Lancaster Environment Centre | Lancaster University

In Vivo and in Vitro Applications of Biospectroscopy: Potential use as a diagnostic tool which can provide insights into cell cycle dependent proliferation

MSc (by Research) Environmental Sciences

Kirsten Esmee Snijders

Lancaster University,
Lancaster Environment Centre

Submitted September 2015

Declaration

This document has been submitted as a final dissertation for the Master of Science degree completed with the Faculty of Science and Technology at Lancaster University. It follows the rules and regulations outlined by the Faculty of Science and Technology. The author declares the originality of this document and that the work is their own. This work has not been submitted elsewhere for any other degree or award.

Acknowledgements

First and foremost I would like to thank Prof. Frank Martin for the all the time and dedication he has put into his role as project supervisor. He has helped me achieve my goals and has given me an amazing opportunity to expand my knowledge in multiple fields. His friendly but professional guidance is something I have cherished and am very thankful for.

An important thank you goes to the PhD students Holly Butler, Kelly Heys, Rebecca Strong, Blessing Obinaju, Junyi Li, George Theophilou, as well as Dr. Valon Llabjani and Dr. Nigel Fullwood. The help they generously offered me, and the time they took to teach me new techniques is greatly appreciated.

My project would not have been possible without the LBC samples provided Dr. Maneesh Singh. Thank you for giving me the opportunity to study endometriosis.

I would also like to thank the research technicians that have assisted me throughout this project. Haydn Morris for contributing his valuable expertise and helping me use the Flow Cytometry machine. Craig Delury for his time and patience during the western blots. Dawn Worrall for her kind help during PCR and Sarah Allison for letting me use the p21 antibody.

A special thank you to my fellow masters students Eleni Chrysostomou, Marina Afami and Antonis Antoniou for their support and the never-ending enthusiasm they brought to the laboratory every day.

A final thank you goes to my family for their continuous support. I greatly value their help and encouragement throughout my studies.

Abstract

Biospectroscopy and its applications characterise in vitro and in vivo cells with abnormal proliferation capacities. Proliferation of endometrial tissues in regions other than the uterus lining is known as endometriosis. It affects around 10% of women in the UK and worldwide. This study tests Fourier-Transform Infrared Attenuated Total Reflectance (FTIR-ATR) and Raman Spectroscopy as a novel approach for early identification and in depth classification of Endometriosis. An implementation of a non-invasive and high-throughput screening method would eliminate bias and thus reduce screening costs. The potential for double screening endometriosis and cervical cancer simultaneously offers a cost effective approach. Each patient spectra can be added to the data base to create a thorough in-depth identification method for varying disease morphologies. Results showed that both spectral techniques successfully detected endometriosis presence when the right pre-processing methods were applied, although Raman spectroscopy carried a higher degree of sensitivity and was able to detect ovarian endometriosis alongside endometriosis.

Thus far this study demonstrates that Benzoapyrene (B[a]P) is capable of inducing alterations in exposed Michigan Cancer Foundation-7 (MCF-7) cell populations even at very low doses of 10^{-9}M. B[a]P gradually reduces cell survival after 7 days exposure and induces long-term inhibition of cell growth. High dose exposure of 10^{-6}M B[a]P causes short-term evasion of G_1 arrest and accumulation in S phase. These findings indicate that growth kinetics in vitro are important pre-determinants of MCF-7 cell susceptibility to genotoxic agents. Biospectroscopy and computational analysis produced biochemical signatures characterising dose and cell cycle related effects of B[a]P that indicated DNA alterations being induced. Overall the cell cycle is an important factor in genotoxic synthesis and cell proliferation. These findings can be linked to abnormal tumour growth in cancer.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AhR	Aryl hydrocarbon Receptor
ATR	Attenuated Total Reflectance
B[a]P	Benzo[a]pyrene
CYP	Cytochrome P450
DLS	Dynamic Light Scattering
DMEM	Dulbecco's Modified Essential Medium
DMSO	Dimethyl Sulfoxide
EtOH	Ethanol
FBS	Foetal Bovine Serum
FTIR	Fourier-Transform Infrared
GnRH	Gonadotropin-Releasing Hormone
IR	Infrared
IRM	Infrared Measurement
LBC	Liquid Based Cytology
LDA	Linear Discriminant Analysis
Low-E	Low-Emissivity
PC	Principal Component
PCA	Principal Component Analysis
PBS	Phosphate Buffered Saline
PI	Propidium Iodide
PVDF	Polyvinylidene Difluoride
ROS	Reactive Oxygen Species
SG	Savitzky-Golay
WDN	Wavelet De-noising

1 Introduction

Spectroscopy has long been a tool for physicists and chemists to obtain a reading of chemical and molecular properties held within samples. The method of obtaining a chemical fingerprint is non-invasive and has the ability to be applied to a very wide range of materials. It has even been applied to ancient art and archaeological findings to uncover underlying properties that have long since been lost (Sandak et al., 2010). The delicate and non-destructive nature of this technique makes it the perfect tool for in depth analysis of samples that are difficult to understand superficially and offers scientists the opportunity to reuse the samples for additional examination. Spectroscopy has also been applied on plant samples to characterise the cellular structures of wood (Tsuchikawa et al., 2015). There has been a rapid shift towards the biological applications of spectroscopy (Carter et al. 2009). This field is referred to as biospectroscopy and there are many different types of spectroscopy techniques that can be used to obtain biological spectra. The most commonly used are currently infrared (IR) and vibrational spectroscopy using a light source in the form of IR light and a photon-emitting laser respectively. This study focuses on Fourier Transform Infrared Attenuated Total Resonance (FTIR-ATR) and Raman spectroscopy. The mechanisms of how these techniques work is explained in more detail in Section 1.1. Spectral measurements acquired through spectroscopy are used as characterisation tools on two tissue types originating from very different environments. One study analyses in vivo biological samples obtained from women suffering from endometriosis. The abnormal proliferation of endometrial tissue in ectopic regions is poorly understood. Examining patient samples through spectroscopy can bring new information to the medical field and can help to better characterise and diagnose the disease. The other study applies spectroscopy to cells grown in vitro. These are Michigan Cancer Foudation-7 (MCF-7) breast cancer cells that are being examined in conjunction with Benzo[a]pyrene (B[a]P). B[a]P is an environmental contaminant mainly found in tobacco

smoke, car exhaust fumes and industrial wastes (Guerreiro et al., 2016). Exposure to B[a]P during different times in the cell cycle have been shown to induce abnormal proliferation (Pang et al., 2012). Spectral analysis on these cells has the potential to uncover underlying mechanisms of genotoxicity and can help further understand the proliferative nature of tumours.

1.1 Spectroscopy

1.1.1 Spectroscopy as a Diagnostic Tool

Spectroscopy has a biomedical application in its ability to distinguish between varying tissue types (Baker et al., 2014). It detects chemical bond vibrations that indicate structural alterations of cellular molecules (Theophilou et al., 2015). It has previously been used in identifying tissues in the cervix and in other neighbouring tissue (Chiriboga et al., 1998). It can help detect signature biomarkers associated with a disease, which in turn can uncover underlying causation and aid treatment options (Gajjar et al., 2013). Endometriosis exhibits a wide variety of symptoms and phenotypes that do not directly correlate to the severity of the disease. Biospectroscopy has the potential to create a biochemical fingerprint of disease variations, (Chiriboga et al., 1998) offering the opportunity to build a new, more detailed disease characterisation method (Kraphy et al. 2009). In turn it can provide a non-invasive and more conclusive diagnostic method for women suffering from endometriosis.

FTIR and Raman spectroscopy have been successfully used as diagnostic tools for the detection of cervical cancer and the papillomavirus in several studies (Taylor et al., 2011) (Harris et al., 2009). Conventional Papanicolaou (Pap) smear screening has a sensitivity rate of 86.6% and specificity rate of 90.5%. FTIR has a sensitivity and specificity rate of 98.6% and 98.8% respectively. Furthermore when compared to the Pap smear, FTIR carries a better false-negative rate (1.4% compared to 13.4%) and a negative predictive value (96.5% compared to 72.3%) (Fung et al., 1997). This means that women suffering from endometriosis will be less

likely to receive a false-negative diagnosis and women receiving a clear diagnosis can be confident that they have been properly diagnosed.

