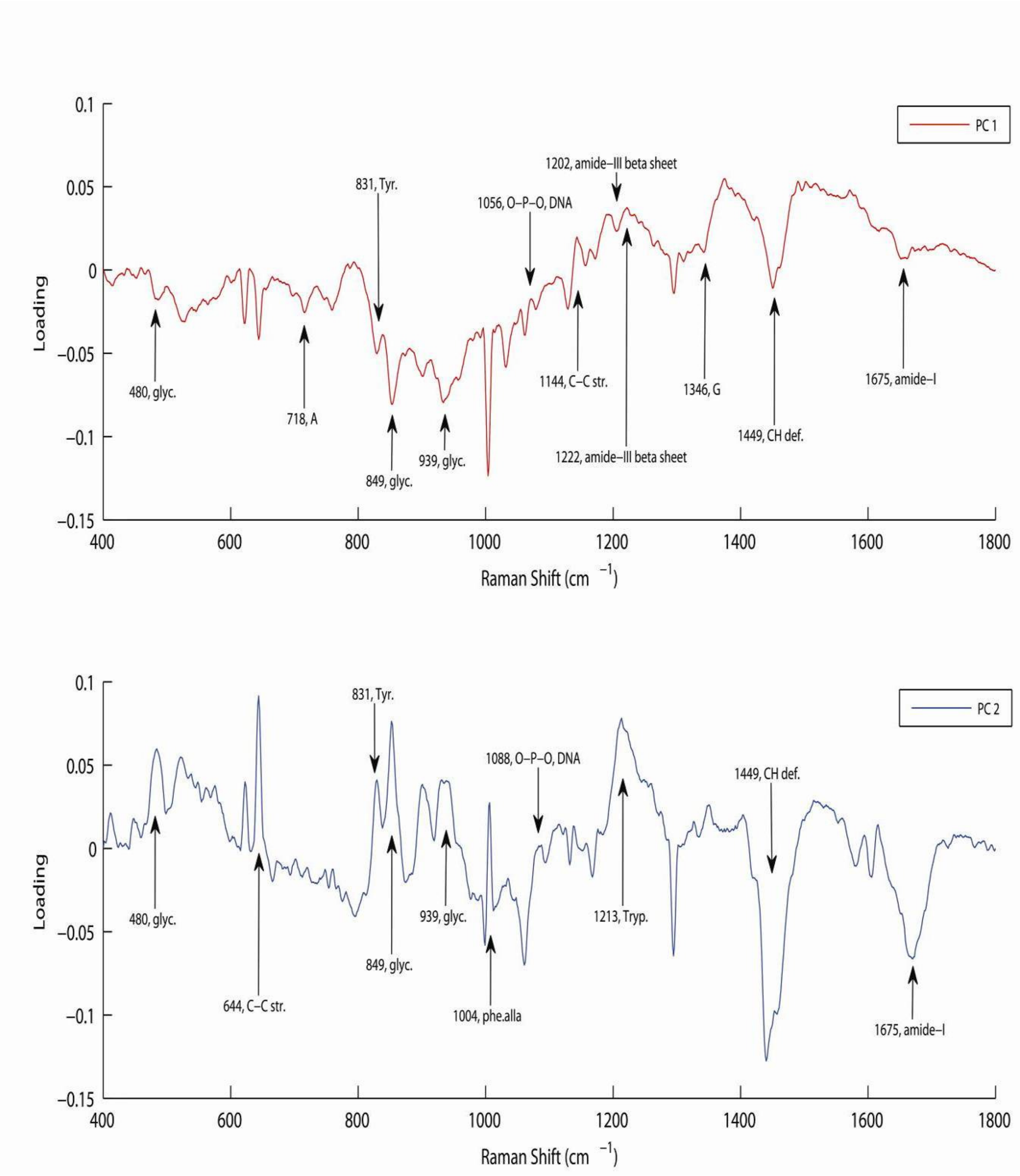


Figure 5.13 PCA loadings for CIN II normal (basal-1) vs basal-2a.

5.5.5 PCA for Raman spectra of the cervical cancer cell lines

HPV infection is accepted as a significant factor associated with cervical cancer, and thus, in an attempt to guide the interpretation of spectral differences associated with biochemical changes in early stage diseased tissue, the Raman profiles of model cell lines were further analyzed in an attempt to establish spectroscopic markers for HPV infection. Raman spectral data from different HPV-infected human cervical cancer cell lines having different numbers of the integrated copies of HPV per cell including HPV 16-positive SiHa (1-2 copies of HPV per cell) and CaSki (60-600 copies), and HPV 18-positive HeLa cells (20-50 copies) along with the HPV negative C33A cells (Choi et al., 2005), from a previous study conducted in DIT (Ostrowska et al., 2010) was used. This was done to find the Raman spectral differences between these model cell lines as a potential spectral marker of HPV infection and to compare with the Raman spectral changes/PCA results of the true normal cervical tissue vs CIN II normal cervical tissue samples, to establish whether the observed differences can be associated with HPV infection in the normal region of the diseased sample. The comparison of the HPV negative cell line C33A with the cell line with the highest HPV content, CaSki, is presented here, although similar and consistent results were seen for the other cell lines.

The scatter plot for the PCA of the *C33A* vs *CaSki* cell lines is presented in **Figure 5.14.** As expected, the cell lines are well differentiated by the first principle component (PC1) which accounts for 71.2 % of the variance. The cell lines are not differentiated by PC2, indicating that the associated variance is independent of cell line and most likely has origin in point to point measurement variations. **Figure 5.15** shows the loading of PC1, which indicates the spectral features which contribute most to the differentiation. Negative loadings are stronger in the negatively scattered data (CaSki), whereas positive loadings are stronger in the positively scattered data (C33A). The features related to the DNA bases, including those at 828 cm^{-1}, 1418 cm^{-1} (O-P-O, DNA), 1181 cm^{-1}, 1318 cm^{-1} (G), 1344 cm^{-1}, 1491 cm^{-1} (A) and 1375 cm^{-1} (T) are contributed by the CaSki cells, while those at 718 cm^{-1} (A), 786 cm^{-1} (C), 1587 cm^{-1} (guanine) are contributed by the C33A cells. Features related to proteins are also observed, including those at 1661 and 1675 cm^{-1} (amide-I) are contributed by the spectra of the C33A cells.

The comparison of the PCA loadings of the cell lines, **Figure 5.15,** with the PCA loadings of the basal-true normal versus basal-CIN II normal, **Figure 5.8**, indicates that there are no strongly matching Raman signatures which could lead us to identify the signatures of HPV infection in the cervical cancer tissue samples. The loading at 1222 cm^{-1} (amide-III beta sheet) in the basal-true normal versus basal-CIN-II normal may be of interest as it is the major difference between the two tissue data sets, but it is not prominent in the loadings which differentiates the cell lines.

Notably, the most prominent bands in **Figure 5.8** are protein related and are related to the physiological effects of HPV infection in cells, rather than the DNA bands associated with the virus itself. Due to the very small genome size of the HPV as compared to the human genome, it has been argued that it is unlikely that the signatures of the virus itself can be identified (Diem et al., 2012). Notably, the physiological effects of HPV infection in tissue, *in vivo*, may be very different from those in cells, *in vitro*. Moreover, the Raman spectra from the cell lines have been acquired by culturing the cells *in vitro* and focusing on the nucleus of the cells, whereas in the case of the tissue samples, the samples have been dewaxed prior to the acquisition of the spectra and data was taken from random points on the tissue rather than specifically cell nuclei. The strong feature at 1222cm^{-1} in **Figure 5.8** is also present in the intermediate and abnormal samples and maybe considered as an early marker of the disease progression and potentially has origin in HPV infection. To further confirm this observation, the PCA of the basal layer of the two normal samples including "true normal-1" and "true normal-2" is presented in the Appendix, **A1.1** and **A1.2** as scatter plot and loadings respectively. In the loadings, the 1222 cm^{-1} is not observed which further can lead to the conclusion that the amide-III beta sheet (1222 cm^{-1}) may be taken as Raman spectral marker of the early progression of the disease.

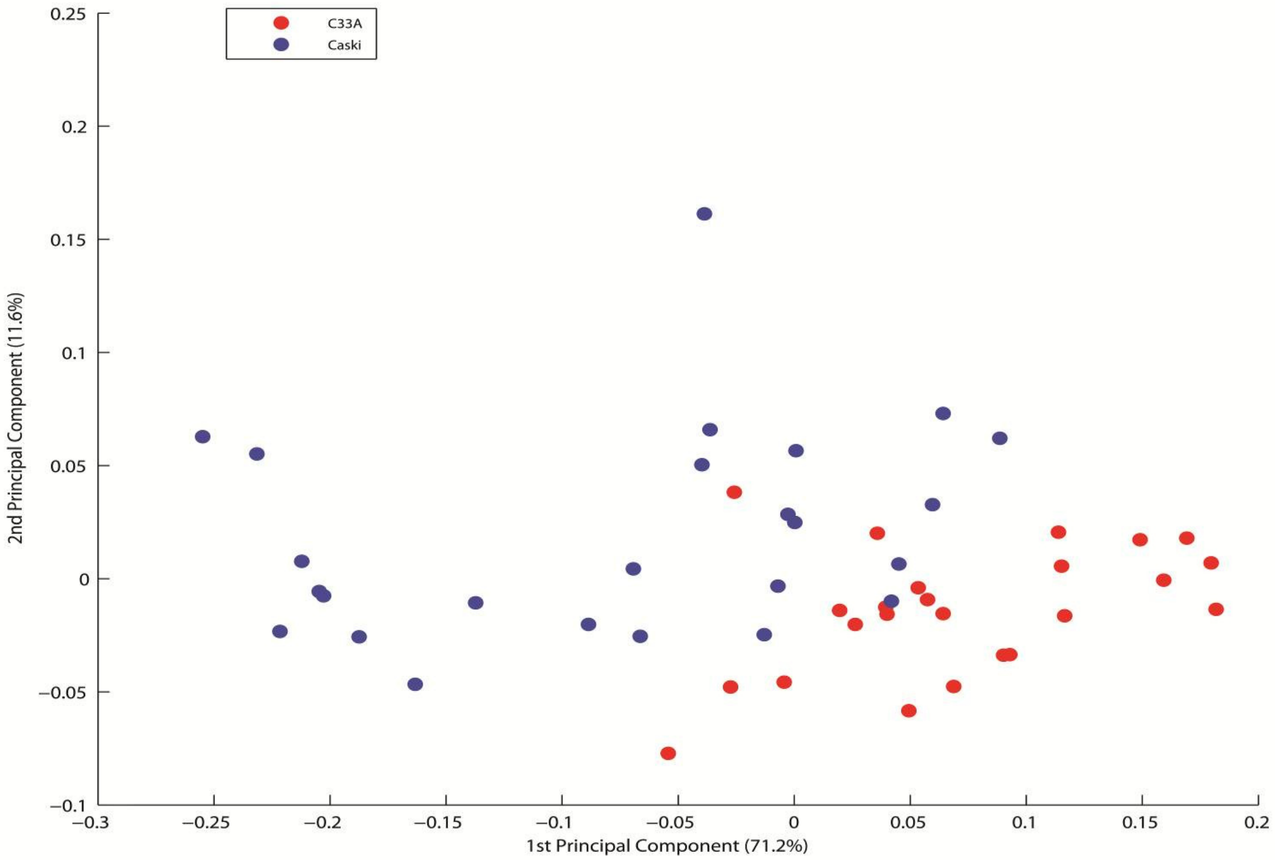


Figure 5.14 PCA scatter plot of C33A vs CaSki cells: C33 A (HPV negative) and CaSki (60-600 copies of HPV 16).

However, although it can be concluded that Raman spectroscopy certainly has potential for detecting the biochemical changes associated with the HPV infection *in vitro*, for tissue *in vivo* or *ex vivo*, further controls to establish signatures of infection are required.

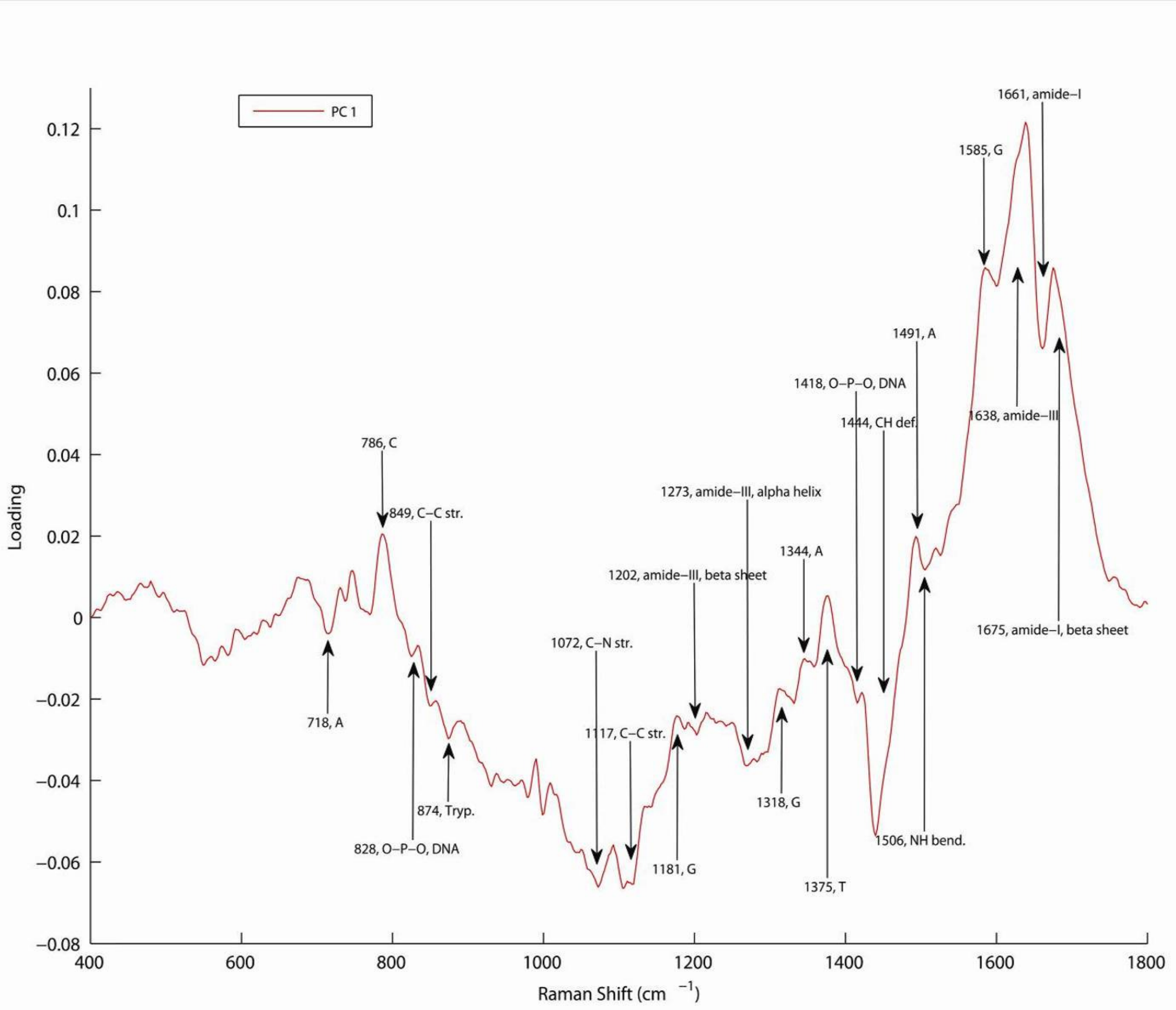


Figure 5.15 PCA loadings of C33A vs CaSki cells; C33 A is HPV negative and CaSki having 60-600 copies of HPV 16.

5.6 Results: Diseased basal and superficial layers

5.6.1 PCA of basal-true normal vs basal-2b-i

PCA analysis of basal-true normal versus basal-2b-i is performed to explore biochemical differences between these layers, as basal-2b-i is expected to be closer in biochemical composition to basal-true normal as compared to basal-2b-ii.

PCA results of the basal-true normal vs basal 2b-i, are presented in **Figure 5.16** and **Figure 5.17** as scatter plot and loadings of the PCA, respectively. It can be seen in the PCA scatter plot that the two groups are different from each other and the Raman spectra of the basal-true normal are clustered in the negative axis and spectra of the basal-2b-i in the positive axis. Notably, the major difference which can be identified at first glance is the presence of glycogen related bands at 482 cm^{-1}, 849 cm^{-1} and 938 cm^{-1} in the negative sense, solely contributed by the Raman spectra of the basal-true normal . This indicates that the basal-true normal layer, as expected, has more glycogen in the cells as compared to the basal-2b-i. Moreover, this also means that the cells of the basal-2b-i are more proliferative in nature as compared to the basal-true normal and hence the glycogen content is reduced due to its consumption by the cells during the process of the cell growth or proliferation. Another important loading is that at 1222 cm^{-1} (amide-III beta sheet) which is contributed by basal-2b-i and differentiates the Raman spectra of this layer from those of the true normal. It should be noted that this band has been observed throughout the PCA results in this chapter, normal versus abnormal regions, apart from the PCA of the cell lines, and remains the exclusive contribution by the diseased parts of the cervical tissue samples but not by the Raman spectra of the true normal sample. This again suggests that this signature could be associated with HPV infection in the cervical tissue samples. Other negative loadings of the PCA including 524 cm^{-1} (S-S stretching), 718 cm^{-1} (Adenine) and 1348 cm^{-1} (tryptophan ring breathing) are contributed by the basal true normal and the positive loadings including 669 cm^{-1} (thymine), 788 cm^{-1}, 1088 cm^{-1} (O-P-O of DNA) 1062 cm^{-1} (C-N stretching) correspond to the basal-2b-i. This can lead us to the conclusion that basal-2b-i does not seem normal or seems to have more dividing /DNA rich cells as compared to the cells of the basal-true normal layer (true normal case).

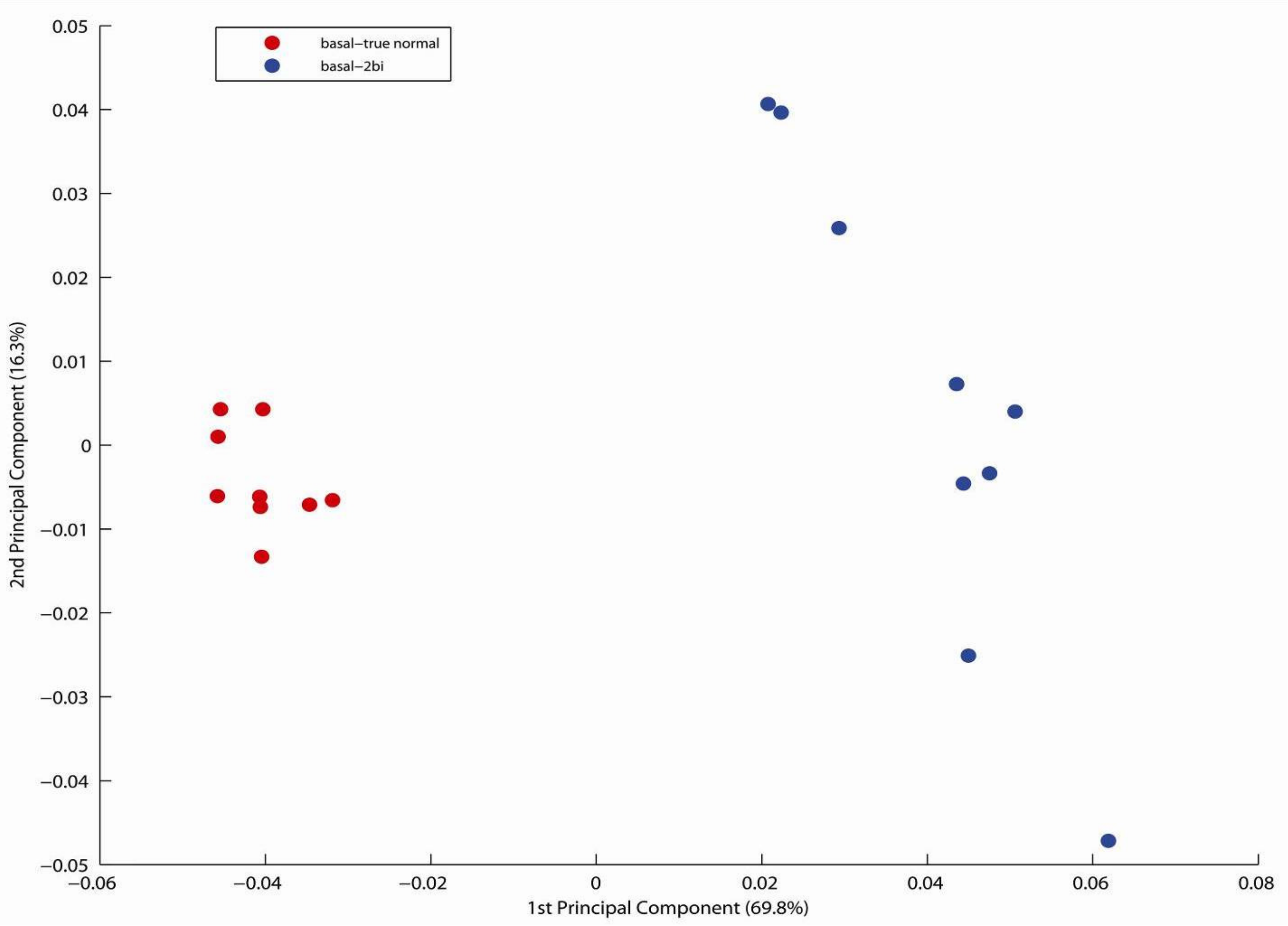


Figure 5.16 Scatter plot of PCA for basal-true normal vs basal-2bi.

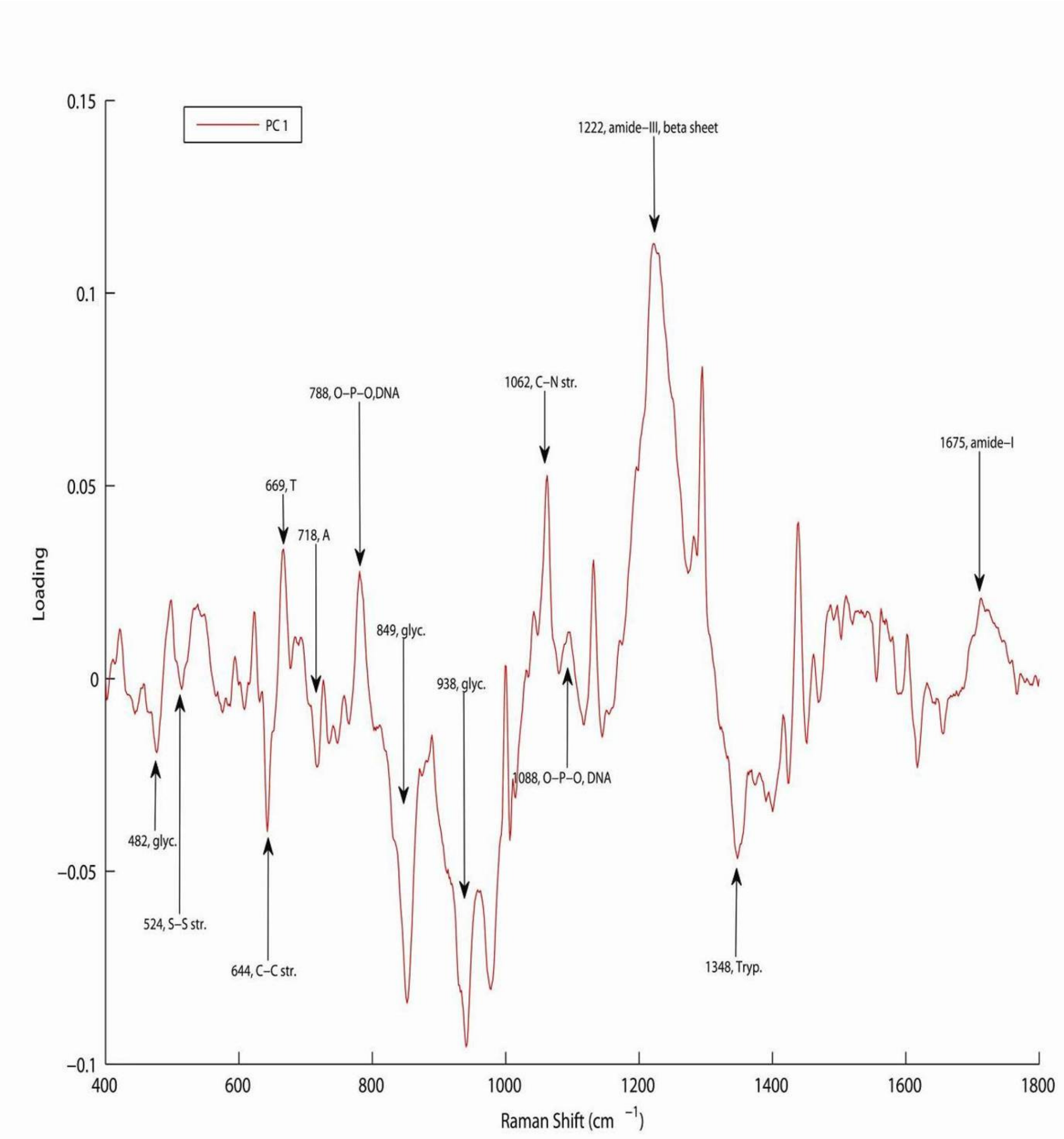


Figure 5.17 Loadings of PCA for basal-true normal vs basal-2bi.

5.6.2 PCA for basal-2a versus basal-2b

Figure 5.18 and **Figure 5.19** represent the PCA scatter plot and loadings of the PCA for basal-2a vs basal-2b, respectively. The basal-2a and basal-2b, are the less diseased part and the more diseased part of the cervical tissue section presented in the stained tissue section, **Figure 5.9**. On the basis of the fact that DNA related Raman features are dominantly observed in the more diseased part as compared to the less diseased/normal part of the tissue, the basal-2b region is expected to give more DNA related features as compared to the basal-2a.

A good separation of the two layers was observed, as indicated by two well distinguished clusters of the spectra in the PCA scatter plot. In the PCA scatter plot, the cluster of the Raman spectra of the basal-2a layer is observed on the negative side of PC1 and Raman spectra of the basal-2b layer in the positive.

It is very interesting and important to mention that the loadings of the PCA give very useful further information here regarding the state of the disease and are consistent with the different degree of disease of the different regions of the cervical tissue section. All the loadings related to the characteristic Raman features of DNA bases, including those at 669 cm^{-1} (T), 718 cm^{-1} (A), 782 cm^{-1} (C), 826 cm^{-1} (O-P-O, DNA) and 888 cm^{-1} (deoxyribose ring breathing) are present on the positive side and hence are associated with the basal-2b part (more diseased) of this cervical tissue section. The negative loadings include 482 cm^{-1} (glycogen), 1158 cm^{-1} (C-N stretching), 1172 cm^{-1} (tyrosine), 1449 cm^{-1} (CH deformation) and 1675 cm^{-1}(amide-I) and are associated with the basal-2a, which is the less diseased side of the basal-2 layer. Thus, although they are grouped by KMCA, basal-2a resembles more a normal tissue state, whereas basal-2b is more characteristic of the diseased state.

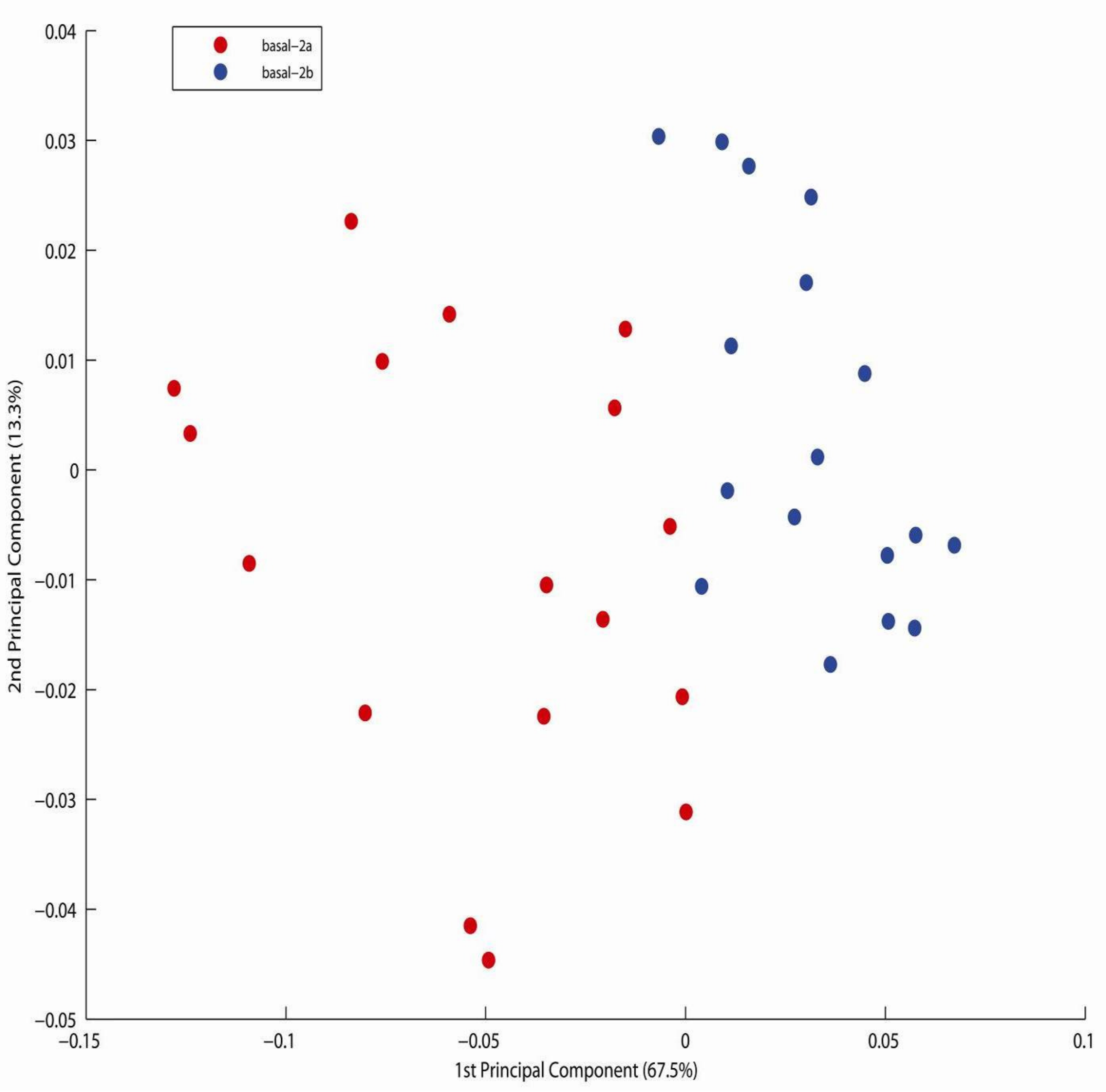


Figure 5.18 Scatter plot of PCA for basal-2a vs basal-2b.

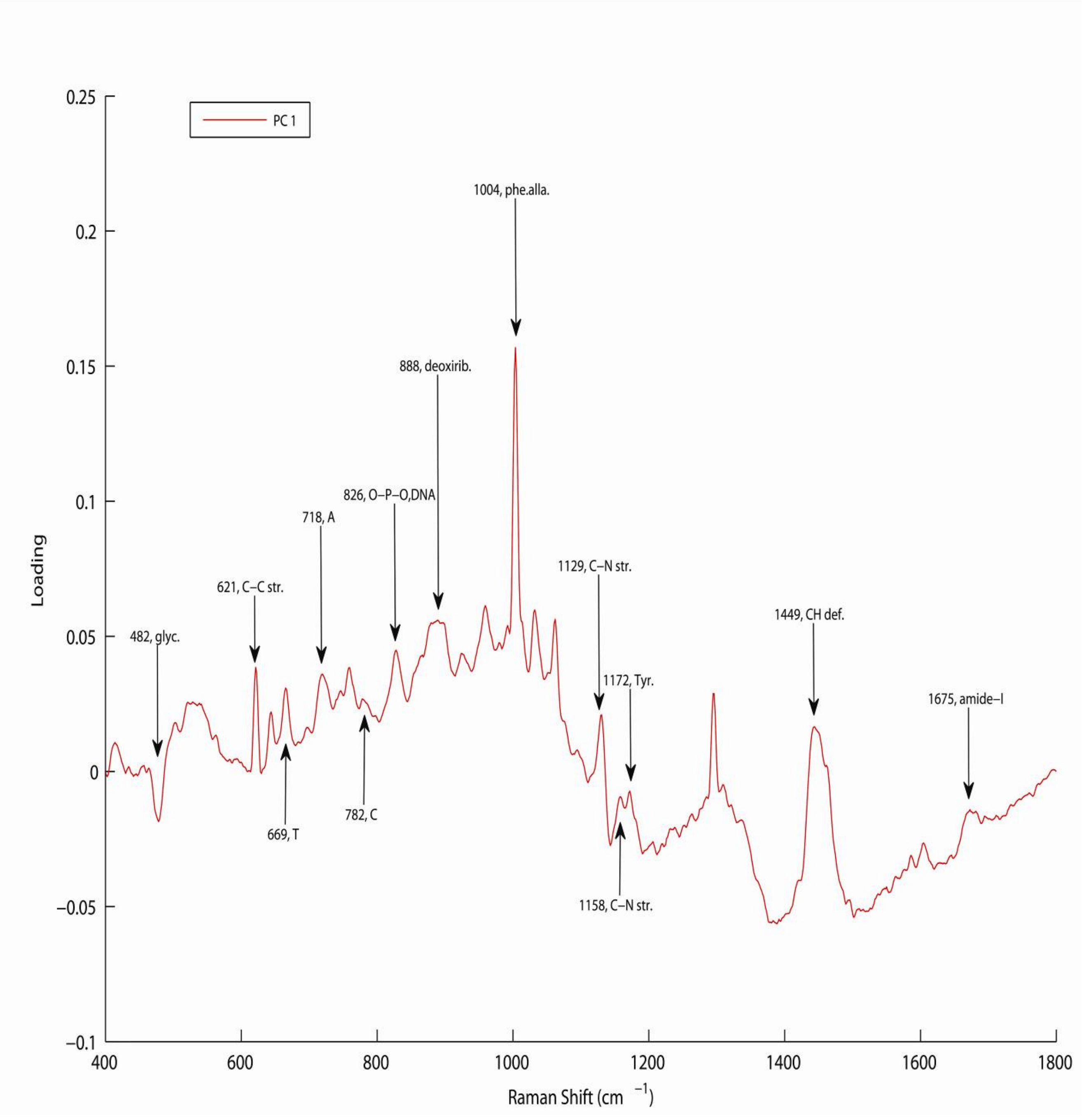


Figure 5.19 Loadings of PCA for basal-2a vs basal-2b.

5.6.3 PCA for basal-2bi vs basal-2bii

PCA results of the basal 2b-i and basal 2b-ii, two subdivisions of the basal 2b, intermediate IV, which is the more diseased part of an intermediate sample, are presented here in **Figure 5.20** and **Figure 5.21** as scatter plot and loadings of the PCA, respectively. The purpose of this analysis was to establish whether these two groups are similar or different in their biochemical composition affected by the progression of the disease, and on what basis the KMCA of the Raman data clustered them.

It is clear from the PCA scatter plot that the two groups are not very different from each other as, although they are partially differentiated, they overlap to some extent. The Raman spectra of basal-2bi are largely clustered in the negative axis and spectra of the basal-2bii in the positive axis. This means that the loadings of the PCA which are negative, including 669 cm^{-1} (thymine), 782 cm^{-1} (cytosine) and 1062 cm^{-1} (O-P-O of DNA), correspond to the basal-2b-i and the positive ones including 718 cm^{-1}, 1238 cm^{-1} (adenine), 1260 cm^{-1} (thymine) and 1278 cm^{-1} (cytosine), correspond to the basal-2b-ii region.

Notably, the results of the KMCA of the intermediate IV sample, presented in **chapter 4**, **Figure 4.17C** & **4.18** as a KMCA pseudo colour map and representative spectra respectively, showed the presence of the characteristic features of the basal layer in the superficial layer, which raised the question of whether these features are developed in the superficial layer itself or are due to the possibility of the invasion of the basal cells into this layer. It should be noted that the loadings of the PCA between basal-2bi and basal-2bii, **Figure 5.21**, are very different than those of the basal versus superficial (**Figure 5.5**) as there are no representations of glycogen in the former as compared to the latter. The differentiating features are rather associated with DNA, characteristic of rapidly proliferating cells. Moreover, under normal conditions, the basal layer provides cells to the superficial layer, but those cells are under controlled growth and have the ability to store glycogen during the process of the maturation, and only matured cells having stored glycogen move into the superficial layer. This may lead to the conclusion that during the progression of the cervical cancer, rapidly proliferating basal cells, rich in DNA and lacking in glycogen, invade the superficial layer due to the progression of the disease. The KMCA clustering of the basal 2bii layer with the basal 2bi layer is a representation of this process.

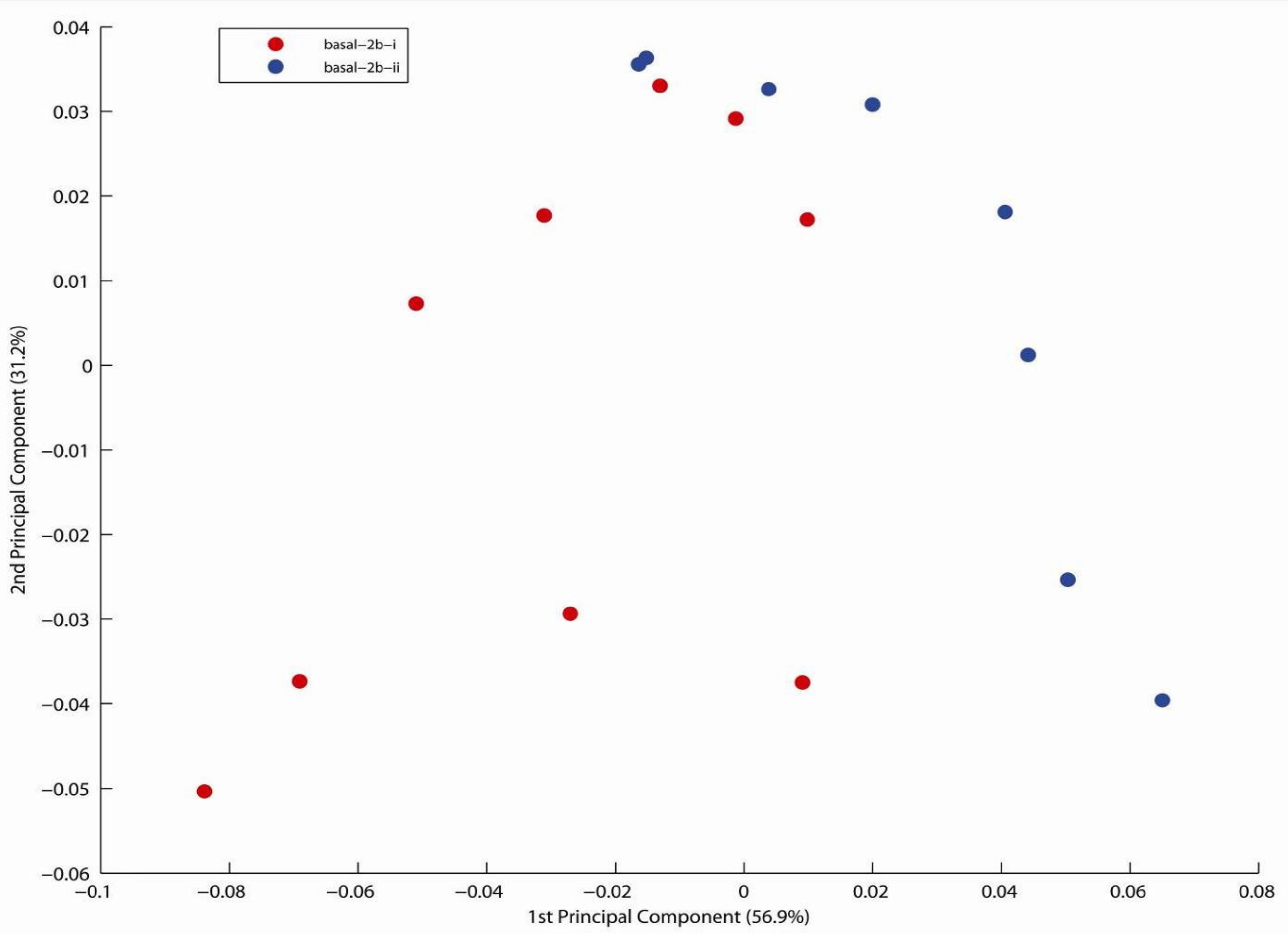


Figure 5.20 Scatter plot of PCA for basal-2bi vs basal-2bii.