Spectroscopy has a high throughput application that eliminates human bias and the need for manual interpretation of slides. A reduction in discarded slides and manual labour proves to be very cost effective. This could allow for a lower screening age to be implemented in the UK (Walsh et al., 2007). A preliminary study showed that FTIR spectroscopy was able to distinguish between eutopic endometrial tissue found in patients with and without endometriosis (Cheung et al., 2011). This shows promise for spectroscopy being used as a diagnostic and screening tool for endometriosis. Spectral analysis on cervical samples offer the potential of multiple disease screenings, as samples can be revisited and spectral data can be screened for endometrial cancer and endometriosis simultaneously.

1.1.2 Using Spectroscopy for Characterisation

Spectroscopy has the potential to detect molecular alterations and differences within samples to uncover their biological fingerprint. FTIR-ATR and Raman spectroscopy have been used to extract biological information from many different human sample types (Trevisan et al., 2012). Biospectroscopy has been successfully been used to distinguish between subtypes of human cancer, and can distinguish between subclasses of ovarian cancer (Theophilou et al., 2016). It has also been applied to corneal epithelial cells to identify putative stem cells, transit-amplifying cells and terminally-differentiated cells (Fogarty et al.,(2013). Many types of cell populations can be characterised through IR biospectroscopy (Martin et al., 2010).

Biospectroscopy is an incredibly sensitive tool that is even able to detect differences between base compositions of oligonucleotides (Kelly et al., 2009). It therefore has very strong characterisation abilities and is an essential tool in many biological investigations.

1.1.3 FTIR-ATR Spectroscopy

FTIR spectroscopy works by shining a beam of infrared light onto a sample. Conventional IR spectroscopy can only shine one monochromatic beam at a time and takes measurements of several frequencies over a period of time (Butler et al., 2014). FTIR has the advantageous ability of exposing a sample to a multitude of frequencies simultaneously with one beam. This results in a high throughput method that acquires multiple wavelength readings over a large spectral range in a short amount of time. FTIR spectroscopy was coupled with an ATR attachment. This piece of equipment is fitted with a crystal that allows the FTIR machine to take measurements directly from solid or liquid samples. The crystal uses pressure to make direct contact with the sample. An IR beam passes through the ATR crystal at a 90° angle and is continuously reflects along the sample and the crystal in a process called total internal reflection. When the IR beam hits the sample an evanescent wave is formed which decays exponentially as it travels further into the sample. This process can be visualised in the schematic in Figure 1.

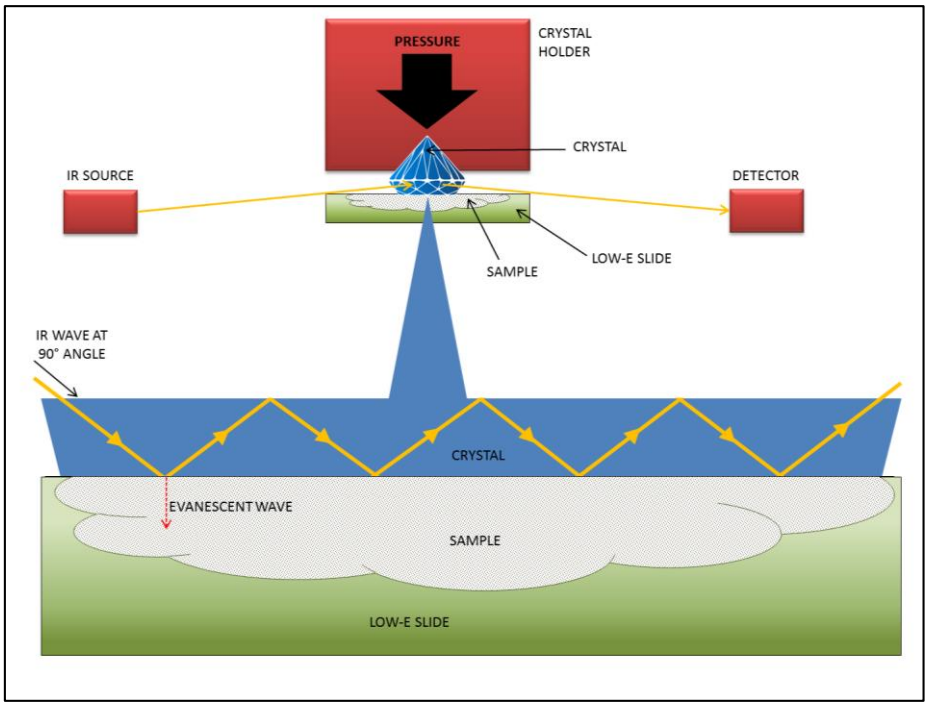


Figure 1 – A visual representation of FTIR spectroscopy with an ATR attachment containing a crystal. An infrared light source shines through a crystal where it continuously reflects between the sample and the diamond until it reaches a detector. An evanescent wave is formed where the IR light comes in contact with the sample and gets absorbed. Pressure is used to create contact between the crystal and the sample mounted onto a low-e slide

Pressure is applied in order for the ATR crystal to make contact with the sample. This ensures that air does not get trapped between the sample and the crystal, as this would result in an inaccurate measurement as the evanescent wave gets distorted by the air (Sneppen et al,. 2010). Light that reaches the samples through the evanescent wave is absorbed in different quantities depending on different molecular and cellular structures within the sample. The absorbance capacity is quantified into an absorbance spectrum containing individual wave numbers (also known as reciprocal centimetres, expressed in cm^{-1}). Proteins and molecular structures within samples have been matched with specific wavelength numbers (Movasaghi et al., 2008). Since ATR relies on contact pressure, the sample could get damaged as it could get removed from the slide if it were to stick to the crystal, however, the damaged area would be minimal as it would be less than 2µm wide.

1.1.4 Raman Spectroscopy

Another spectroscopy machine is the Raman photometer. This spectroscopy technique relies on a laser to beam photons onto the sample source. The laser is projected through a series of mirrors in order to reach the sample. Once they do so the photons are absorbed by the sample's molecules and are subsequently reflected back; a process known as Raman scattering (Butler et al., 2016). During photon absorption, sample molecules get excited and vibrate at different rates depending on their chemical compositions. The absorbed photons experience a shift in their energy levels as some of the energy is retained in the excited sample. The reflected photon beam is directed by mirrors through a filter and a focusing lens or grating, before reaching a photometer. The described mechanism of Raman spectroscopy can be visualised in the schematic shown in Figure 2.

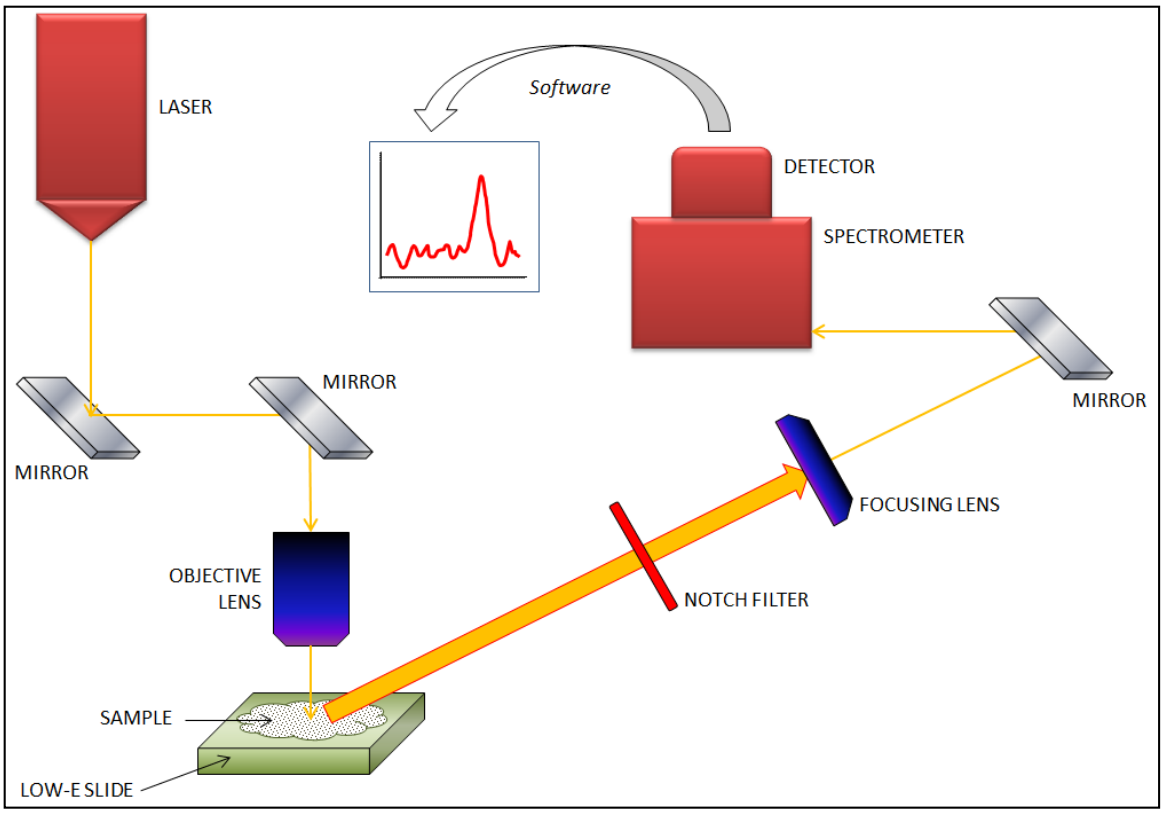


Figure 2 – A visual representation of the mechanism employed during Raman spectroscopy. A laser beam is guided by mirrors to a sample mounted on a low-e slide. The laser is reflected off the sample and concentrated through a notch filter and focusing lens towards a spectrometer where sample data is detected and converted by software for analysis

Only around 1% of the laser emissions result in Raman scattering as the majority of the photons do not get absorbed but instead directly reflect off the sample in a process known as elastic or Raleigh scattering. Filters and grating techniques are often used to optimise the amount of pure Raman scattering being received and reflected. Surface enhanced Raman scattering has also been used to increase Raman scattering efficiency. This utilises nanoparticles embedded on top of the sample slides to capture background fluorescence and significantly reduces the signal to noise ratio. Raman scattering is detected in the form of Raman shifts that are converted into wavenumbers (cm^{-1}) and allow us to obtain information about the underlying molecular structures of samples. Biological molecules such as nucleic acids, lipids and proteins can all be paired with specific wavenumber to create a biochemical fingerprint of the sample (Movasaghi et al in 2007). The laser used in Raman spectroscopy has the potential to burn through the sample and result in sample damage. The sample can't be reused and the reading obtained with too high a laser power will be inaccurate as the sample morphology changes when it gets burnt. It is important to test the machine's acquisition parameters on all new sample types to determine the appropriate custom setting.

1.1.5 Sample Preparations

In order to expose samples to spectral beams they need to be mounted onto low-e slides. These resemble conventional glass microscope slides but are coated with Ag/SnO_2. The slides provide a fantastic tool for carrying the samples without contributing any chemical signals to the readings. They are almost entirely transparent to visible light (from $4000cm^{-1}$ to $400\ cm^{-1}$) and therefore do not produce any interfering absorptions. Whilst low-e slides do not obscure sample readings, they carry the capacity of reflecting mid infrared radiation efficiently (Severcan & Haris, 2012). This makes them crucial components in FTIR-ATR spectroscopy. Samples are regularly fixed directly onto the slides in fixative mediums such as Thin Prep, ethanol (EtOH) or paraformaldehyde.

1.2 Endometriosis

1.2.1 Sample Origin and Study Participants

Biospectroscopy was applied to patient samples to test its application as a diagnostic tool for detecting endometriosis. The women participating in the study were being treated for pelvic pain or sub fertility induced by endometriosis. A total of 43 patient samples were collected from women aged 16 to 67 years suffering from various degrees of endometriosis. Disease stages ranged from 1-4 (according to the American Society for Reproductive Medicine's grading system shown in Figure 3) and showed a manifestation of both red and black lesions. Ovarian endometriosis was present in 7 patients; all of whom were in stage ≥ 3. This was an anonymous blind study that included 11 endometriosis free patients as negative controls.

Spectral data obtained from endometriosis samples was divided into subclasses for in depth analysis, with the aim to uncover biochemical trends associated with different aspects of the disease.

These subclasses are as follows:

- Endometriosis presence (no endometriosis vs. endometriosis-free control)
- Disease stage (1-4)
- Disease morphology (red or black lesions)
- Ovarian endometriosis presence

Cervical cytology samples were obtained through the use of a Tao Brush on the cervix. The collected cell material was stored at 37°C in thin prep solution. Full consent forms were completed and samples were obtained with local Institutional Review Board (IRB) approval from the Gold Coast, Australia.

1.2.2 Endometriosis Pathology

Endometriosis is defined as the occurrence of endometrial-like tissue, specifically glands and stroma, growing outside the uterus. Whilst the tissue greatly resembles endometrium, it remains histologically different. Its presence causes chronic inflammation, scar tissue, adhesions and lesions that grow and bleed in response to hormonal changes. Endometriosis is usually located within the pelvic region and has been found in all the pelvic organs such as the bladder, uterus, ovaries, fallopian tubes, bowel and peritoneum. It can also present itself outside the pelvic cavity and has been documented in women post hysterectomy and in extreme cases has been discovered behind the knee and in the brain (Machairiotis et al., 2013).

1.2.3 Causation

Whilst the true cause of endometriosis still remains unclear, several proposed theories exist. Sampson was the first to name and outline the pathogenesis of endometriosis (Sampson, 1927). The first proposed cause of endometriosis examined the theory of retrograde menstruation (Speert, 1958). The retrograde flow of menstrual content through the fallopian tubes causes the dissemination of endometrial tissue to the peritoneum and ovaries. Since retrograde flow is a common occurrence in most menstruations and endometrial cells are not histologically identical to endometriosis lesions, this theory has since been disputed (Brosens & Benagiano, 2011).

Currently the most widely acknowledged theory focuses on oestrogen dominance, where the level of progesterone is greatly outweighed by oestrogen. This correlates to the treatment for relieving endometriosis symptoms through the reduction of oestrogen production (Leyendecker at al., 1998).

It has also been suggested that Metaplasia is in some way related. This is the process where cells have the ability to change into tissue types that are different to their original state. Adult

endometrial cells could have retained the ability to transform into reproductive tissue; a trait that was in place in the embryonic stage during uterus development (Latham Macer & Taylor, 2012.)

Endometriosis may be an inherited disease. First-degree relatives of women suffering from endometriosis have a higher chance of being diagnosed with the disease. The suspected genetic pre-disposition has been attributed to a loci near chromosome 7p13-15 (Zondervan et al., 2007).

The dispersal of endometrial fragments through lymphatic or vascular pathways could allow endometrial tissue to establish itself around the body (Sourial, Tempest & Hapangama, 2014). This accounts for the presence of endometriosis outside the pelvic cavity, such as the eyes, brain and lungs.

1.2.4 Symptoms and Treatment

The American Society for Reproductive Medicine has established a classification chart to help quantify the progression of endometriosis (Figure 3). It is based on the presence and severity of endometriomas or adhesions, the extent and depth of the disease, as well as its location. Through a numerical scoring system, based on physical pathogenesis, endometriosis can be categorised into Stage 1 (minimal), Stage 2 (mild), Stage 3 (moderate) and Stage 4 (severe). Black lesions, ovarian endometriosis and infertility are associated with the later more severe stages (American Society for Reproductive Medicine, 1997).