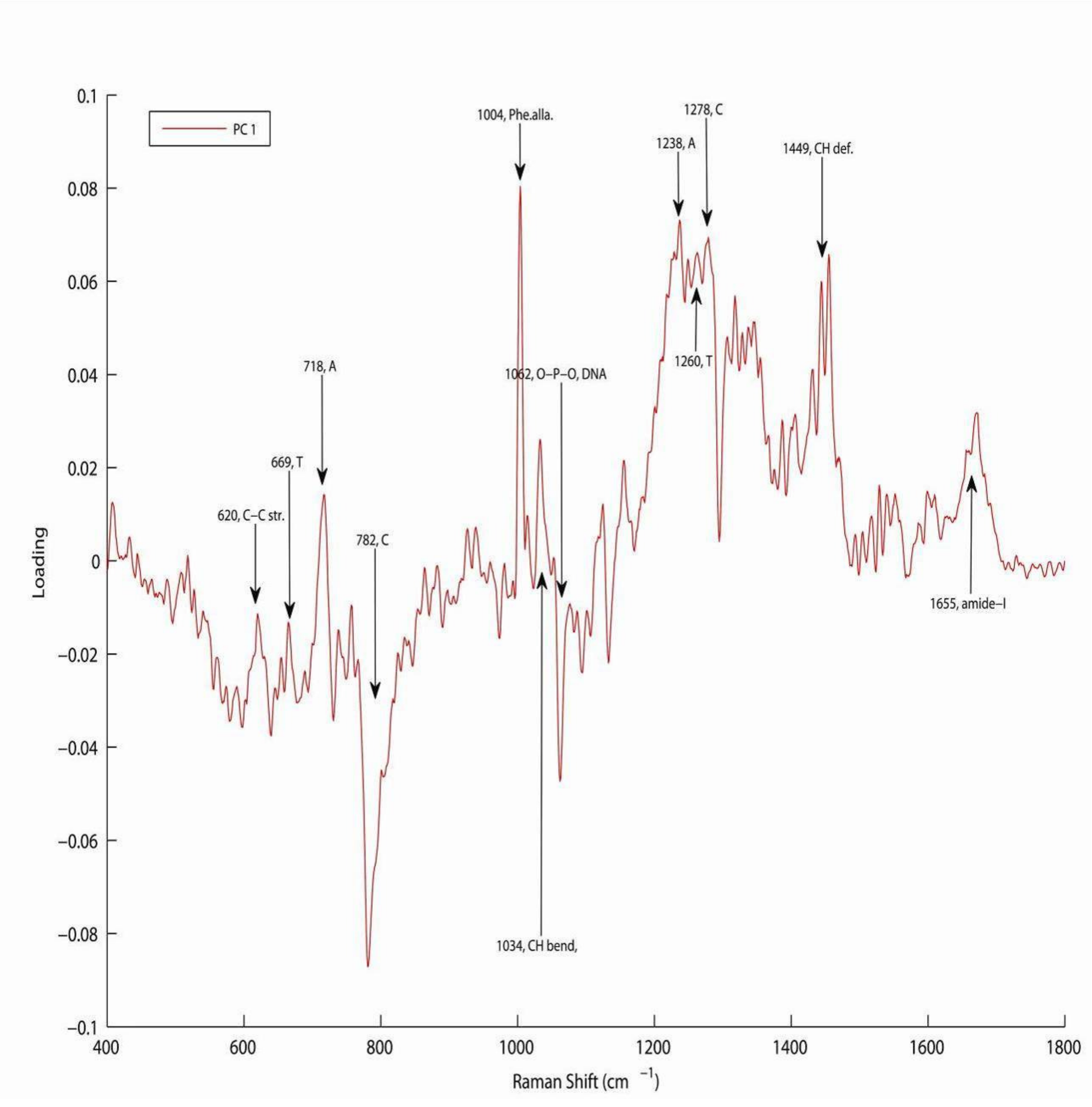


Figure 5.21 Loadings of PCA for basal-2bi vs basal-2bii.

5.6.4 PCA of basal-true normal vs basal-2b-ii

PCA analysis of basal-true normal versus basal-2b-ii was carried out to explore biochemical differences between these layers, as basal-2b-ii was expected to be different from the basal-true normal and resemble the superficial layer of the true normal sample. It should be noted that the basal-2b-ii derives from the superficial layer and has been named basal due to the classification of KMCA on the basis of the observation of the DNA related bands, in **chapter 4**.

PCA results of the basal-true normal vs basal 2b-ii are presented in **Figure 5.22** and **Figure 5.23** as scatter plot and loadings of the PCA, respectively. It can be seen in the PCA scatter plot that the two groups are different from each other and the Raman spectra of the basal-true normal are clustered in the negative axis and spectra of the basal-2b-ii in the positive axis. Once again, the major difference between the two groups of the Raman spectra appearing in the form of the loadings is the presence of the glycogen related bands at 482 cm^{-1}, 849 cm^{-1} and 938 cm^{-1}, which are solely contributed by the Raman spectra of the basal-true normal (in the negative). This confirms, as described earlier, that the basal-true normal layer of the cells is rich in glycogen as compared to the basal-2b-ii. In addition, the cells of the basal-2b-ii appear to be more proliferative in nature as compared to the basal-true normal and hence glycogen content is reduced due to its consumption by the cells during the process of the cell growth or proliferation. The loading observed at 1222 cm^{-1} (amide-III beta sheet), contributed by basal-2bii, is again the predominant difference between these two groups of spectra and is consistent with the previous PCA results, further indicating that it may be associated with HPV infection, making the tissue abnormal and differentiating it from the Raman spectra of the basal layer of the true normal sample. Other positive loadings contributed by the basal-2b-ii include those at 669 cm^{-1} , 759 cm^{-1} (thymine), 1062 cm^{-1} (C-N stretching) and by the basal true normal include 644 cm^{-1}and 1108 cm^{-1} (C-C stretching) and 1348 cm^{-1} (tryptophan ring breathing).

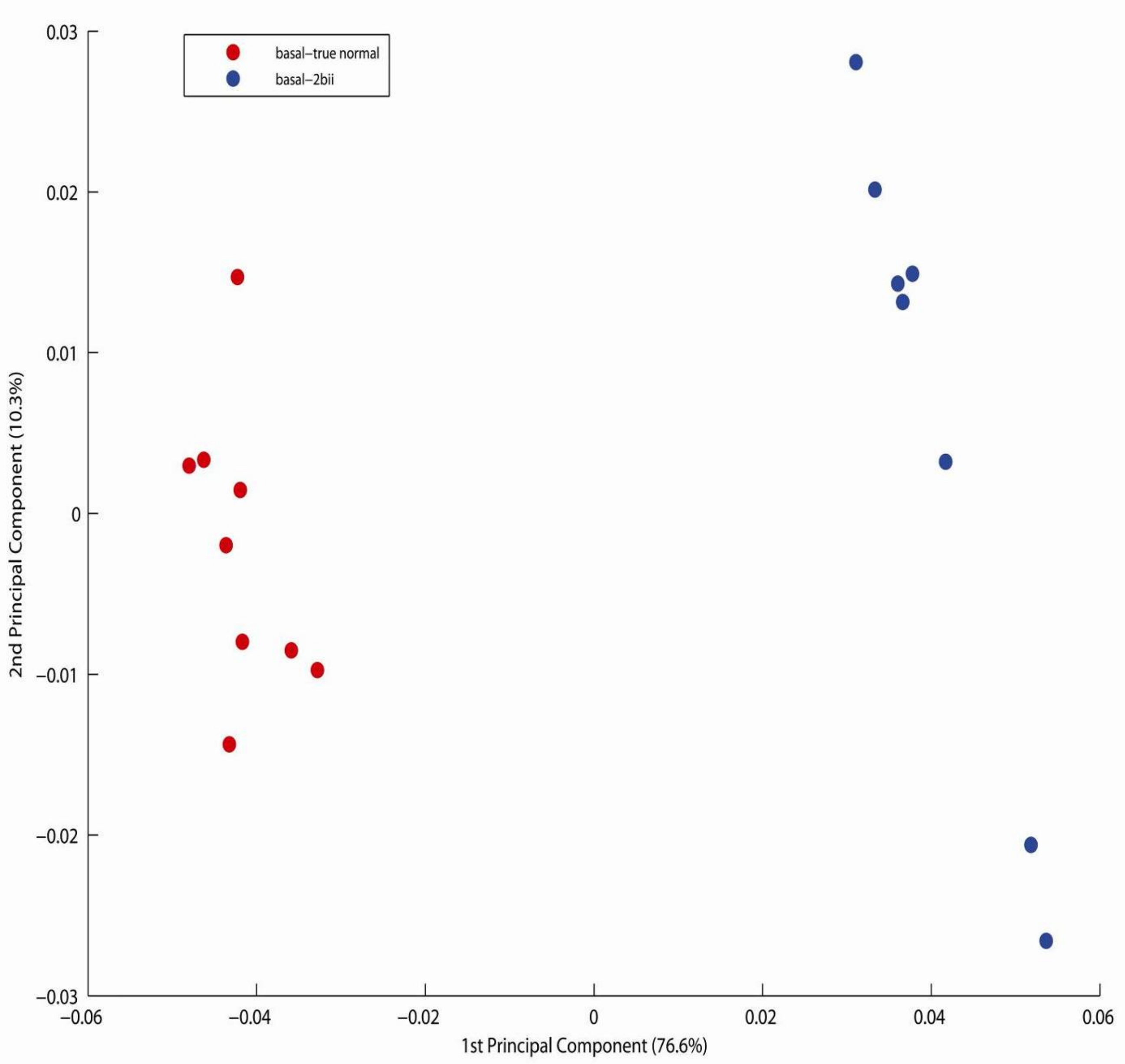


Figure 5.22 Scatter plot of PCA for basal-true normal vs basal-2b-ii.

Thus the basal-2b-ii is well differentiated from the true normal basal layer and seems to have more dividing /DNA rich cells as compared to the cells of the basal-normal layer (normal case). In addition, no positive loadings associated with glycogen are seen, which are normally contributed by the superficial layer. It can be concluded that basal-2b-ii is not similar or close to the biochemical composition of the superficial layer. To further confirm this, PCA of basal-2b-ii vs superficial is performed and presented in the following section.

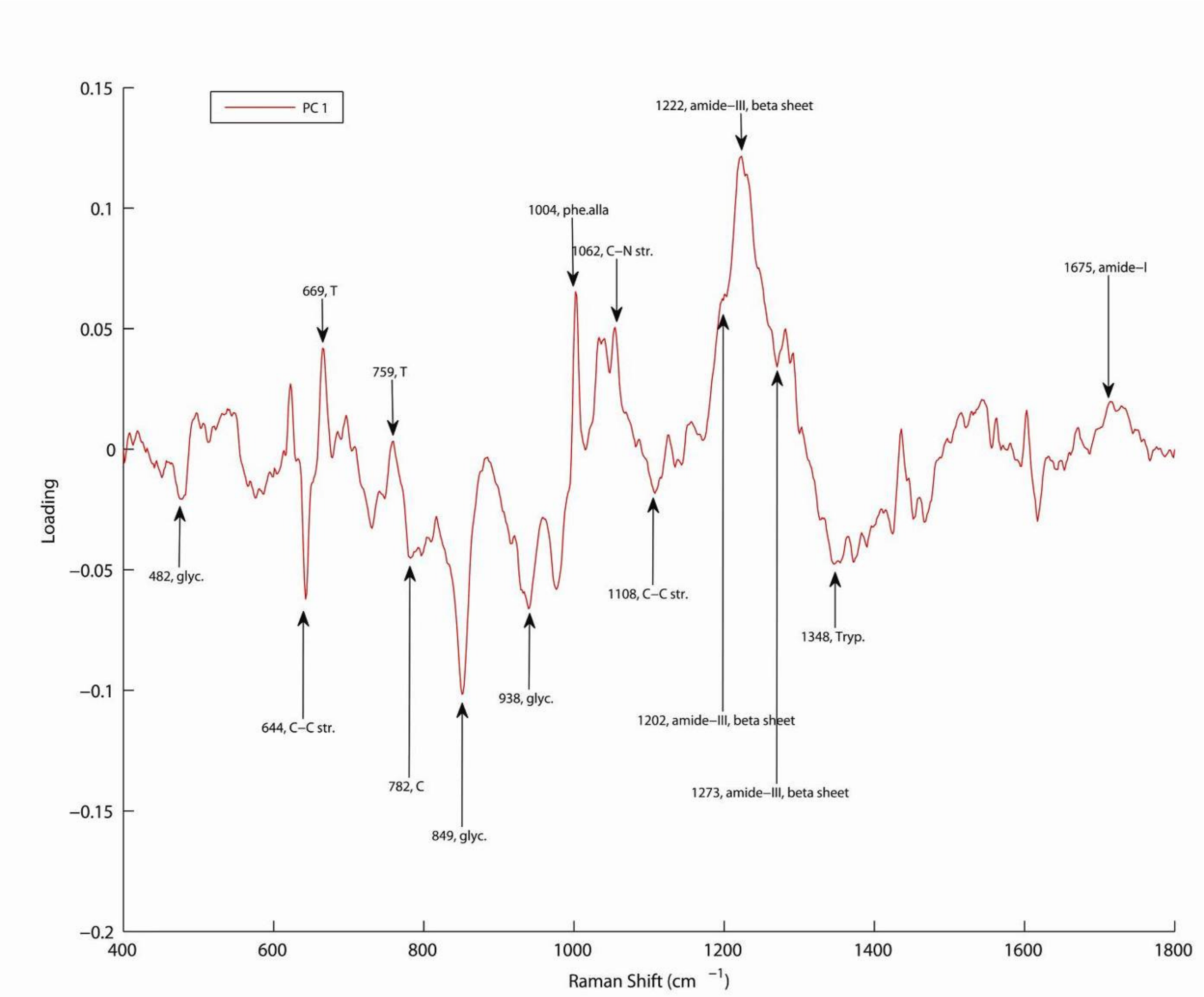


Figure 5.23 Loadings of PCA for basal-true normal vs basal-2b-ii.

5.6.5 PCA of basal-2b-ii versus superficial

To elucidate the expected similarities between the basal-2b-ii layer and the superficial layer of the true normal (superficial-true normal) cervical tissue section, PCA is performed on these groups of Raman spectra. Notably, the basal-2b-ii layer was originally superficial layer and has been named basal due to the presence of DNA bands (shown by KMCA in previous chapter).

It can be seen in the PCA scatter plot, presented in **Figure 5.24**, that the two groups are differentiated from each other and the Raman spectra of the superficial layer are clustered in the positive axis and spectra of the basal-2b-ii layer in the negative axis. This means that the loadings of the PCA, shown in **Figure 5.25**, which are positive, including those at 480 cm^{-1}, 853 cm^{-1} and 938 cm^{-1} (glycogen), 1136 cm^{-1} (C-C skeletal stretching), 1355 cm^{-1} (tryptophan ring breathing), 1386 cm^{-1} (CH_3 bending), derive from the superficial layer. The negative loadings contributed by basal-2b-ii, including those at 669 cm^{-1}, 749 cm^{-1}, 1235 cm^{-1} (thymine) 728 cm^{-1}, 1208 cm^{-1} (adenine), 783 cm^{-1} (cytosine), and 1182 cm^{-1}, 1313 cm^{-1} (guanine) are related to DNA and at 1044 cm^{-1} (CH bending), 1228 cm^{-1} (amide-III beta sheet), 1449 cm^{-1} (CH deformation) and 1675 cm^{-1} (amide-I) are attributed to the proteins.

It can be concluded that the cells of the superficial and basal-2b-ii are not similar, as shown by the scatter plot and translated into the loadings of the PCA. Moreover, the loadings contributed by the Raman spectra of the basal-2b-ii are attributed to DNA and no glycogen related loadings are contributed by the cells in this layer. These changes, including the absence of the glycogen contents in the cells of the basa-2b-ii layer as compared to the superficial-true normal, differentiate the samples and lead to the conclusion that the hypothesis made in the **chapter 4** on the basis of the KMCA results of “invasion of the basal cells into the superficial layer during the progression of the disease” is true and in fact the changes are not occurring in the superficial layer itself.

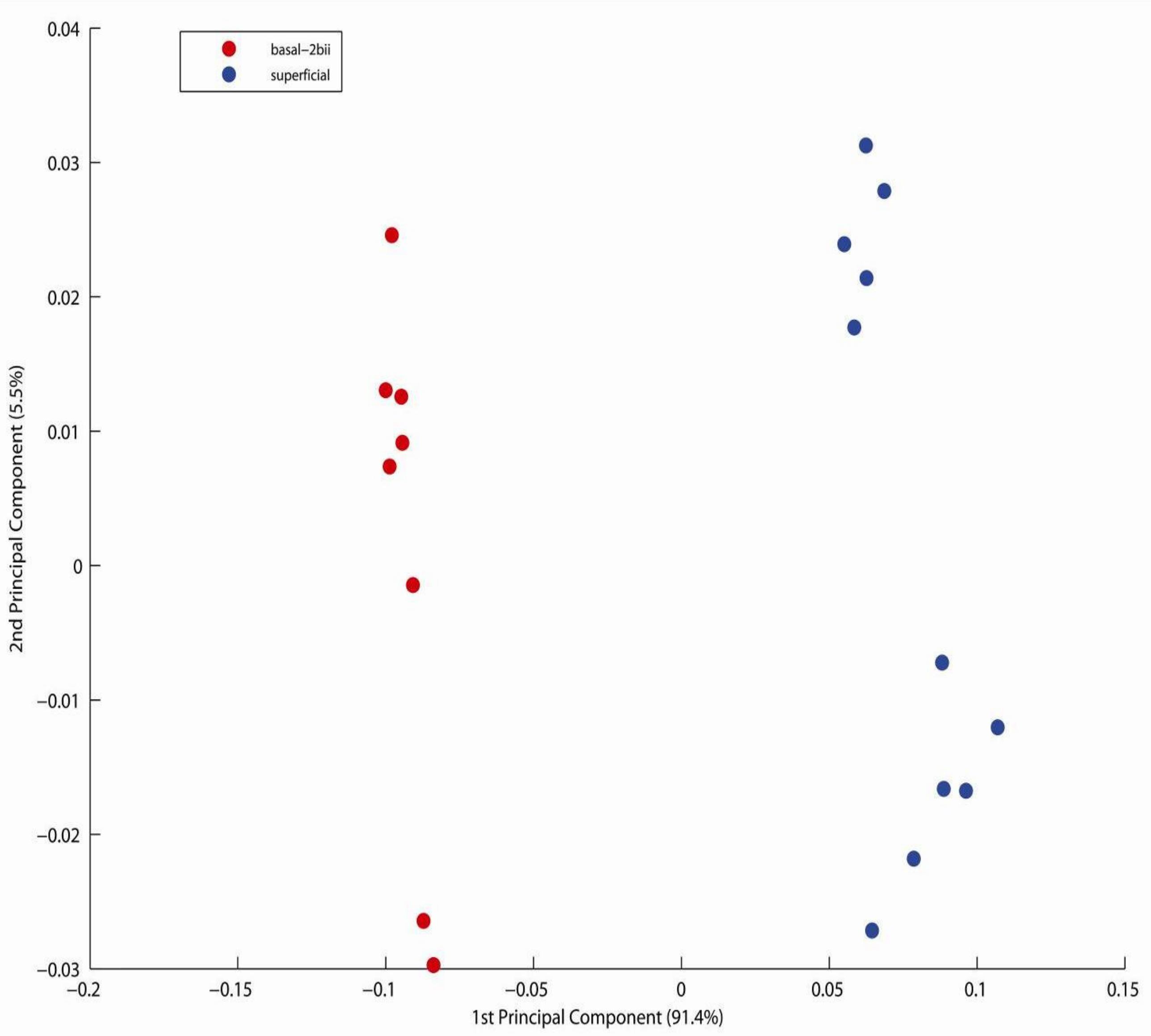


Figure 5.24 PCA scatter plot for basal-2b-ii versus superficial.

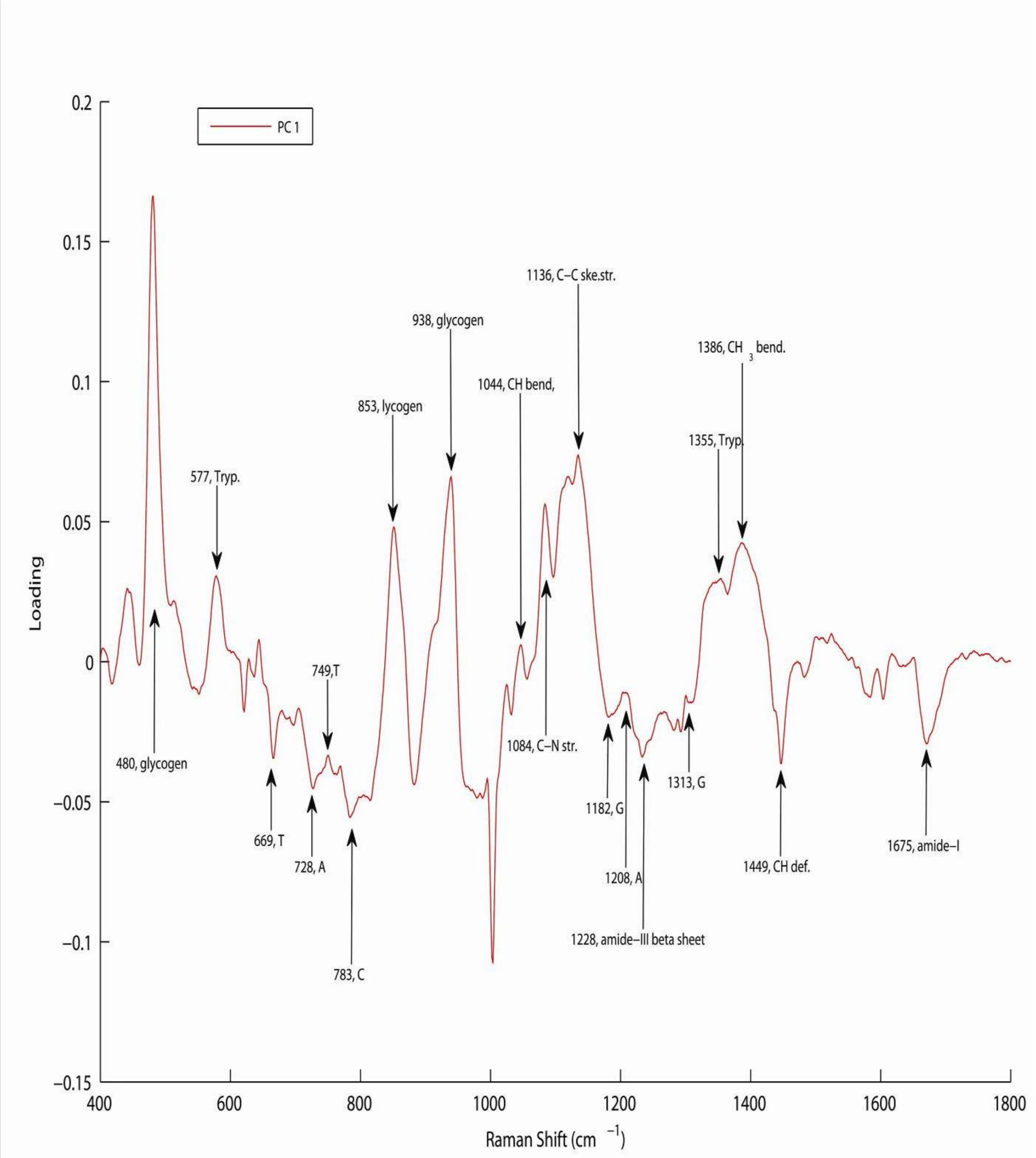


Figure 5.25 PCA loadings for basal-2b-ii versus superficial.

5.6.6 PCA of basal-true normal vs basal-3

PCA of the basal layer of the normal cervical tissue sample (basal-true normal) versus the basal layer of an intermediate sample (basal-3) is presented in **Figure 5.26** and **Figure 5.27** as scatter plot and loadings of the PCA, respectively. The PCA scatter plot shows a good separation of the clusters of the spectra and indicates that the two groups are significantly different from each other. The Raman spectra of the basal-true normal layer are clustered in the negative axis and the spectra of the basal-3 layer are clustered in the positive axis. The basal-true normal layer contributes the negative loading of the PCA at 480 cm^{-1}, 849 cm^{-1}, 939 cm^{-1} (glycogen), related to the glycogen and some other loadings including those at 644 cm^{-1} (C-C stretching), 718 cm^{-1} (adenine), 831 cm^{-1} (tyrosine), 1056 cm^{-1} and 1088 cm^{-1} (O-P-O, DNA). The positive loadings include those at 1102 cm^{-1} (C-C stretching), 1202 cm^{-1} and 1222 cm^{-1} (amide-III beta sheet), 1346 cm^{-1} (guanine) and are contributed by the basal-3 layer. These features, particularly related to amide-III beta sheets of proteins, are consistent with the results of the basal-true normal versus basal-2 and its sub-regions, discussed in section 5.7 and maybe indicative of the .HPV infection as these are consistently observed in the PCA loadings of the basal true normal versus the diseased samples but require further investigation.

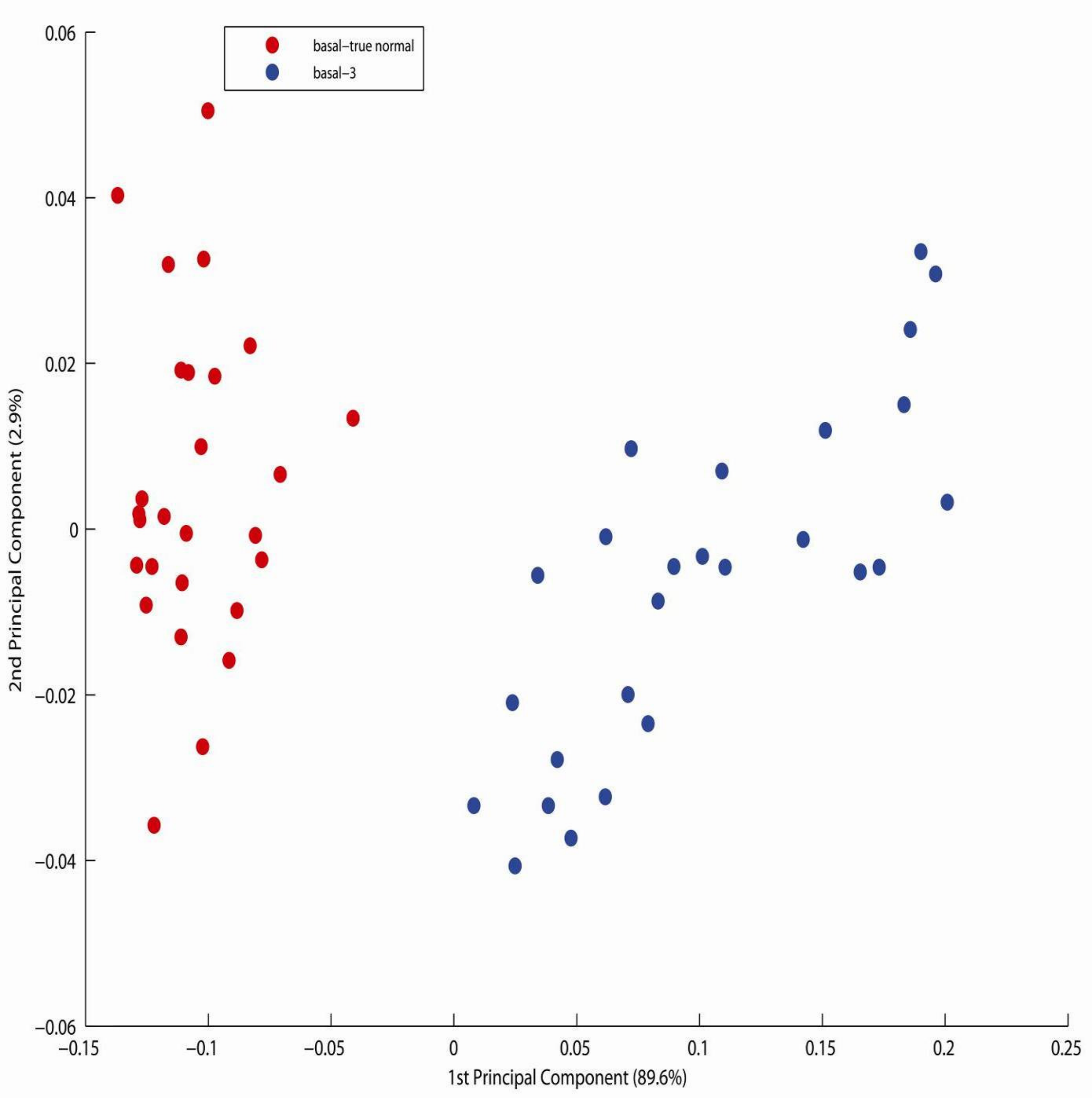


Figure 5.26 PCA Scatter plot for basal-true normal vs basal-3.

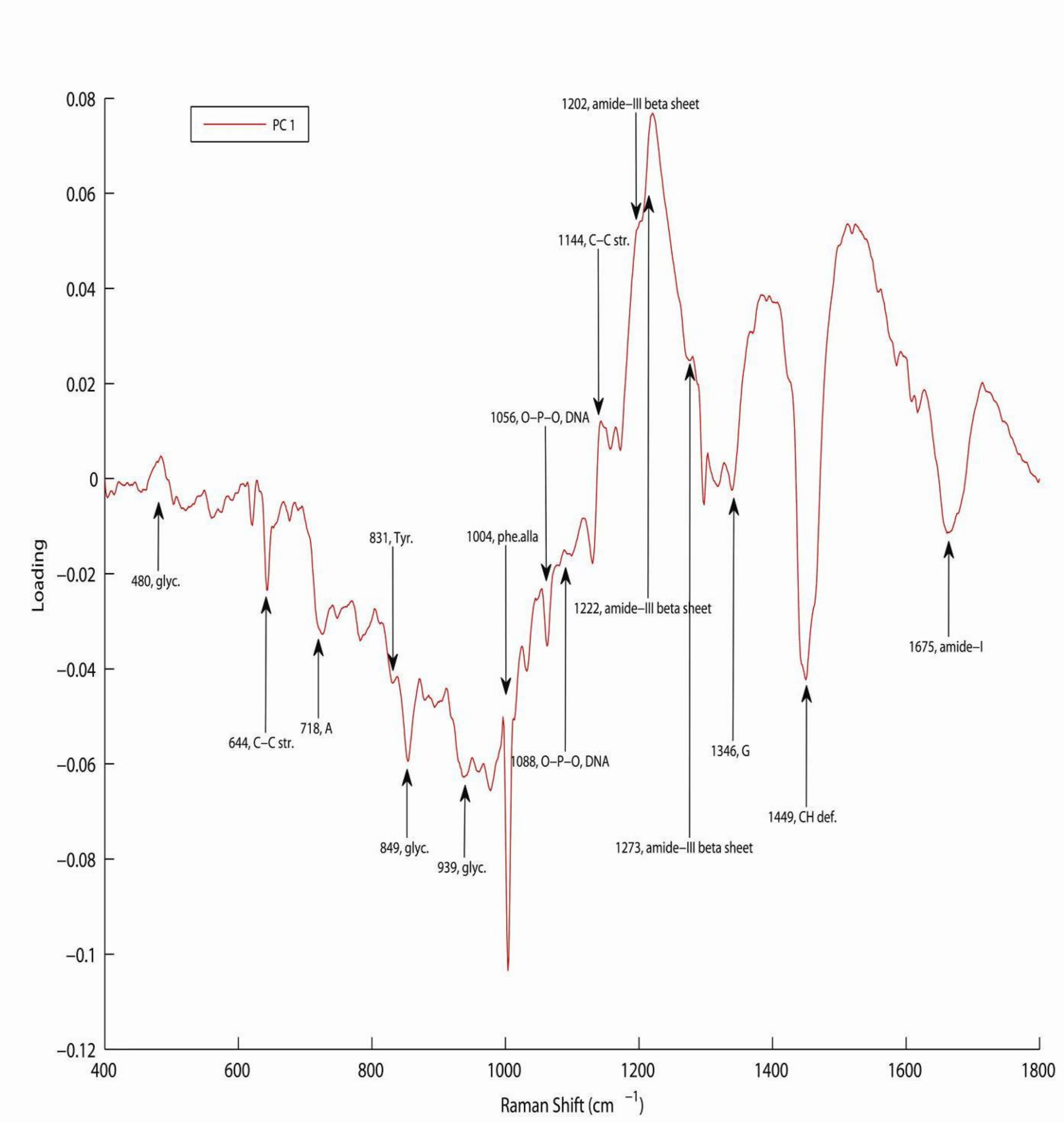


Figure 5.27 PCA loading (PC-1) for basal-true normal vs basal-3.

5.6.7 PCA results of basal-3-i vs basal-3-ii

PCA results of the two parts of the, basal layer (basal-3) of an intermediate sample III, presented in chapter 4 in **Figure 4.15,** which are called basal 3b-i and basal 3b-ii as explained in **Figure 5.28**, are presented here in **Figure 5.29** and **Figure 5.30** as scatter plot and loadings of the PCA, respectively. The purpose of this analysis was to establish whether these two groups are similar or different in their biochemical composition affected by the progression of the disease and to confirm the findings of the PCA of basal-2bi-vs-basal-2bii (**Figure 5.19**) where it was concluded that during the progression of the cervical cancer, rapidly proliferating basal cells, rich in DNA and lacking in glycogen, invade the superficial layer due to the progression of the disease.

The basal-3b-i layer was expected to be similar in biochemical composition to the basal-true normal layer (normal case) while the basal-3b-ii layer was expected to be different from the basal-true normal layer and to resemble the superficial layer of the true normal sample. It should be noted that basal-3b-ii was originally superficial layer and has been called basal due to the observation of the DNA bands (shown by KMCA in previous chapter) which can be associated with the progression of the cervical cancer into the superficial layer from the basal layer of the intermediate tissue section.

The PCA scatter plot shows that the two groups are different from each other but some overlap is observed. The Raman spectra of the basal-3b-i layer are clustered in the positive axis and spectra of the basal-3b-ii layer in the negative axis. This indicates that the positive loadings of the PCA, including 669 cm^{-1}, 755 cm^{-1} (thymine), 577 cm^{-1}, 783 cm^{-1} (cytosine) correspond to the basal-3b-i layer and the negative loadings, including those at 1248 cm^{-1}, 1340 cm^{-1} (adenine), 1375 cm^{-1} (thymine) and 1092 cm^{-1} (O-P-O of DNA) are contributed by the basal-3-ii layer. Notably, the 480 cm^{-1} (glycogen) loading is negative, contributed by basal-3ii but other glycogen related loadings including 849 cm^{-1}, 938 cm^{-1} are contributed by the basal-3i and not by the basal-3ii. This indicates that there is depletion of the glycogen in the basal-3ii, although not completely depleted as indicated by observation of only one feature of the glycogen at 480 cm^{-1} only, and increase in the DNA related features. These

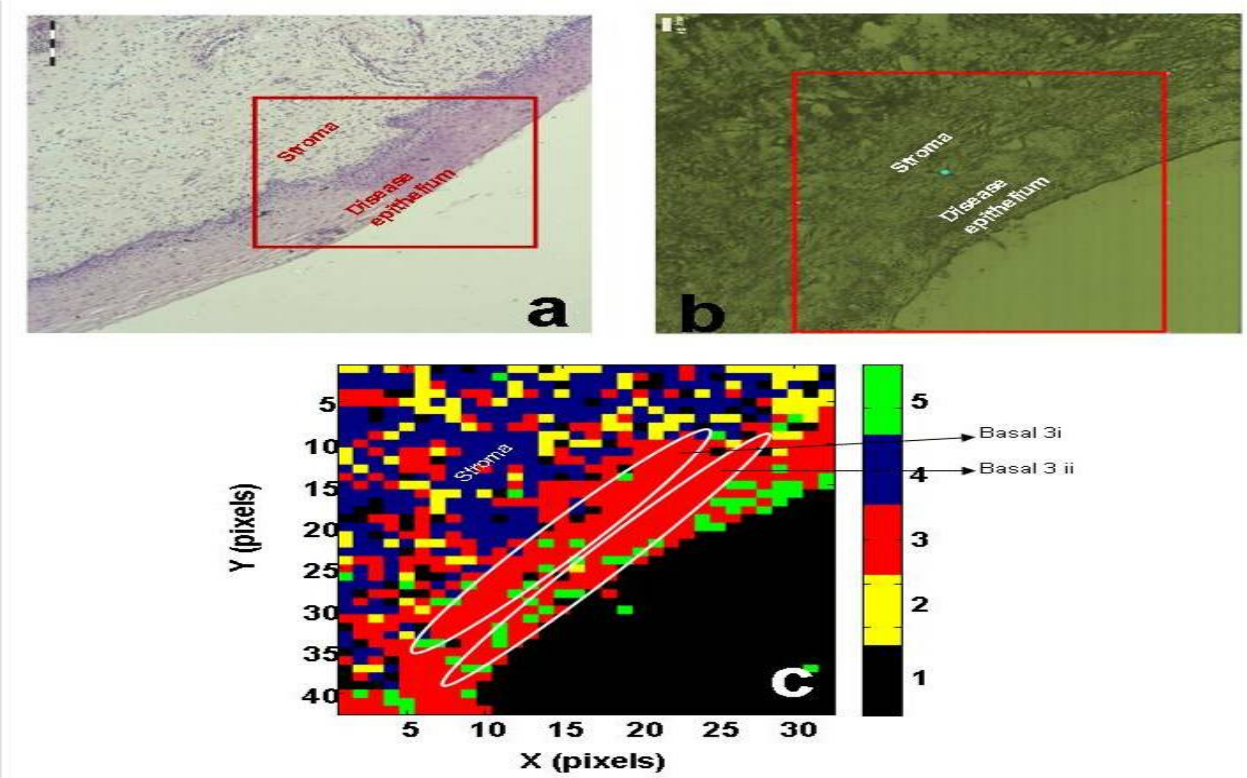


Figure 5.28 H&E stained disease side of CIN II marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope, 40µm (b) Five cluster KMCA map generated from the FTIR map (c).

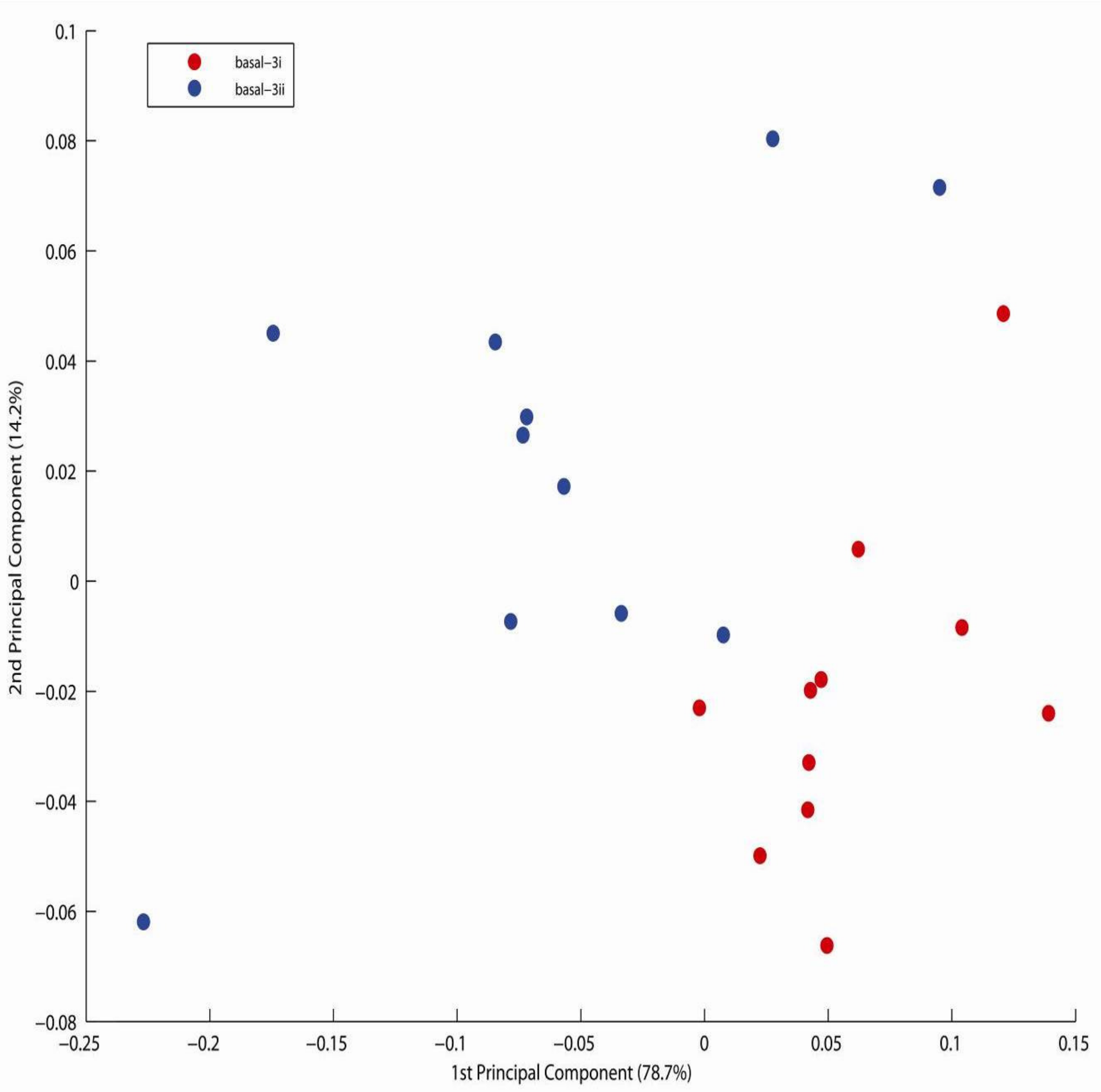


Figure 5.29 Scatter plot of PCA for basal-3i vs basal-3ii.

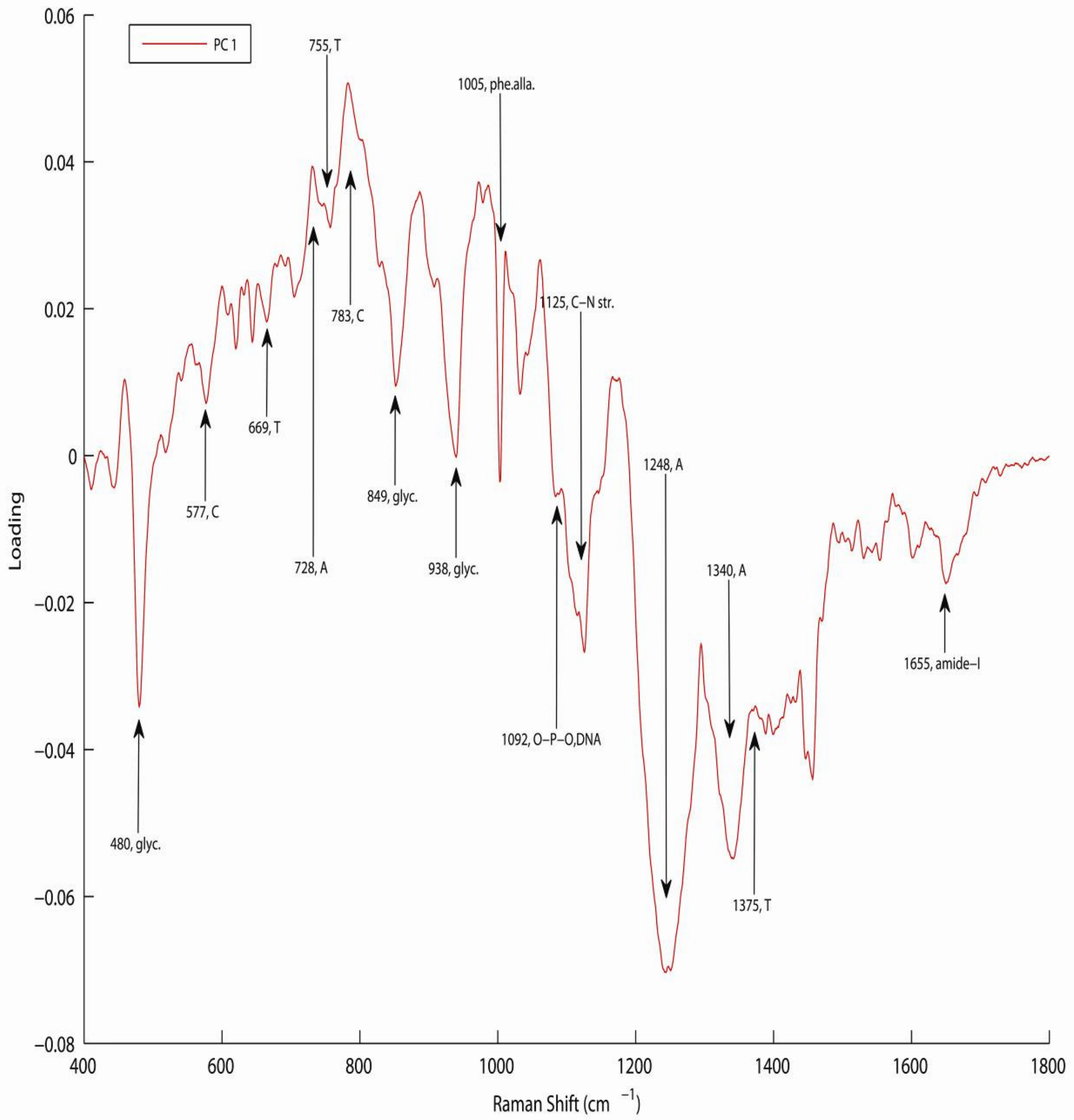


Figure 5.30 Loadings of PCA for basal-3i vs basal-3ii.

5.6.8 PCA results of basal-true normal vs basal-carcinoma

Figure 5.32 and **Figure 5.33** represent the scatter plot and loadings of the PCA respectively, for the Raman spectra from the basal layer of the normal case (basal-true normal) against the basal layer of an abnormal sample called basal-carcinoma (carcinoma *in situ*) presented in **Figure 5.31** has been rproduced here from chapter 4 (**Figure 4.9**). The differentiation of the two groups can be seen as the clustering of the Raman spectra of the basal-true normal in the negative axis and those of the basal-carcinoma in the positive axis of the scatter plot.

The loadings of the PCA contributed by the basal-true normal include 482 cm^{-1}, 849 cm^{-1} and 938 cm^{-1} (glycogen), 718 cm^{-1} (Adenine) and 644 cm^{-1}, 1144 cm^{-1} (C-C stretching). The basal-carcinoma contributes the loadings including 669 (thymine), 782 (cytosine), 781 (cytosine) and 1062 (C-N stretching) and 1116 (C-C stretching). This shows that there are more DNA related loadings contributed by the abnormal case as compared to the true normal case which is due to the presence of more dividing cells and in turn more DNA content in the abnormal sample. The features associated with the amide-III including 1222 is very much reduced in their intensity. Notably, this is the most diseased sample (abnormal most) and similar reduction in the intensity of these bands is observed in the PCA of the basal-true normal against the basal-7 sample as presented in the respective loadings in the **Appendix** (**Figure A1.12**). The reduction of the amide-III related features may indicate the integration of the HPV virus into the host genome instead of its present with the protein coat independently in the host cell.

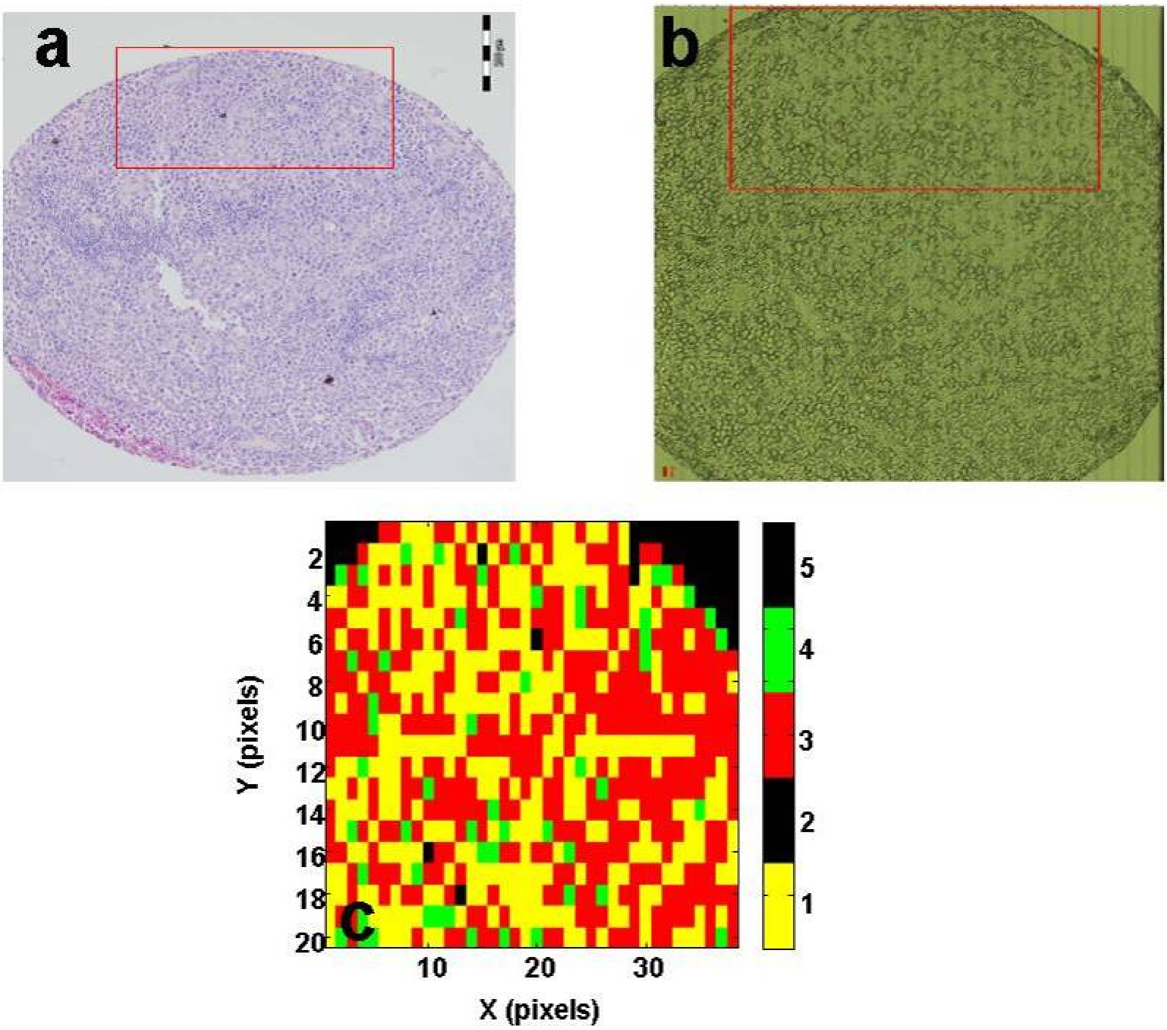


Figure 5.31 H&E stained carcinoma *in situ* marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster KMCA map generated from the Raman map (c).

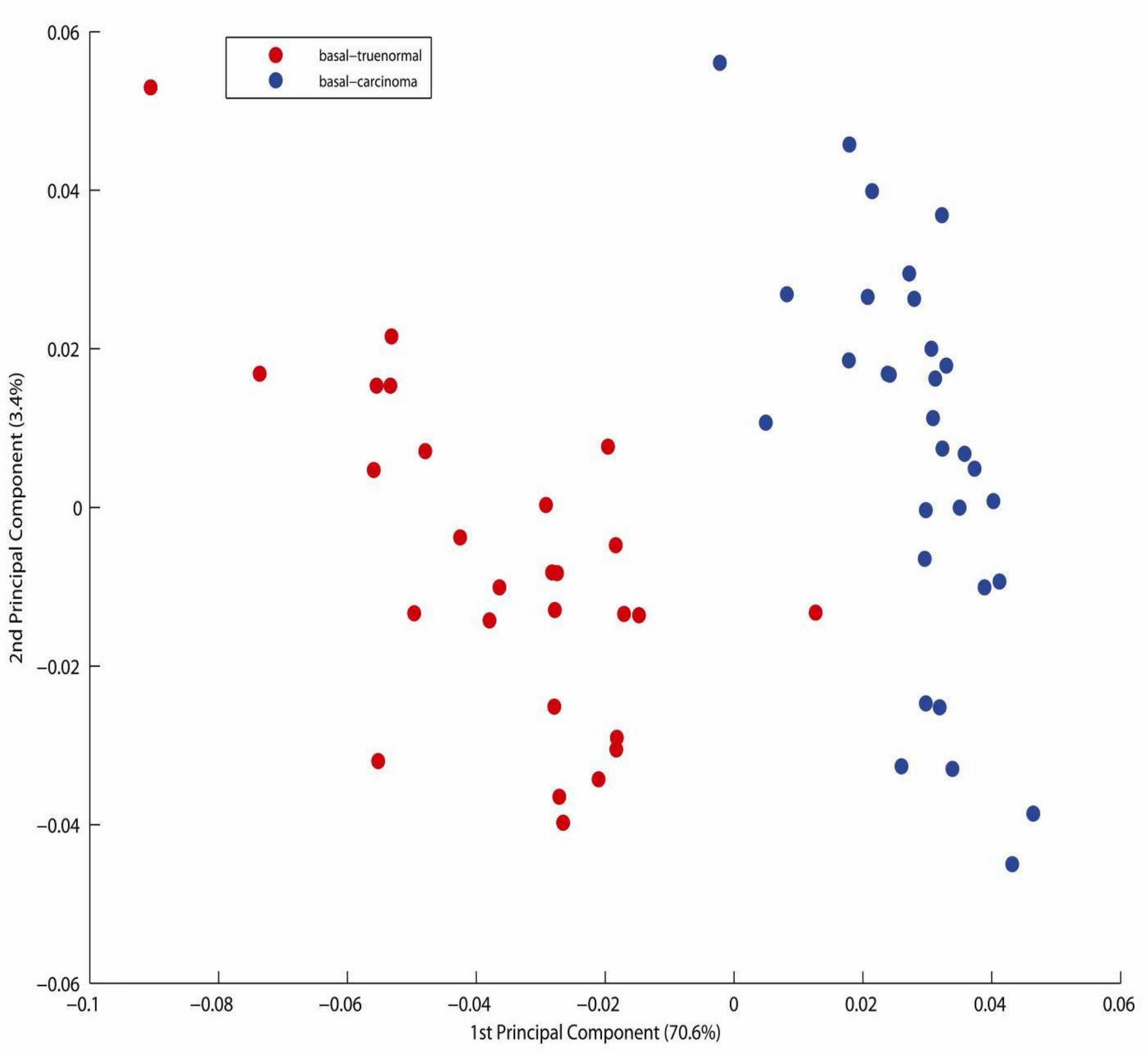


Figure 5.32 PCA scatter plot for basal-true normal versus basal-carcinoma.

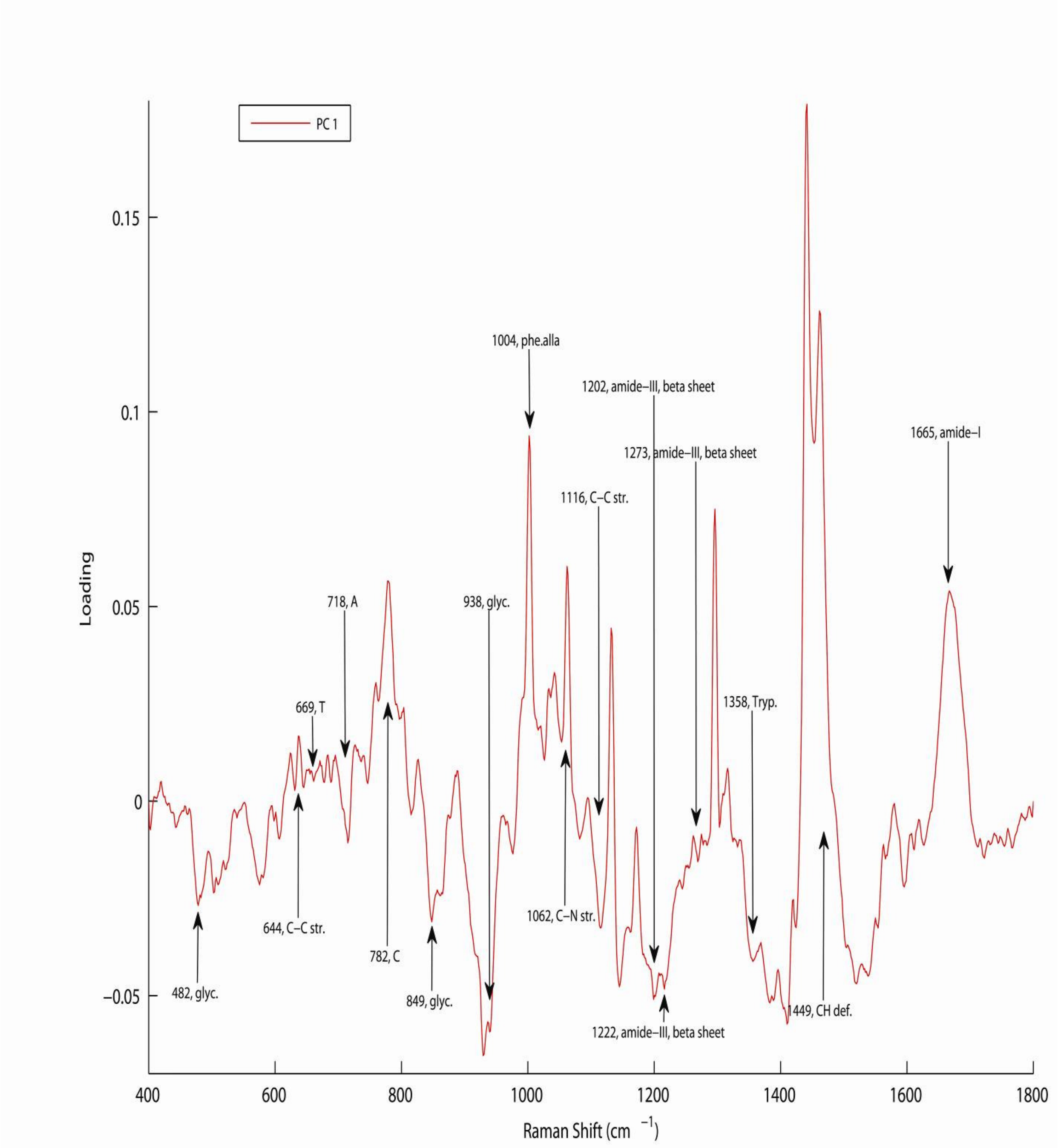


Figure 5.33 PCA loading (PC-1) for basal-true normal versus basal-carcinoma.

5.6.9 Summary

The major goals of the above described analyses were to evaluate the ability of Raman micro spectroscopy to determine the biochemical changes occurring in the different parts of the tissue structure, to establish the signatures of HPV infection in the cervical tissue samples by comparing them with the cell line data, and to elucidate the biochemical changes which lead to the results of the KMCA which indicate invasion of the basal layer/characteristics of the basal layer into the superficial layer.

The PCA results presented so far in this chapter, including comparison of different basal layers of the true normal, CIN II normal, sub-regions of the basal-2 and basal-3 layers along with the cell lines representative of cervical cancer, have led to different conclusions. The comparison of the PCA loadings of the cell lines, **Figure 5.15** with the PCA loadings of the true normal basal versus CIN II normal basal, **Figure 5.8**, have indicated that there are no strongly matching Raman signatures which could lead us to identify the signatures of HPV infection in the cervical cancer tissue samples. The loading at 1222 cm^{-1} (amide-III beta sheet) in the basal-true normal versus basal-CIN-II normal may be of interest as it is the only major difference between the two data sets reflected in the form of loading of the differentiating PC1, but it is absent in the loadings of the cell lines. The amide-III feature (1222 cm^{-1}) is consistently present in the loadings of the PCA contributed by the basal layers of the intermediate and abnormal samples in the PCA with the basal-true normal. Thus this feature may be considered as an early marker of disease progression and could potentially indicate HPV infection. Notably, this loading is much reduced in intensity in the carcinoma *in situ* sample shown in **Figure 5.33** (**Figure 4.8** in chapter 4 called abnormal

III) and most extreme abnormal sample (**Figure 4.7** in chapter 4 called abnormal II) presented in the **Appendix**, **Figure A1.XII**. As, these are the most diseased samples and the reduction of the intensity of amide-III related feature may indicate that at this stage of the disease, integration of the HPV virus into the host genome has taken place instead of its presence with the protein coat independently in the host cell.

PCA of the basal-true normal against the basal layers of the intermediate and abnormal samples (basal 4-7) has also been carried out to differentiate the biochemical changes which take place due to the progression of the disease from the normal case to the intermediate and abnormal stages and the results are presented in the **Appendix**.

The following conclusions can be made from the PCA results:

a) The basal layer of the CIN II normal is different to the basal layer of the true normal.
b) The CIN II normal basal region, far from the diseased area, is similar in biochemical content to the basal region close to the diseased area.
c) The basal region in the diseased section is already low in glycogen and rapidly proliferating.
d) The changes in the superficial layer are due to rapidly proliferating cells invading into this layer.
e) The amide-III feature (1222 cm^{-1}) which is consistently observed in the loadings of the PCA contributed by the basal layers of the intermediate and abnormal samples

in the PCA with the basal-true normal may be considered as an early marker of disease progression and may be indicative of the HPV infection.

5.7 Results: Diseased basal and stroma layers

5.7.1 PCA of stroma-true normal versus sub-regions of the stroma-3

One of the indications from **chapter 4** was that cervical cancer progression from the basal layer into the stromal layer is shown by the results of the KMCA, based on the observation of more DNA bands/basal like characteristics in the stromal layer of the intermediate or abnormal cervical tissue samples as compared to that of the normal sample. The question is therefore raised as to whether this progression takes place due to the invasion of the basal cells into the stroma or whether there are some biochemical changes which have taken place in the cells of the stroma layer itself, due to the changes in the biochemistry of the basal layer. For this purpose, the stromal layer (stroma-3) of the sample called intermediate III (**Figure 4.15**) in **chapter 4**, has been compared with the stroma-true normal, **Figure 4.1** in **chapter 4** by using PCA. Moreover, stroma-3 has been divided into two sub-regions, stroma-3i and stroma-3ii, as identified by the KMCA in **chapter 4**. These subgroups of stroma-3 are made on the basis of the DNA bands/basal like characteristics in some parts, stroma-3i, and other parts having typical characteristics of the stroma (collagen related bands), stroma-3ii, as indicated in **Figure 5.34**. The red and blue clusters in the stromal layer correspond to stroma-3i and stroma-3ii, respectively. The PCA of the Raman spectra, comparing these sub-regions, as well as stroma-true normal, is presented in the following sections.

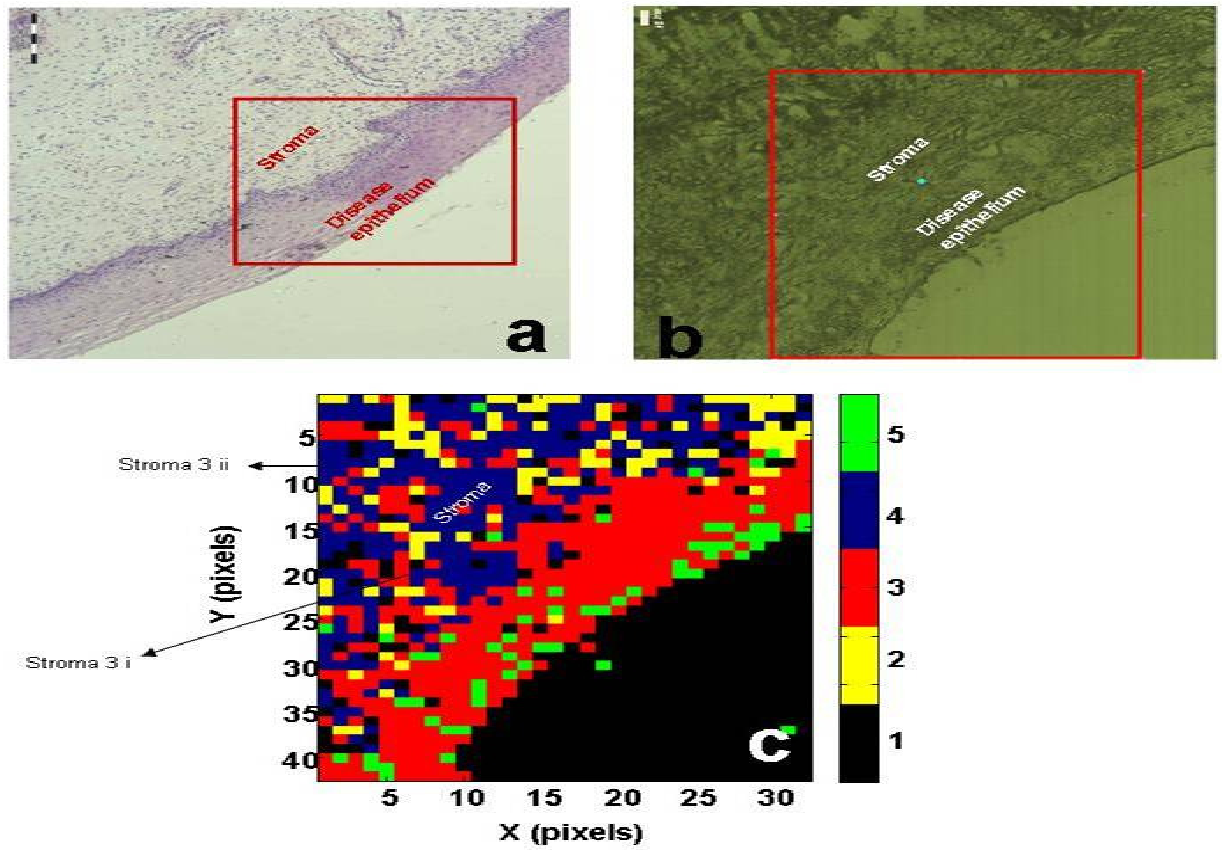


Figure 5.34 H&E stained disease side of CIN II marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope, 40µm (b) Five cluster KMCA map generated from the FTIR map (c).

5.7.2 PCA of stroma-true normal versus stroma-3i

The PCA results of the stroma-true normal versus stroma-3i are presented in **Figure 5.35** and **Figure 5.36** as PCA scatter plot and loadings, respectively. The differentiation of the two groups can be seen, although not very distinctively, as the clustering of the Raman spectra of the stroma-true normal in a negative sense with respect to PC1 and the stroma-3i in the positive sense. This means that the loadings appearing in the negative are contributed by the stroma-true normal and those in the positive are contributed by the stroma-3i.

The major differentiating factors between the two groups of Raman spectra reflected in the form of the loadings are the features associated with collagen, including those at 853 cm^{-1}, 877 cm^{-1}, 920 cm^{-1} and 939 cm^{-1}, which are contributed by the stroma-true normal and are completely absent in the stroma-3i. This indicates a depletion of the collagen content of the identified regions of the stroma-3i layer due to the development and progression of the cervical cancer. The other differentiating factor between the two groups is the features related to the DNA, which are mainly contributed by the stroma-3i, and include those at 1218 cm^{-1} (T), 1245 cm^{-1} (A), 1272 cm^{-1} (C) and 1308 cm^{-1}, 1340 cm^{-1} and 1360 cm^{-1} (G), apart from the two bands at 765 cm^{-1} (T) and 815 cm^{-1} (O-P-O, DNA) which are contributed by the stroma-true normal.

There are some features associated with proteins, including those at 570 cm^{-1} (S-S stretching), 902 cm^{-1} (C-C stretching), 1031 cm^{-1} (C-N stretching), 1449 cm^{-1} (CH deformation) and 1665 cm^{-1} (amide-I), contributed by the stroma-true normal and those contributed by the stroma-3i include 1118 cm^{-1} (C-C stretching) and 1145 cm^{-1} (C-N stretching).

Notably, the two observations of significant depletion of the collagen content and increase in the DNA content in the stroma-3i as compared to the stroma-true normal layer, may lead to the conclusion that due to the depletion of the collagen during disease progression, relative to the DNA, collagen has a diminished contribution at some points in the Raman map of the stroma, and KMCA in **chapter 4** recognises the greater similarities with the DNA rich cells of the basal layer. This depletion of the collagen and increase in the DNA features may be the result of changes to the stromal cells themselves or associated with the invasion of the basal cells into the stroma and requires further analysis including PCA of the stroma and basal layers of the abnormal and intermediate samples consisting of stroma-4 versus basal-4 and stroma-6 versus basal-6, presented in this section.

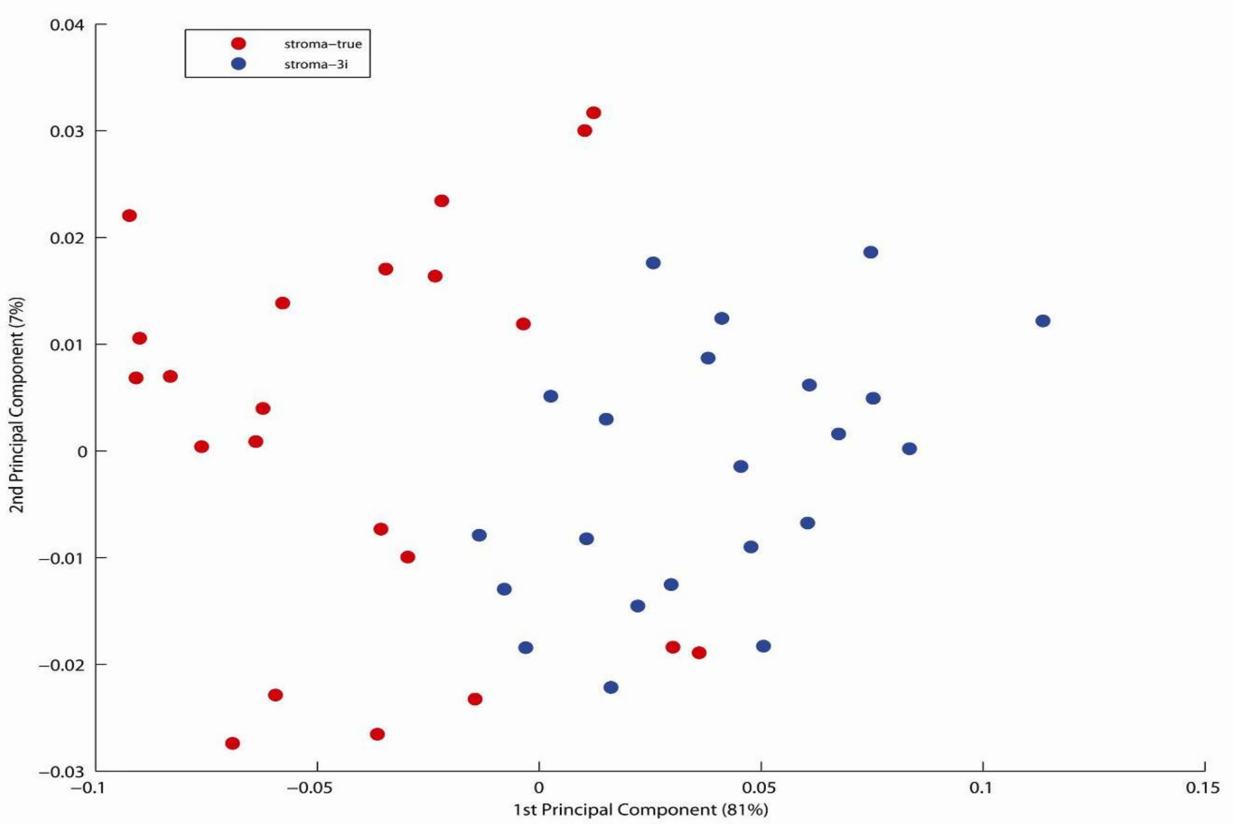


Figure 5.35 PCA scatter plot for stroma-true normal vs stroma 3-i.

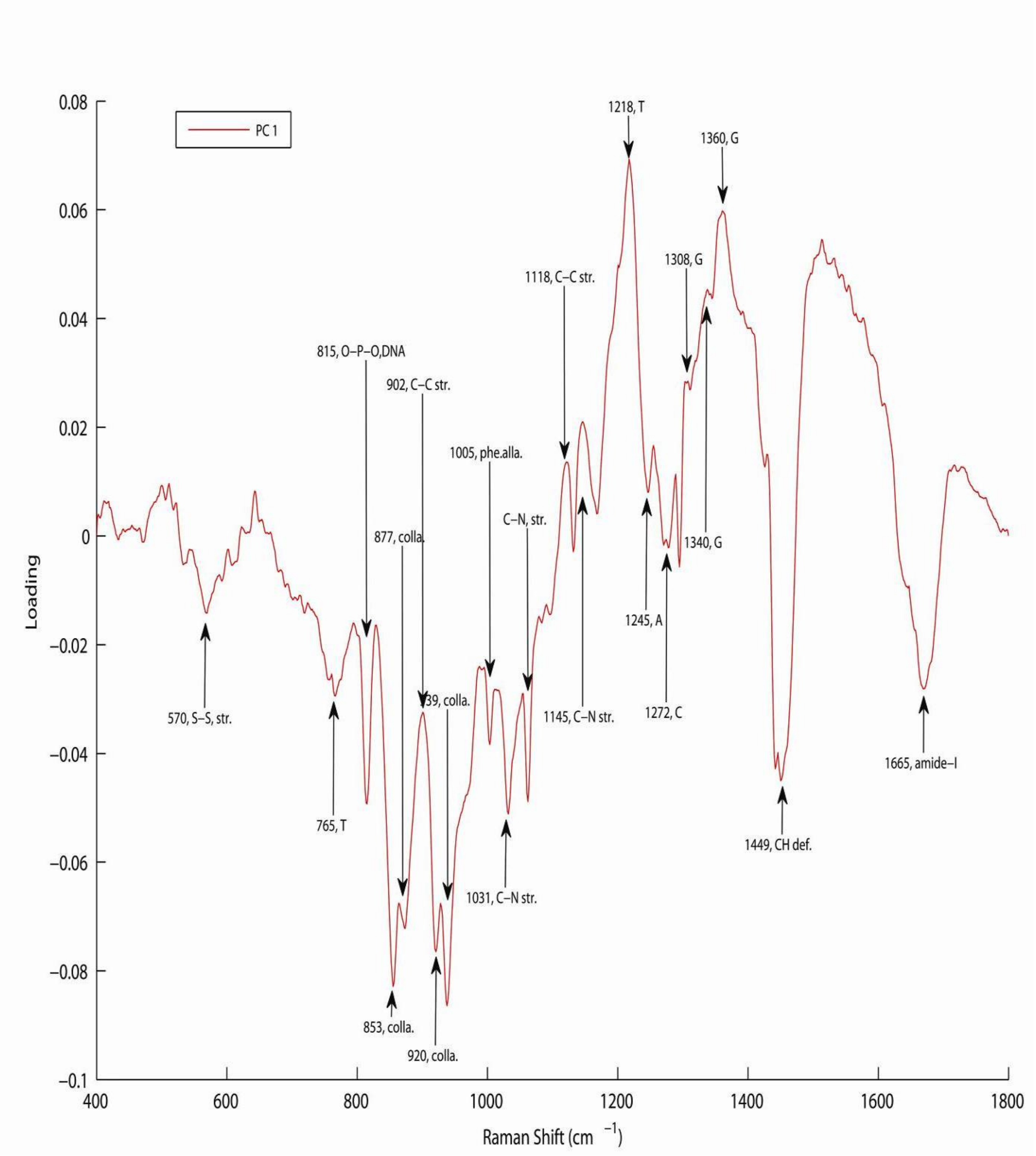


Figure 5.36 PCA loadings for stroma-true normal vs stroma 3-i.

5.7.3 PCA of stroma-true normal versus stroma-3ii

The PCA results of the stroma-true normal versus stroma-3ii are presented in **Figure 5.37** and **Figure 5.38** as PCA scatter plot and loadings, respectively. The spectra of the stroma-true normal are clustered in the negative axis with respect to PC1 and the stroma-3ii in the positive axis, showing the differentiation of the two groups. This indicates that the loadings in the negative sense are contributed by the stroma-true normal and those in the positive are contributed by the stroma-3ii dataset. The differentiation between the two groups of the spectra seems to be due to collagen and DNA related features, reflected in the form of the loadings. The features associated with collagen, including those at 853 cm^{-1}, 877 cm^{-1}, 920 cm^{-1} and 939 cm^{-1}, are contributed by the stroma-true normal and are completely absent in the stroma-3ii. The other differentiating factor between the two groups is the features related to DNA, which are predominantly contributed by the stroma-3ii, including those at 1218 cm^{-1} (T), 1245 cm^{-1} (A), 1272 cm^{-1} (C) and 1320 cm^{-1}, 1340 cm^{-1} and 1360 cm^{-1} (G), although two bands including 765 cm^{-1} (T) and 815 cm^{-1} (O-P-O, DNA) are contributed by the stroma-true normal.

Some features associated with proteins are observed, including those at 902 cm^{-1} (C-C stretching), 1031 cm^{-1} (C-N stretching), 1449 cm^{-1} (CH deformation) and 1665 cm^{-1} (amide-I), contributed by the stroma-true normal and those contributed by the stroma-3ii include 1118 cm^{-1} (C-C stretching) and 1145 cm^{-1} (C-N stretching).

As mentioned in the case of the PCA loadings of the stroma-true normal versus stroma-3i, the two observations, including significant depletion of the collagen content and the increase in the DNA content in the stroma-3ii as compared to the stroma-true normal layer may indicate that the depletion of the collagen during the disease progression and increased levels of DNA maybe the results of changes to the stromal cells themselves or associated with the invasion of the basal cells into the stroma.

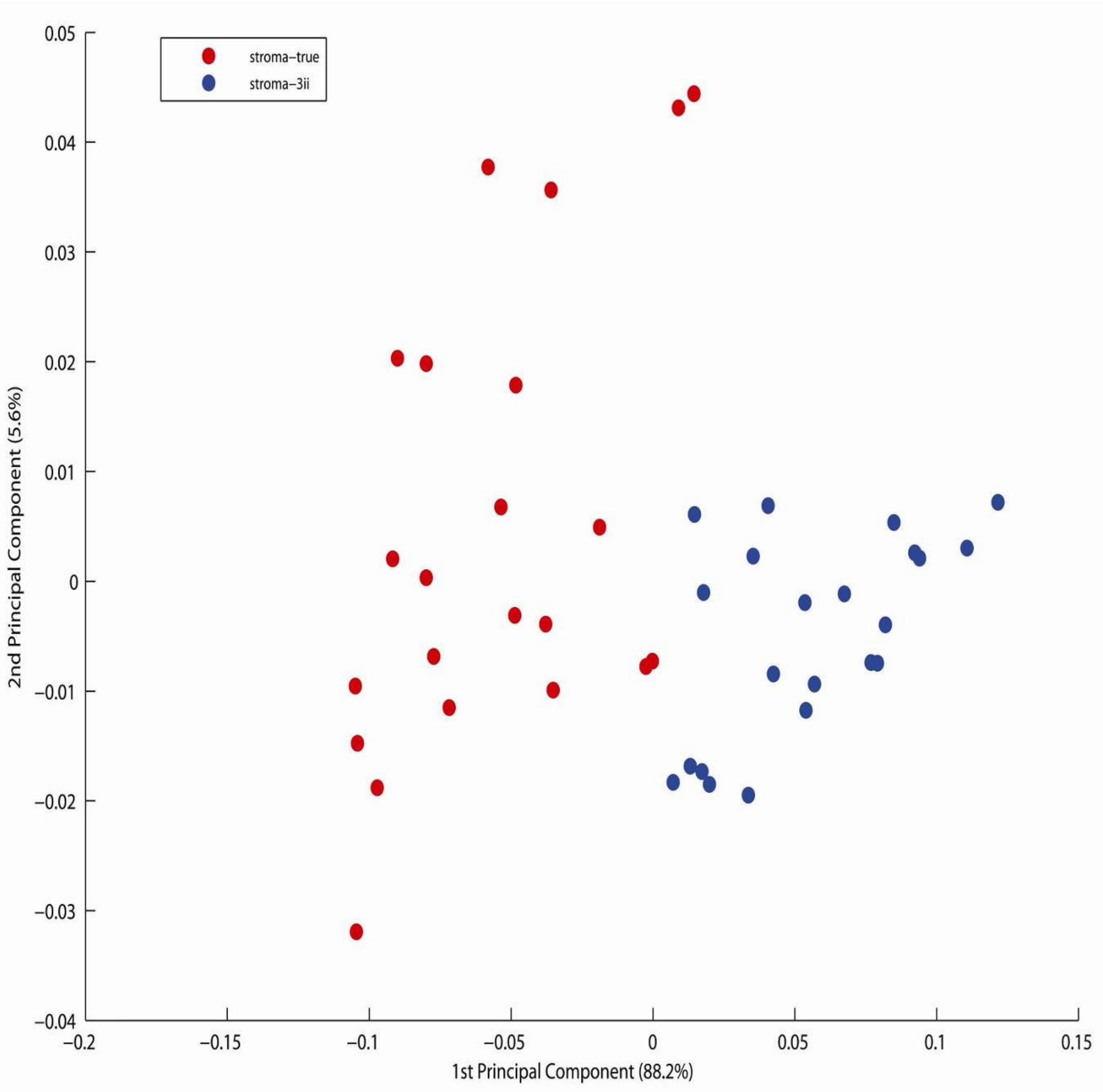


Figure 5.37 PCA scatter plot for stroma-true normal vs stroma 3-ii.