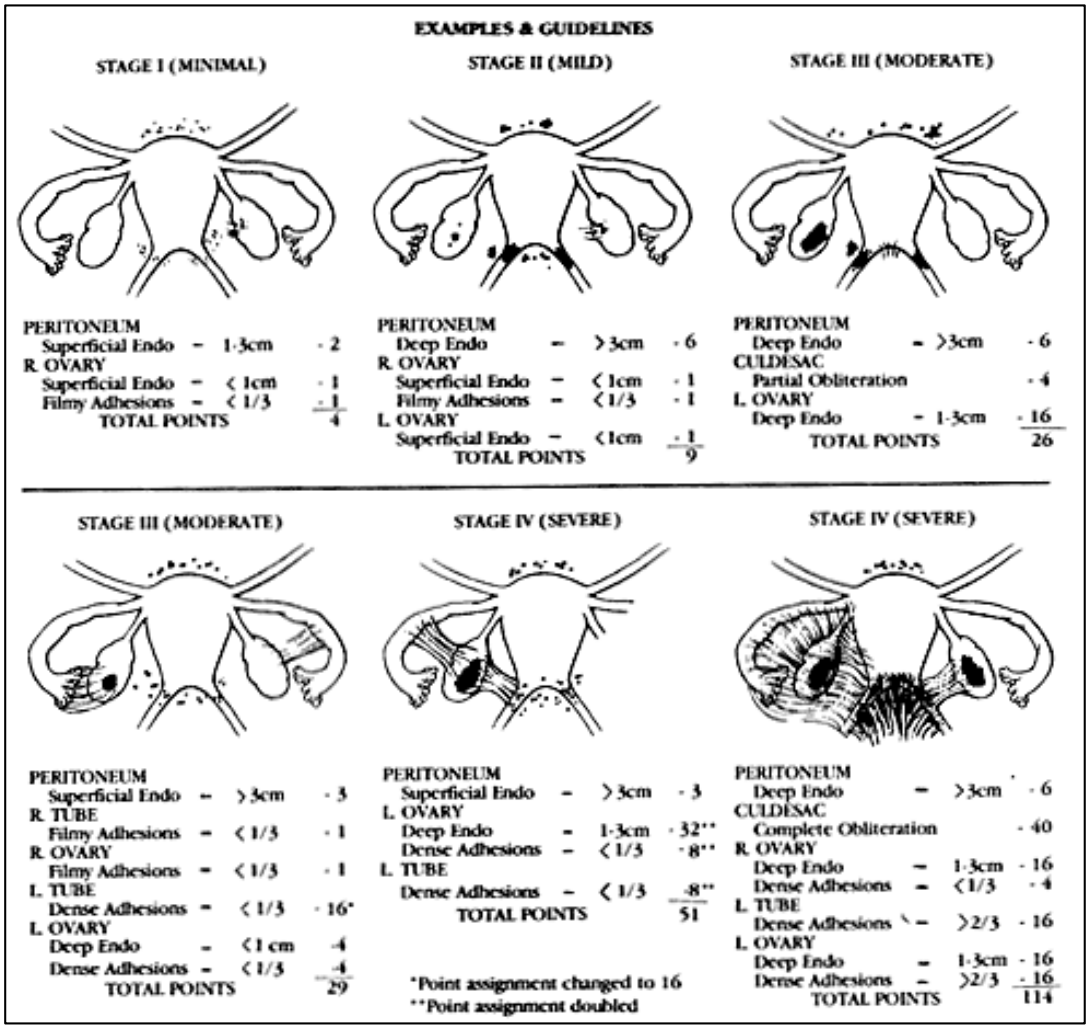


Figure 3 – A visual guide to classify endometriosis stage following a numerical scoring system, by the American Society for Reproductive Medicine, 1997

There is currently no conclusive cure for endometriosis but there are several options aimed at managing and reducing its symptoms. Removal of endometriosis through laproscopic excision or ablation can be successful, although the latter works best on red lesions found in the earlier stages. Triggering a temporary medical menopause through the administration of hormonal medication can help to shrink lesions over time (Muñoz-Hernando et al., 2015). Contraceptives containing gonadotropin-releasing hormone (GnRH) analogues or progestins are widely used. These can be released through intra uterine devices, injections, tablets or nasal sprays.

Even though the endometriosis tissue is benign, it causes great discomfort and is accompanied by symptoms of pelvic pain, dysmenorrhea, and dyspareunia. 10% of the female population

has been diagnosed with endometriosis (Gruppo italiano per lo studio dell'endometriosi, 1994). 20–25% of these patients are asymptomatic and many women's symptoms are misdiagnosed (Bérubé, Marcoux & Maheux, 1998). Disease prevalence is therefore likely to be greater than 10%.

The probability of a successful diagnosis increases with age, as the symptoms grow more severe. If the presence of endometriosis is suspected, it can only be confirmed through laparoscopic surgery. This is currently the only conclusive diagnostic method for the disease.

1.2.5 Current Diagnostic Methods

Unlike endometrial cancer, there are currently no active screening methods for the detection of endometriosis. MRI scans, ultrasounds, and gynaecological examinations can be suggestive of the disease but will not detect smaller lesions. Elevated CA-125 serum levels can be an indicator of endometriosis, however this method of diagnosis only has a sensitivity of 61.1% and a specificity of 87.5% (Chen et al., 1998).

Diagnosing endometriosis by laparoscopic surgery involves a keyhole surgery through the abdomen to examine the affected region with a light-bearing microscope. Surgery is very invasive and carries the risk of complications and side effects (Kennedy et al., 2005). Spectral analysis on LBC samples avoids the need for surgery as a diagnostic tool and offers a non-invasive and safe screening method that would allow women to detect endometriosis at early stages when the disease is easier to maintain.

1.2.6 Liquid Based Cytology

Endometrial samples are collected through several devices made up of varying materials and shapes. These include Novak's curette, Vabra, Pipelle, Masterson's curette, Accurette, Endorette, Gynoscann, Leicester, endometrial needle, Cornier Pipelle, Explora and the Tao Brush. The majority of these have a diagnostic accuracy for detecting endometrial cancer of 80% (Rullo et al., 2015).

There is however a vast difference in patient discomfort associated with these different collection methods. The Pipelle is the most commonly used sampling method and involves the collection of a biopsy. This process can be very painful and is concentrated on one location. Therefore small localised spots of cancer or endometriosis are likely to be excluded from the sample. The Tao Brush is gaining increased popularity in replacing the conventional smear test for cervical cancer screening (Zhu et al., 2007).

The Tao Brush is protected from vaginal and endocervival contamination through a protective sheath and it covers a wide collection area. This results in a sensitivity rate of 95.5%, and up to 100% specificity when detecting endometrial carcinoma (Del Priore et al., 2001). The method of preparing and processing cervical samples in suspension is known as liquid based cytology (LBC). This offers a clearer representation of the samples as it prevents blood and mucus from concealing disease presence during microscopic examination (Williams et al., 2008). The cost effectiveness, increased sensitivity, and the option to re-use samples are encouraging a switch to LBC (Cox, 2004). This movement allows for biospectrocopy techniques to be applied more readily in the medical field. It offers the potential to provide doctors with a multiple screening tool that is non-invasive and carries a higher accuracy of diagnosis.

1.3 MCF-7 Cells

1.3.1 Study Overview

MCF-7 cells were treated with a range of B[a]P concentrations to understand its genotoxic synthesis and to uncover the mechanisms behind B[a]P induced G_0/G_1 arrest evasion. MCF-7 cells have previously shown an evasion of G_0/G_1 arrest following a B[a]P treatment (Davis et al,. 2002).

MCF-7 cells grown in culture were screened for protein expression of Cytochrome P450 (CYP) isoforms; CYP1A1 and CYP1B1. These are genes encoding for catalytic enzymes involved in the

conversion of pro-carcinogens into carcinogens, or genotoxins. They have been found to be present at elevated levels in many cancer patients. Previous studies have shown that the cell cycle plays in the homeostasis of proliferating tissues (Davidson et al., 2012) and the susceptibility of B[a]P (Hamouchene et al., 2011).

Cells were grown and arrested in different phases of the cell cycle to identify how G_1 arrest and quiescence can be evaded, whilst demonstrating how expression of CYP is altered at different stages of the cell cycle. The cell cycle concentration will follow the method close to the one used by Jiao et al., (2007), where flow cytometry was performed in order to confirm that the cells are distributed in the correct phases. Biospectroscopy was applied to determine whether there are apparent spectral differences between different culture and treatment methods of MCF-7 cells, and whether these differences can be categorised with a specific biochemical fingerprint.

1.3.2 Benzo[a]pyrene

Benzo[a]pyrene (B[a]P) is an environmental contaminant mainly found in tobacco smoke, car exhaust fumes and industrial wastes (Guerreiro et al., 2016). It has pro-mutagenic and pro-carcinogenic properties and damages DNA through the formation of DNA adducts (Genies et al., 2016). It has been found to be non-toxic until it enters the plasma membrane and binds to the Aryl hydrocarbon receptor (AhR). Once bound, it activates AhR causing it to translocate into the nucleus where it dissociates (Figure 4). Active AhR is responsible for the regulation of transcriptional responses and therefore involved in controlling the expression of many genes (Bersten et al., 2013).

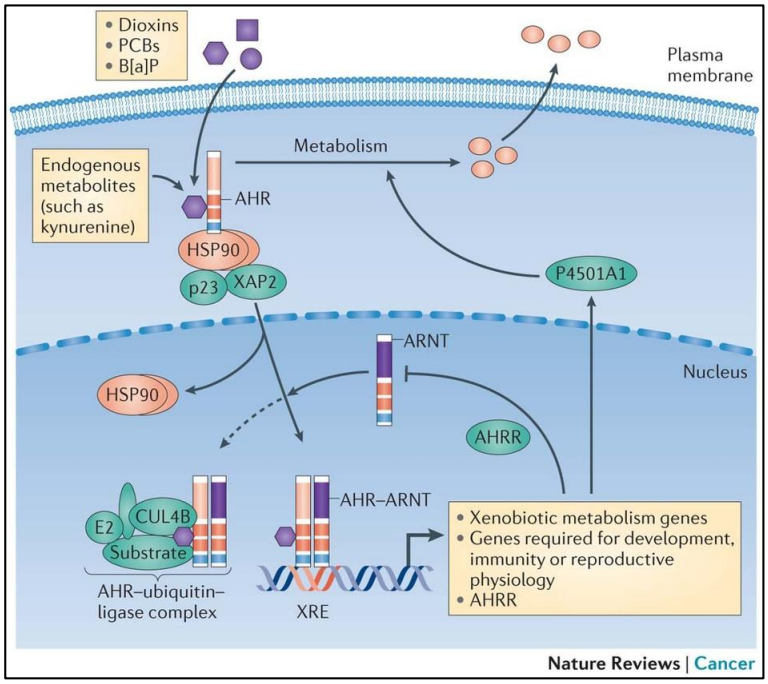


Figure 4 – A representation of B[a]P uptake at the cell membrane. Binding of B[a]P activates the AhR receptor allowing for xenobiotic metabolism of genes. (Figure obtained from Bersten et al., 2013)

1.3.3 Cytochrome P450s

This particular research project will focus on the expression of CYP isoforms *CYP1A1* and *CYP1B1*. These enzymes belong to a large superfamily containing over 11500 protein variations; they have catalytic activity and mediate oxidation of organic substances. The number following CYP refers to the gene family (CYP1) whilst the adjacent capital letter refers to the subfamily (CYP1A/ CYP1B). The final number indicates which specific gene is being referred to (CYP1A1). Italics are used when discussing the gene with regards to the enzyme it encodes for; *CYP1A1* gene encoding CYP1A1 enzyme. The Human Genome Project identified 57 individual genes coding for CYP enzymes (2004). In humans CYPs are often found in the cell's endoplasmic reticulum or by the inner membrane of mitochondria (Berka et al. 2011). Past research has indicated that CYP isoforms are involved in cancer formation and treatment. This concerns prostate, breast, ovarian, kidney and many other forms of cancer (Rodriguez-Antona, 2006). This is due to CYP's metabolic activity of exogenous and endogenous substrates, often aiding the oxidative conversion of pro-carcinogens into genotoxic

metabolites (Jiao H, 2007). It appears that a variation of CYP expression influences the susceptibility that the target cells have towards pro-carcinogens. Gajjar et al (2012) performed a meta-analysis to determine whether there was indeed a link between the up-regulation of CYP polymorphisms and an increased risk of ovarian cancer. It focused on hormone-induced carcinogenesis since CYP1B1 catalytic activity (converting non-toxic E2 into 4OH2 E2 which has active genotoxic activity) is affected by the levels of oestrogen present. The meta-analysis showed a weak trend where homozygous Leu432Val (CYP1B1 Polymorphism) expression brought reduced risk, but not enough significance was found overall and the study remained inconclusive.

1.3.4 MCF-7 Cell Morphology

The types of cells that will be cultured in this research project are MCF-7 cells. They are an immortal breast adenocarcinoma cell line originating from Detroit and established by Soule et al., in 1973. They are ideal for research use because of their resilience, and they are also easy to grow and hard to kill. They also incorporate oestrogen receptors in their cell cytoplasm and can process oestrogen in the form of oestradiol (Lykkesfeldt & Briand, 1986). Oestrogen receptors are known to interact with the AhR during gene regulation (Matthews & Gustafsson, 2006). Since B[a]P is an exogenous ligand for the AhR, using an oestrogen receptor positive cell line to study B[a]P genotoxicity is crucial.

MCF-7 cells grow in a monolayer and exhibit strong cell-cell adhesion. Their morphological phenotype has been described as resembling cobblestones (Lombardo et al 2014). Figures 5 and 6 show cells grown for 24 hours and 96 hours respectively. For imaging purposes 1 x 10^5 cells were grown directly on low-e slides submerged in a 60 x 15mm petri dish filled with media.

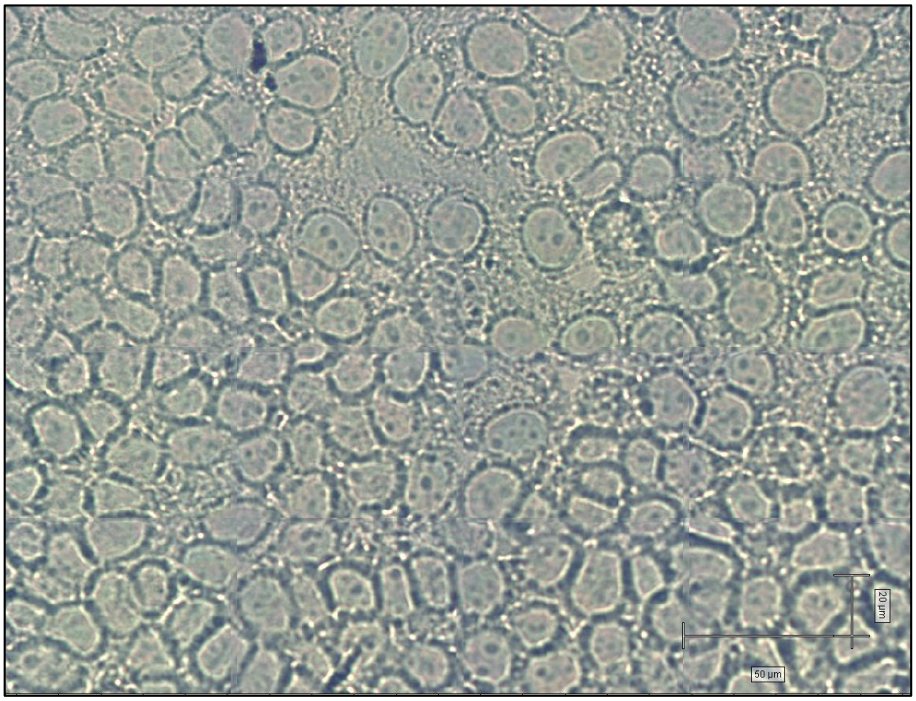


Figure 5 – S phase concentrated cells grown for 24 hours imaged on a low-e slide. The scale bar shows 50µm and 20µm