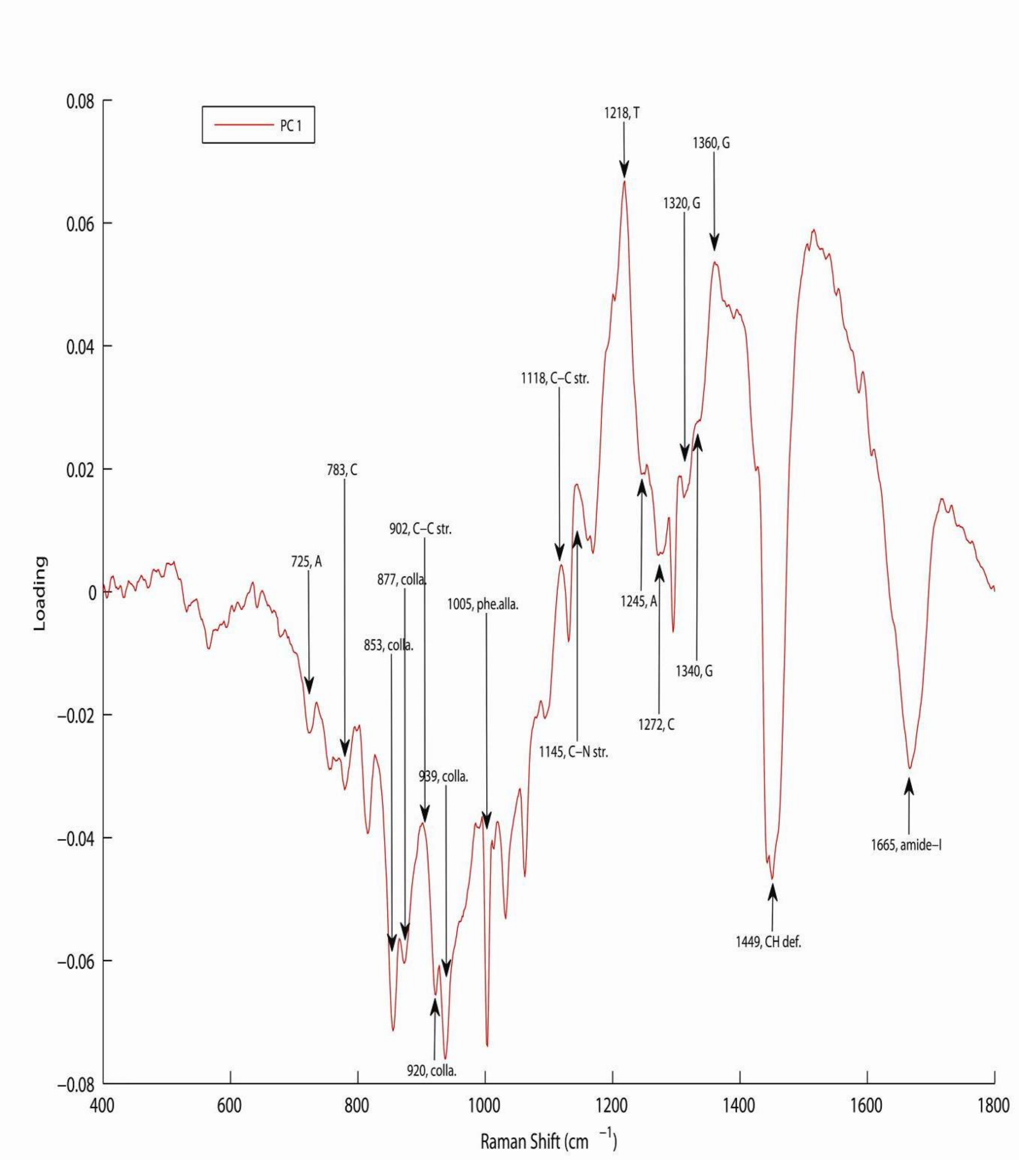


Figure 5.38 PCA loadings for stroma-true normal vs stroma 3-ii.

5.7.4 PCA of stroma-3i versus stroma-3ii

As discussed in the previous sections, the two types of the stromal layers, namely the stromal 3-i and stromal-3ii regions, are differentiated from the stroma-true normal, by PCA. Notably, KMCA showed that the stroma-3i corresponds to the pixels which do not show any features associated with collagen and was considered to be due to the development of the biochemical changes in the stroma consisting of the basal like characteristics. The stromal-3ii region was categorised by KMCA to be the diseased stromal layer, indicated by the presence of some collagen related bands. However, in the previous section, PCA loadings contributed by the stroma-3ii, did not show any features associated with collagen. In order to establish whether these two sub-regions/groups of the spectra are biochemically different from each other and the origin of that difference, PCA of these two groups of spectra was performed and the results are presented in **Figure 5.39** and **Figure 5.40** as PCA scatter plot and loadings, respectively.

The PCA scatter plot does not show any differentiation of the two groups with respect to PC1 (78.2%) and only a slight differentiation according to PC2 (10.4%). According to PC2, the stroma-3i spectra are clustered in the positive and stroma-3ii in the negative sense. The loadings contributed by the stroma-3i include those at 509 cm^{-1} (S-S stretching), 642 cm^{-1} (C-C stretching), 783 cm^{-1} (C), 1218 cm^{-1} (T) and 1340 cm^{-1} (G) while stroma-3ii contributes the features associated with collagen including 853 cm^{-1}, 877 cm^{-1}, 920 cm^{-1} and 939 cm^{-1}. Notably, there are no features of collagen contributed by the stroma-3i data, which can lead to the conclusion that the pixels identified by the KMCA as separate clusters in the stromal layer of the intermediate-III sample in **chapter 4** (**Figure 4.15**), are actually associated with the biochemical changes in the cells of the stromal layer.During cervical cancer progression, collagen can have a diminished contribution, relative to the DNA contribution, at some points in the Raman map of the stroma of the above mentioned sample, and KMCA recognises the greater similarities with the DNA rich cells of the basal layer. Moreover, the presence of the collagen related bands in the loadings contributed by stroma-3ii confirms the findings of the KMCA, whereby this sub-region was shown to have some collagen content, albeit diminished, as indicated by PC1 of the PCA of the stroma-true normal versus stroma-3ii (**Figure 5.35**).

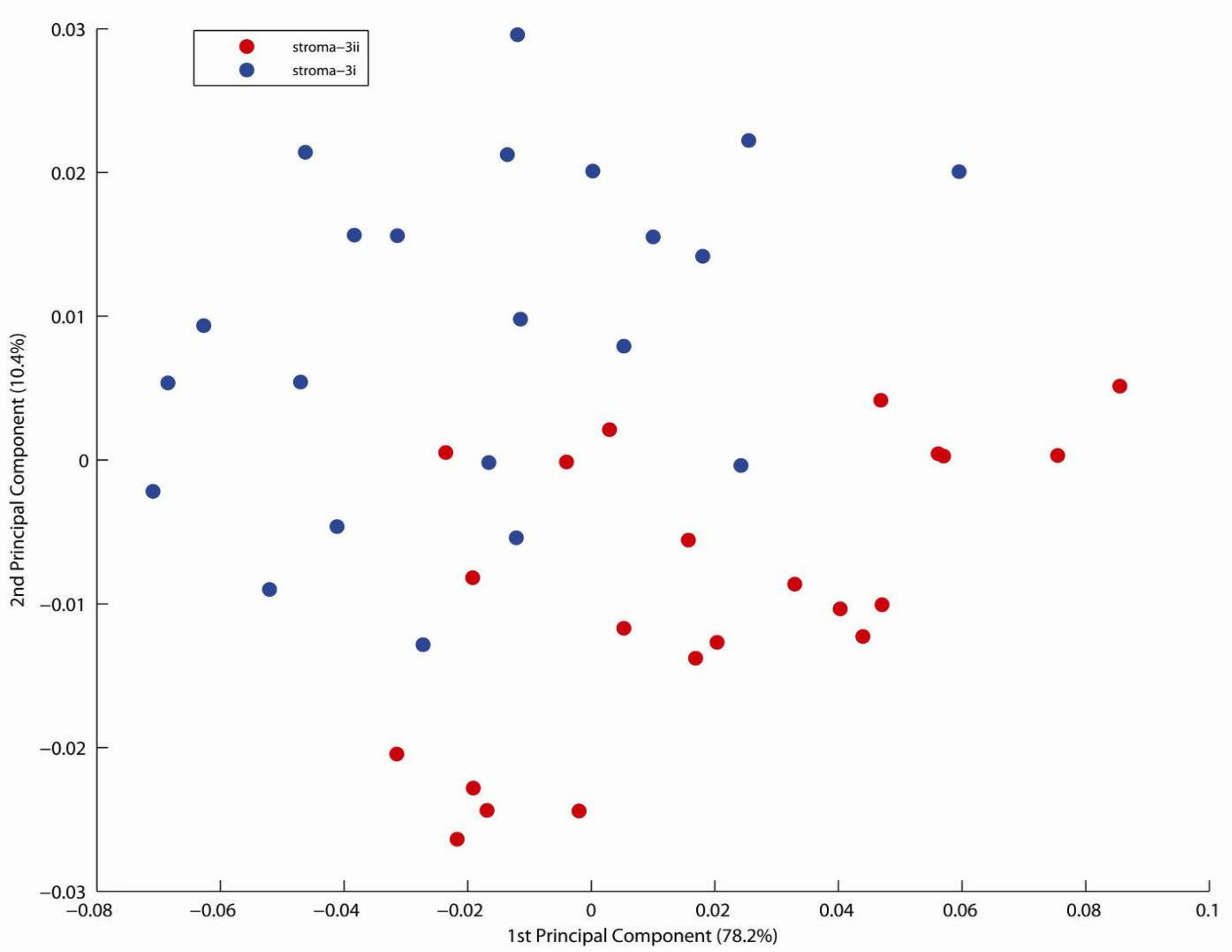


Figure 5.39 PCA scatter plot for stroma-3i vs stroma 3-ii.

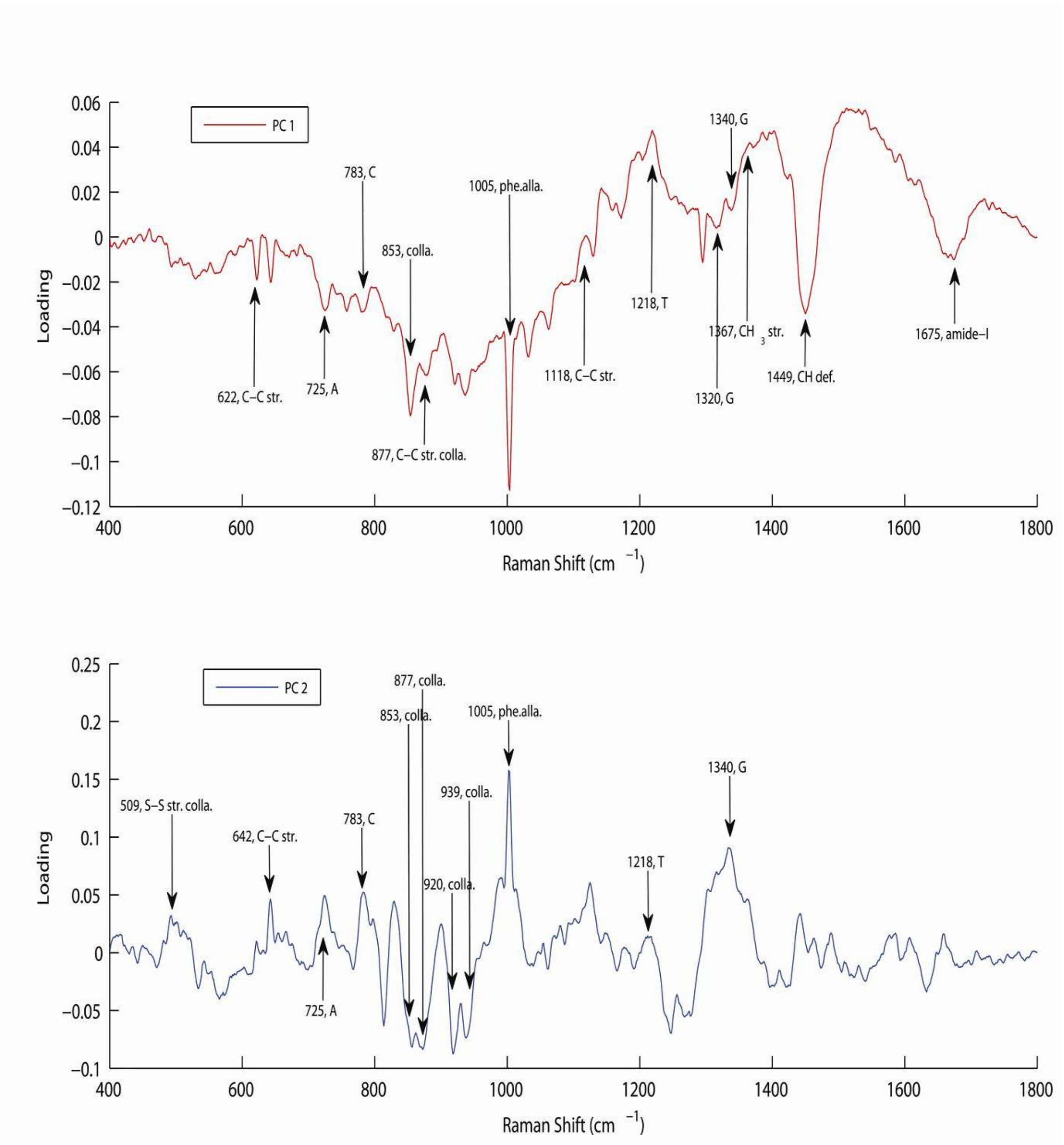


Figure 5.40 PCA loadings (PC1 & PC2) for stroma-3i vs stroma 3-ii.

5.7.5 PCA results of the stroma versus basal layers of abnormal and intermediate cervical samples

PCA of the Raman spectral data of the stromal layers of abnormal and intermediate samples (stroma-4 and stroma-6) versus their respective basal layers (basal-4 and basal-6) is presented in the following sections. PCA of (a) stroma of the abnormal (stroma-4) versus basal-5 and (b) stroma of the intermediate (stroma-6) versus basal-6 as well as their comparison may lead to the elucidation of the biochemical changes occurring in the stromal layers of these samples and also how these changes can bring them close in their biochemical composition to their respective basal layer.

5.7.5.1 PCA of Stroma vs Basal of the abnormal sample/ Stroma-4 vs basal-4

As confirmed in the previous sections the finding of the KMCA indicates that the increased basal layer-like characteristics in the stromal layer are associated with disease progression and biochemical changes in the stromal layer itself. To further confirm these findings, PCA of stroma-4, the stromal layer of the abnormal-I sample in **chapter 4**, **Figure 4.5**, with the basal-4 data set, which belongs to the same sample, was carried out.

Figure 5.41 and **Figure 5.42** represent the scatter plot and loadings of the PCA for the Raman spectra from stroma-4, the stromal layer of the abnormal-I characterised sample in chapter 4, **Figure 4.5**, with the basal-4 data set, which belongs to the same sample. The differentiation of the two groups can be seen as the clustering of the Raman spectra of the stroma-4 in a positive sense with respect to PC1 and the basal-4 in the negative sense. **Figure 5.42** indicates collagen related loadings of PC1, including those at 853 cm^{-1}, 874 cm^{-1}, 921 cm^{-1} and 939 cm^{-1}, which are associated with the stromal layer. The presence of these collagen related features indicates that here, collagen has not been depleted completely. Other positive loadings associated with the stroma-4 layer are associated with DNA and include those at 429 cm^{-1}, 751 cm^{-1} (thymine), 721 cm^{-1} (adenine), 775 cm^{-1}, 781 cm^{-1} (cytosine), 827 cm^{-1} (O-P-O of DNA) and 898 cm^{-1} (deoxyribose ring breathing). DNA related loadings are in fact seen in both negative and positive senses, indicating the presence of the features of the DNA in the basal layer and the stromal layer. Negative loadings associated with DNA which are contributed by the basal-4 layer include those at 1220 cm^{-1} (adenine) and 1318 cm^{-1}, 1340 cm^{-1} (guanine). Notably, these findings can be confirmed from the PCA loadings of the basal-true normal versus stroma-normal presented in the **Figure 5.3**, where the features contributed by the stroma-normal were only related to collagen including those at 853 cm^{-1}, 921 cm^{-1}, 939 cm^{-1} and 1242 cm^{-1}, along with some protein related features at 1165 cm^{-1}, 1195 cm^{-1} (C-N stretching) and 1655 cm^{-1} (amide-I) and there were no DNA features observed. The loadings associated with proteins, including those at 1003 cm^{-1} (phenyl alanine), 1129 cm^{-1} (C-N stretching), 1449 cm^{-1} (CH deformation) and 1655 cm^{-1} (amide-I) are contributed only by the basal-4 layer (in negative) and differentiate this layer from the stromal layer.

The observation of the collagen related bands contributed by stroma-4 indicates that there is still some collagen present, although the DNA features are increased as compared to the normal stroma. This confirms the observations of the previous section that, due to the depletion of the collagen during disease progression, the cells of the stromal layer exhibit a more DNA rich Raman spectrum. This can explain why the KMCA in **chapter 4** clusters the basal and stromal groups as one cluster, as relative to the DNA, collagen has a diminished contribution at some points in the Raman map of the stroma, and KMCA recognises greater similarities with the DNA rich cells of the basal layer.

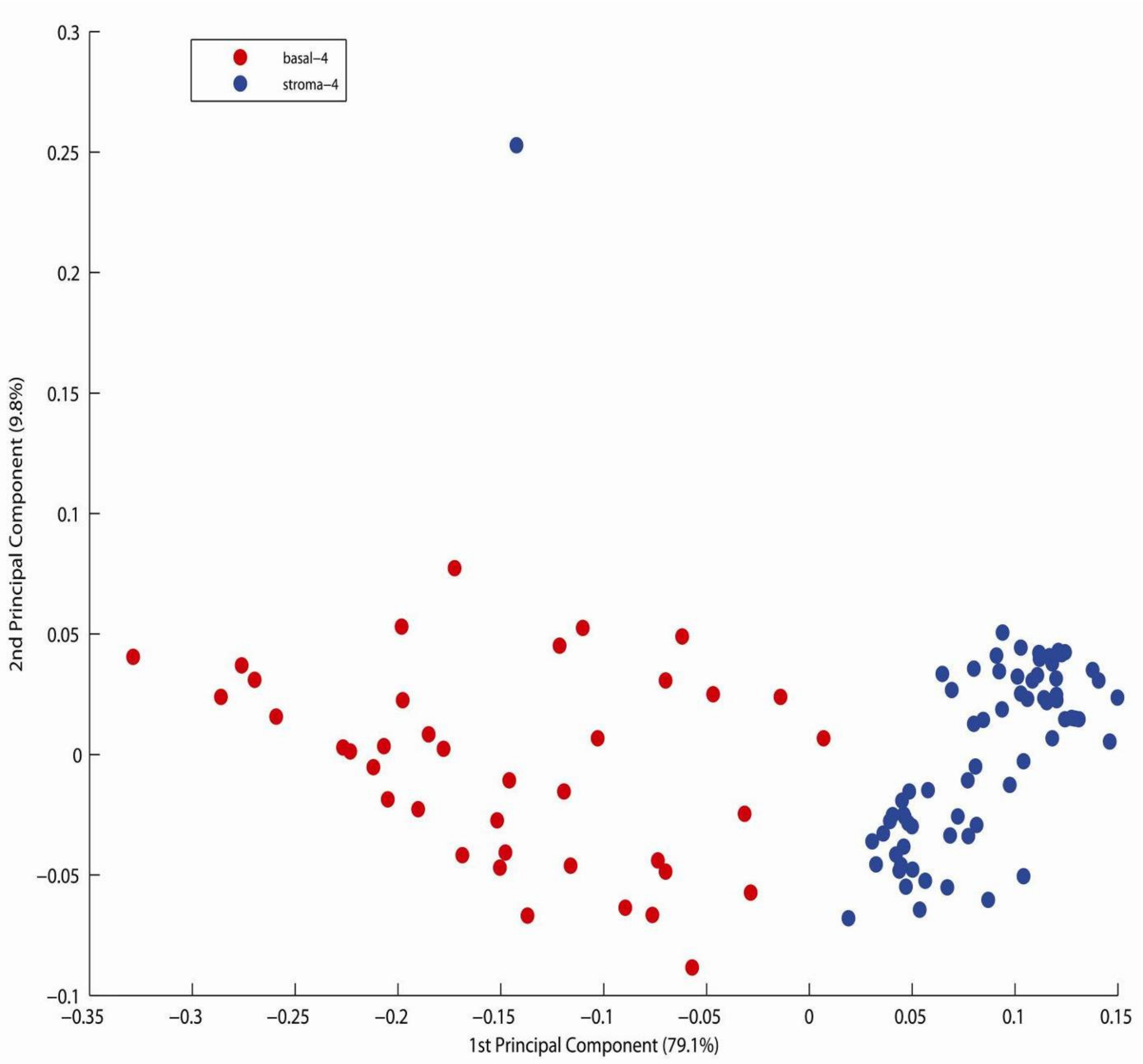


Figure 5.41 PCA scatter plot for stroma-4 vs basal-4.

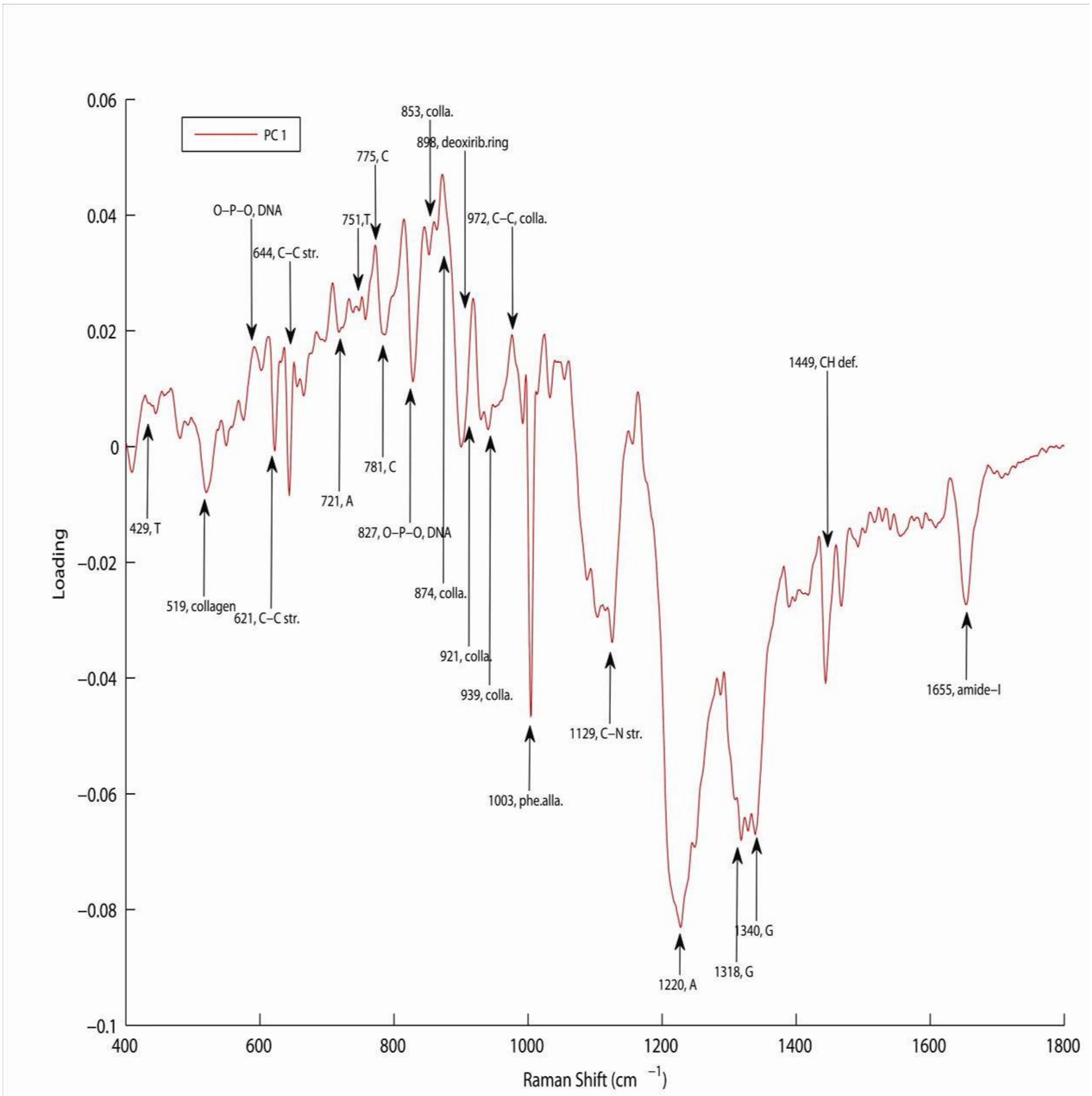


Figure 5.42 PCA loading (PC-1) for stroma-4 vs basal-4.

5.7.5.2 PCA of Stroma vs Basal of the Intermediate sample/ Stroma-6 vs basal-6

Figure 5.43 and **Figure 5.44** represent the scatter plot and loadings of the PCA for the Raman spectra from the stromal layer of the intermediate sample and its basal layer, termed basal-6. The differentiation of the two groups can be seen as the clustering of the Raman spectra of the stroma of the intermediate sample (stroma-6) in the positive sense of PC1 and those from the respective basal layer (basal-6) in the negative sense. The positive loadings correspond to the stroma and the negative loadings correspond to the basal layer. Collagen related loadings including those at 853 cm^{-1} and 939 cm^{-1} are contributed only by the stromal layer in the positive and are characteristic features of this layer and differentiate it from the basal-6 layer. Other loadings contributed by the stromal layer are associated with DNA and include those at 815 cm^{-1} (O-P-O of DNA), 898 cm^{-1} (deoxyribose ring breathing), 1243 cm^{-1} (thymine) and 1272 cm^{-1} (guanine). The loadings associated with DNA which are contributed by the basal-6 layer include those at 1218 cm^{-1} (adenine) and 1340 cm^{-1} (guanine). The negative loadings associated with proteins, including those at 1003 cm^{-1} (phenyl alanine), 1449 cm^{-1} (CH deformation) and 1655 cm^{-1} (amide-I), are contributed by the basal-6 layer and differentiate this layer from the stromal layer.

The observation of the collagen related bands contributed by stroma-6 indicates that there is still some collagen, although the DNA features are increased as compared to the normal stroma. This supports the conclusion that the biochemical changes observed are associated with the cells of the stromal layer themselves.

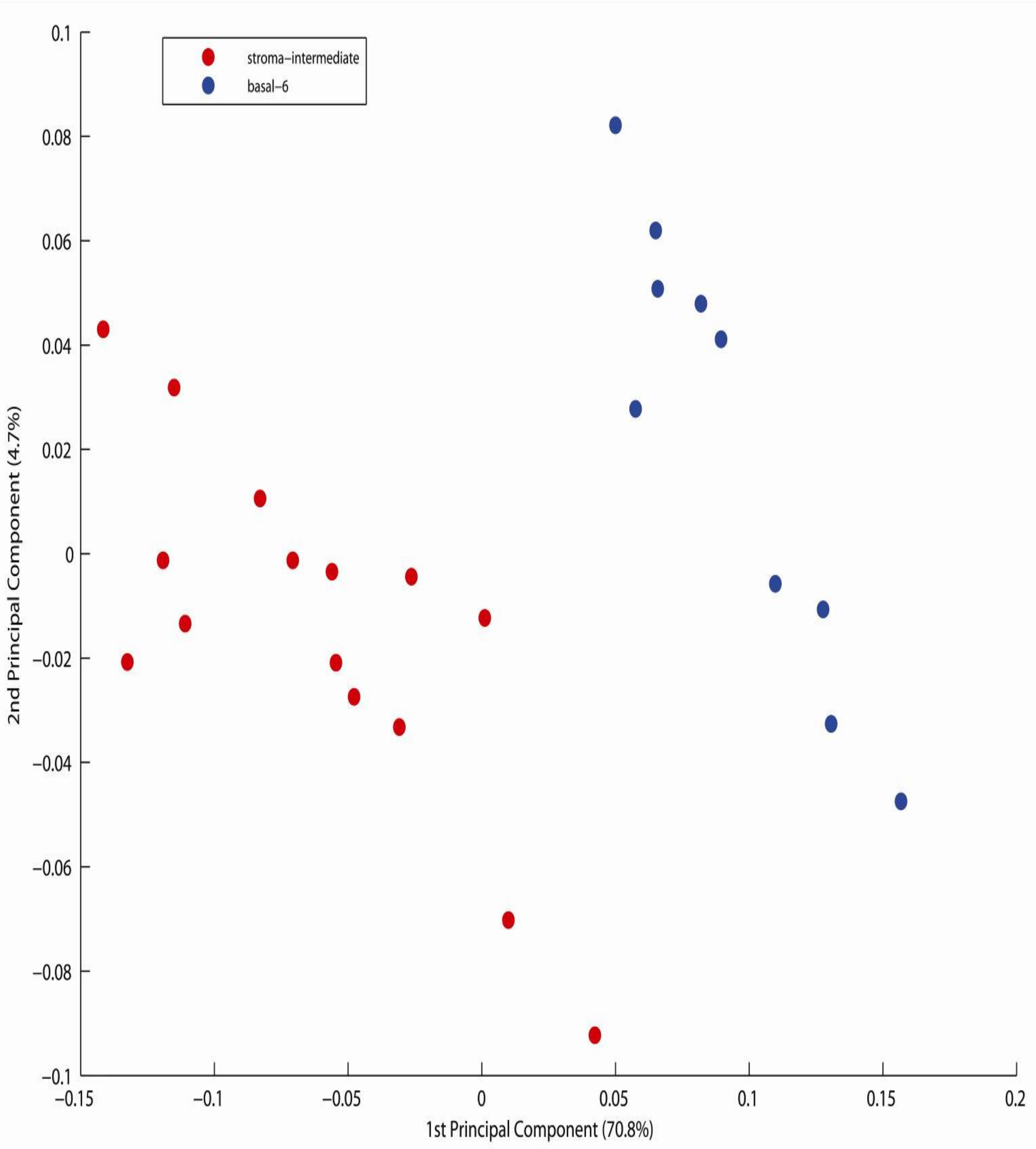


Figure 5.43 PCA scatter plot for stroma-6 vs basal-6.

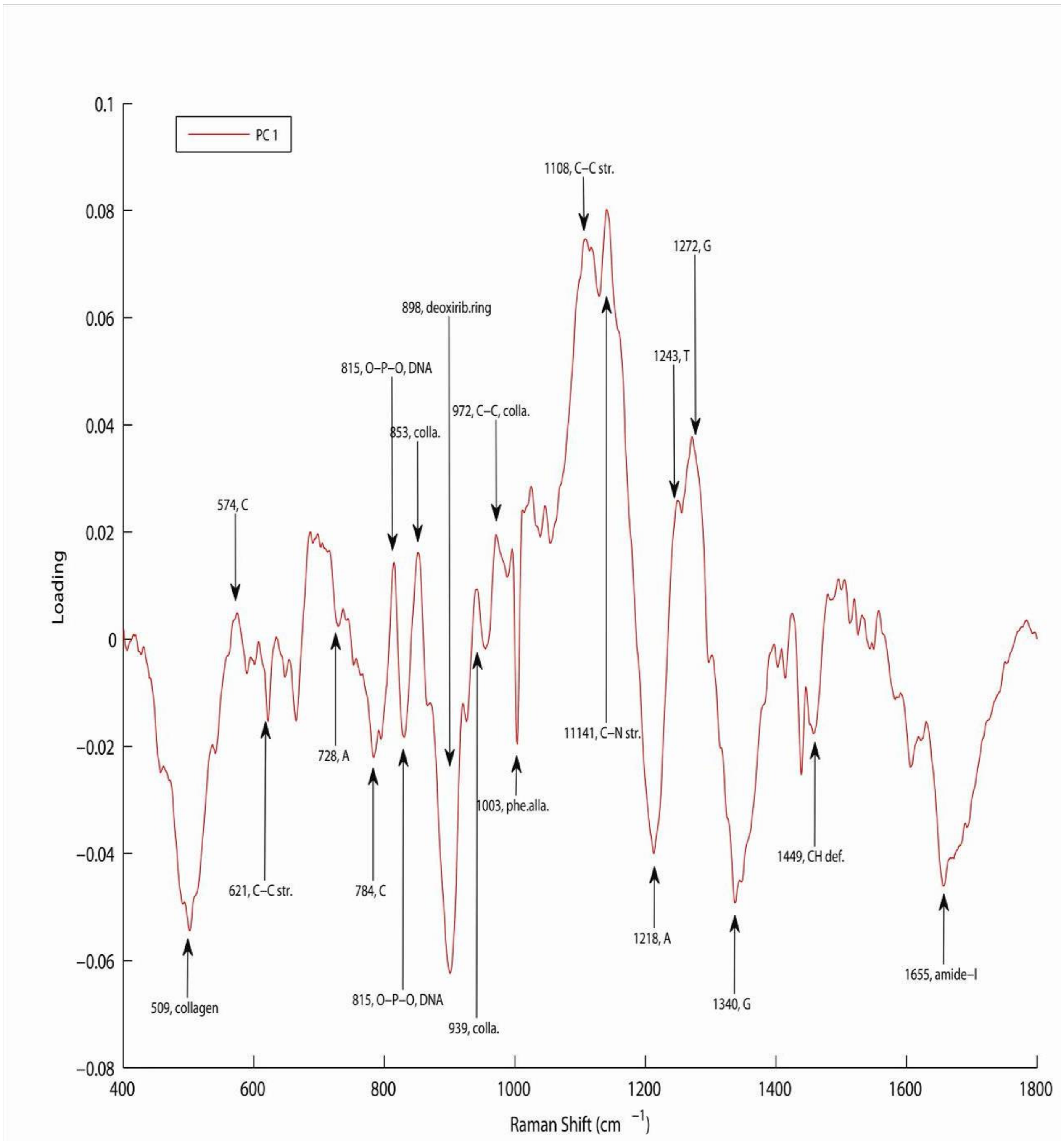


Figure 5.44 PCA loading (PC-1) for stroma-6 vs basal-6.

5.7.5.3 Comparison of the PCA results of stroma of abnormal (stroma-4) vs basal-4 and stroma of intermediate (stroma-6) vs basal-6

Comparison of the PCA scatter plots and loadings for (a) stroma of abnormal (stroma-4) vs basal-4 and (b) stroma of intermediate (stroma-6) vs basal-6 was made. The PCA scatter plot of stroma-4 vs basal-4 shows that the cluster of the Raman spectra of the stromal layer of this sample is closer to the cluster of the basal layer of the same sample (**Figure 5.41**) as compared to the clusters of the stroma-6 and basal-6 shown in the scatter plot (**Figure 5.43**).

The collagen loadings contributed by stroma-4 (stroma-abnormal), shown in **Figure 5.42,** including those at 853 cm^{-1} and 939 cm^{-1}, are weaker as compared to those contributed by stroma-6 (stroma-intermediate) shown in **Figure 5.44**.

The comparison of the 400-1000 cm^{-1} region of the PCA loadings for both types of the PCA analysis clearly shows that the major difference in this part is based on increased DNA related loadings in stroma-4 as compared to stroma-6, in which the major loadings observed derive from collagen.

Consideration of the 1000-1800 cm^{-1} region shows that, in the loadings of stroma-4, there are no protein related features but in the stroma of the intermediate there are two strong loadings which are related to proteins at 1108 cm^{-1} and 1141 cm^{-1}.

5.8 Conclusion

In this chapter, Raman micro spectroscopy, with the aid of PCA, has been employed to determine the biochemical changes occurring in the different layers of the cervical tissue samples, namely the basal, stromal and superficial layers. PCA has been found to be very useful to further interpret the results of the KMCA and to elucidate the differentiation of the basal-true normal and basal-CIN II normal (far from diseased), both of which have been found different than the basal layers of the diseased samples. This means that in order to predict the presence of the disease it is not required to essentially go for the region of disease and suggests that Raman micro spectroscopy can identify onset of disease before the morphological changes.

The comparison of the PCA loadings of the HPV negative and positive cell lines (*C33A* and *CaSki*), with the PCA loadings of the basal-true normal versus basal-CIN II normal indicated that there are no strongly matching Raman signatures which could lead us to identify the signatures of the HPV infection in the cervical cancer tissue samples. However, there is a feature associated with the amide-III beta sheet (1222 cm^{-1}) observed in the PCA of the basal-true normal versus basal-CIN-II normal which may be of interest, as it is the major difference between the two data sets reflected in the form of loading by PCA but it is absent in the loadings of the cell lines. The amide-III feature (1222 cm^{-1}) is consistently present in the loadings of the PCA contributed by the basal layers of the intermediate and abnormal samples in the PCA with the basal-true normal. Thus this feature may be considered as an early marker of disease progression and could potentially indicate HPV infection. In this regard, interestingly, this band is much reduced in intensity of the PCA loadings of their basal layers with the basal-true normal presented in the carcinoma *in situ* sample shown in **Figure 5.33** (**Figure 4.8** in chapter 4 called abnormal III) and most extreme abnormal sample (**Figure 4.7** in chapter 4 called abnormal II) presented in the **Appendix**, **Figure A1.XII**. Notably, these are the most diseased samples and the reduction of the amide-III related feature may indicate that at this stage of the disease, integration of the HPV virus into the host genome has taken place instead of its presence with the protein coat independently in the host cell. The PCA of the basal layer of the two normal samples including "true normal-1" and "true normal-2" is presented in the Appendix, **A1.1** and **A1.2** as scatter plot and loadings respectively. The scatter plot shows some differentiation of the two clusters of the Raman spectra which is attributed to the inter-sample variability but the Raman spectral features which has been the differentiating features, particularly the band at 1222 cm^{-1}, between the true-normal and diseased samples are not observed here in the loadings which further support the conclusion that the amide-III beta sheet (1222 cm^{-1}) may be taken as Raman spectral marker of the early progression of the disease.

It can be concluded that Raman spectroscopy certainly has potential for detecting the biochemical changes associated with HPV infection but as there are no strongly matching Raman signature loadings of the HPV negative and positive cell lines (*C33A* and *CaSki*), with the PCA loadings of the basal- true normal versus CIN II normal basal, further studies of spectral markers of HPV in tissue samples rather than cell lines maybe helpful in this regard.

Regarding the evaluation of the KMCA observation on the invasion of the basal cells/characteristics of the basal layer into the superficial layer, Raman spectroscopy along with PCA have led to the conclusion that, during the progression of cervical cancer, there are the cells of the basal layer which are not normal indicated by their rapidly proliferating behaviour (DNA rich content) and lack of glycogen content,. These cells are progressing to the superficial layer due to the progression of the disease. The PCA confirms these findings made on the basis of the KMCA in the **chapter 4**.

For the evaluation of the KMCA observation on the appearance of the basal cells/characteristics of the basal layer in the stromal layer, Raman spectroscopy along with the PCA have led to the conclusion that these are actually associated with the biochemical changes in the cells of the stromal layer and there is no migration of the basal cells into the stroma. In addition, during cervical cancer progression, relative to the DNA, collagen has a diminished contribution at some points in the Raman map of the stroma and KMCA recognises the greater similarities with the DNA rich cells of the basal layer. This has been supported by the observation of the collagen related loadings contributed by the stroma-4 and stroma-6 indicating that there is still some collagen present, although the DNA features are increased as compared to the normal stroma. Raman spectroscopy has identified significant changes in the stroma which may be useful for diagnostics and also can be important for disease progression and understanding mechanisms for treatment of medium stage progression.

In order to further confirm the findings from this chapter, including progression of the basal cells into the superficial layer but not into the stroma, immunohistochemical staining for the evaluation of the expression of p16 protein is performed and presented in **chapter 6**. The expression of p16 is a biomarker for the presence of HPV and is expected to be enhanced in the basal and superficial cells/layers rather than in the stroma/stromal cells.

Chapter 6

Immmunohistochemical staining of cervical tissue samples

6.1 Immunohistochemistry

Immunohistochemistry is a method of detecting specific macromolecules (e.g. proteins) in cells and tissues due to their specific interactions. The techniques can be used to detect, for example, target proteins by using antigen-antibody interactions, nucleic acids by using hybridization of the complementary segments of the nucleic acid and carbohydrate moieties by using lectin binding (Junqueira, 1992).

6.2 Background

6.2.1 Marker proteins associated with HPV infection

In 99.7% cases of cervical intraepithelial neoplasia, infection with human papillomavirus (HPV) is considered to be a key factor. HPVs are non-enveloped viruses with an icosahedral protein coat and double stranded circular DNA of genome size approximately 8kb (Longworth and Laimins, 2004).

The life cycle of HPV has two stages or phases; the first is the productive phase which involves the replication of viral DNA in the basal cells of the host to maintain a copy number of 50-100 genomes per cell. Viruses achieve this with the help of early expression of genes/proteins including E1, E2, E4, E5, E6 and E7.

The E6 and E7 genes/proteins affect the functioning of tumour suppressor protein p16 and pRB (Jedpiyawongse et al., 2008). The p16 protein is only expressed when the hypophosphorylated retinoblastoma (Rb) protein (pRb), which controls it negatively, is made dysfunctional. The oncoproteins encoded by the E6 and E7 genes have the ability to bind host cell regulatory proteins, especially the tumor suppressor gene products p53 and pRb. These changes may lead to functional inactivation of Rb protein through binding to the E7 gene product (Dyson et al., 1989, Werness et al., 1990). In several studies of cancer, the p16 protein is over-expressed when the Rb gene is mutated, deleted or inactivated and, conversely, p16 protein is markedly reduced or absent in cell lines and clinical specimens that contain an intact Rb gene. As p16 is controlled or regulated by pRb, which is regulated by the E7 gene product from the HPV infection, the expression analysis of p16 in cervical tissue samples may help to identify the HPV infection level during the progression of cervical cancer.

In eukaryotic cells, a complex pattern of controlled expression and post-translational modifications (*e.g.* phosphorylation) of the cell-cycle regulating proteins affects the progression of the cell cycle. The p16INK4a protein plays a crucial role in the regulation of the mechanism of the cell cycle. It is an important part of the retinoblastoma protein (pRB)-mediated control of the G1-S-phase, and triggers the cell cycle arrest in the course of cellular differentiation processes. In terminally differentiated epithelial cells, p16INK4a expression is reduced to the levels which are not detectable by immunohistochemistry. In various tumor samples, the tumor suppressor gene p16INK4a has been found to be functionally inactivated by gene mutation or promoter hypermethylation. In replication competent cervical epithelial cells, however, where high risk type HPV (HR-HPV) oncoproteins have initiated the cellular transformation process, p16INK4a has been shown to be highly up-regulated and thus strongly over expressed.

In the current study, immunohistochemical staining of the cervical tissue samples for p16 protein is performed to be used as a marker of the presence of the cervical cancer and its progression. The objective of this study is to identify the severity of the cervical cancer by evaluating the expression profile of the p16 protein, which is considered as a marker protein for cervical cancer progression. This is also done to confirm and compare the findings including the migration of the basal cells/characteristics of the basal layer into the superficial and stromal layer which has been made on the basis of the results of Raman micro spectroscopy, described in **chapters 4 & 5**. The p16 expression is expected to be enhanced in the basal and superficial cells/layers as this protein is expressed only in these layers as a result of the onset of the disease and notably, not in the stroma/stromal cells.