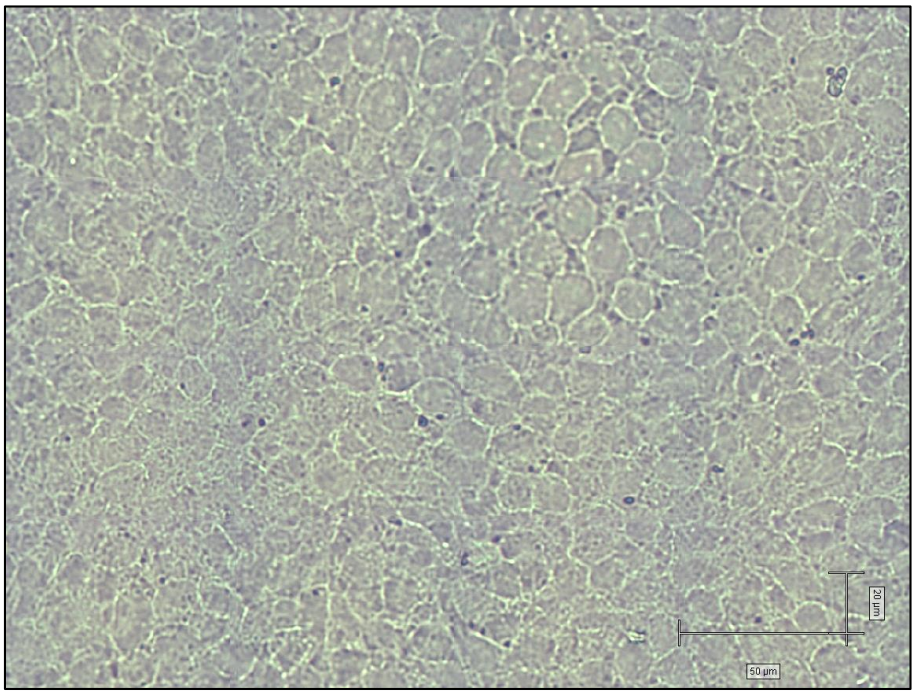


Figure 6 – G_0/G_1 concentrated cells grown for 96 hours imaged on a low-e slide. The scale bar shows 50µm and 20µm

1.3.5 Importance of Cell Cycle Stage in Experimental Studies

This research project aims to observe the molecular delineation of CYP isoforms (CYP1A1 and CYP1B1) by administering B[a]P at different concentrations in conjunction with the manipulation of the cell cycle. In the cell cycle the G_1 phase is where cell growth occurs whilst the S phase controls replication and DNA synthesis. Cells have to pass the restriction point, between G_1 and S phase, in order to undergo replication and later division. If the restriction point is not passed then cells remain in G_1 arrest, or quiescence, and are unable to pass into the subsequent phases. An illustration of the cell cycle can be seen in Figure 7.

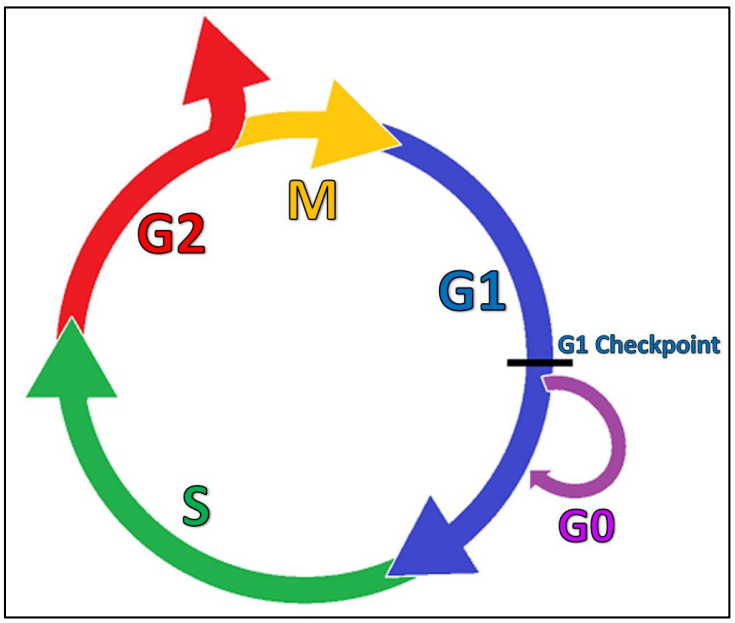


Figure 7 – An illustration showing the progression of cell cycle phases starting at G1

Failure to regulate the cell cycle restriction point accordingly results in the avoidance of arrest by cells (Foster et al., 2011). This characteristic has been linked to the uncontrolled growth of tumours, as the cells are able to continuously re-enter the cell cycle (Zölzer, Mußfeldt & Streffer, 2014). The choice to manipulate the cell cycle in this research project relates to a previous studies conducted by Jiao in 2007, which also incorporated CYP1A1 expression. He indicated that tissue turnover is a potential important factor influencing the susceptibility of target cells to genotoxins. Jiao (2007) hypothesised that avoidance of G_1 phase arrest

influences transforming activity of genotoxins (DNA-adduct forming carcinogens). His findings showed that the majority of cells exposed to non genotoxins would avoid G_1 arrest and accumulate in S phase. After exposure genotoxic B[a]P induced the up-regulation of CYP1A1 expression, this effect was significantly highest in S phase concentrated cells. This shows that the CYP expressing cells are able to avoid quiescence and become highly replicated in S phase. The consequences of this are that an increased cell population is grown which expresses a CYP gene which will go on to catalyse pro-carcinogens and cause DNA damage. These experiments show that the growth kinetics of cells play an important role in the expression of CYPs. Being able to observe the effect that different expression levels of CYP isoforms have, in conjunction with their place in the cell cycle, gives insight to the way in which genotoxic synthesis is driven and in what ways quiescence is avoided.

Another important regulator of the cell cycle is p21 (WAF1/CIP). This is a p53-dependent gene and is strongly involved in controlling the cell cycle arrest in G_1 phase (Overton et al., 2014). A relationship has been observed between p21 and bcl-2, where p21 suppression can be induced when bcl-2 is over expressed (Dai et al., 2016). This results in an evasion of arrest and bcl-2 over expression also inhibits apoptosis. It is not surprising that bcl-2 is often found at elevated levels in tumours, as these cells have overridden the natural cell cycle defence mechanisms to halt cell proliferation.

1.4 Aims and Objectives of Current Study

Overall this research project aims to bring new insights into ways in which spectral techniques can be applied. Other aims and objectives of this study are as follows:

- Use FTIR-ATR and Raman spectroscopy as a novel tool to screen patients for endometriosis.
- Provide a non-invasive procedure that could increase the detection rate and eliminate surgeon bias during surgical diagnosis.

- Reduce the number of undiagnosed women due to the varied symptoms associated with the disease.
- Identify the different spectral regions to lead the way in uncovering the aetiology of the disease.
- Gain a better understanding of the mechanism of avoiding quiescence to get the closer to the goal of providing new cancer treatments and cures.
- Demonstrate that B[a]P synthesis is cell cycle dependant.
- Outline the proliferative function of CYP isoforms during B[a]P genotoxic synthesis.