In addition, **chapter 4** & **5** also led to the conclusion that there is no migration of the basal cells into the stroma. In addition, during cervical cancer progression, relative to the DNA, collagen has a diminished contribution at some points in the Raman map of the stroma and KMCA recognises the greater similarities with the DNA rich cells of the basal layer. If invasion of the basal cells into the stroma is occuring, one would expect the expression of the p16 protein to be detectable in those cells, originating from the basal layer, which have migrated into the stroma.

6.3 Methodology

The staining of formalin-fixed, paraffin-embedded tissue sections was done using the CINtec® p16^{INK4a} Histology kit (Roche), which uses the DAB (3, 3-diaminobenzidine) detection system. As the samples in use for the current study are obtained as formalin-fixed and paraffin-embedded sections, in order to undertake immunohistochemical staining on them, they are required to be dewaxed. The tissue processing and detection of the p16 protein in these samples was done by using following protocol:

6.3.1 Deparaffinization and Rehydration

Prior to deparaffinization, slides were placed in a drying oven at a temperature of 60 °C for at least 20 minutes to remove water, thereby improving adherence of the tissue section to the slide (“baking”), as well as to melt the paraffin. Tissue slides were deparaffinized to remove embedding medium and then rehydrated before the staining procedure. The slides were processed at ambient temperature (20-25 °C) according to the following steps:

a) Immersion for 5 minutes in a xylene bath; this step was repeated, once with a fresh bath.
b) Excess liquid was removed.
c) Immersion for 3 minutes in 95% ethanol.
d) This step was repeated once with a fresh bath.
e) Excess liquid was removed.
f) Immersion for 3 minutes in 70% ethanol.
g) This step was repeated once with a fresh bath.
h) Excess liquid was removed.
i) Slides were rinsed for 30 seconds in deionized water.

6.3.2 Staining Protocol

6.3.2.1 Epitope Retrieval

Plastic Coplin staining jars were filled with the diluted Epitope retrieval solution (provided with the kit) and placed in the heated water bath (95-99 °C) to heat the Epitope Retrieval Solution. At this step, it was important to adjust the level of the water in the water bath to make sure that the jar was immersed in the water to a level of 80%.

To stabilize the temperature and avoid evaporation, the jars were covered. Further details of the epitope retrieval step are:

a) The deparaffinised sections were immersed in the preheated epitope retrieval solution in the staining jars; this step lowered the temperature in the jars to less than 90°C. The temperature of the water bath and the Epitope Retrieval Solution was brought to 95-99 °C.
b) The sections were kept for 10 minutes at 95-99 °C.
c) The entire jar with slides was removed from the water bath.
d) The slides were left to cool in the Epitope Retrieval Solution for 20 minutes at room temperature.
e) The Epitope Retrieval Solution was removed and the sections were rinsed in diluted wash buffer (provided with the kit) for 5 minutes.

6.3.2.2 Peroxidase-Blocking Reagent

a) 200 μL Peroxidase-blocking reagent was applied to cover the samples.
b) The sections were incubated for 5 minutes;
c) The excess liquid was tapped off and slides were placed in a fresh wash buffer bath for 5 minutes.

6.3.2.3 Antibody or Negative Reagent Control

The excess buffer was removed and tissue sections were covered with 200 μl of the antibody (Mouse Anti-Human p16INK4a). For negative controls to confirm the specificity of the staining, tissue sections were not incubated with the primary antibody but instead were incubated with the Negative Reagent Control which includes monoclonal mouse anti-Rat oxytocin-related neurophysin antibody, supplied in 50 mmol/L Tris buffer pH 7.2, containing 15 mmol/L sodium azide (NaN_3) and stabilizing protein. Rat oxytocin-related neurophysin is not present in human tissues and thus no staining would be expected in these negative control samples. After this step, the samples were incubated for 30 minutes. This was followed by the removal of the excess liquid and the slides were placed in a fresh wash buffer bath for 5 minutes.

6.3.2.4 Visualization Reagent

a) Excess buffer was removed.

b) The sections were covered with 200 µl of the visualisation reagent (DAB Chromogen added to 2 mL of DAB Buffered Substrate) which is provided with the kit.

c) The sections were incubated for 30 minutes.

d) Excess liquid was decanted and the slides were placed in a fresh wash buffer bath for 5 minutes.

e) This step was repeated twice with a fresh wash buffer bath.

6.3.2.5 Substrate-Chromogen Solution (DAB)

a) The sections were covered with 200 µl of Substrate-chromogen solution (DAB).

b) They were then incubated for 10 minutes.

c) Excess liquid was decanted and the sections were rinsed gently with deionized water.

d) The substrate-chromogen solution (DAB) waste was collected in a hazardous materials container for safe disposal.

6.3.2.6 Counterstain with Hematoxylin

a) The slides were immersed in a bath of hematoxylin and incubated for 5 minutes.

b) They were then placed in a tap water bath and rinsed gently with running tap water until all the residual hematoxylin was cleared. This gave a blue colour.

6.3.2.7 Mounting

a) The sections were immersed for 3 min in 70% Ethanol.

b) They were then immersed for 3 min 99% Ethanol.

c) The sections were mounted with a glass coverslip using DPX.

d) The slides were protected from light and stored at room temperature.

6.4 Results of Immunohistochemical staining of cervical tissue samples

The immunohistochemical staining for p16 causes the development of a brown colour in the nucleus and cytoplasm of the epithelial cells of cervical tissue (Hoshikawa et al., 2010). To confirm the specificity of the staining, expression evaluation of p16 protein, the immunostaining of p16 of a CIN II diagnosed cervical tissue section recorded with X10 and X40 objective lenses, has been shown in **Figure 6.1 A & B** respectively. This sample was used as a negative control for immuno histochemical staining of p16 protein. It was not incubated with the primary antibody and hence it did not show any expression detection or brown colour due to antibody-antigen attachment.

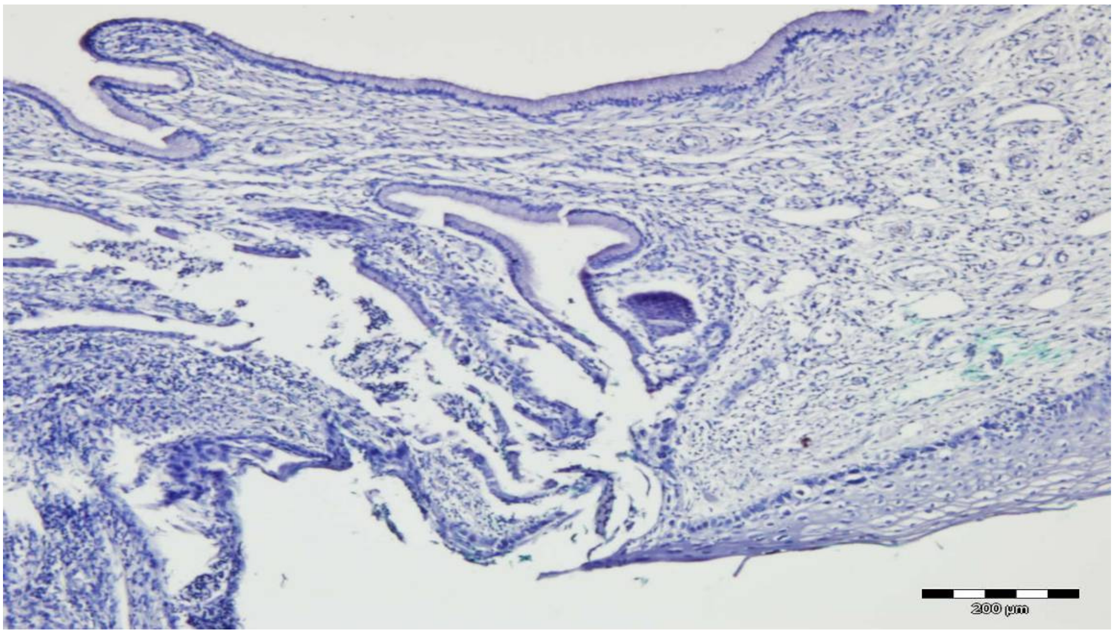


Figure 6.1 A Micrograph with X10 objective showing negative immunohistochemical staining for p16 protein in CIN II cervical tissue section with normal and diseased sides, used as a negative control (no antibody).

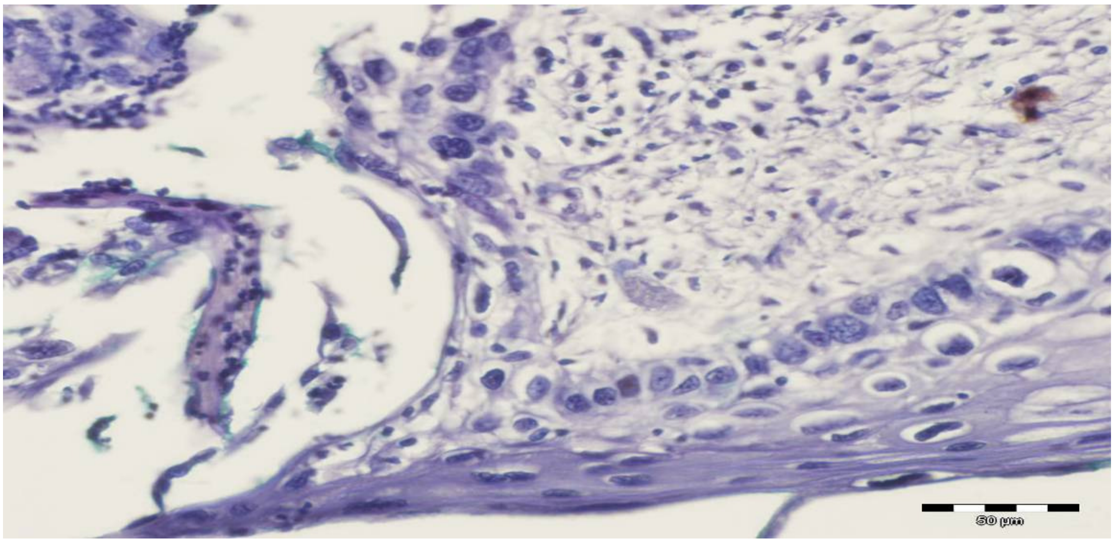


Figure 6.1 B Micrograph with X40 objective showing negative immunohistochemical staining for p16 protein in CIN II cervical tissue section with diseased side, used as a negative control (no primary antibody).

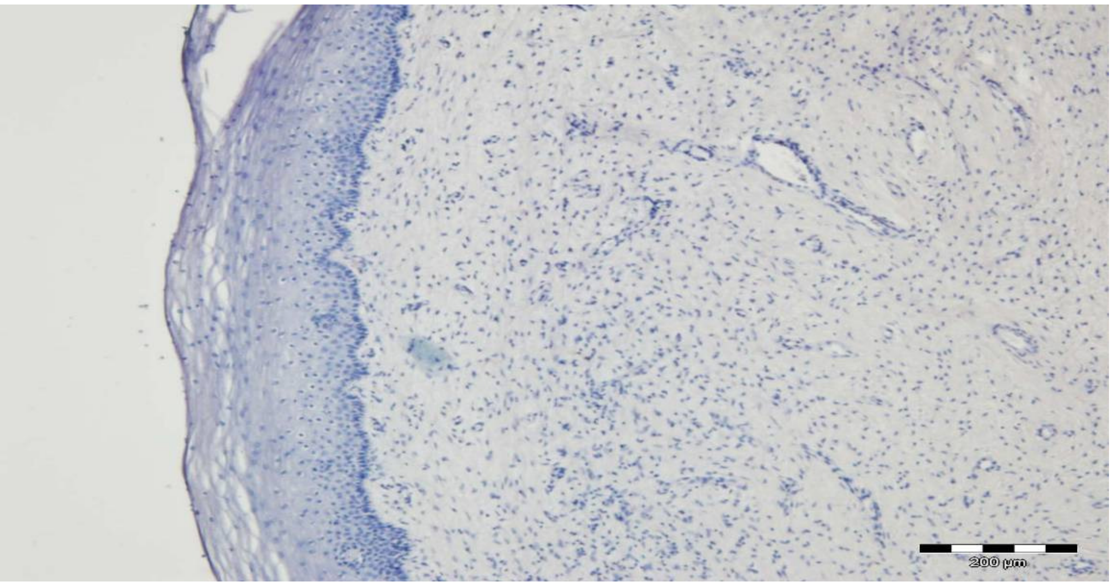


Figure 6.2 A Micrograph with X10 objective showing negative immunohistochemical staining for p16 protein in normal cervical tissue section (chapter 4 reference: Figure 4.3).

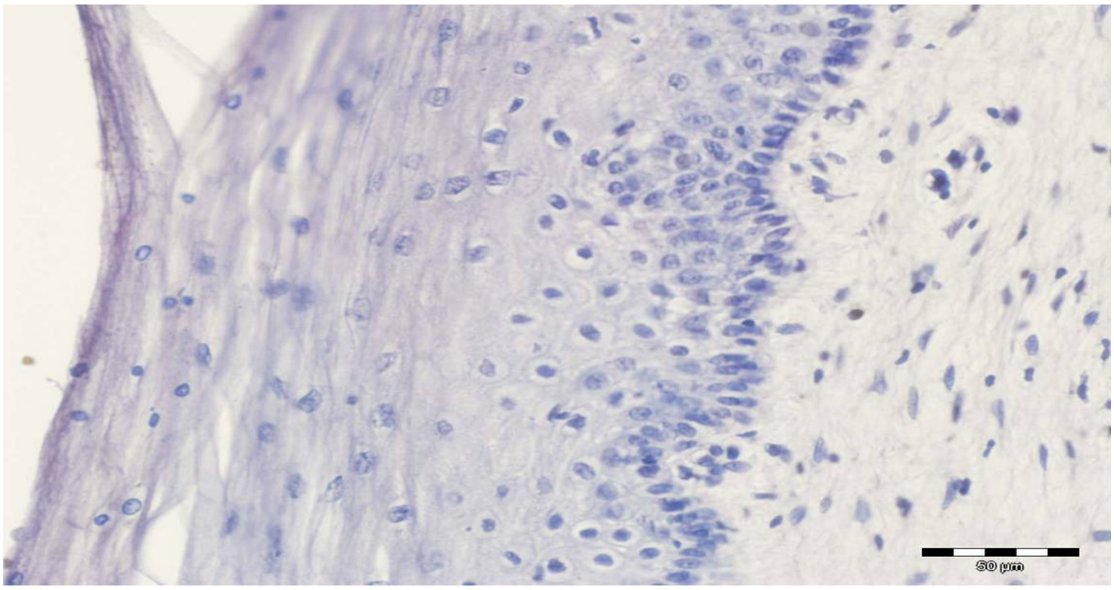


Figure 6.2 B Micrograph recorded with X40 objective showing negative immunohistochemical staining for p16 protein in normal cervical tissue section (chapter 4 reference: Figure 4.3).

In **Figure 6.3 A & B** the p16 staining of the CIN III cervical tissue section recorded with X10 and X40 objective lenses respectively, is shown, revealing extensive staining of the cervical epithelial cells. In chapter 4, the Raman analysis of this sample is shown in **Figure 4.7**. The Raman analysis characterised it as the most abnormal sample, as KMCA indicated no discrimination of the basal, stroma and superficial layers. This analysis is confirmed by immunohistochemical staining of the p16 protein. Notably, however, there is no evidence of p16 expression in the stromal layer, supporting the information acquired from chapter 5 regarding the absence of the invasion of the basal cells into the stroma and that changes in the Raman features associated with the progression of the disease are related to the cells of the stroma themselves.

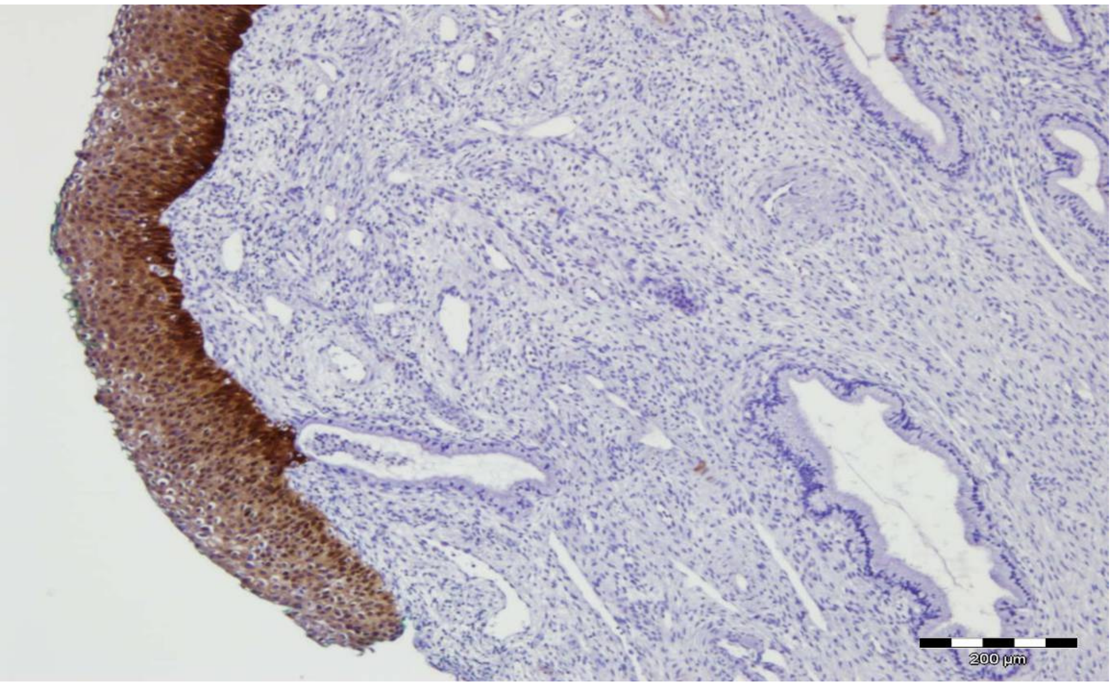


Figure 6.3 A Micrograph recorded with X10 objective showing positive immunohistochemical staining for p16 protein in CIN-III cervical tissue section (chapter 4 reference: Figure 4.7).

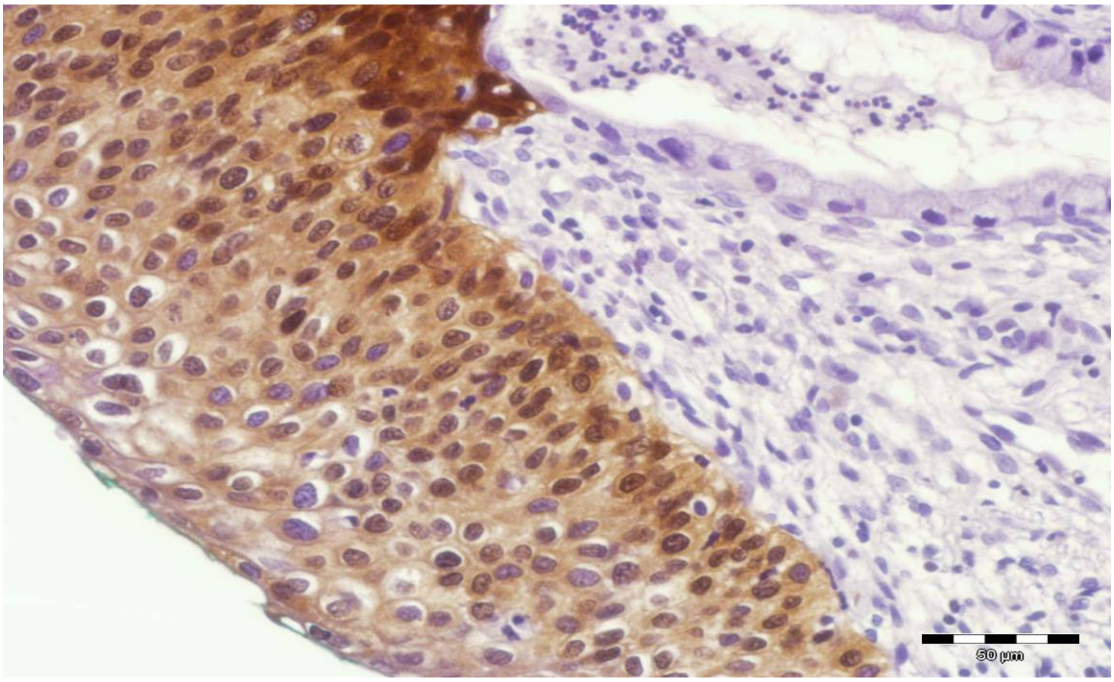


Figure 6.3 B Micrograph recorded with X40 objective showing positive immunohistochemical staining for p16 protein in CIN-III cervical tissue section (chapter 4 reference: Figure 4.7).

.

Figure 6.4 A & B shows the p16 staining of a CIN II cervical tissue section recorded with X10 and X40 objective lenses respectively, indicating no staining of the cervical epithelial cells. In chapter 4, this sample corresponds to **Figure 4.15**, and Raman characterised it as an intermediate sample as indicated by the KMCA map where discrimination of the basal, stroma and superficial was observed, but the latter two layers were observed to have some characteristics of the basal layer. The negative immunohistochemical staining for p16 protein found here indicates that the expression of this protein is not detectable. However, Raman micro spectroscopy could detect the biochemical changes taking place in this sample due to the progression of the cervical cancer independent of the expression of p16 protein.

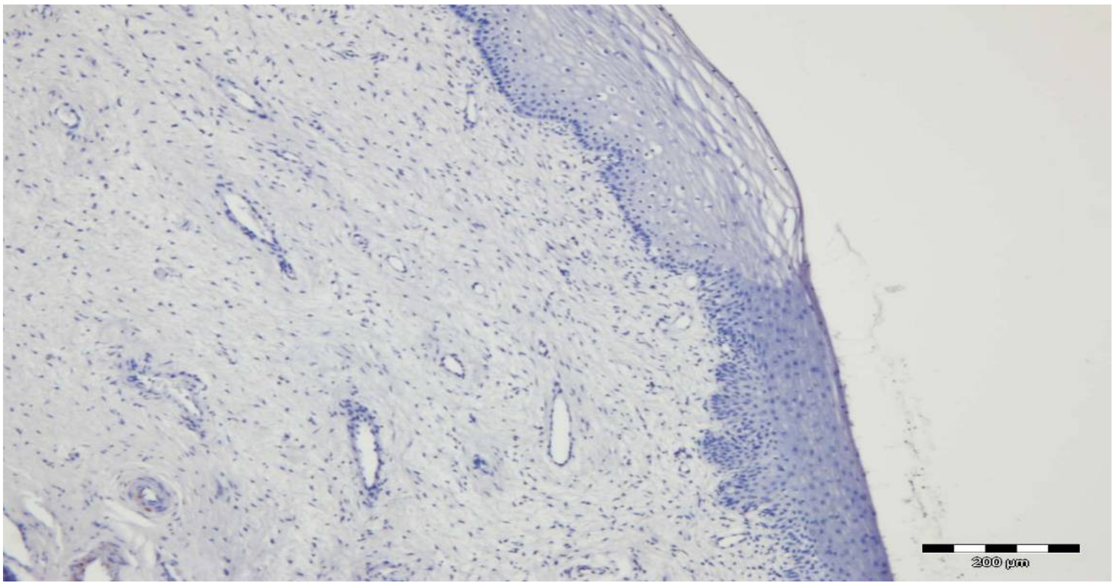


Figure 6.4 A Micrograph recorded with X10 objective showing negative immunohistochemical staining for p16 protein in CIN II marked tissue (in Figure 4.15) cervical tissue section.

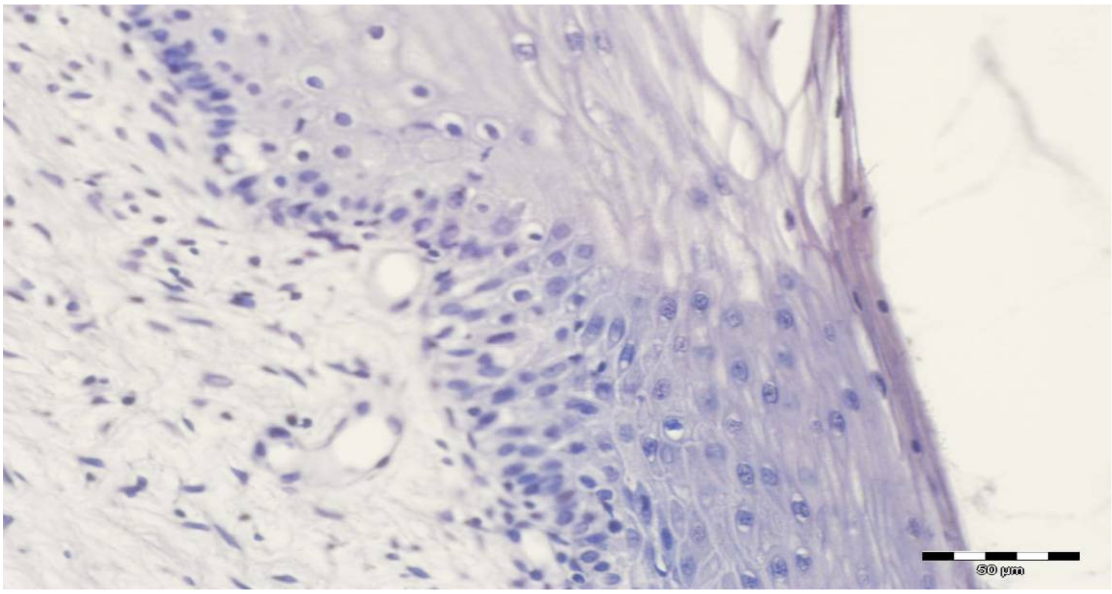


Figure 6.4 B Micrograph recorded with X40 objective showing negative immunohistochemical staining for p16 protein in CIN II marked tissue (in Figure 4.15) cervical tissue section.

Figure 6.5 A & B shows the immunohistochemical staining for p16 protein in a CIN II cervical tissue section recorded with X10 and X40 objective lenses respectively. The staining/brown colour is evident in some parts of the cervical epithelium. Here the results of the immunohistochemical staining for p16 protein are consistent with the Raman micro spectral mapping of the CIN II sample as Raman micro spectroscopy characterised it as intermediate sample indicated by the KMCA colour map, presented in chapter 4, **Figure 4.17**.

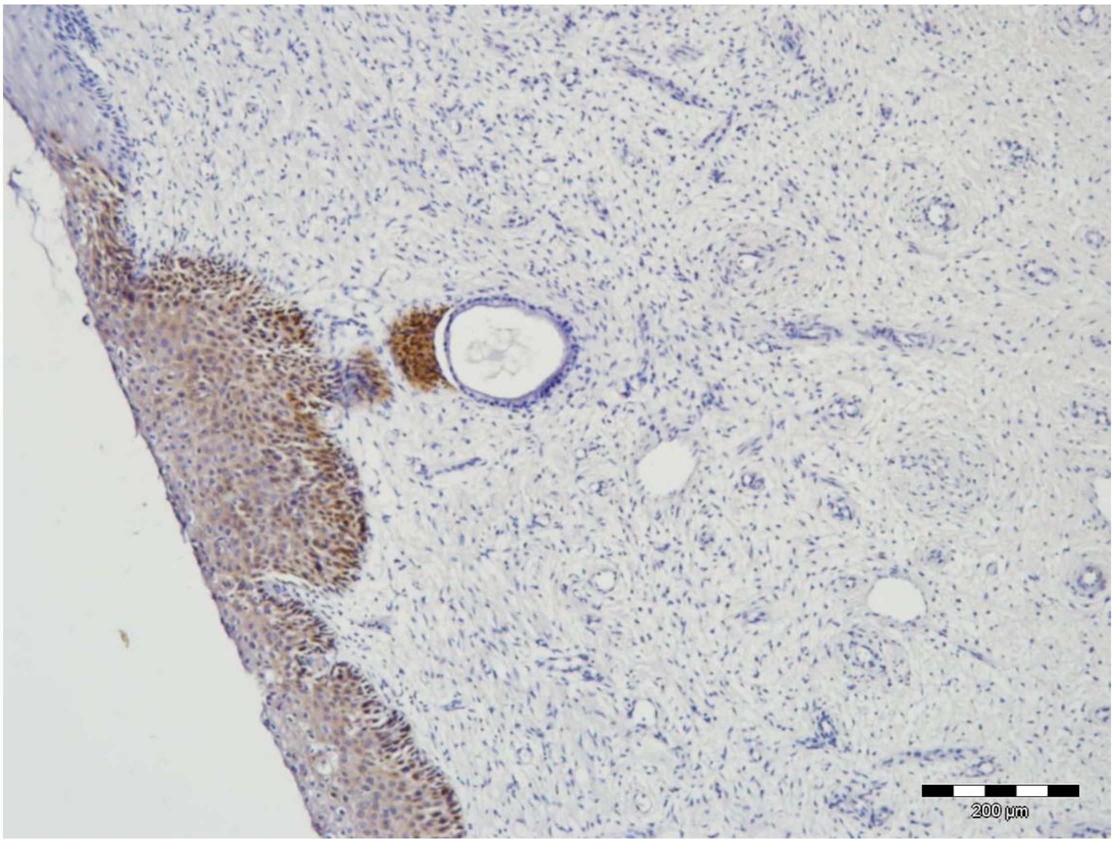


Figure 6.5 A Micrograph recorded with X10 objective showing positive immunohistochemical staining for p16 protein in CIN II marked tissue (chapter 4 reference; Figure 4.17) cervical tissue section.

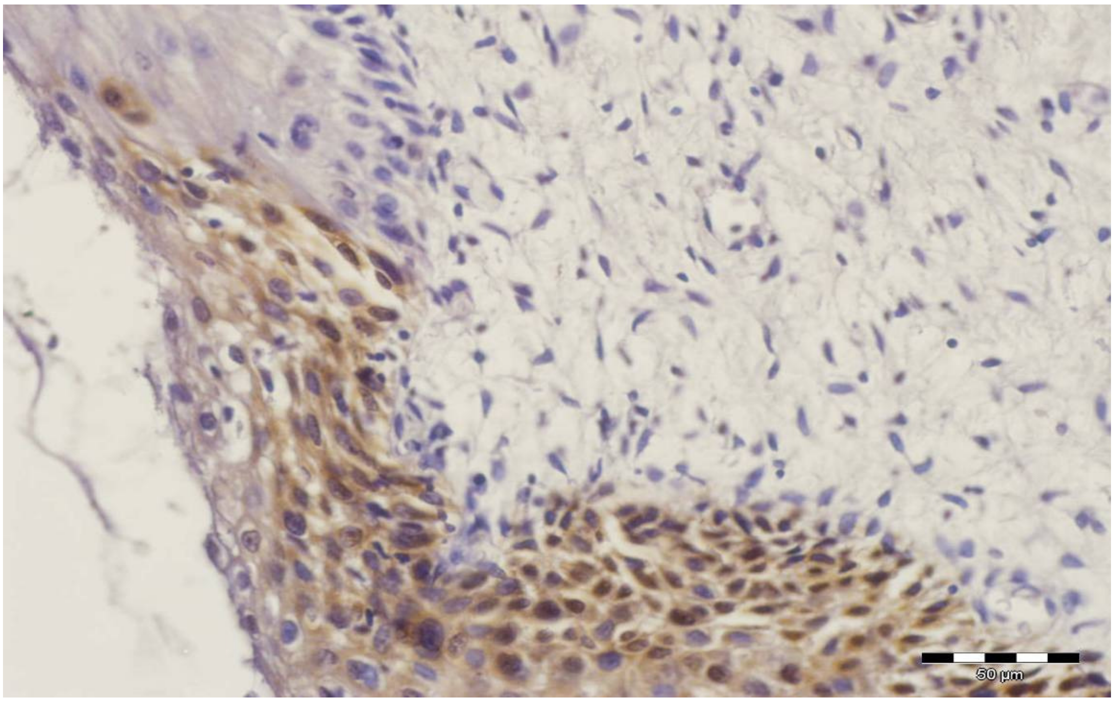


Figure 6.5 B Micrograph recorded with X40 objective showing positive immunohistochemical staining for p16 protein in CIN II marked (chapter 4 reference; Figure 4.17) cervical tissue section.

Figure 6.6 A & B shows the immunohistochemical staining for p16 protein in the normal side of the CIN III cervical tissue section recorded with X10 and X40 objective lenses respectively. No positive staining/brown colour was seen in the cervical epithelium. Here, the results of the immunohistochemical staining for p16 protein are not consistent with the Raman micro spectral mapping of the same sample as Raman micro spectroscopy characterised it as intermediate sample indicated by the KMCA colour map, presented in chapter 4, **Figure 4.13**. The Raman spectral mapping indicated that the separation of the stroma and basal layer is very clear but the superficial layer showed evidence of disease based on strong DNA features throughout this layer, indicating rapidly proliferating cells.

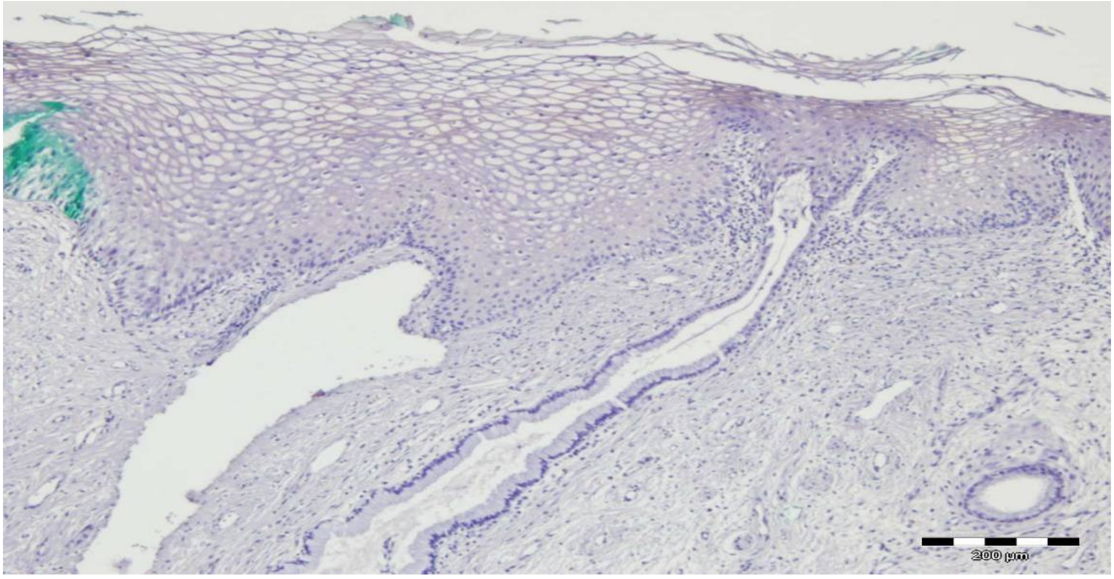


Figure 6.6 A Micrograph recorded with X10 objective showing negative immunohistochemical staining for p16 protein in normal side of the CIN III cervical tissue section (chapter 4 reference: fig 4.13).

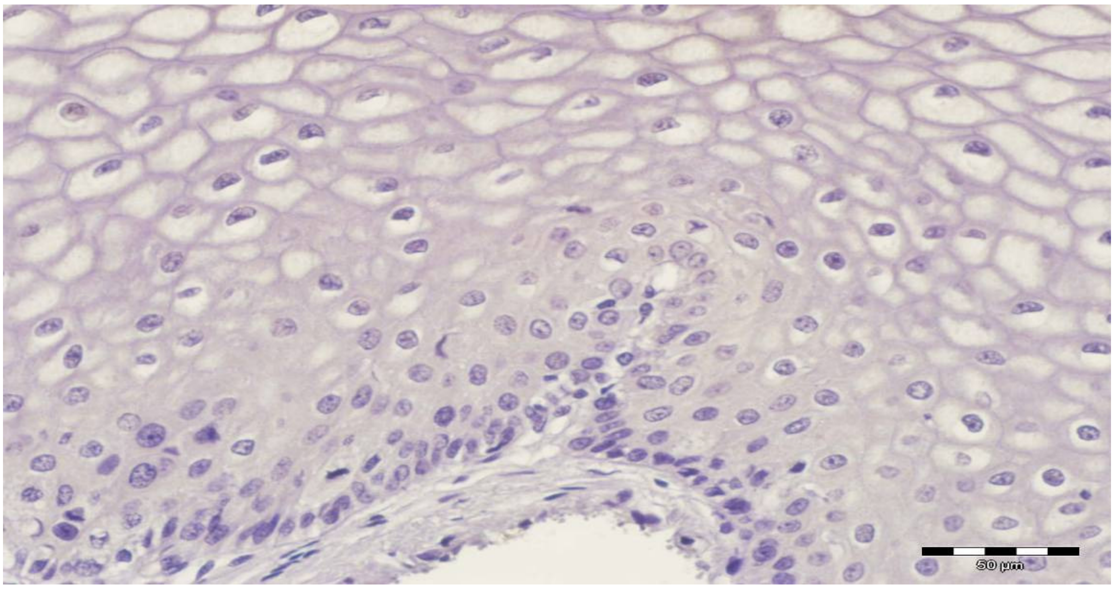


Figure 6.6 B Micrograph recorded with X40 objective showing negative immunohistochemical staining for p16 protein in normal side of the CIN III cervical tissue section (chapter 4 reference: Figure 4.13).

6.5 Conclusion

The immunohistochemical staining for p16 causes the development of a brown colour in the nucleus and cytoplasm of the epithelial cells of cervical tissue. In the current study, control samples including a negative control, CIN II diagnosed cervical tissue section which was not incubated with the primary antibody, did not show any expression detection or brown colour due to antibody-antigen attachment. This confirmed the specificity of the staining for the expression evaluation of p16 protein.

The CIN III cervical tissue section (chapter 4, **Figure 4.7**) , which was shown by the Raman analysis as the most abnormal sample where KMCA indicated no discrimination of the basal, stroma and superficial layers, has shown extensive staining of the cervical epithelial cells. Notably, however, there is no evidence of p16 expression in the stromal layer, consistent with the PCA analysis from **chapter 5** indicating that basal cells do not migrate into the stroma and changes in the Raman features associated with the progression of the disease are related to the cells of the stroma themselves.

The p16 staining of a CIN II cervical tissue section indicated no staining of the cervical epithelial cells. This sample corresponds to **Figure 4.15** chapter 4, whereby Raman characterised it as an intermediate sample as indicated by the KMCA map where discrimination of the basal, stroma and superficial layers was observed but the latter two layers were observed to have some characteristics of the basal layer. The negative immunohistochemical staining for p16 protein found here leads to the conclusion that the expression of this protein is not detectable. However, Raman micro spectroscopy could detect the biochemical changes taking place in this sample due to the progression of the cervical cancer at an earlier stage than the expression of p16 protein.

In another CIN II sample (chapter 4, **Figure 4.17**) which was also characterised as an intermediate sample by Raman micro spectroscopy as indicated by the KMCA colour map, the staining/brown colour is evident in some parts of the cervical epithelium and the results of the immunohistochemical staining for p16 protein are consistent with the Raman micro spectral mapping.

The staining for p16 protein in the normal side of the CIN III cervical tissue section (chapter 4, **Figure 4.13**) is found to be negative. Again, the results of the immunohistochemical staining for p16 protein are consistent with the Raman micro spectral mapping of the same sample as Raman micro spectroscopy characterised it as intermediate sample indicated by the KMCA colour map. The Raman spectral mapping indicated that the separation of the stroma and basal layer is very clear, but the superficial layer showed evidence of disease based on strong DNA features throughout this layer, indicating rapidly proliferating cells and early detection of the disease by Raman spectroscopy as compared to the immunostaining method.

Chapter 7

Summary and Conclusions

The major purpose of this study is to evaluate the potential of the technique of vibrational micro spectroscopy in elucidating the biochemical changes occurring during the progression of Cervical Intraepithelial Neoplasia (CIN) to cancer. The pathologist, in general, divides the cervical epithelium into basal, parabasal, intermediate and superficial layers. The cervical epithelium is supported by the underlying connective tissue known as the stroma. Common currently employed diagnostic techniques rely on morphological changes to the tissue of cellular structures, rather than the underlying biochemical changes. Microspectroscopy promises a rapid, label free, automated technique for disease diagnostics based on these underlying biochemical changes.

FTIR and Raman micro spectroscopy were employed to visualize the layer separation in cervical tissue sections and also how this separation is affected due to the biochemical changes caused by the progression of the disease. The comparison of the results of both the techniques was made to choose one for the further studies based on the biochemical information attained. Although, both Raman micro spectroscopy and FTIR spectroscopy were able to differentiate the normal region of the cervical tissue sample into three layers, including stroma, basal/para-basal and superficial layers, on the basis of the spectral features of the collagen, DNA bases and glycogen respectively, Raman spectroscopy due to its higher spatial and spectral resolution compared to FTIR spectroscopy revealed more detailed biochemical changes related to the DNA.