2 Materials and Methods

2.1 Biospectroscopy Sample Preparation

In order to apply spectroscopy techniques to samples, the samples need to be fixed and suspended onto reflective low-e slides. The LBC samples were collected and stored in Thinprep fixative, eliminating the need for a fixation step. 5ml of the LBC samples were vortexed then centrifuged at 1500rpm for 5 minutes. The supernatant was aspirated and the pellets underwent 3 washes using 3ml of autoclaved double distilled water and were spun at 1500rpm for 5 minutes. After aspirating the supernatant the pellet was re-suspended in 1ml of autoclaved double distilled water and pipetted onto previously cut low-e slides (1cm by 1cm). These were left to dry overnight and were then stored in a desiccator awaiting analysis (Figure 8).

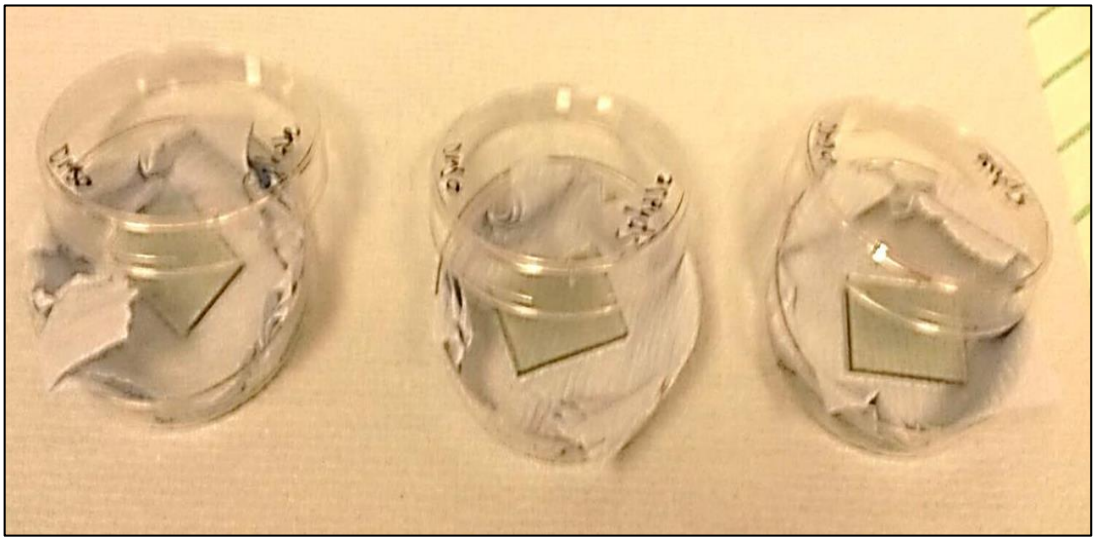


Figure 8 – Photograph showing EtOH fixed MCF-7 samples air-drying on the bench after being suspended onto 1cm by 1cm low-e slides

For MCF-7 spectral analysis 1 x 10^5 cells were grown in triplicate T25 flasks (25cm^2 surface area) and concentrated in S phase (24h) or G_0/G_1 phase (96h) before being treated with 25µl of B[a]P at concentrations of 10^{-9}M, 10^{-6}M and 0M (Dimethyl Sulfoxide (DMSO) control) for 24 hours. Media was refreshed 5 hours before treatment in G_0/G_1 phase concentrated flasks.

Trypsin was added to the flasks to lift the cells for collection. The collected cells were washed three times with 5ml 70% EtOH centrifuged at 1500rpm for 5 minutes. The cells were left to fix in EtOH at room temperature for one hour. After fixing cells were again centrifuged and re-suspended in 1ml of EtOH, this was gently pipetted onto low-e slides, left to dry overnight and subsequently stored in a desiccator.

2.2 FTIR-ATR Spectral Acquisition

Sample readings were obtained using a Bruker Vector 27 FTIR spectrometer with a Helios ATR attachment containing a diamond crystal (Bruker Optics Ltd., Coventry, UK). 10 measurements were taken from independent points across the sample. The ATR crystal was cleaned using distilled water before every measurement to ensure no cross contamination occurred between sample points. A background spectrum was taken between every sample to account for environmental variations. Bruker OPUS software was used to convert spectra into absorbance values.

2.3 Raman Spectral Acquisition

Measurements were taken using an InVia Renishaw Raman spectrometer (Renishaw plc, Gloucestershire, UK) within the 400cm^{-1} to 2000cm^{-1} wavelength region. A total of 1075 spectral measurements were obtained from the endometriosis samples. 25 random points were mapped on each of the 43 sample slides. Measurements were obtained using 10% laser power with a 13 second exposure period and three repeat acquisitions. Epithelial like samples have a tendency to be more delicate, so low laser power was used to avoid burning through the sample material to ensure the slide could be revisited. A total of 504 spectral measurements were obtained from the MCF-7 sample slides. 28 random points were mapped across each slide. Measurements were taken using 100% laser power with a 15 second exposure and three repeat accumulations. All of the spectra were individually examined for

cosmic rays. These present themselves as spurious sharp peaks and were excluded from the spectra using Renishaw's WiRE (Windows-based Raman Environment) software.

2.4 Pre-Processing Spectra

Data obtained through Raman and FTIR-ATR spectroscopy was analysed using Matlab (Mathworks, Natick, MA, USA) coupled with the IRootlab toolbox specifically designed for biospectroscopy data analysis (Trevisan et al., 2013).

It is of great importance to apply pre-processing methods and to Spectra were cut to focus on the wavelengths falling between 1800cm-1 and 900cm^{-1} for FTIR-ATR (Figure 9) and between 2000cm^{-1} and 400cm^{-1} for Raman (Figure 10). This is the region associated with the biological spectral fingerprints.

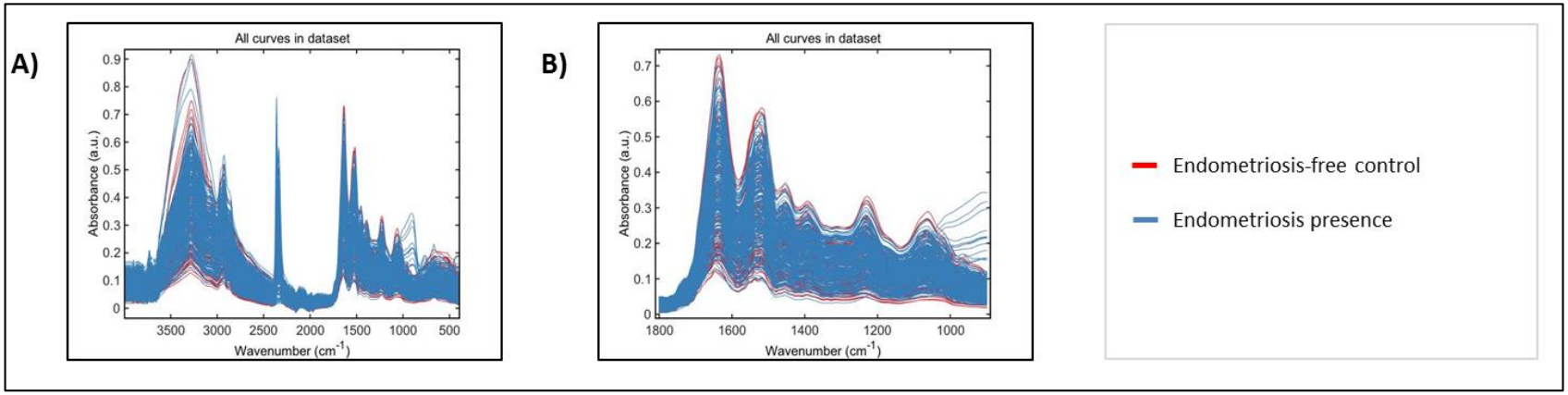


Figure 9 – FTIR-ATR raw spectral data from LBC samples of endometriosis and endometriosis free patients (control). All curves contained in the raw data set are shown as Absorbance (a.u) peaks over Wavenumber cm^{-1}. A) FTIR-ATR raw uncut spectra ranging from 4000 cm^{-1} and 500 cm^{-1}. B) FTIR-ATR spectra cut between 1800cm^{-1} and 900cm^{-1}