Before further exploring the chosen technique, Raman micro spectroscopy, for the analysis of the cervical tissue samples, different dewaxing agents including hexane, cyclohexane, xylene, histolene, ether and dish washing liquid were utilised to determine their ability to remove wax from rat skin tissue samples in order to optimize sample processing protocols for the cervical tissue samples. On the basis of the spectral quality obtained with dewaxed rat skin, xylene and hexane were found to be equally good for the removal of the wax. The dishwashing liquid resulted in the inefficient removal of wax and Raman spectra exhibited the paraffin bands even after the application of the dewaxing agent. Cyclohexane itself contributed to the Raman spectra as indicated by the bands present at 787 and 802 cm^{-1}. The use of ether did not appear to be suitable as a dewaxing agent due to its interference with the proteins in the tissue section as indicated by the observation of deformation of the amide I band. Xylene and hexane were found to be the best dewaxing solvents, followed by histolene, as compared to the other dewaxing agents employed in the current study. Xylene was adopted as solvent of choice to keep the protocol consistent with the one used in the hospital histopathology laboratory.

The selection of the most suitable laser line was made from three laser lines, including 532 nm, 660 nm and 785 nm. The Raman bands were better defined for the 785nm laser line as compared to the 532nm laser line, and a poorly defined spectrum was obtained for 660 nm laser line, which showed it to be the least suitable of the three for the current study. The 785nm laser line was preferred over the 532nm laser line due to the major concern of photo degradation of the tissue section.

For the selection of the most suitable substrate, three different substrates including Mirr-IR, quartz and calcium fluoride were used and results were compared. The Mirr-IR substrate gave rise to a broad background under 785nm laser illumination and as such was not considered suitable. The quartz and CaF_2 substrates did not cause fluorescence with the 785nm laser line. The CaF_2 contribution in the biological spectrum was negligible as compared to a prominent band from quartz which led to the selection of the CaF_2 as the most substrate to be used for the further experiments with Raman micro spectroscopy.

The optimization of the objective lens was done by employing X50 dry, X100 dry and X60, X100 water immersion objectives for the Raman spectral acquisition. Similar spectral features were obtained with X50 dry, X100 dry and X100 immersion objectives but the spectra acquired by the use of the X100 immersion objective were cleaner as compared to the X50 and X100 dry objectives and thus less spectral preprocessing was required for this objective. This is due to the reason that the immersion of the tissue section in the water makes the tissue surface less scattering and results in better defined peaks. However, tissue immersion does not readily allow for repetition of the experiment if needed for further investigations. Use of the dry objective avoids the problem of immersing the tissue sections in water repeatedly. So, from these two objective lenses, the X100 dry objective lens was selected for further experiments of Raman mapping.

After optimization of the above mentioned parameters for Raman micro spectroscopy, the technique was employed for spectral mapping of the cervical tissue sections. The results of the Raman micro spectroscopy have proven that the technique, in combination with K-Means Clustering Analysis, is very useful in the discrimination of the diseased areas from the normal areas and particularly in the detailed identification of the biochemical changes in the diseased samples. The differentiated clusters of Raman spectra contain the signatures of the DNA bases and can be used as a discriminating factor for the differentiation between the layers and between normal and abnormal regions. On moving from normal to abnormal regions of the cervical tissue sample, the characteristic Raman features of the basal layer were observed in the superficial layer and in the stromal layer. Notably, the region of the CIN III sample, which was pathologically classified as normal, has been recognised as having biochemical information similar to the abnormal region. This has been indicated by the diminution of the collagen Raman spectral bands in the stromal layer as well as absence or minute presence of the glycogen bands in the superficial layer.

The results, based on KMCA of Raman spectroscopic maps of different cervical tissue samples, led to the conclusion that Raman micro spectral mapping of the cervical tissue sections has been able to identify the biochemical changes occurring during disease progression. The information and conclusion made on the basis of the biochemical changes would not be possible only on the basis of the morphological observations/changes occurring in the tissue.

PCA of the Raman spectral data was performed and was found to be very useful to further interpret the results of the KMCA and to elucidate the basis of differentiation of the basal-true normal and basal-CIN II normal (far from diseased region), both of which have been found to be different to the basal layers of the diseased samples. This means that, in order to predict the presence of the disease, it is not necessary to investigate the region of disease directly and suggests that Raman micro spectroscopy can identify the onset of disease before the morphological changes to the tissue structure are observable.

The comparison of the PCA loadings of the HPV negative and positive cell lines (*C33A* and *CaSki*), with the PCA loadings of the basal-true normal versus basal-CIN II normal indicated that there were no strongly matching Raman signatures which could lead to the identification of signatures of HPV infection in the cervical cancer tissue samples. However, a feature associated with the amide-III beta sheet (1222 cm^{-1}) has been consistently observed in the PCA of the basal-true normal versus basal-CIN II normal which might be of interest, as it is the major difference between the two data sets reflected in the form of a loading by PCA, although it is not prominent in the loadings of the cell lines. This amide-III protein feature (1222 cm^{-1}) is also consistently present in the loadings of the PCA contributed by the basal layers of the intermediate and abnormal samples in the PCA with the basal-true normal. Thus, this feature may be considered as an early marker of disease progression and could potentially indicate HPV infection. Notably, this band is much reduced in intensity in the loadings contributed by the most extreme abnormal sample and the carcinoma *in situ* samples which are the most diseased samples. Referring to Figure 1.5 (Woodman et al., 2007), this feature may therefore be a signature of episomal HPV, in which the virus is present in the cytoplasm, surrounded by proteins and the reduction of the amide-III related feature as the disease progresses may be an indicator for integration of the HPV virus into the host genome. It is also notable that the HPV infected cervical cell lines are derived from cancer patients, and thus the HPV should exist in its full integrated form, and thus, if it is a marker for episomal HPV, the absence of the strong protein band at 1222cm^{-1} is to be expected. Further investigation is needed however to confirm this suggestion. It can be concluded that Raman spectroscopy certainly has potential for detecting the biochemical changes associated with HPV infection but as there are no strongly matching Raman signature loadings of the HPV negative and positive cell lines (*C33A* and *CaSki*), with the PCA loadings of the basal- true normal versus CIN II normal basal, further studies of spectral markers of HPV in tissue samples rather than cell lines are required.

As the KMCA (in chapter 4) suggested the migration of the characteristics of the basal layer into the superficial layer, Raman spectroscopy along with PCA have led to the conclusion that, during the progression of cervical cancer, cells of the basal layer are characterised by high DNA content and lack of glycogen content. These cells progress to the superficial layer during the progression of the disease. PCA confirms these findings made on the basis of the KMCA.

For the evaluation of the KMCA observation on the appearance of the basal cells/characteristics of the basal layer in the stromal layer, Raman spectroscopy along with the PCA, have led to the conclusion that these characteristics are actually associated with the biochemical changes in the cells of the stromal layer and there is no migration of the basal cells into the stroma. In addition, during cervical cancer progression, relative to the DNA, collagen has a diminished contribution at some points in the Raman map of the stroma and KMCA recognises the greater similarities with the DNA rich cells of the basal layer. This has been supported by the observation of the collagen related loadings contributed by the stroma-4 and stroma-6 (chapter 5) indicating that there is still some collagen present, although the DNA features are increased as compared to the normal stroma. Raman spectroscopy has identified significant changes in the stroma which may be useful for diagnostics and also can be important for tracking disease progression and understanding mechanisms for treatment of medium stage progression.

In order to further confirm the findings from **chapter 5**, including the classification of the cervical tissue samples and the progression of the basal cells into the superficial layer but not into the stroma, immunohistochemical staining for the evaluation of the expression of p16 protein was performed. The expression of p16 is a biomarker for the presence of HPV and is expected to be enhanced in the basal and superficial cells/layers rather than in the stroma/stromal cells. The immunohistochemical staining for p16 causes the development of a brown colour in the nucleus and cytoplasm of the epithelial cells of cervical tissue. For this purpose, control samples, including a negative control, CIN II diagnosed cervical tissue section which was not incubated with the primary antibody, did not show any expression or brown colour due to antibody-antigen attachment. This confirmed the specificity of the staining for the expression evaluation of p16 protein.

The CIN III cervical tissue section (chapter 4, **Figure 4.7**), which was shown by the Raman analysis as the most abnormal sample, whereby KMCA indicated no discrimination of the basal, stroma and superficial layers, showed extensive staining of the cervical epithelial cells. Notably, however, there is no evidence of p16 expression in the stromal layer, consistent with the PCA analysis from **chapter 5,** indicating that basal cells do not migrate into the stroma and changes in the Raman features associated with the progression of the disease are related to the cells of the stroma themselves.

The p16 staining of a CIN II cervical tissue section indicated no staining of the cervical epithelial cells. This sample corresponds to **Figure 4.15** in chapter 4, whereby Raman analysis characterised it as an intermediate sample as indicated by the KMCA map where discrimination of the basal, stroma and superficial layers was observed, but the latter two layers were observed to have some characteristics of the basal layer. The negative immunohistochemical staining for p16 protein found here leads to the conclusion that the expression of this protein, if present, is not detectable. However, Raman micro spectroscopy could detect the biochemical changes taking place in this sample due to the progression of the cervical cancer at an earlier stage than the expression of p16 protein.

In another CIN II sample (chapter 4, **Figure 4.17**) which was also characterised as an intermediate sample by Raman micro spectroscopy, as indicated by the KMCA colour map, the staining/brown colour is evident in some parts of the cervical epithelium and the results of the immunohistochemical staining for p16 protein are consistent with the Raman spectral mapping.

The staining for p16 protein in the normal side of the CIN III cervical tissue section (chapter 4, **Figure 4.13**) was found to be negative. The results of the immunohistochemical staining for p16 protein are not consistent with the Raman micro spectral mapping of the same sample as Raman micro spectroscopy characterised it as intermediate sample indicated by the KMCA colour map. The Raman spectral mapping indicated that the separation of the stroma and basal layer is very clear, but the superficial layer showed evidence of disease based on strong DNA features throughout this layer, indicating rapidly proliferating cells and early detection of the disease by Raman spectroscopy as compared to the immunohistochemical method.

Raman spectroscopy has been found to be a powerful technique to investigate the biochemical changes associated with progression of cervical cancer from normal to abnormal stages and carcinoma *in situ* and has shown the potential to produce clinically important data for the characterisation of the disease. Raman spectral mapping of the cervical tissue samples of various stages of the disease along with the multivariate data analysis techniques including KMCA and PCA have been shown to be very promising for the discrimination of these cervical tissue samples and extracting Raman markers associated with the biochemical changes occurring in different regions of the tissue including basal, superficial and stromal layers. In addition, different sub-regions of these layers of the cervical tissue have also been characterised and differentiated on the basis of the changes in the spectral features observed which has proved the diagnostic capabilities of the Raman microspectral imaging with respect to cervical cancer and early markers of the disease.

The technique promises a "Raman based classification system" which depends on the identification of markers associated with early stage biochemical changes rather than the morphological ones which appear at a later stage of disease progression. Moreover, the technique is objective, no prior labeling is required, it can be cost effective and can be automated. To bring the technique to the clinical environment, it is essential to make the technique acceptable to the clinicians by making it easy to use for them, affordable and adaptable from laboratory to clinic by excluding the requirement of personal trained for acquisition of the Raman spectra and implementation of complex algorithms for data analysis. These improvements will enable the technique to produce final results in a single visit to the “Raman clinic” and reduce the workload on the clinicians as well as the anxiety of the patient in relation to the outcome of the results associated with long waits and repeated visits, a particularly significant big problem in third world countries where patients are reluctant to follow up initial visits.

References

AGGARWAL, P., BATRA, S., GANDHI, G. & ZUTSHI, V. (2010) Comparison of Papanicolaou Test With Visual Detection Tests in Screening for Cervical Cancer and Developing the Optimal Strategy for Low Resource Settings. *International Journal of Gynecological Cancer,* 20, 862-868.

AGORASTOS, T., MILIARAS, D., , A. F. L., CHRISAFI, S., KOTSIS, A., MANTHOS, A. & , J. B. (2005) Detection and typing of human papillomavirus DNA in uterine cervices with coexistent grade I and grade III intraepithelialneoplasia: biologic progression or independent lesions? *European Journal of Obstetrics & Gynecology and Reproductive Biology* 121, 99–103.

AGRAWAL, A., UTZINGER, U., BROOKNER, C., PITRIS, C., MITCHELL, M. F. & RICHARDS-KORTUM, R. (1999) Fluorescence spectroscopy of the cervix: Influence of acetic acid, cervical mucus, and vaginal medications. *Lasers in Surgery and Medicine,* 25, 237-249.

ANDERSON, J. R., PARAMSOTHY, P., HEILIG, C., JAMIESON, D. J., SHAH, K. & DUERR, A. (2006) Accuracy of Papanicolaou test among HIV-infected women. *Clinical Infectious Diseases,* 42, 562-568.

ANDRUS, P. G. L. & STRICKLAND, R. D. (1998) Cancer grading by Fourier transform infrared spectroscopy. *Biospectroscopy,* 4, 37-46.

APGAR, B. S., ZOSCHNICK, L. & WRIGHT, T. C. (2003) The 2001 Bethesda System terminology. *American Family Physician,* 68, 1992-1998.

BALL, D. W. (2001) Theory of Raman Spectroscopy. *SPECTROSCOPY,* 16.

BARTON, A. F. M. (1983) *Handbook of Solubility Parameters end Other Cohesion Parameters.*, Boca Raton, Florida: CRC Press, Inc.

BAZANT-HEGEMARK, F., EDEY, K., SWINGLER, G. R., READ, M. D. & STONE, N. (2008) Optical Micrometer Resolution Scanning for Non-invasive Grading of Precancer in the Human Uterine Cervix. *Technology in Cancer Research & Treatment,* 7, 483-496.

BLOUT, E. R. & MELLORS, R. C. (1949) Infrared Spectra of Tissues. *Science,* 110, 137-138.

BONNIER, F. & BYRNE, H. J. (2012) Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems. *Analyst,* 137, 322-332.

BONNIER, F., KNIEF, P., LIM, B., MEADE, A. D., DORNEY, J., BHATTACHARYA, K., LYNG, F. M. & BYRNE, H. J. (2010a) Imaging live cells grown on a three dimensional collagen matrix using Raman microspectroscopy. *Analyst,* 135, 3169-3177.

BONNIER, F., MEHMOOD, A., KNIEF, P., MEADE, A. D., HORNEBECK, W., LAMBKIN, H., FLYNN, K., MCDONAGH, V., HEALY, C., LEE, T. C., LYNG, F. M. & BYRNE, H. J. (2010b) In vitro analysis of immersed human tissues by Raman microspectroscopy. *Journal of Raman Spectroscopy,* 42, 888-896.

BONNIER, F., MEHMOOD, A., KNIEF, P., MEADE, A. D., HORNEBECK, W., LAMBKIN, H., FLYNN, K., MCDONAGH, V., HEALY, C., LEE, T. C., LYNG, F. M. & BYRNE, H. J. (2011) In vitro analysis of immersed human tissues by Raman microspectroscopy. *Journal of Raman Spectroscopy,* 42, 888-896.

BOPPART, S. A., BREZINSKI, M. E., PITRIS, C. & FUJIMOTO, J. G. (1998) Optical coherence tomography for neurosurgical imaging of human intracortical melanoma. *Neurosurgery,* 43, 834-841.

BOSCH, F. X., MANOS, M. M., MUNOZ, N., SHERMAN, M., JANSEN, A. M., PETO, J., SCHIFFMAN, M. H., MORENO, V., KURMAN, R., SHAH, K. V., ALIHONOU, E., BAYO, S., MOKHTAR, H. C., CHICAREON, S., DAUDT, A., DELOSRIOS, E., GHADIRIAN, P., KITINYA, J. N., KOULIBALY, M., NGELANGEL, C., TINTORE, L. M. P., RIOSDALENZ, J. L., SARJADI, SCHNEIDER, A., TAFUR, L., TEYSSIE, A. R., ROLON, P. A., TORROELLA, M., TAPIA, A. V., WABINGA, H. R., ZATONSKI, W., SYLLA, B., VIZCAINO, P., MAGNIN, D., KALDOR, J., GREER, C. & WHEELER, C. (1995) Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *Journal of the National Cancer Institute,* 87, 796-802.

BOSHART, M., GISSMANN, L., IKENBERG, H., KLEINHEINZ, A., SCHEURLEN, W. & HAUSEN, H. Z. (1984) A NEW TYPE OF PAPILLOMAVIRUS DNA, ITS PRESENCE IN GENITAL CANCER BIOPSIES AND IN CELL-LINES DERIVED FROM CERVICAL-CANCER. *Embo Journal,* 3, 1151-1157.

BOYKO, E. J. (1996) Meta-analysis of Pap test accuracy. *American Journal of Epidemiology,* 143, 406-407.

BOYLE, W. & SMITH, G. (1970) Charge coupled devices. *Bell Syst. Tech. J.,,* 49, 587-593.

BROOKNER, C. K., UTZINGER, U., STAERKEL, G., RICHARDS-KORTUM, R. & MITCHELL, M. F. (1999) Cervical fluorescence of normal women. *Lasers in Surgery and Medicine,* 24, 29-37.

BURKE, J. (1984) *Solubility parameters: Theory and applications.*, The Book and Paper group Annual.

BYRNE, H. J., D.SOCKALINGUM, G. & STONE, N. (2011) *Biomedical Applications of Synchrotron Infrared Microspectroscopy.*

CARLSON, K., PAVLOVA, I., COLLIER, T., DESCOUR, M., FOLLEN, M. & RICHARDS-KORTUM, R. (2005) Confocal microscopy: Imaging cervical precancerous lesions. *Gynecologic Oncology,* 99, S84-S88.

CARTER, E. A. & EDWARD, H. G. M. (2001) Biological Applications of Raman Spectroscopy. IN GREMLICH, H.-U. & YAN, B. (Eds.) *Infrared and raman spectroscopy of biological materials.* Marcel Dekker.

CHAN, J. W., TAYLOR, D. S., ZWERDLING, T., LANE, S. M., IHARA, K. & HUSER, T. (2006) Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophysical Journal,* 90, 648-656.

CHIDANANDA, S. M., SATYAMOORTHY, K., RAI, L., MANJUNATH, A. P. & KARTHA, V. B. (2006) Optical diagnosis of cervical cancer by fluorescence spectroscopy technique. *International Journal of Cancer,* 119, 139-145.

CHOI, J. W., AHN, W. S., BAE, S. M., LEE, D. B. & KIM, Y. W. (2005) Adenoviral p53 effects and cell-specific E7 protein-protein interactions of human cervical cancer cells. *Biosensors & Bioelectronics,* 20, 2236-2243.

CHOWDHURY, T. R. & CHOWDHURY, J. R. (1981) Significance of the occurence and distribution of glycogen in cervical cells exfoliated under different physiologic and pathologic conditions. *Acta Cytologica,* 25, 557-565.

CLAD, A., REUSCHENBACH, M., WEINSCHENK, J., GROTE, R., RAHMSDORF, J. & FREUDENBERG, N. (2011) Performance of the Aptima High-Risk Human Papillomavirus mRNA Assay in a Referral Population in Comparison with Hybrid Capture 2 and Cytology. *Journal of Clinical Microbiology,* 49, 1071-1076.

CLARK, A. L., GILLENWATER, A., ALIZADEH-NADERI, R., EL-NAGGAR, A. K. & RICHARDS-KORTUM, R. (2004) Detection and diagnosis of oral neoplasia with an optical coherence microscope. *Journal of Biomedical Optics,* 9, 1271-1280.

COHENFORD, M. A. & RIGAS, B. (1998) Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: Implications for tumor biology. *Proceedings of the National Academy of Sciences of the United States of America,* 95, 15327-15332.

COLLINS, S., ROLLASON, T. P., YOUNG, L. S. & WOODMAN, C. B. J. (2010) Cigarette smoking is an independent risk factor for cervical intraepithelial neoplasia in young women: A longitudinal study. *European Journal of Cancer,* 46, 405-411.

CREMER, M., BULLARD, K., MAZA, M., PERALTA, E., MOORE, E., GARCIA, L., MASCH, R., LERNER, V., ALONZO, T. A. & FELIX, J. (2010) Cytology versus visual inspection with acetic acid among women treated previously with cryotherapy in a low-resource setting. *International Journal of Gynecology & Obstetrics,* 111, 249-252.

CUZICK, J., ARBYN, M., SANKARANARAYANAN, R., TSU, V., RONCO, G., MAYRAND, M. H., DILLNER, J. & MEIJER, C. (2008) Overview of Human Papillomavirus-Based and Other Novel Options for Cervical Cancer Screening in Developed and Developing Countries. *Vaccine,* 26, K29-K41.

DENG, J. L., WEI, Q., ZHANG, M. H., WANG, Y. Z. & LI, Y. Q. (2005) Study of the effect of alcohol on single human red blood cells using near-infrared laser tweezers Raman spectroscopy. *Journal of Raman Spectroscopy,* 36, 257-261.

DIEM, M., MILJKOVIC, M., BIRD, B., CHERNENKO, T., SCHUBERT, J., MARCSISIN, E., MAZUR, A., KINGSTON, E., ZUSER, E., PAPAMARKAKIS, K. & LAVER, N. (2012)

Applications of Infrared and Raman Microspectroscopy of Cells and Tissue in Medical Diagnostics: Present Status and Future Promises. *Spectroscopy-an International Journal,* 27, 463-496.

DOCKTER, J., SCHRODER, A., HILL, C., GUZENSKI, L., MONSONEGO, J. & GIACHETTI, C. (2009) Clinical performance of the APTIMAÂ® HPV Assay for the detection of high-risk HPV and high-grade cervical lesions. *Journal of Clinical Virology,* 45, Supplement 1, S55-S61.

DOVBESHKO, G. I., GRIDINA, N. Y., KRUGLOVA, E. B. & PASHCHUK, O. P. (2000) FTIR spectroscopy studies of nucleic acid damage. *Talanta,* 53, 233-246.

DUFRESNE, S., SAUTHIER, P., MAYRAND, M. H., PETIGNAT, P., PROVENCHER, D., DROUIN, P., GAUTHIER, P., DUPUIS, M. J., MICHON, B., OUELLET, S., HADJERES, R., FERENCZY, A., FRANCO, E. L. & COUTLEE, F. (2011) Human Papillomavirus (HPV) DNA Triage of Women with Atypical Squamous Cells of Undetermined Significance with Amplicor HPV and Hybrid Capture 2 Assays for Detection of High-Grade Lesions of the Uterine Cervix. *Journal of Clinical Microbiology,* 49, 48-53.

DURST, M., GISSMANN, L., IKENBERG, H. & ZURHAUSEN, H. (1983) A PAPILLOMAVIRUS DNA FROM A CERVICAL-CARCINOMA AND ITS PREVALENCE IN CANCER BIOPSY SAMPLES FROM DIFFERENT GEOGRAPHIC REGIONS. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 80, 3812-3815.

DYSON, N., HOWLEY, P. M., MUNGER, K. & HARLOW, E. (1989) THE HUMAN PAPILLOMA VIRUS-16 E7-ONCOPROTEIN IS ABLE TO BIND TO THE RETINOBLASTOMA GENE-PRODUCT. *Science,* 243, 934-937.

EKALAKSANANAN, T., PIENTONG, C., THINKHAMROP, J., KONGYINGYOES, B., EVANS, M. F. & CHAIWONGKOT, A. (2010) Cervical cancer screening in north east Thailand using the visual inspection with acetic acid (VIA) test and its relationship to high-risk human papillomavirus (HR-HPV) status. *Journal of Obstetrics and Gynaecology Research,* 36, 1037-1043.

ESCOBAR, P. F., BELINSON, J. L., WHITE, A., SHAKHOVA, N. M., FELDCHTEIN, F. I., KARETA, M. V. & GLADKOVA, N. D. (2004) Diagnostic efficacy of optical coherence tomography in the management of preinvasive and invasive cancer of uterine cervix and vulva. *Int J Gynecol Cancer,* 14, 470-4.

ESCOBAR, P. F., ROJAS-ESPAILLAT, L. & BELINSON, J. L. (2005) Optical diagnosis of cervical dysplasia. *Int J Gynaecol Obstet,* 89, 63-4.

FABIAN, H., JACKSON, M., MURPHY, L., WATSON, P. H., FICHTNER, I. & MANTSCH, H. H. (1995) A comparative Infrared spectroscopic study of human breast-tumors and breast-tumor cell xenografts. *Biospectroscopy,* 1, 37-45.

FAOLAIN, E. O., HUNTER, M. B., BYRNE, J. M., KELEHAN, P., LAMBKIN, H. A., BYRNE, H. J. & LYNG, F. M. (2005) Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. *Journal of Histochemistry & Cytochemistry,* 53, 121-129.

FARQUHARSON, S., GIFT, A. D., SHENDE, C., MAKSYMIUK, P., INSCORE, F. E. & MURRAN, J. (2005a) Detection of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy. *Vibrational Spectroscopy,* 38, 79-84.

FARQUHARSON, S., SHENDE, C., INSCORE, F. E., MAKSYMIUK, P. & GIFT, A. (2005b) Analysis of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy. *Journal of Raman Spectroscopy,* 36, 208-212.

FELDCHTEIN, F., GELIKONOV, G., GELIKONOV, V., KURANOV, R., SERGEEV, A., GLADKOVA, N., SHAKHOV, A., SHAKHOVA, N., SNOPOVA, L., TERENT'EVA, A., ZAGAINOVA, E., CHUMAKOV, Y. & KUZNETZOVA, I. (1998) Endoscopic applications of optical coherence tomography. *Opt Express,* 3, 257-70.

FRANK, C. J., MCCREERY, R. L. & REDD, D. C. B. (1995) RAMAN-SPECTROSCOPY OF NORMAL AND DISEASED HUMAN BREAST TISSUES. *Analytical Chemistry,* 67, 777-783.

FUJIMOTO, J. G., PITRIS, C., BOPPART, S. A. & BREZINSKI, M. E. (2000) Optical coherence tomography: An emerging technology for biomedical imaging and optical biopsy. *Neoplasia,* 2, 9-25.

GAFFIKIN, L., BLUMENTHAL, P. D., EMERSON, M., LIMPAPHAYOM, K., LUMBIGANON, P., RINGERS, P., SRISUPUNDIT, S., WARAKAMIN, S., LEWIS, R., CHUMWORATHAYEE, B., KANAVACHARAKUL, S. & GRP, R. J. C. C. C. (2003) Safety, acceptability, and feasibility of a single-visit approach to cervical-cancer prevention in rural Thailand: a demonstration project. *Lancet,* 361, 814-820.

GALLWAS, J., TURK, L., FRIESE, K. & DANNECKER, C. (2010) Optical coherence tomography as a non-invasive imaging technique for preinvasive and invasive neoplasia of the uterine cervix. *Ultrasound in Obstetrics & Gynecology,* 36, 624-629.

GEORGAKOUDI, I., SHEETS, E. E., MULLER, M. G., BACKMAN, V., CRUM, C. P., BADIZADEGAN, K., DASARI, R. R. & FELD, M. S. (2002) Trimodal spectroscopy for the detection and characterization of cervical precancers in vivo. *Am J Obstet Gynecol,* 186, 374-82.

GREMLICH, H. U. & YAN, B. (2001) *Infrared and raman spectroscopy of biological materials*, Marcel Dekker.

GROSSMAN, N., ILOVITZ, E., CHAIMS, O., SALMAN, A., JAGANNATHAN, R., MARK, S., COHEN, B., GOPAS, J. & MORDECHAI, S. (2001) Fluorescence spectroscopy for detection of malignancy: H-ras overexpressing fibroblasts as a model. *Journal of Biochemical and Biophysical Methods,* 50, 53-63.

HAGGITT, R. (1994) Barrett's esophagus, dysplasia, and adenocarcinoma. *Hum Pathol.* , 25, 982-93.

HAUSEN, H. Z., MEINHOF, W., SCHEIBER, W. & BORNKAMM, G. W. (1974a) ATTEMPTS TO DETECT VIRUS-SPECIFIC DNA IN HUMAN TUMORS .1. NUCLEIC-ACID HYBRIDIZATIONS WITH COMPLEMENTARY RNA OF HUMAN WART VIRUS. *International Journal of Cancer,* 13, 650-656.

HAUSEN, H. Z., SCHULTEH.H, WOLF, H., DORRIES, K. & EGGER, H. (1974b) ATTEMPTS TO DETECT VIRUS-SPECIFIC DNA IN HUMAN TUMORS .2. NUCLEIC-ACID HYBRIDIZATIONS WITH COMPLEMENTARY RNA OF HUMAN HERPES GROUP VIRUSES. *International Journal of Cancer,* 13, 657-664.

HOLLAS (2004).

HOSHIKAWA, S., SANO, T., YOSHIDA, T., ITO, H., OYAMA, T. & FUKUDA, T. (2010) Immunohistological analysis of HPV L1 capsid protein and p16 protein in low-grade dysplastic lesions of the uterine cervix. *Pathology Research and Practice,* 206, 816-820.

HUANG, Z., MCWILLIAMS, A., LUI, H., MCLEAN, D. I., LAM, S. & ZENG, H. (2003a) Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int J Cancer,* 107, 1047-52.

HUANG, Z. W., MCWILLIAMS, A., LAM, S., ENGLISH, J., MCLEAN, D. I., LUI, H. & ZENG, H. (2003b) Effect of formalin fixation on the near-infrared Raman spectroscopy of normal and cancerous human bronchial tissues. *International Journal of Oncology,* 23, 649-655.

HUBMANN, M. R., LEINER, M. J. P. & SCHAUR, R. J. (1990) ULTRAVIOLET FLUORESCENCE OF HUMAN SERA .1. SOURCES OF CHARACTERISTIC DIFFERENCES IN THE ULTRAVIOLET FLUORESCENCE-SPECTRA OF SERA FROM NORMAL AND CANCER-BEARING HUMANS. *Clinical Chemistry,* 36, 1880-1883.

HULEIHEL, M., SALMAN, A., ERUKHIMOVITCH, V., RAMESH, J., HAMMODY, Z. & MORDECHAI, S. (2002) Novel spectral method for the study of viral carcinogenesis in vitro. *Journal of Biochemical and Biophysical Methods,* 50, 111-121.

HUTCHINGS, J., KENDALL, C., SMITH, B., SHEPHERD, N., BARR, H. & STONE, N. (2009) The potential for histological screening using a combination of rapid Raman mapping and principal component analysis. *journal of biophotonics,* 2, 91-103.

INSINGA, R. P., DASBACH, E. J. & ELBASHA, E. H. (2009) Epidemiologic natural history and clinical management of Human Papillomavirus (HPV) Disease: a critical and systematic review of the literature in the development of an HPV dynamic transmission model. *BMC Infect Dis,* 9, 119.

JASTREBOFF, A. M. & CYMET, T. (2002) Role of the human papilloma virus in the development of cervical intraepithelial neoplasia and malignancy. *Postgraduate Medical Journal,* 78, 225-228.

JEDPIYAWONGSE, A., HOMCHA-EM, P., KARALAK, A. & SRIVATANAKUL, P. (2008) Immunohistochemical Overexpression of p16 Protein Associated with Cervical Cancer in Thailand. *Asian Pacific Journal of Cancer Prevention,* 9, 625-630.

JEMAL, A., BRAY, F., CENTER, M. M., FERLAY, J., WARD, E. & FORMAN, D. (2011) Global Cancer Statistics. *Ca-a Cancer Journal for Clinicians,* 61, 69-90.

JESS, P. R. T., SMITH, D. D. W., MAZILU, M., DHOLAKIA, K., RICHES, A. C. & HERRINGTON, C. S. (2007) Early detection of cervical neoplasia by Raman spectroscopy. *International Journal of Cancer,* 121, 2723-2728.

JUNQUEIRA, L. C. (1992) *Basic Histology.*

KACHI, S., HIRANO, K., TAKESUE, Y. & MIURA, M. (2000) Unusual corneal deposit after the topical use of cyclosporine as eyedrops. *American Journal of Ophthalmology,* 130, 667-669.

KAMINAKA, S., YAMAZAKI, H., ITO, T., KOHDA, E. & HAMAGUCHI, H. O. (2001) Near-infrared Raman spectroscopy of human lung tissues: possibility of molecular-level cancer diagnosis. *Journal of Raman Spectroscopy,* 32, 139-141.

KANTER, E. M., MAJUMDER, S., VARGIS, E., ROBICHAUX-VIEHOEVER, A., KANTER, G. J., SHAPPELL, H., JONES, H. W. & MAHADEVAN-JANSEN, A. (2009a) Multiclass discrimination of cervical precancers using Raman spectroscopy. *Journal of Raman Spectroscopy,* 40, 205-211.

KANTER, E. M., VARGIS, E., MAJUMDER, S., KELLER, M. D., WOESTE, E., RAO, G. G. & MAHADEVAN-JANSEN, A. (2009b) Application of Raman spectroscopy for cervical dysplasia diagnosis. *Journal of Biophotonics,* 2, 81-90.

KAUFMAN, R. H. & ADAM, E. (1999) Is human papillomavirus testing of value in clinical practice? *American Journal of Obstetrics and Gynecology,* 180, 1049-1053.

KELLER, M. D., KANTER, E. M., LIEBER, C. A., MAJUMDER, S. K., HUTCHINGS, J., ELLIS, D. L., BEAVEN, R. B., STONE, N. & MAHADEVAN-JANSEN, A. (2008) Detecting temporal and spatial effects of epithelial cancers with Raman spectroscopy. *Disease Markers,* 25, 323-337.

KENDALL, C., ISABELLE, M., BAZANT-HEGEMARK, F., HUTCHINGS, J., ORR, L., BABRAH, J., BAKER, R. & STONE, N. (2009) Vibrational spectroscopy: a clinical tool for cancer diagnostics. *Analyst,* 134, 1029-1045.

KINNEY, W., STOLER, M. H. & CASTLE, P. E. (2010) Special Commentary Patient Safety and the Next Generation of HPV DNA Tests. *American Journal of Clinical Pathology,* 134, 193-199.

KNIEF, P. (2010a) Interactions of Carbon Nanotubes with Biological Systems assessed by Raman Spectroscopy. Dublin Institute of Technology.

KNIEF, P. (2010b) Interactions of carbon Nanotubes with human lung epithelial cells *in vitro*, assesed by Raman spectroscopy., Dublin Institute of Technology.

KNIEFE, P. (2010).

KOLLIAS, N., GILLIES, R., MORAN, M., KOCHEVAR, I. E. & ANDERSON, R. R. (1998) Endogenous skin fluorescence includes bands that may serve as quantitative markers of aging and photoaging. *Journal of Investigative Dermatology,* 111, 776-780.

KOSS, L. G. (1992) *Diagnostic Cytology and its Histopathology.*, (Lipincott, Philadelphia), 4th Ed.

KRAFFT, C., CODRICH, D., PELIZZO, G. & SERGO, V. (2008a) Raman and FTIR imaging of lung tissue: Methodology for control samples. *Vibrational Spectroscopy,* 46, 141-149.

KRAFFT, C. & SERGO, V. (2006) Biomedical applications of Raman and infrared spectroscopy to diagnose tissues. *Spectroscopy-an International Journal,* 20, 195-218.

KRAFFT, C., STEINER, G., BELEITES, C. & SALZER1, R. (2008b) Disease recognition by infraredand Raman spectroscopy. *Journal of Biophotonics.*

KRISHNA, C. M., PRATHIMA, N. B., MALINI, R., VADHIRAJA, B. M., BHATT, R. A., FERNANDES, D. J., KUSHTAGI, P., VIDYASAGAR, M. S. & KARTHA, V. B. (2006) Raman spectroscopy studies for diagnosis of cancers in human uterine cervix. *Vibrational Spectroscopy,* 41, 136-141.

KRISHNA, C. M., SOCKALINGUM, G. D., KEGELAER, G., RUBIN, S., KARTHA, V. B. & MANFAIT, M. (2005) Micro-Raman spectroscopy of mixed cancer cell populations. *Vibrational Spectroscopy,* 38, 95-100.

KUHNERT, N. & THUMSER, A. (2004) An investigation into the use of Raman microscopy for the detection of labelled compounds in living human cells. *Journal of Labelled Compounds & Radiopharmaceuticals,* 47, 493-500.

LATIMER, P. (1967) ABSOLUTE ABSORPTION AND SCATTERING SPECTROPHOTOMETRY. *Archives of Biochemistry and Biophysics,* 119, 580-&.

LIU, C. H., DAS, B. B., GLASSMAN, W. L. S., TANG, G. C., YOO, K. M., ZHU, H. R., AKINS, D. L., LUBICZ, S. S., CLEARY, J., PRUDENTE, R., CELMER, E., CARON, A. & ALFANO, R. R. (1992) RAMAN, FLUORESCENCE, AND TIME-RESOLVED LIGHT-

SCATTERING AS OPTICAL DIAGNOSTIC-TECHNIQUES TO SEPARATE DISEASED AND NORMAL BIOMEDICAL MEDIA. *Journal of Photochemistry and Photobiology B-Biology,* 16, 187-209.

LIU, G., LIU, J. H., ZHANG, L., YU, F. & SUN, S. Z. (2005) Raman spectroscopic study of human tissues. *Spectroscopy and Spectral Analysis,* 25, 723-725.

LONGWORTH, M. S. & LAIMINS, L. A. (2004) Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiology and Molecular Biology Reviews,* 68, 362-+.

LORD, R. C. & YU, N. T. (1970) LASER-EXCITED RAMAN SPECTROSCOPY OF BIOMOLECULES .2. NATIVE RIBONUCUEASE AND ALPHA-CHYMOTRYPSIN. *Journal of Molecular Biology,* 51, 203-&.

LYNG, F. M., FAOLAIN, E. O., CONROY, J., MEADE, A. D., KNIEF, P., DUFFY, B., HUNTER, M. B., BYRNE, J. M., KELEHAN, P. & BYME, H. J. (2007) Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool. *Experimental and Molecular Pathology,* 82, 121-129.

MADRID-MARINA, V., TORRES-POVEDA, K., LÓPEZ-TOLEDO, G. & GARCÍA-CARRANCÁ, A. (2009) Advantages and Disadvantages of Current Prophylactic Vaccines Against HPV. *Archives of Medical Research,* 40, 471-477.

MAHADEVAN-JANSEN, A., MITCHELL, M. F., RAMANUJAM, N., MALPICA, A., THOMSEN, S., UTZINGER, U. & RICHARDS-KORTUM, R. (1998) Near-infrared Raman spectroscopy for in vitro detection of cervical precancers. *Photochemistry and Photobiology,* 68, 123-132.

MAITLAND, K. C., GILLENWATER, A. M., WILLIAMS, M. D., EL-NAGGAR, A. K., DESCOUR, M. R. & RICHARDS-KORTUM, R. R. (2008) In vivo imaging of oral neoplasia using a miniaturized fiber optic confocal reflectance microscope. *Oral Oncology,* 44, 1059-1066.

MARTENS, H. & NAES, T. (1994) *Multivariate Calibration*, John Wiley and Sons.

MCCLUGGAGE, W. G., WALSH, M. Y., THORNTON, C. M., HAMILTON, P. W., DATE, A., CAUGHLEY, L. M. & BHARUCHA, H. (1998) Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system. *British Journal of Obstetrics and Gynaecology,* 105, 206-210.

MERA, S. L. (1997) *pathology and understanding disease preventation*, Stanley Thornes.

MIN, Y. K., YAMAMOTO, T., KOHDA, E., ITO, T. & HAMAGUCHI, H. (2005) 1064 nm near-infrared multichannel Raman spectroscopy of fresh human lung tissues. *Journal of Raman Spectroscopy,* 36, 73-76.

MITCHELL MD, M. S. M. F., SCHOTTENFELD MD, M. S. D., TORTOLERO-LUNA MD, P. G., CANTOR PHD, S. B. & RICHARDS-KORTUM PHD, R. (1998) Colposcopy For The Diagnosis of Squamous Intraepithelial Lesions: A Meta-Analysis. *Obstetrics & Gynecology,* 91, 626-631.

MITCHELL, M. F. (1994) Accuracy of colposcopy. *Clin. Consult. Obstetrics Gynecol.,* 6, 70-73.

MITCHELL, M. F., SCHOTTENFELD, D., TORTOLERO-LUNA, G., CANTOR, S. B. & RICHARDS-KORTUM, R. (1998) Colposcopy for the diagnosis of squamous intraepithelial lesions: A meta-analysis. *Obstetrics and Gynecology,* 91, 626-631.

MORDECHAI, S., SAHU, R. K., HAMMODY, Z., MARK, S., KANTAROVICH, K., GUTERMAN, H., PODSHYVALOV, A., GOLDSTEIN, J. & ARGOV, S. (2004) Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma. *Journal of Microscopy-Oxford,* 215, 86-91.

MOSCICKI, A. B. (1999) Human papillomavirus infection in adolescents. *Pediatric Clinics of North America,* 46, 783-+.

MURILLO, R., LUNA, J., GAMBOA, O., OSORIO, E., BONILLA, J., CENDALES, R. & STUDY, I. N. C. C. C. S. (2010) Cervical cancer screening with naked-eye visual inspection in Colombia. *International Journal of Gynecology & Obstetrics,* 109, 230-234.

MUWONGE, R., MANUEL, M. D., FILIPE, A. P., DUMAS, J. B., FRANK, M. R. & SANKARANARAYANAN, R. (2010) Visual screening for early detection of cervical neoplasia in Angola. *International Journal of Gynecology & Obstetrics,* 111, 68-72.

NA, R. H., STENDER, I. M., HENRIKSEN, M. & WULF, H. C. (2001) Autofluorescence of human skin is age-related after correction for skin pigmentation and redness. *Journal of Investigative Dermatology,* 116, 536-540.

NANDA, K., MCCRORY, D. C., MYERS, E. R., BASTIAN, L. A., HASSELBLAD, V., HICKEY, J. D. & MATCHAR, D. B. (2000) Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: A systematic review. *Annals of Internal Medicine,* 132, 810-819.

NESSA, A., HUSSAIN, M. A., RAHMAN, J. N., RASHID, M. H. U., MUWONGE, R. & SANKARANARAYANAN, R. (2010) Screening for cervical neoplasia in Bangladesh using visual inspection with acetic acid. *International Journal of Gynecology & Obstetrics,* 111, 115-118.

NGOMA, T., MUWONGE, R., MWAISELAGE, J., KAWEGERE, J., BUKORI, P. & SANKARANARAYANAN, R. (2010) Evaluation of cervical visual inspection screening in Dar es Salaam, Tanzania. *International Journal of Gynecology & Obstetrics,* 109, 100-104.

NOTINGHER, I. (2007) Raman Spectroscopy cell-based Biosensors. *Sensors,* 7, 1343-1358.

NOTINGHER, I., VERRIER, S., HAQUE, S., POLAK, J. M. & HENCH, L. L. (2003) Spectroscopic study of human lung epithelial cells (A549) in culture: Living cells versus dead cells. *Biopolymers,* 72, 230-240.

O FAOLAIN, E., HUNTER, M. B., BYRNE, J. M., KELEHAN, P., MCNAMARA, M., BYRNE, H. J. & LYNG, F. M. (2004) A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. *3rd International Conference on Shedding Light on Disease - Optical Diagnostics for the New Millenium (SPEC 2004).* Newark, NJ, Elsevier Science Bv.

O FAOLAIN, E., HUNTER, M. B., BYRNE, J. M., KELEHAN, P., MCNAMARA, M., BYRNE, H. J. & LYNG, F. M. (2005) A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. *Vibrational Spectroscopy,* 38, 121-127.

OOI, G. J., FOX, J., SIU, K., LEWIS, R., BAMBERY, K. R., MCNAUGHTON, D. & WOOD, B. R. (2008) Fourier transform infrared imaging and small angle x-ray scattering as a combined biomolecular approach to diagnosis of breast cancer. *Medical Physics,* 35, 2151-2161.

OSTROWSKA, K. M., MALKIN, A., MEADE, A., O'LEARY, J., MARTIN, C., SPILLANE, C., BYRNE, H. J. & LYNG, F. M. (2010) Investigation of the influence of high-risk human papillomavirus on the biochemical composition of cervical cancer cells using vibrational spectroscopy. *Analyst,* 135, 3087-3093.

OVESTAD, I. T., GUDLAUGSSON, E., SKALAND, I., MALPICA, A., KRUSE, A. J., JANSSEN, E. A. M. & BAAK, J. P. A. (2010) Local immune response in the microenvironment of CIN2-3 with and without spontaneous regression. *Modern Pathology,* 23, 1231-1240.

OVESTAD, I. T., VENNESTROM, U., ANDERSEN, L., GUDLAUGSSON, E., MUNK, A. C., MALPICA, A., FENG, W. W., VOORHORST, F., JANSSEN, E. A. M. & BAAK, J. P. A. (2011) Comparison of different commercial methods for HPV detection in follow-up cytology after ASCUS/LSIL, prediction of CIN2-3 in follow up biopsies and spontaneous regression of CIN2-3. *Gynecologic Oncology,* 123, 278-283.

PAPANICOLAOU, G. N. & TRAUT, H. F. (1941) The diagnostic value of vaginal smears in carcinoma of the uterus. *American Journal of Obstetric and Gynecology,* 42, 193-206.

PAPANICOLAOU, G. N. & TRAUT, H. F. (1997) The diagnostic value of vaginal smears in carcinoma of the uterus (Reprinted from American Journal of Obstetrics and Gynecology, vol 42, 1941). *Archives of Pathology & Laboratory Medicine,* 121, 211-223.

PARKIN, D. M., BRAY, F., FERLAY, J. & PISANI, P. (2005) Global cancer statistics, 2002. *Ca-a Cancer Journal for Clinicians,* 55, 74-108.

PAVLOVA, I., SOKOLOV, K., DREZEK, R., MALPICA, A., FOLLEN, M. & RICHARDS-KORTUM, R. (2003) Microanatomical and biochemical origins of normal and precancerous cervical autofluorescence using laser-scanning fluorescence confocal microscopy. *Photochemistry and Photobiology,* 77, 550-555.

PEARSON, K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. *Philiosophical Magazine,* 2, 559-572.

PETO, J., GILHAM, C., FLETCHER, O. & MATTHEWS, F. E. (2004) The cervical cancer epidemic that screening has prevented in the UK. *Lancet,* 364, 249-256.

PINTO, A. P. & CRUM, C. P. (2000) Natural history of cervical neoplasia: Defining progression and its consequence. *Clinical Obstetrics and Gynecology,* 43, 352-362.

PITRIS, C., GOODMAN, A., BOPPART, S. A., LIBUS, J. J., FUJIMOTO, J. G. & BREZINSKI, M. E. (1999) High-resolution imaging of gynecologic neoplasms using optical coherence tomography. *Obstet Gynecol,* 93, 135-9.

PUPPELS, G. J., GARRITSEN, H. S. P., KUMMER, J. A. & GREVE, J. (1993) CAROTENOIDS LOCATED IN HUMAN LYMPHOCYTE SUBPOPULATIONS AND NATURAL-KILLER-CELLS BY RAMAN MICROSPECTROSCOPY. *Cytometry,* 14, 251-256.

QIAN, X. M., PENG, X. H., ANSARI, D. O., YIN-GOEN, Q., CHEN, G. Z., SHIN, D. M., YANG, L., YOUNG, A. N., WANG, M. D. & NIE, S. M. (2008) In vivo tumor targeting and spectroscopic detection with surface-enhanced Raman nanoparticle tags. *Nature Biotechnology,* 26, 83-90.

QUINN, M. A. (1998) Adenocarcinoma of the Cervix. *Ann Acad Med Singapore* 27, 662-665.

RAMANUJAM, N., MITCHELL, M. F., MAHADEVAN, A., WARREN, S., THOMSEN, S., SILVA, E. & RICHARDSKORTUM, R. (1994) IN-VIVO DIAGNOSIS OF CERVICAL INTRAEPITHELIAL NEOPLASIA USING 337-NM-EXCITED LASER-INDUCED FLUORESCENCE. *Proceedings of the National Academy of Sciences of the United States of America,* 91, 10193-10197.

RAMANUJAM, N., MITCHELL, M. F., MAHADEVANJANSEN, A., THOMSEN, S. L., STAERKEL, G., MALPICA, A., WRIGHT, T., ATKINSON, N. & RICHARDSKORTUM, R.

(1996) Cervical precancer detection using a multivariate statistical algorithm based on laser-induced fluorescence spectra at multiple excitation wavelengths. *Photochemistry and Photobiology,* 64, 720-735.

RAMANUJAM, N., RICHARDS-KORTUM, R., THOMSEN, S., MAHADEVAN-JANSEN, A., FOLLEN, M. & CHANCE, B. (2001) Low temperature fluorescence imaging of freeze-trapped human cervical tissues. *Optics Express,* 8, 335-343.

REAGAN, J. W., SEIDEMANN, I. L. & SARACUSA, Y. (1953) THE CELLULAR MORPHOLOGY OF CARCINOMA INSITU AND DYSPLASIA OR ATYPICAL HYPERPLASIA OF THE UTERINE CERVIX. *Cancer,* 6, 224-235.

REHMAN, I., SMITH, R., HENCH, L. L. & BONFIELD, W. (1995) STRUCTURAL EVALUATION OF HUMAN AND SHEEP BONE AND COMPARISON WITH SYNTHETIC HYDROXYAPATITE BY FT-RAMAN SPECTROSCOPY. *Journal of Biomedical Materials Research,* 29, 1287-1294.

RICHART, R. M. (1968) Natural history of cervical intraepithelial neoplasia. *Clinical Obstetrics and Gynecology,* 5.

RICHART, R. M. (1990) A MODIFIED TERMINOLOGY FOR CERVICAL INTRAEPITHELIAL NEOPLASIA. *Obstetrics and Gynecology,* 75, 131-133.

ROBICHAUX-VIEHOEVER, A., KANTER, E., SHAPPELL, H., BILLHEIMER, D., JONES, H. & MAHADEVAN-JANSEN, A. (2007) Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia. *Applied Spectroscopy,* 61, 986-993.

RONCO, G., GIORGI-ROSSI, P., CAROZZI, F., DALLA PALMA, P., DEL MISTRO, A., DE MARCO, L., DE LILLO, M., NALDONI, C., PIEROTTI, P., RIZZOLO, R., SEGNAN, N., SCHINCAGLIA, P., ZORZI, M., CONFORTINI, M. & CUZICK, J. (2006) Human papillomavirus testing and liquid-based cytology in primary screening of women younger than 35 years: results at recruitment for a randomised controlled trial. *Lancet Oncology,* 7, 547-555.

RUIZ-CHICA, A. J., MEDINA, M. A., SANCHEZ-JIMENEZ, F. & RAMIREZ, F. J. (2004) Characterization by Raman spectroscopy of conformational changes on guanine-cytosine and adenine-thymine oligonucleotides induced by aminooxy analogues of spermidine. *Journal of Raman Spectroscopy,* 35, 93-100.

SALZER, R. & SIESLER, H. W. (2009) *Infrared and Raman spectroscopic imaging*, Wiley-VCH.

SANKARANARAYANAN, J. W. S. A. R. (Ed.) (2004) *Colposcopy and treatment of cervical intraepithelial neoplasia: A beginner's manual.*

SANKARANARAYANAN, R., ESMY, P. O., RAJKUMAR, R., MUWONGE, R., SWAMINATHAN, R., SHANTHAKUMARI, S., FAYETTE, J. M. & CHERIAN, J. (2007) Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial. *Lancet,* 370, 398-406.

SANKARANARAYANAN, R., WESLEY, R., SOMANATHAN, T., DHAKAD, N., SHYAMALAKUMARY, B., AMMA, N. S., PARKIN, D. M. & NAIR, M. K. (1998) Visual inspection of the uterine cervix after the application of acetic acid in the detection of cervical carcinoma and its precursors. *Cancer,* 83, 2150-2156.

SASIC, S. (2008) *Pharmaceutical Applications of Raman Spectroscopy*, John Wiley & Sons. .

SASLOW, D., SOLOMON, D., LAWSON, H. W., KILLACKEY, M., KULASINGAM, S. L., CAIN, J. M., GARCIA, F. A. R., MORIARTY, A. T., WAXMAN, A. G., WILBUR, D. C., WENTZENSEN, N., DOWNS, L. S., SPITZER, M., MOSCICKI, A. B., FRANCO, E. L., STOLER, M. H., SCHIFFMAN, M., CASTLE, P. E. & MYERS, E. R. (2012) American

Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology Screening Guidelines for the Prevention and Early Detection of Cervical Cancer. *Journal of Lower Genital Tract Disease,* 16, 175-204.

SAUVAGET, C., FAYETTE, J. M., MUWONGE, R., WESLEY, R. & SANKARANARAYANAN, R. (2011) Accuracy of visual inspection with acetic acid for cervical cancer screening. *International Journal of Gynecology & Obstetrics,* 113, 14-24.

SCHRADER, B. (1995) *Infrared and Raman spectroscopy: methods and applications*, VCH.

SCHUBERT, J. M., MAZUR, A. I., BIRD, B., MILJKOVIC, M. & DIEM, M. (2010) Single point vs. mapping approach for spectral cytopathology (SCP). *Journal of Biophotonics,* 3, 588-596.

SELLORS, W. & SANKARANARAYANAN, R. (2003/4) *Colposcopy and Treatment of Cervical Intraepithelial Neoplasia:*
A Beginner's Manual, Published by the International Agency for Research on Cancer.

SERGEEV, A., GELIKONOV, V., GELIKONOV, G., FELDCHTEIN, F., KURANOV, R., GLADKOVA, N., SHAKHOVA, N., SNOPOVA, L., SHAKHOV, A., KUZNETZOVA, I., DENISENKO, A., POCHINKO, V., CHUMAKOV, Y. & STRELTZOVA, O. (1997) In vivo endoscopic OCT imaging of precancer and cancer states of human mucosa. *Opt Express,* 1, 432-40.

SHEPHERD, L. J. & BRYSON, S. C. P. (2008 Nov) Human papillomavirus--lessons from history and challenges for the future. *J Obstet Gynaecol Can,* 30, 1025-33.

SHORT, M. A., LAM, S., MCWILLIAMS, A., ZHAO, J. H., LUI, H. & ZENG, H. S. (2008) Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers. *Optics Letters,* 33, 711-713.

SIGURDSSON, K. (1999) Cervical cancer, Pap smear and HPV testing: An update of the role of organized Pap smear screening and HPV testing. *Acta Obstetricia Et Gynecologica Scandinavica,* 78, 467-477.

SMITH, J. S., GREEN, J., DE GONZALEZ, A. B., APPLEBY, P., PETO, J., PLUMMER, M., FRANCESCHI, S. & BERAL, V. (2003) Cervical cancer and use of hormonal contraceptives: a systematic review. *The Lancet,* 361, 1159-1167.

SOLOMON, D., DAVEY, D., KURMAN, R., MORIARTY, A., O'CONNOR, D., PREY, M., RAAB, S., SHERMAN, M., WILBUR, D., WRIGHT, T., JR. & YOUNG, N. (2002) The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA,* 287, 2114-9.

STONE, N., KENDALL, C., SHEPHERD, N., CROW, P. & BARR, H. (2002) Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy,* 33, 564-573.

STONE, N., KENDALL, C., SMITH, J., CROW, P. & BARR, H. (2004) Raman spectroscopy for identification of epithelial cancers. *Faraday Discussions,* 126, 141-157.

TAN, K. M., HERRINGTON, C. S. & BROWN, C. T. (2011) Discrimination of normal from pre-malignant cervical tissue by Raman mapping of de-paraffinized histological tissue sections. *J Biophotonics,* 4, 40-8.

THEKKEK, N. & RICHARDS-KORTUM, R. (2008) Optical imaging for cervical cancer detection: solutions for a continuing global problem. *Nature Reviews Cancer,* 8, 725-731.

UTZINGER, U., HEINTZELMAN, D. L., MAHADEVAN-JANSEN, A., MALPICA, A., FOLLEN, M. & RICHARDS-KORTUM, R. (2001) Near-infrared Raman spectroscopy for in vivo detection of cervical precancers. *Applied Spectroscopy,* 55, 955-959.

VIEHOEVER, A. R., ANDERSON, D., JANSEN, D. & MAHADEVAN-JANSEN, A. (2003) Organotypic raft cultures as an effective in vitro tool for understanding Raman spectral analysis of tissue. *Photochemistry and Photobiology,* 78, 517-524.

VILLA, L. L. & DENNY, L. (2006) Methods for detection of HPV infection and its clinical utility. *International Journal of Gynecology & Obstetrics,* 94, S71-S80.

WALBOOMERS, J. M. M., JACOBS, M. V., MANOS, M. M., BOSCH, F. X., KUMMER, J. A., SHAH, K. V., SNIJDERS, P. J. F., PETO, J., MEIJER, C. & MUNOZ, N. (1999) Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology,* 189, 12-19.

WANG, L. Q. & MIZAIKOFF, B. (2008) Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. *Analytical and Bioanalytical Chemistry,* 391, 1641-1654.

WERNESS, B. A., LEVINE, A. J. & HOWLEY, P. M. (1990) ASSOCIATION OF HUMAN PAPILLOMAVIRUS TYPE-16 AND TYPE-18 E6 PROTEINS WITH P53. *Science,* 248, 76-79.

WONG, P. T. T., WONG, R. K., CAPUTO, T. A., GODWIN, T. A. & RIGAS, B. (1991) INFRARED-SPECTROSCOPY OF EXFOLIATED HUMAN CERVICAL CELLS - EVIDENCE OF EXTENSIVE STRUCTURAL-CHANGES DURING CARCINOGENESIS. *Proceedings of the National Academy of Sciences of the United States of America,* 88, 10988-10992.

WOOD, B. R., QUINN, M. A., TAIT, B., ASHDOWN, M., HISLOP, T., ROMEO, M. & MCNAUGHTON, D. (1998) FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies. *Biospectroscopy,* 4, 75-91.

WOODMAN, C. B. J., COLLINS, S. I. & YOUNG, L. S. (2007) The natural history of cervical HPV infection: unresolved issues. *Nature Reviews Cancer* 7.

WOODMAN, C. B. J., COLLINS, S. I. & YOUNG, L. S. (January 2007) The natural history of cervical HPV infection: unresolved issues. *Nature Reviews Cancer* 7.

WRIGHT, T. C., MASSAD, S., DUNTON, C. J., SPITZER, M., WILKINSON, E. J. & SOLOMON, D. (2007) 2006 consensus guidelines for the management of women with abnormal cervical cancer screening tests. *American Journal of Obstetrics and Gynecology,* 197, 346-355.

YAZDI, Y., RAMANUJAM, N., LOTAN, R., MITCHELL, M. F., HITTELMAN, W. & RICHARDS-KORTUM, R. (1999) Resonance Raman spectroscopy at 257 nm excitation of normal and malignant cultured breast and cervical cells. *Applied Spectroscopy,* 53, 82-85.

ZHOU, Y., CAO, H., HU, J. X., YU, S. J., ZHAO, Q., LI, X. Z. & JU, L. C. (2011) Spectrum purification augmented Classical Least -Squares (CLS): A Robust multivariate calibration method for compound quantitative determination. *Instrumentation Science & Technology,* 39, 394-405.

ZLATKOV, V. (2009) [Possibilities of the TruScreen for screening of precancer and cancer of the uterine cervix]. *Akush Ginekol (Sofiia),* 48, 46-50.

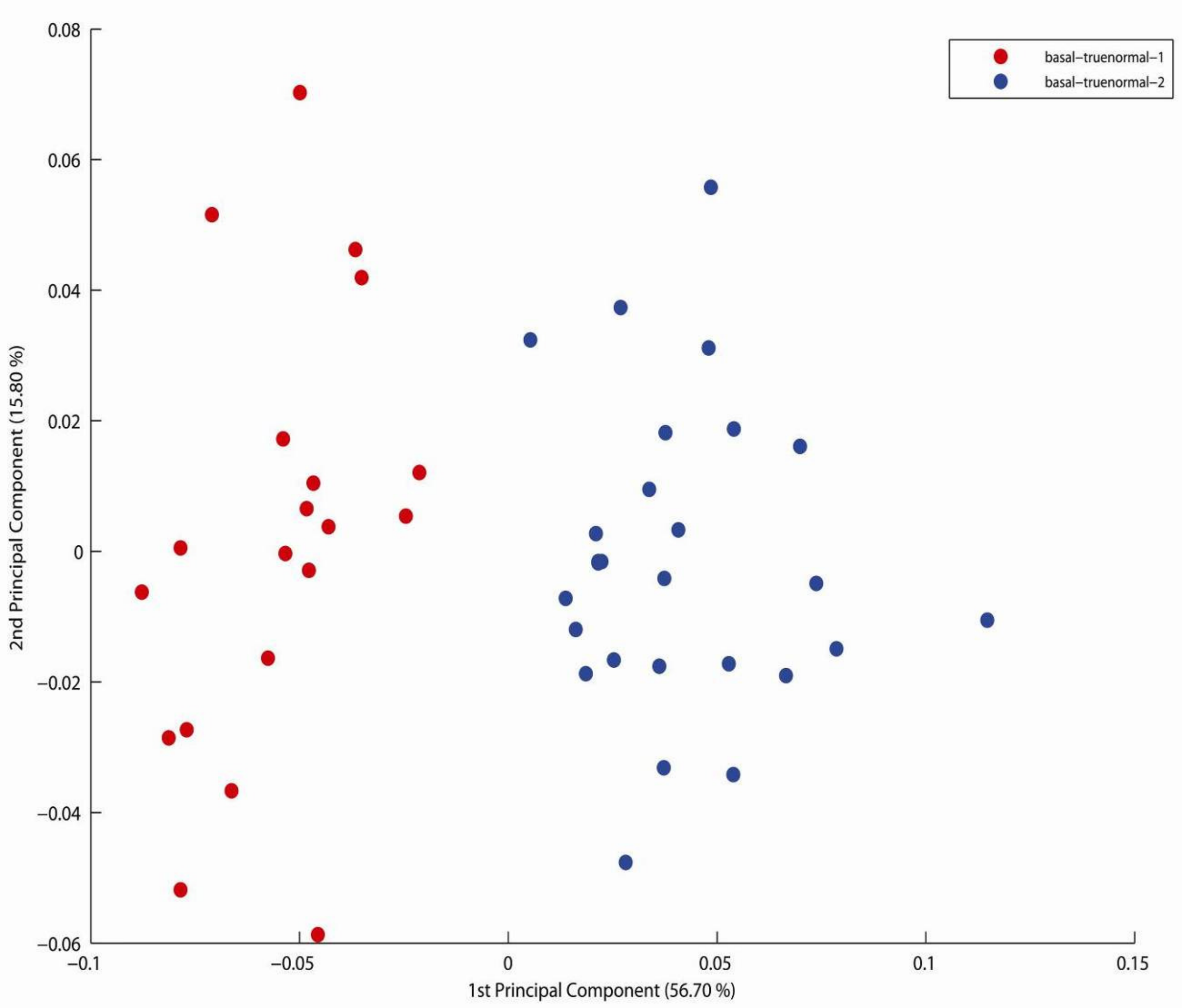
ZULUAGA, A. F., UTZINGER, U., DURKIN, A., FUCHS, H., GILLENWATER, A., JACOB, R., KEMP, B., FAN, J. & RICHARDS-KORTUM, R. (1999) Fluorescence excitation emission matrices of human tissue: A system for in vivo measurement and method of data analysis. *Applied Spectroscopy,* 53, 302-311.

ZURHAUSEN, H. (1976) Condylomata acuminata and human genital cancer. *Cancer Research,* 36, 794-794.

Appendix

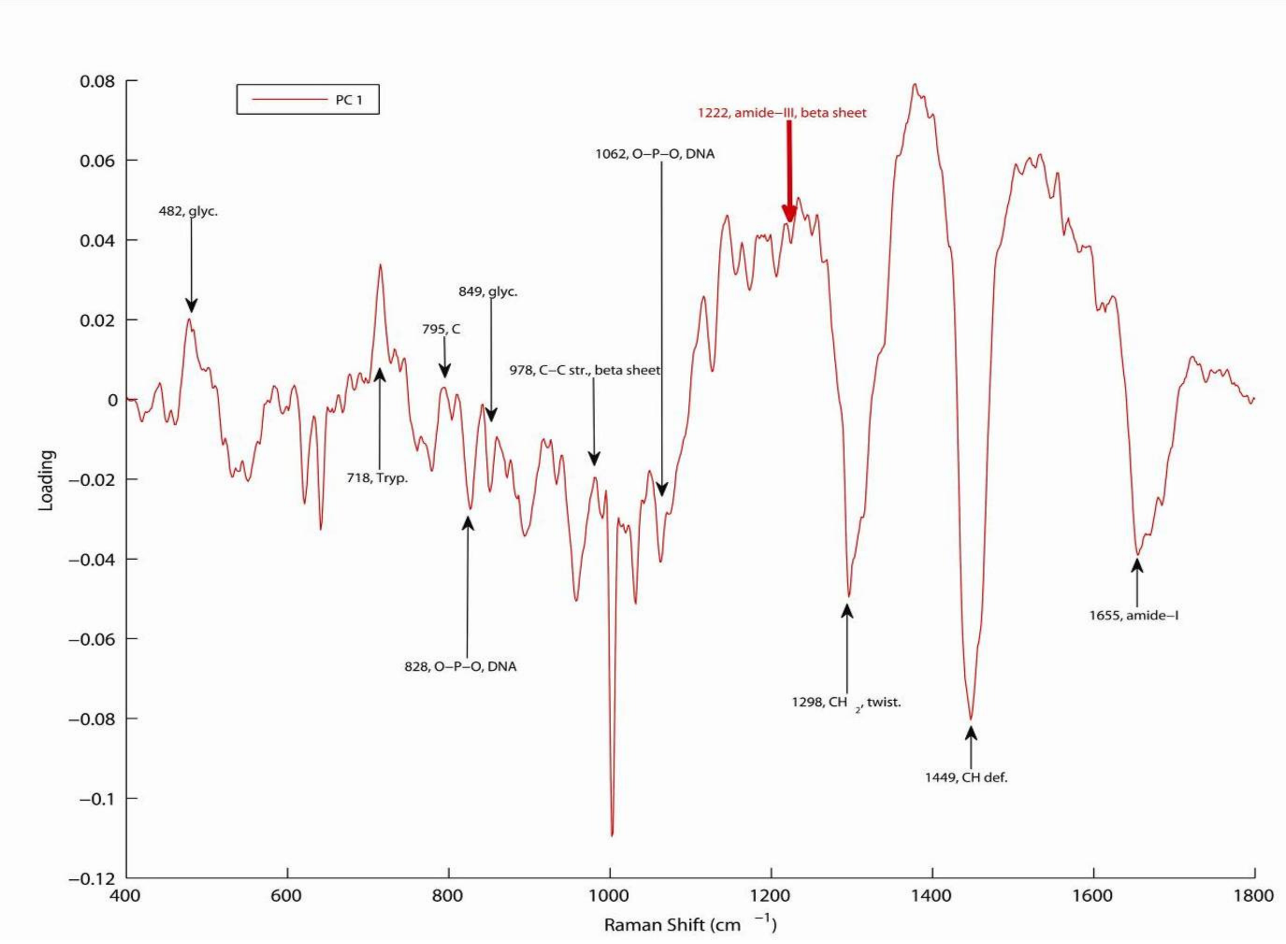
PCA results of basal layer of normal-1 versus normal-2

PCA results of the basal-true normal-1 versus basal-true normal-2 are discussed here. The purpose of this analysis was to differentiate the inter-patient/inter-sample variability. The PCA results are presented in **A1.1** and **A1.2** as scatter plot and loadings respectively.



A1.1 PCA scatter plot of basal true normal-1 vs basal true normal-2.

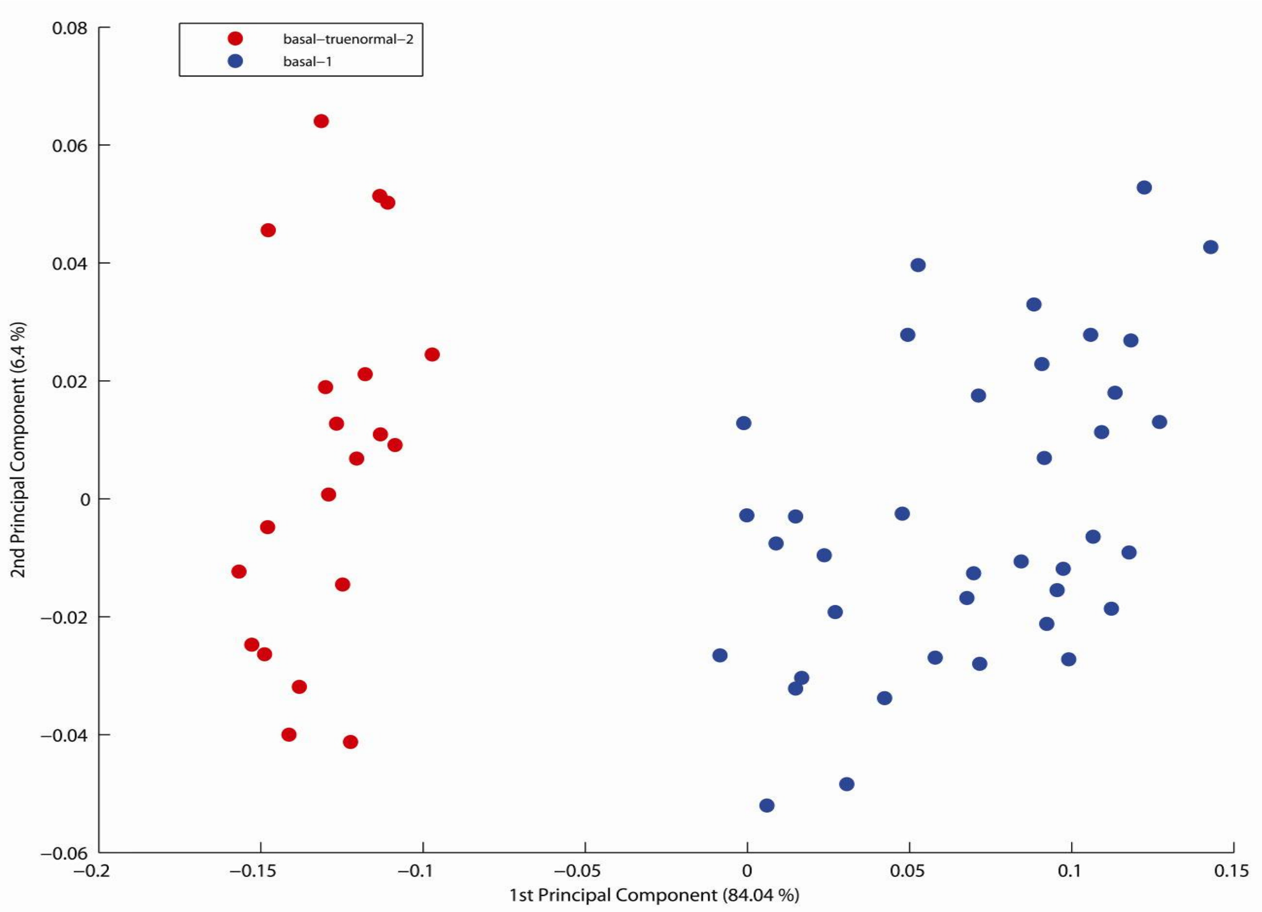
The PCA scatter plot shows some differentiation of the two clusters of the Raman spectra which is attributed to the inter-sample variability but the Raman spectral features which has been the differentiating features, particularly the band at 1222 cm^{-1}, between the true-normal and diseased samples are not observed here in the loadings. Moreover, major differentiating features include 1298, 1449 and 1655 cm^{-1} associated with proteins and lipids which are not taken, in the current studies discussed in chapter 5, as marker of the disease progression and HPV infection. This indicates the differentiation of the two clusters is solely on the basis of the intensity difference of these spectral features.



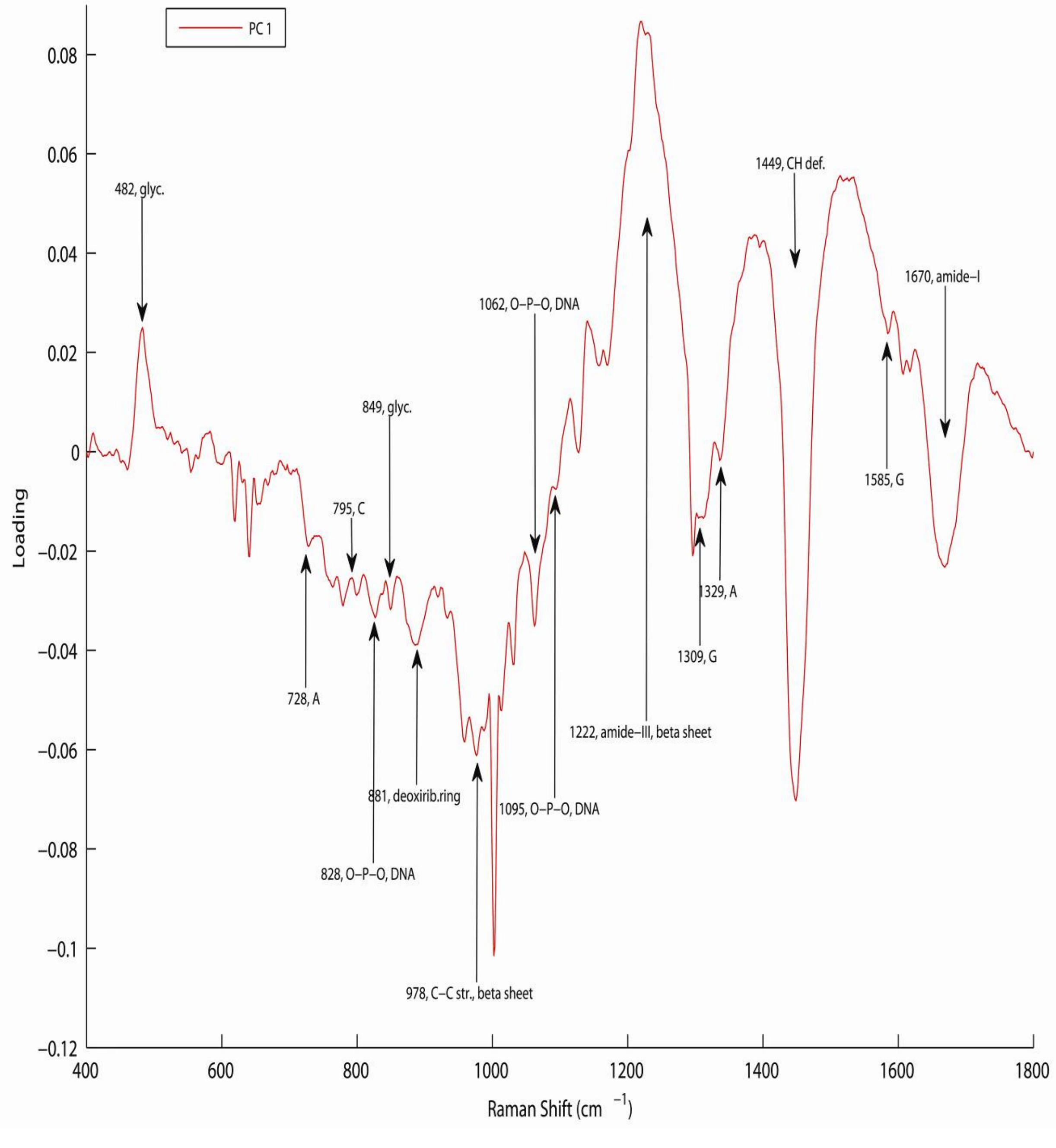
A1.2 PCA loadings (PC-1) of basal true normal-1 vs basal true normal-2.

PCA results of basal layer of normal-2 versus basal-1 (CIN II normal

PCA of the basal-1 (CIN II normal basal layer) with the second true normal sample "true normal-2" was also done and is presented in the Appendix (**A1.3** and **A1.4**) as scatter plot and loadings respectively. Notably, the loadings are the same as observed in case of "basal-true normal-1" versus "basal-1 (CIN II normal)" presented in chapter 5, **Figure 5.7** and **5.8**, and the 1222 cm^{-1} feature is also observed in these loadings which further confirms that the changes in the amide-III beta sheet (1222 cm^{-1}) may be taken as a Raman spectral marker of the early progression of the disease.



A1.3 PCA scatter plot of basal true normal-2 vs basal-1(basal-CIN II normal).



A1.4 PCA loadings (PC-1) of basal true normal-2 vs basal-1(basal-CIN II normal).

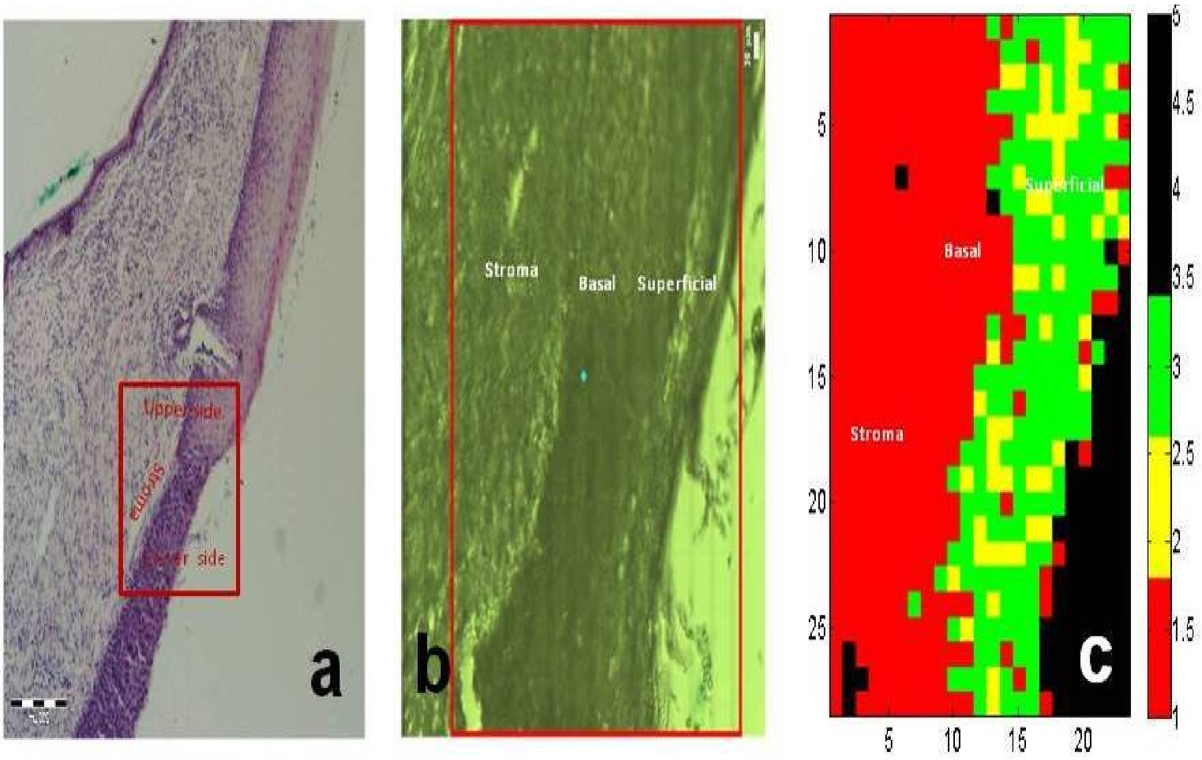
PCA results of basal layer of normal case (basal-true normal) vs intermediate, abnormal (basal-4 to basal-7) samples

PCA results of the basal-true normal versus other basal layers of different samples of different stages of the disease including intermediate and abnormal cases are discussed in the following sections. The purpose of this analysis was to differentiate the biochemical changes which take place due to the progression of the disease from the normal case to the intermediate and abnormal stages.

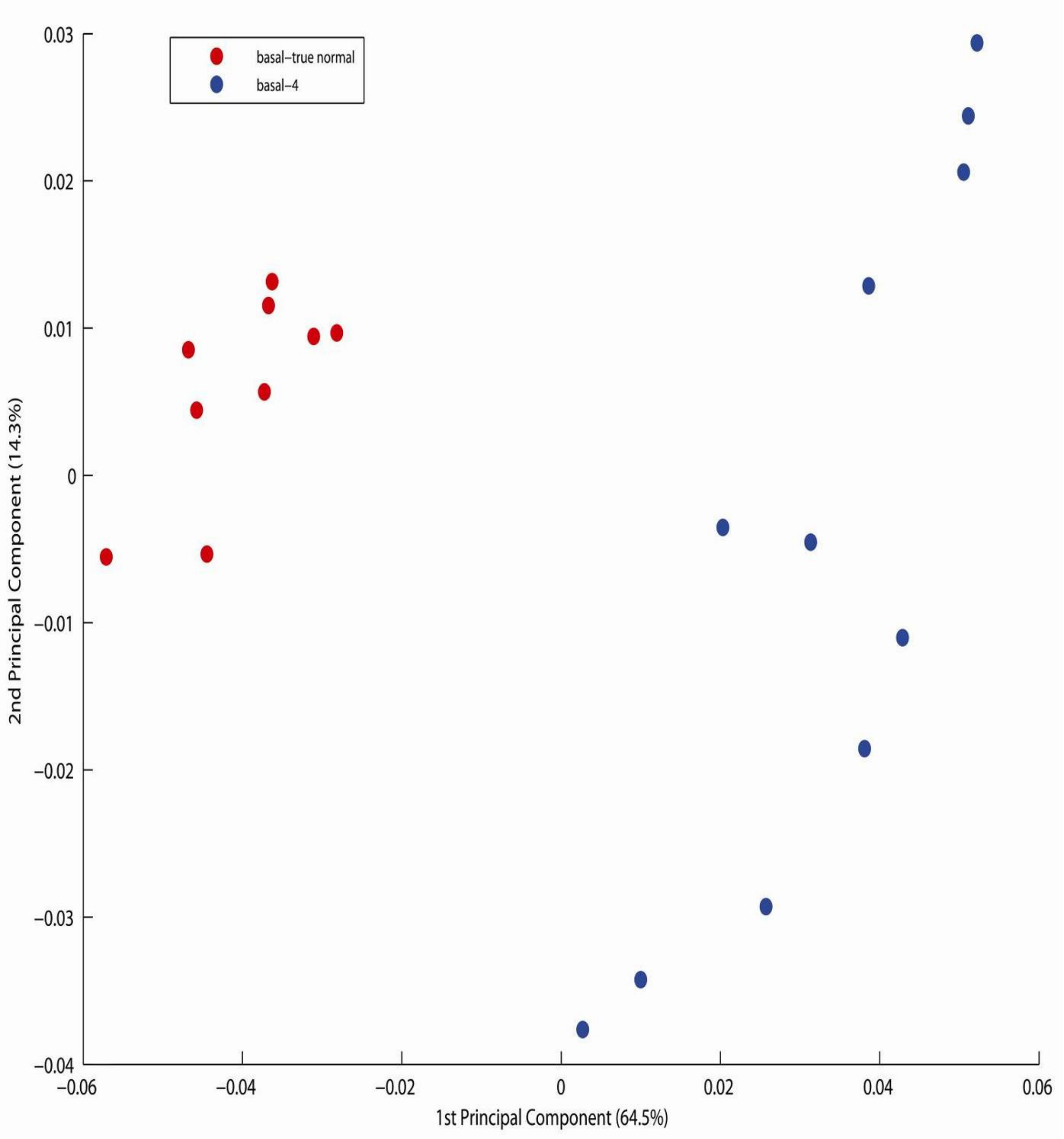
PCA results of basal-true normal vs basal-4

PCA of the Raman spectra from the basal layer of the normal case (basal-true normal) against the basal layer of an abnormal sample called basal-4, shown in **A1.3**, are presented here in **A1.4** and **A1.5** as scatter plot and loadings of the PCA, respectively. PCA scatter plot shows that two groups are different from each other. The Raman spectra of the basal-true normal layer are clustered in the negative axis and spectra of the basal-4 layer in the positive axis of the scatter plot. The loadings contributed by the basal-true normal include those which are associated with glycogen at 849 cm-1, 939 cm-1 (glycogen) and some others related to proteins including 644 cm-1 (C-C stretching), 831 cm-1 (tyrosine) and some related to the DNA including 487 cm-1 (adenine), 749 cm-1 (thymine) and 1346 cm-1 (guanine).

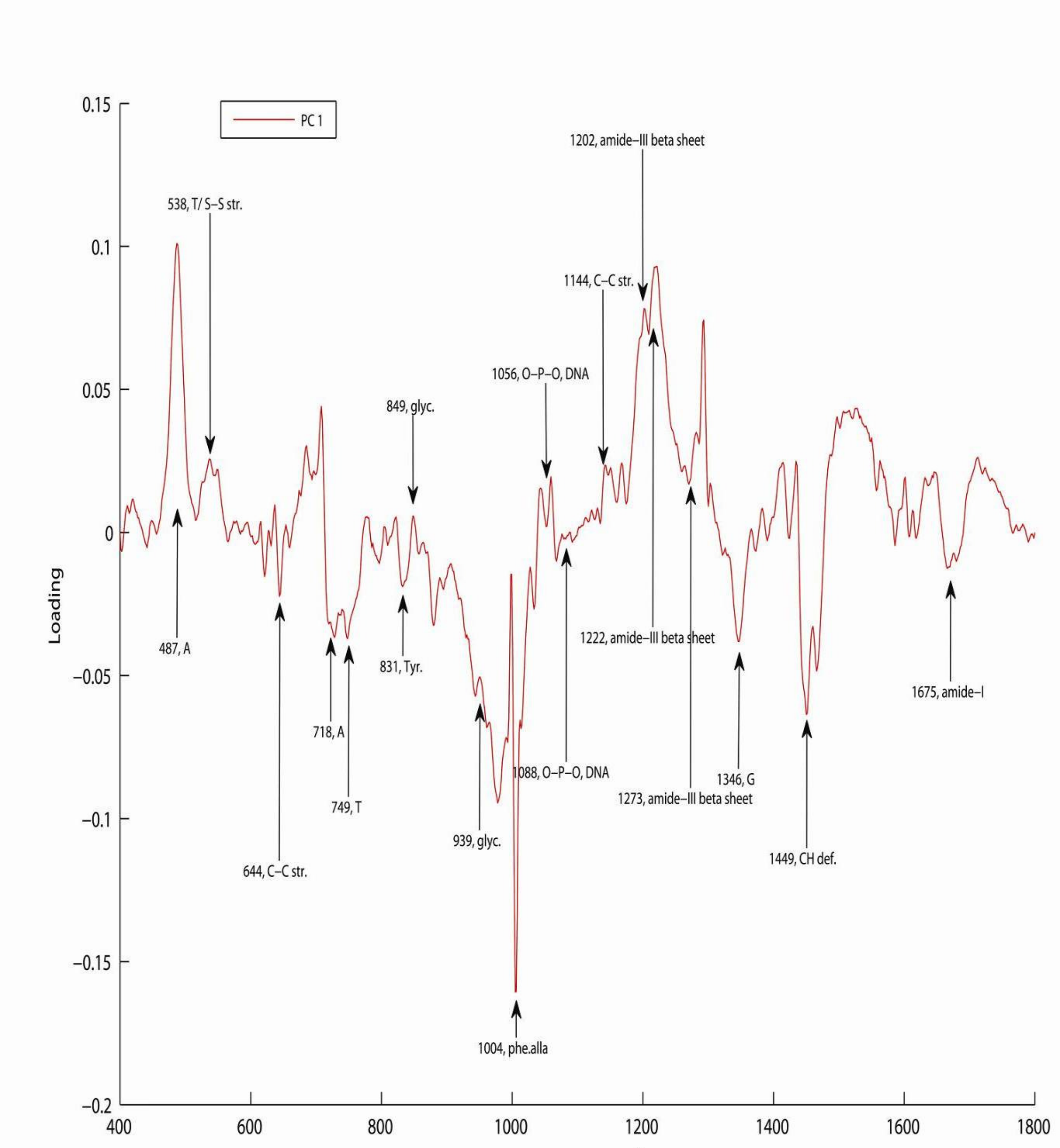
PCA loadings contributed by the basal-4 include those related to DNA at 487 cm^{-1} (adenine), 538 (thymine/S-S stretching of proteins, 1056 cm^{-1}, 1088 cm^{-1} (O-P-O, DNA) and others related to proteins including 1144 cm^{-1} (C-C stretching of proteins) and 1202 cm^{-1} and 1222 cm^{-1} (amide-III beta sheet). The contribution of these loadings by the basal-4 layer is consistent with the previously discussed samples and indicates towards the biochemical changes related to proteins due to the progression of cervical cancer, from basal layer of the true normal case to the abnormal cases.



A1.5 H&E stained diseased side of CIN III marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 20µm (b) Five cluster KMCA map generated from the Raman map (c).



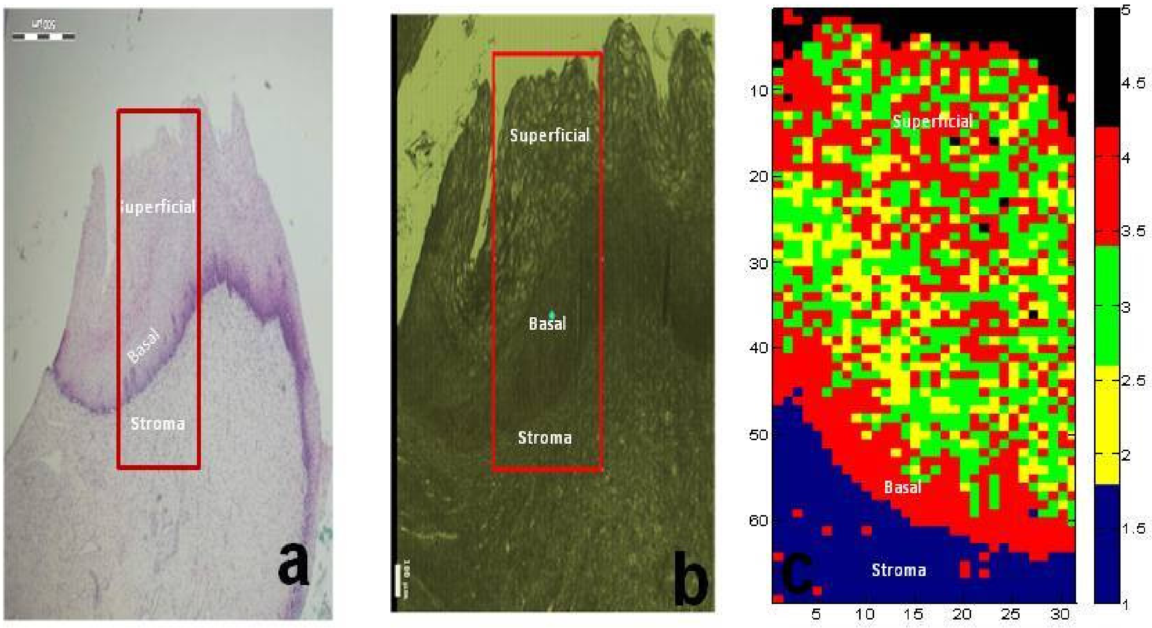
A1.6 PCA Scatter plot for basal-true normal versus basal-4.



A1.7 PCA loading (PC-1) for basal-true normal versus basal-4.

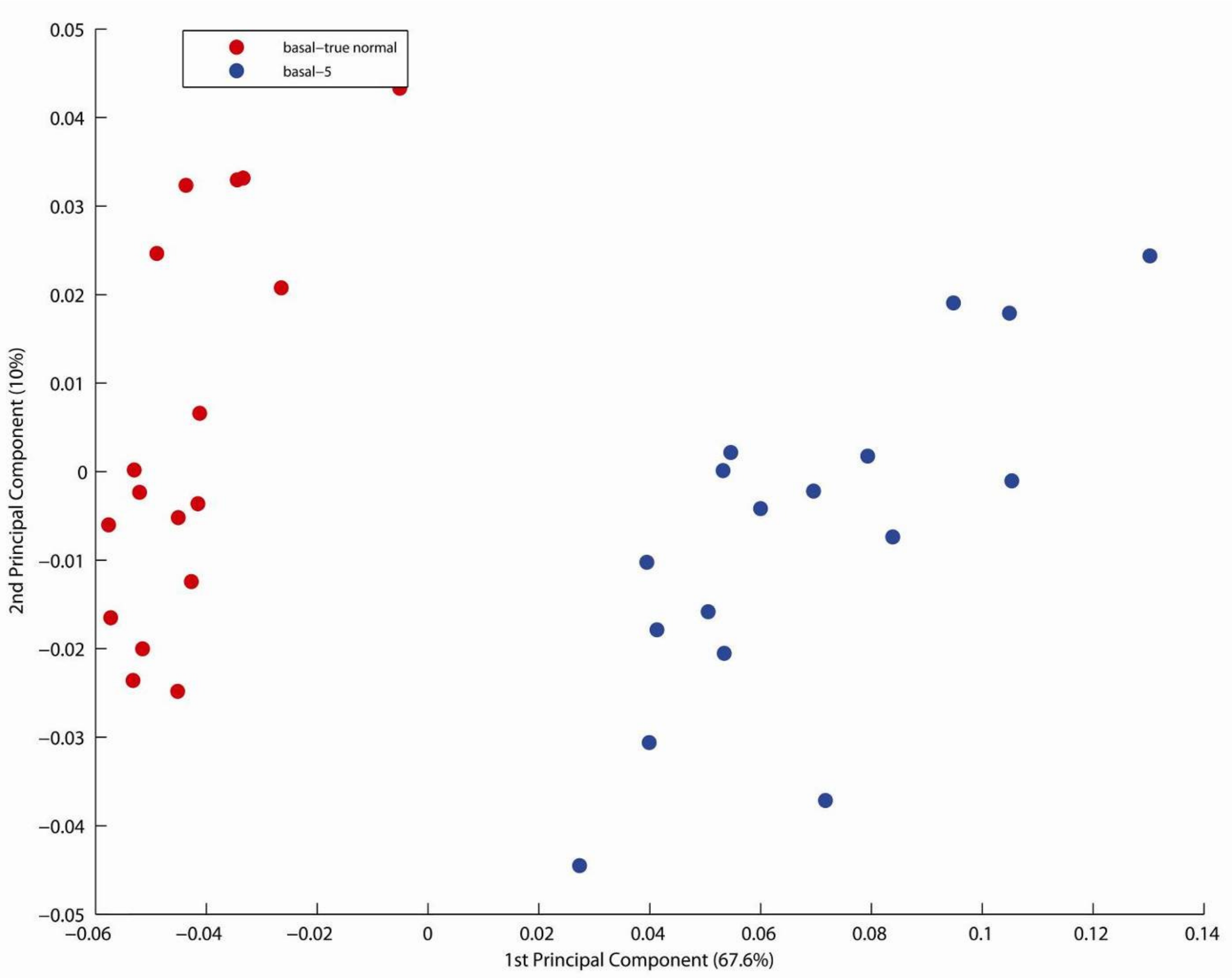
PCA results of basal-true normal vs basal-5

PCA of the Raman spectra from the basal layer of the normal case (basal-true normal) against the basal layer of an intermediate sample (basal-5) shown in Figure **A1.6**, are shown in Figure **A1.7** and **A1.8** as scatter plot and loadings of the PCA, respectively. The PCA scatter plot shows clear differentiation of the two groups by clustering the Raman spectra of the basal-true normal layer in the negative axis and the spectra of the basal-5 layer in the positive axis of the scatter plot and hence negative and positive loadings are associated with them respectively.



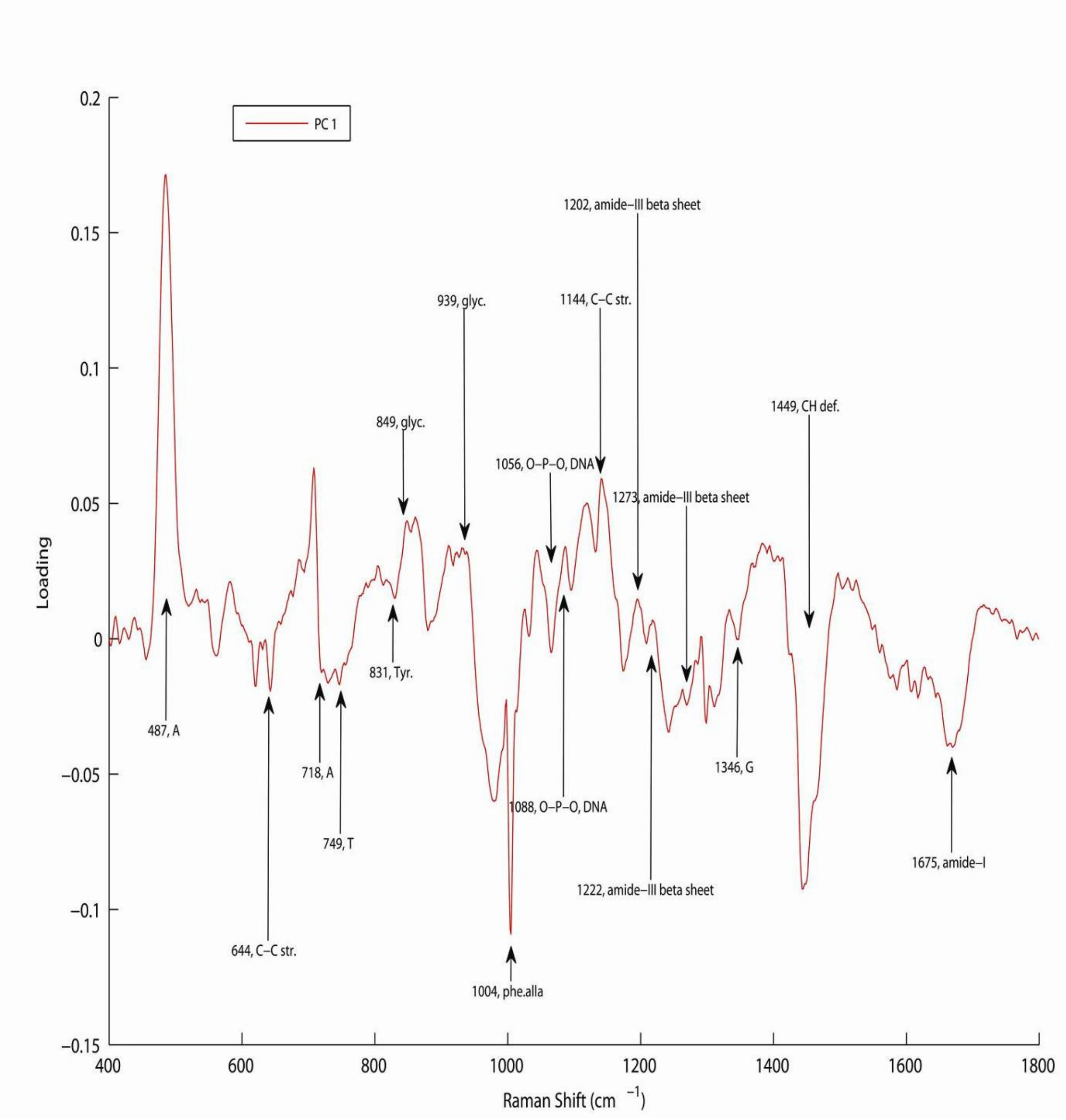
A1.8 H&E stained normal region of CIN III diagnosed tissue with scale bar of 500µm (a) unstained tissue under Raman spectroscope, with scale bar of 100µm (b) Five cluster KMCA map generated from the Raman map (c).

The loadings of the PCA associated with glycogen at 939 cm^{-1} and 849 cm^{-1} (glycogen) are positive but are not very sharp and dominant and hence do not seem to differentiate these layers. The PCA loadings related to DNA including 487 (adenine), 1346 (guanine), 1056, 1088 cm^{-1} (O-P-O of DNA) and those associated with the proteins including 831 cm^{-1} (tyrosine), 1202, 1222 and 1273 cm^{-1} (amide-III beta sheet) are contributed by the basal-5 layer.



A1.9 PCA Scatter plot for basal-true normal vs basal-5.

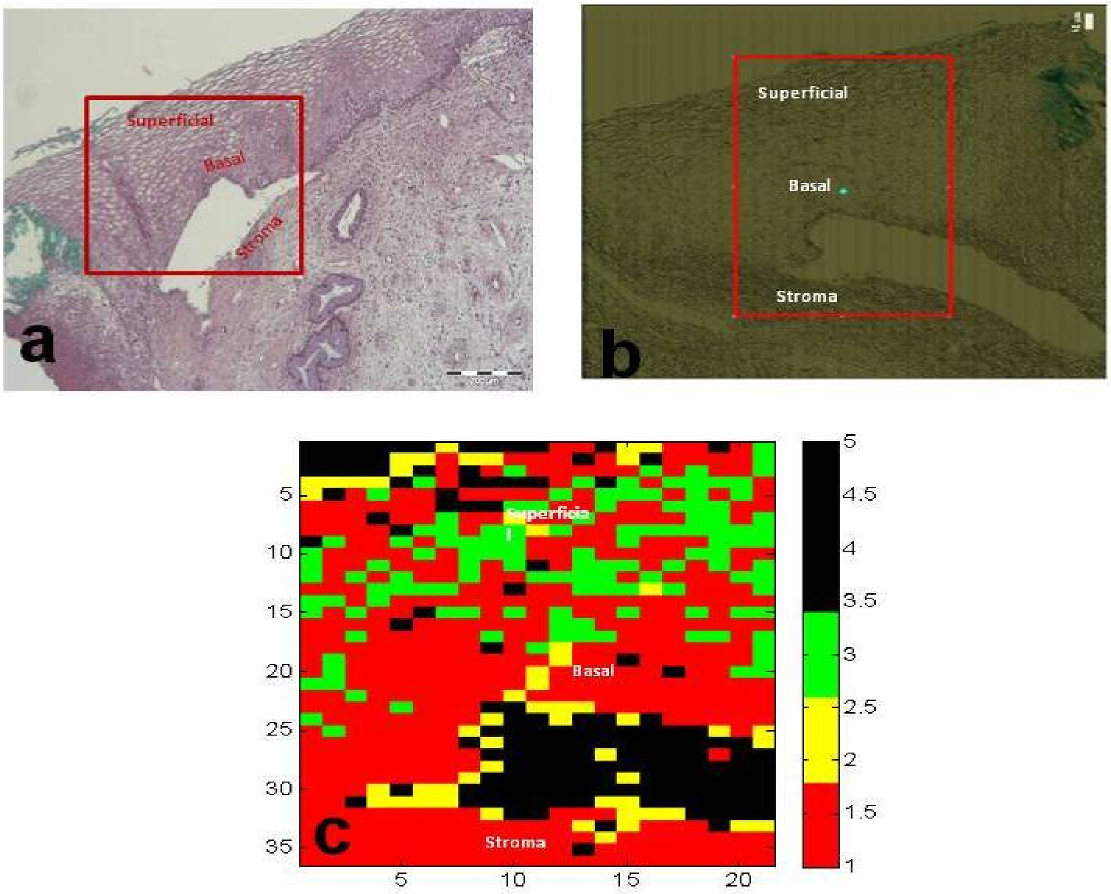
The basal-true normal layer contributed the loadings related to DNA including 718 cm-1 (adenine), 749 (thymine) and those which are related to proteins including 644 (C-C stretching), 1449 (CH deformation) and 1675 (amide-I).



A1.10 PCA loading (PC-1) for basal-true normal vs basal-5.

PCA results of basal-true normal vs basal-6

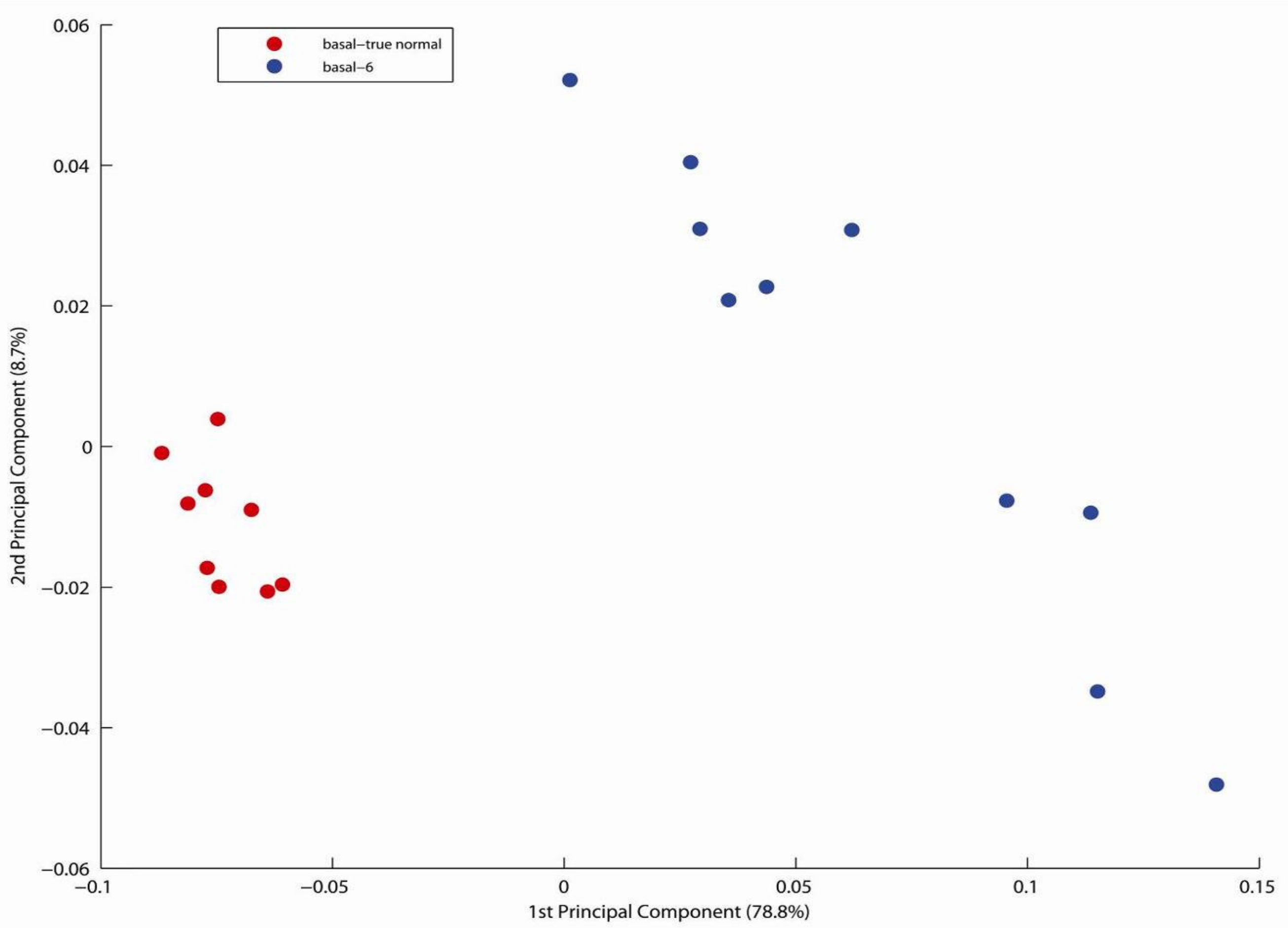
PCA of the Raman spectra from the basal layer of the normal case (basal-true normal) against the basal layer of an intermediate sample called basal-6 (**Figure A1.9**), are shown in **Figure A1.10** and **Figure A1.11** as scatter plot and loadings of the PCA, respectively.



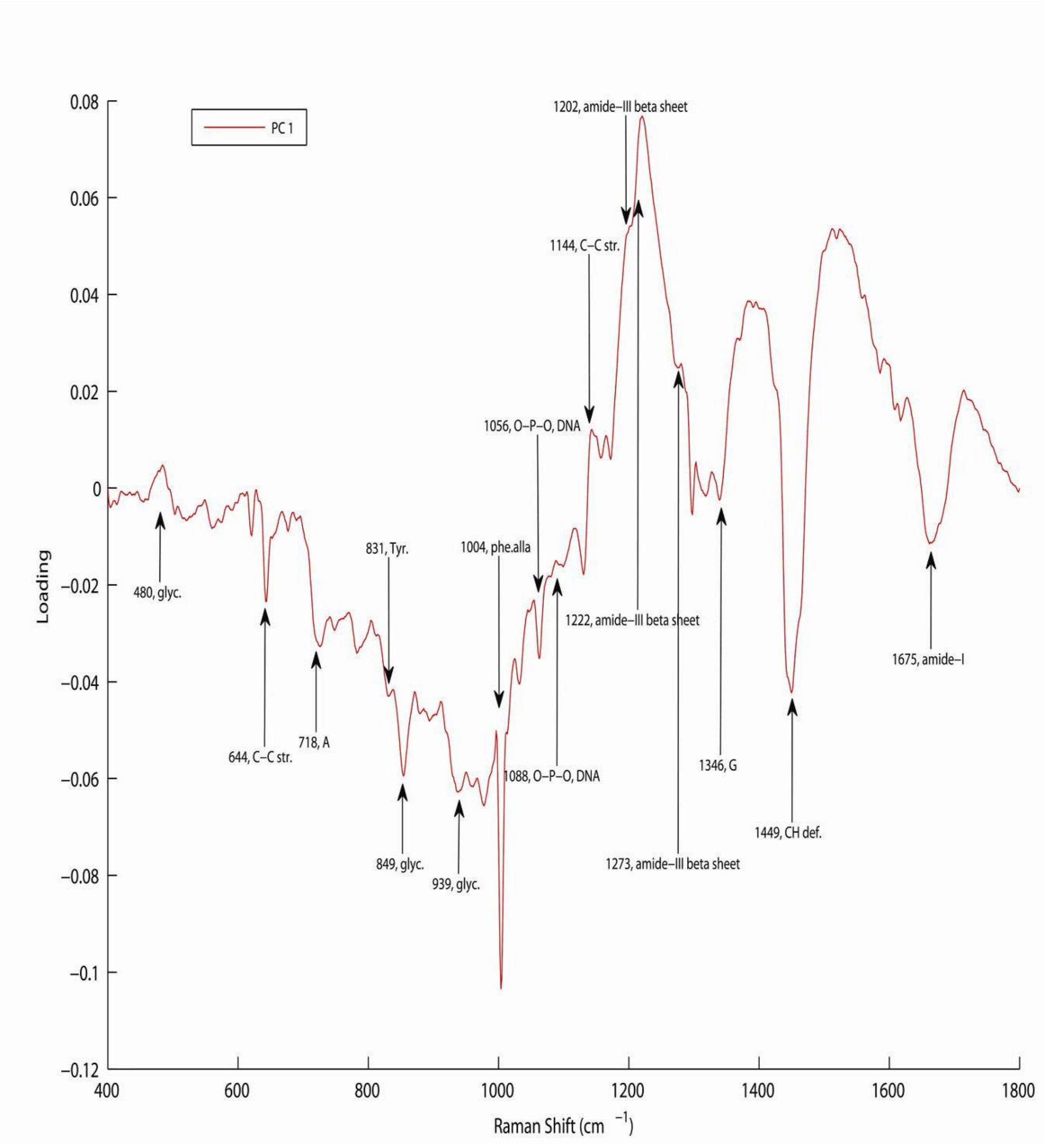
A1.11 H&E stained normal side of CIN III marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster K mean map generated from the Raman map (c).

The PCA scatter plot shows differentiation of the two groups by clustering the Raman spectra of the basal-true normal layer in the negative axis and spectra of the basal-6 layer in the positive axis of the scatter plot which indicates that negative and positive loadings are contributed by the basal-true normal layer and the basal-6 layer respectively.

The characteristic features of glycogen, 480 cm^{-1}, 849 cm^{-1} and 939 cm^{-1} are observed as negative loadings contributed by the basal-true normal layer along with other loadings including 718 cm^{-1} (adenine), 1056 cm^{-1} and 1088 cm^{-1} (O-P-O, DNA). The basal-6 contributes the loadings including 1303 (adenine), 1318 and 13456 (guanine) as well as 1202 cm^{-1} 1222 cm^{-1} and 1273 cm^{-1} (amide-III beta sheet).



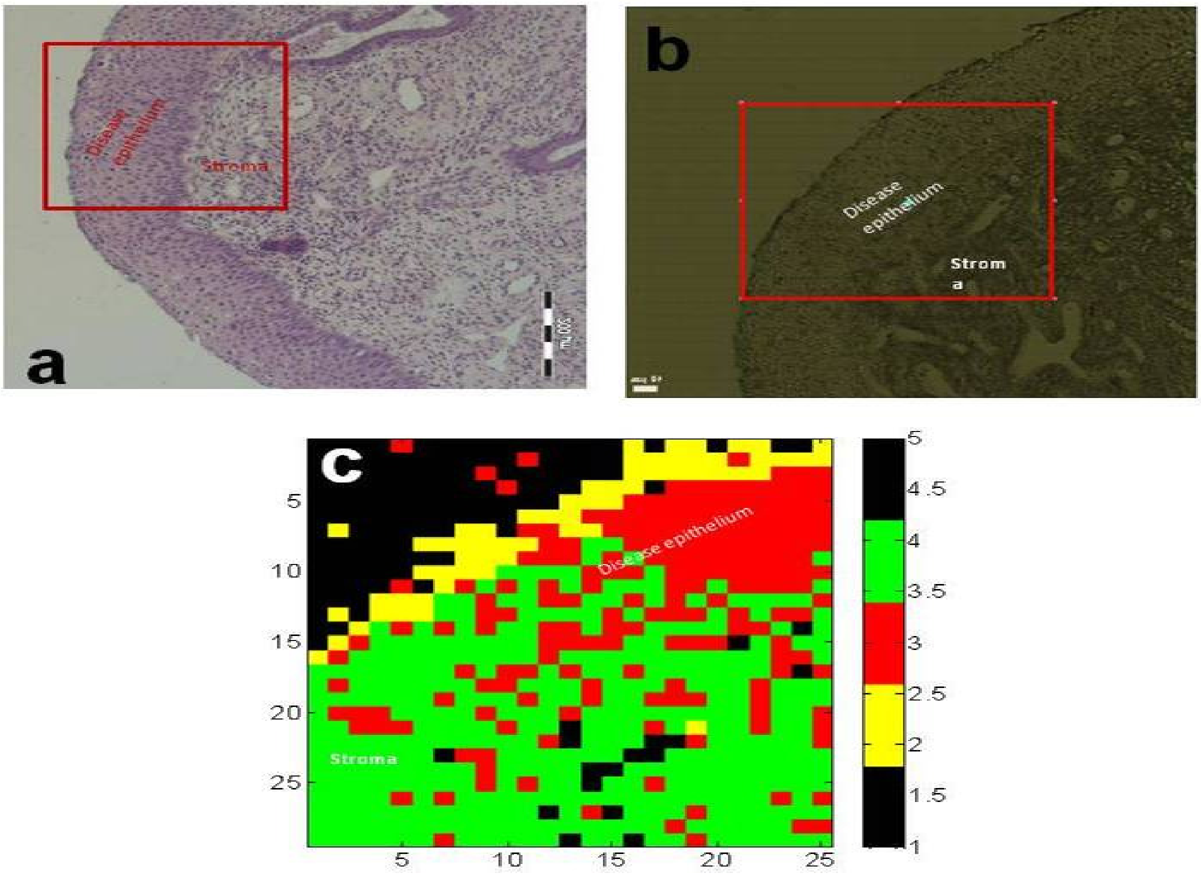
A1.12 PCA Scatter plot for basal-normal vs basal-6.



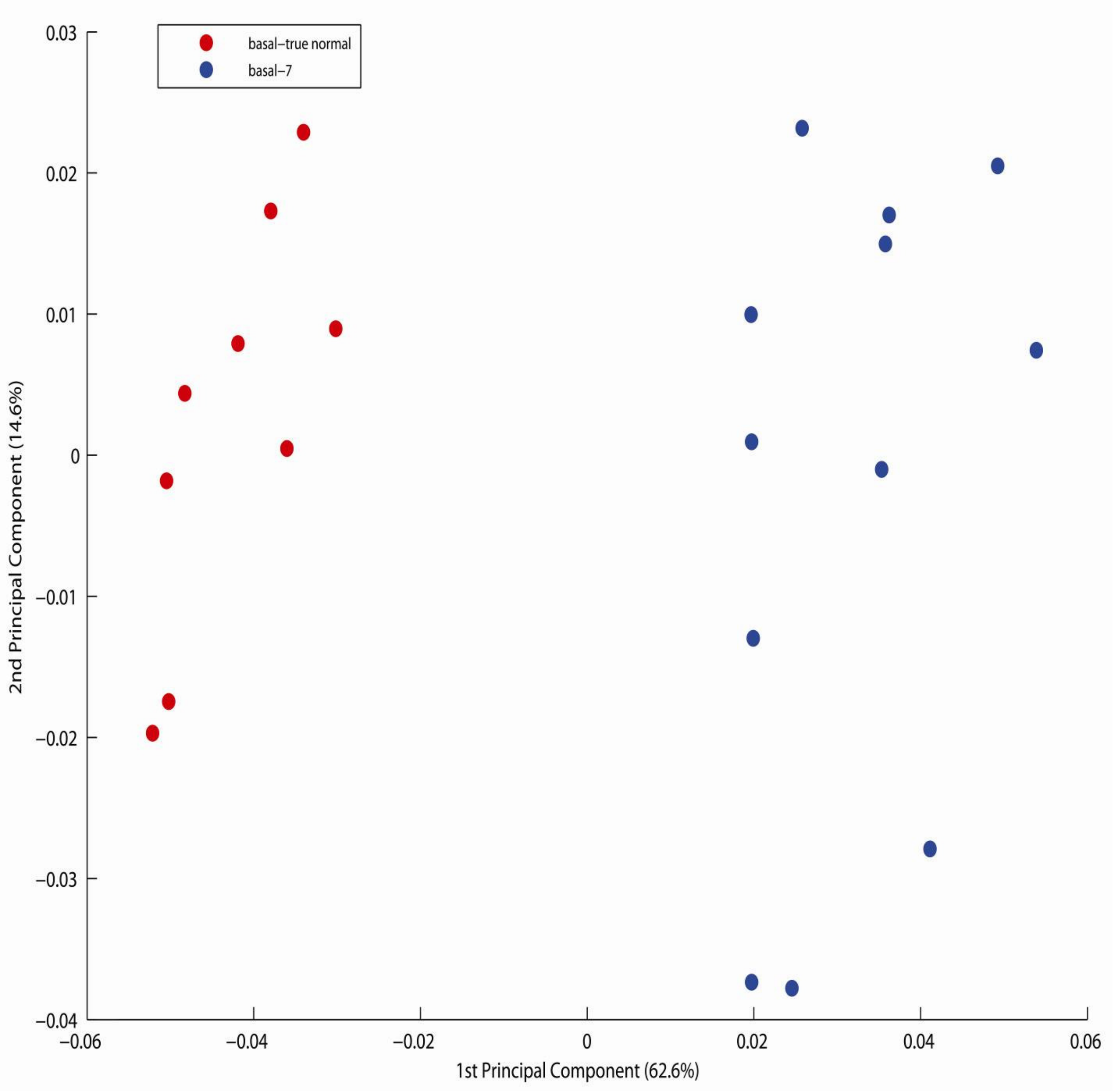
A1.13 PCA loading (PC-1) for basal-true normal vs basal-6.

PCA results of basal-true normal vs basal-7

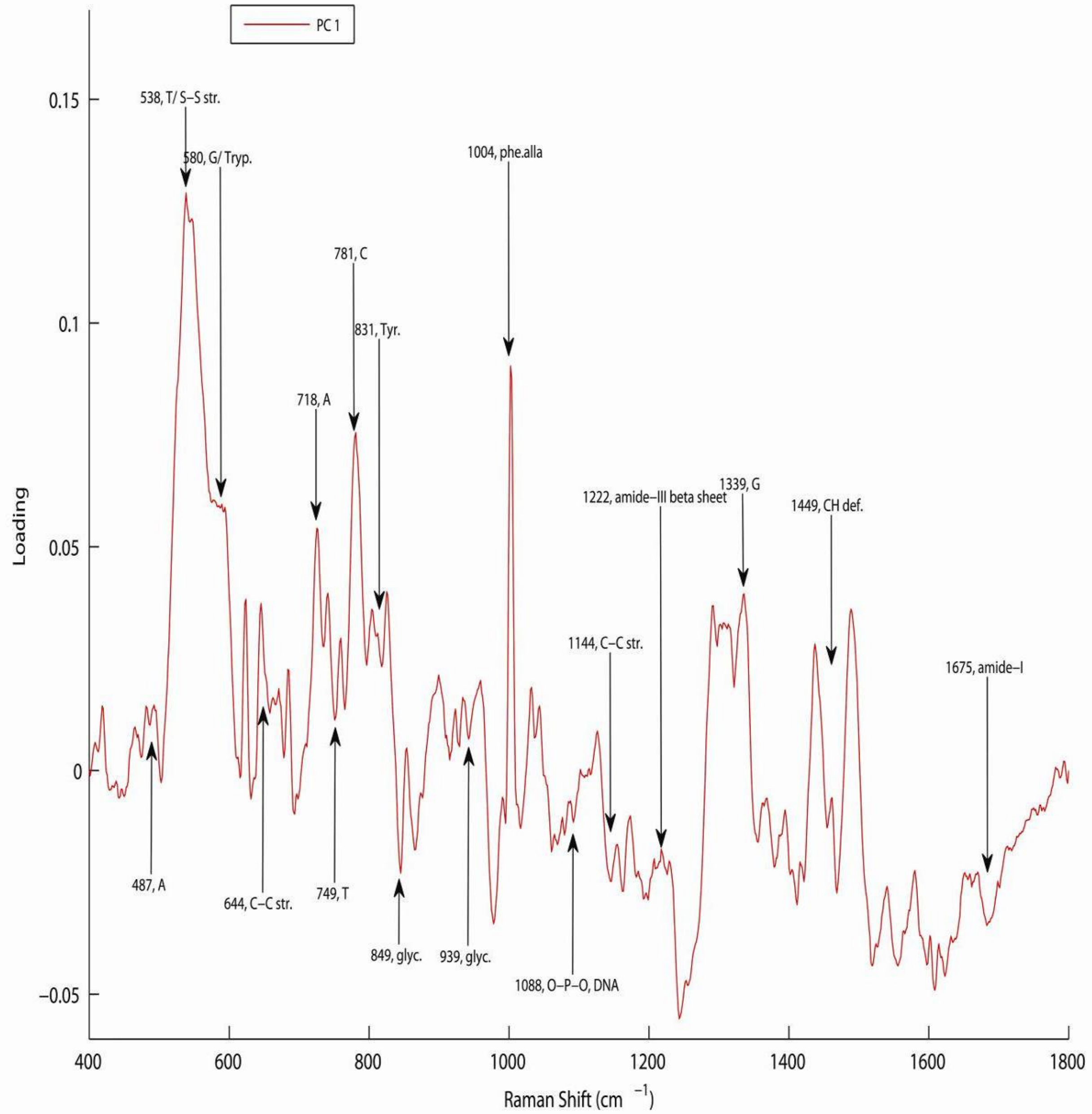
Figure A1.13 and **Figure A1.14** represent the scatter plot and loadings of the PCA respectively, for the Raman spectra from the basal layer of the normal case (basal-true normal) against the basal layer of an abnormal sample called basal-7, **Figure A1.12**. The differentiation of the two groups can be seen as the clustering of the Raman spectra of the basal-true normal in the negative axis and those of the basal-7 in the positive axis of the scatter plot.



A1.14 H&E stained disease side of CIN III marked tissue with scale bar of 200µm (a) unstained tissue under Raman spectroscope with scale bar of 40µm (b) Five cluster KMCA map generated from the Raman map (c).



A1.15 PCA Scatter plot for basal-true normal versus basal-7.



A1.16 PCA loading (PC-1) for basal-true normal versus basal-7.

List of presentations/Publications

Publications:

1. K Means Cluster Analysis for the Characterisation of Cervical Cancer Progression by Raman Spectroscopy "*In preparation*".

2. Raman Spectroscopy and Principal Components Analysis for the Analysis of Cervical Cancer Progression "*In preparation*".

Presentations (Poster/Oral)

1. Bio PIC 2010 "Compatibility of histological stains with Raman spectroscopy" (*Poster*).
2. 1st annual, DIT postgraduate research symposium (2010) "Characterization of cervical cancer with Raman spectroscopy" (*Poster*).
3. Biospec (2010), "Characterisation of Cervical Cancer Progression by Raman Spectroscopy" (*Poster*).
4. MSI (2010), Raman spectroscopy for the mapping of cervical cancer tissue samples. (*Poster*).
5. MSI (2011), "New technologies for assessment of cervical cancer progression based on vibrational spectroscopy" (*Poster*).
6. MSI (2012) New technologies for the assessment of cervical cancer progression based on Vibrational spectroscopy. (Oral presentation).