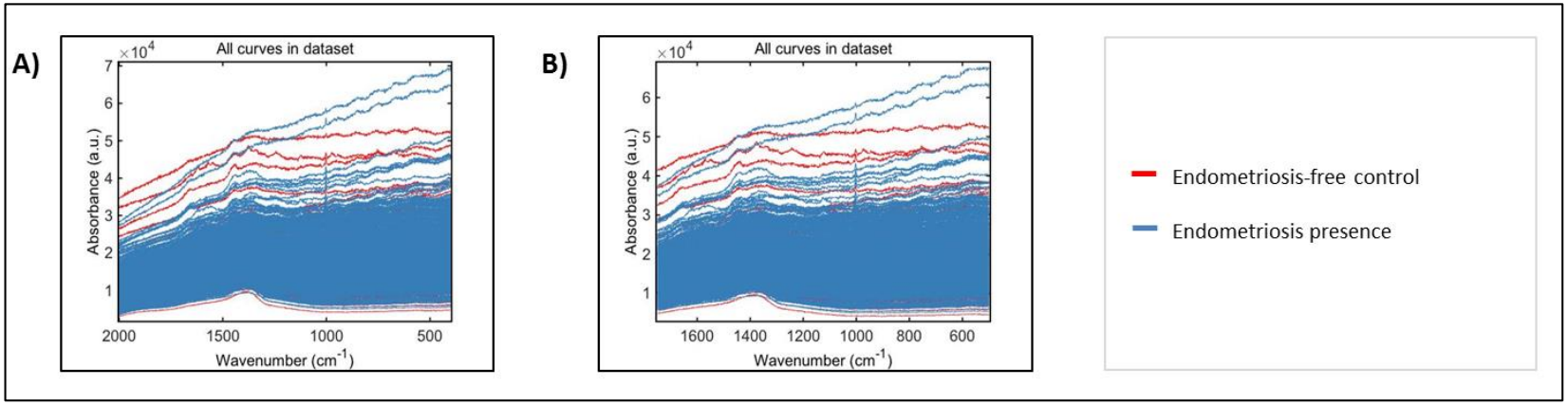


Figure 10 – Raman raw spectral data from LBC samples of endometriosis and endometriosis free patients (control). All curves contained in the data set are shown as Absorbance (a.u) peaks over Wavenumber cm^{-1}. A) All raw uncut Raman spectra ranging from 2000 cm^{-1} to 300cm^{-1}. B) Raman spectra cut to 2000cm^{-1} to 400cm^{-1}

In order to compress and smooth out the data sets the following algorithms were applied first: rubber band, Savitzky-Golay (SG) differentiation (for FTIR-ATR), and polynomial (for Raman). Spectra were then normalised by alignment to the amide I peak at 1650 cm^{-1}, or through vector normalisation. Finally, background noise and scatter was removed through trained mean centring (for FTIR-ATR) and a de-noising function (for Raman).

The FTIR-ATR and Raman spectra were each subjected to three different combinations of pre-processing methods. These included the following:

- Rubber band-Amide I (for both FTIR-ATR and Raman)
- Rubber band-Vector normalisation (for both FTIR-ATR and Raman)
- SG differentiation-Vector normalisation (for FTIR-ATR only)
- Polynomial-Vector normalisation (for Raman only)

In Figures 11 and 12 (endometriosis spectra) and Figures 13 and 14 (MCF-7 spectra) these pre-processing methods can be visualised as a guide to understand how the pre-processing algorithms mould and align the spectral acquisitions.

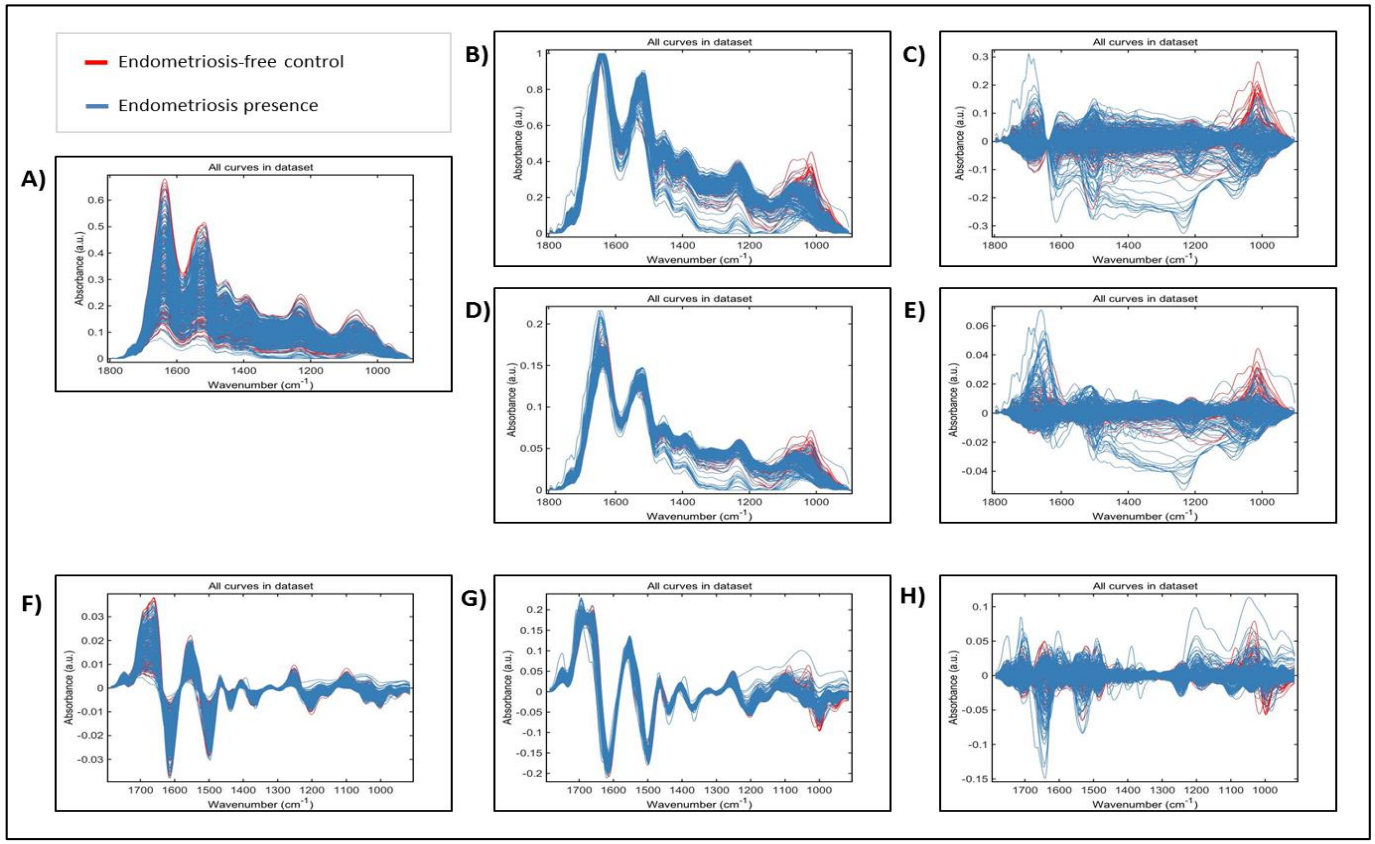


Figure 11 – FTIR-ATR raw data from LBC samples of endometriosis and endometriosis free patients underwent several pre-processing steps in order to facilitate analysis. A combination of algorithms were applied including: A) Rubber band. B) Rubber band-Amide I. C) Rubber band-Amide I-Trained mean centering. D) Rubber band-Vector normalisation. E) Rubber band-Vector normalisation-Trained mean centering. F) SG differentiation. G) SG differentiation-Vector normalisation. H) SG differentiation-Vector normalisation-Trained mean centering

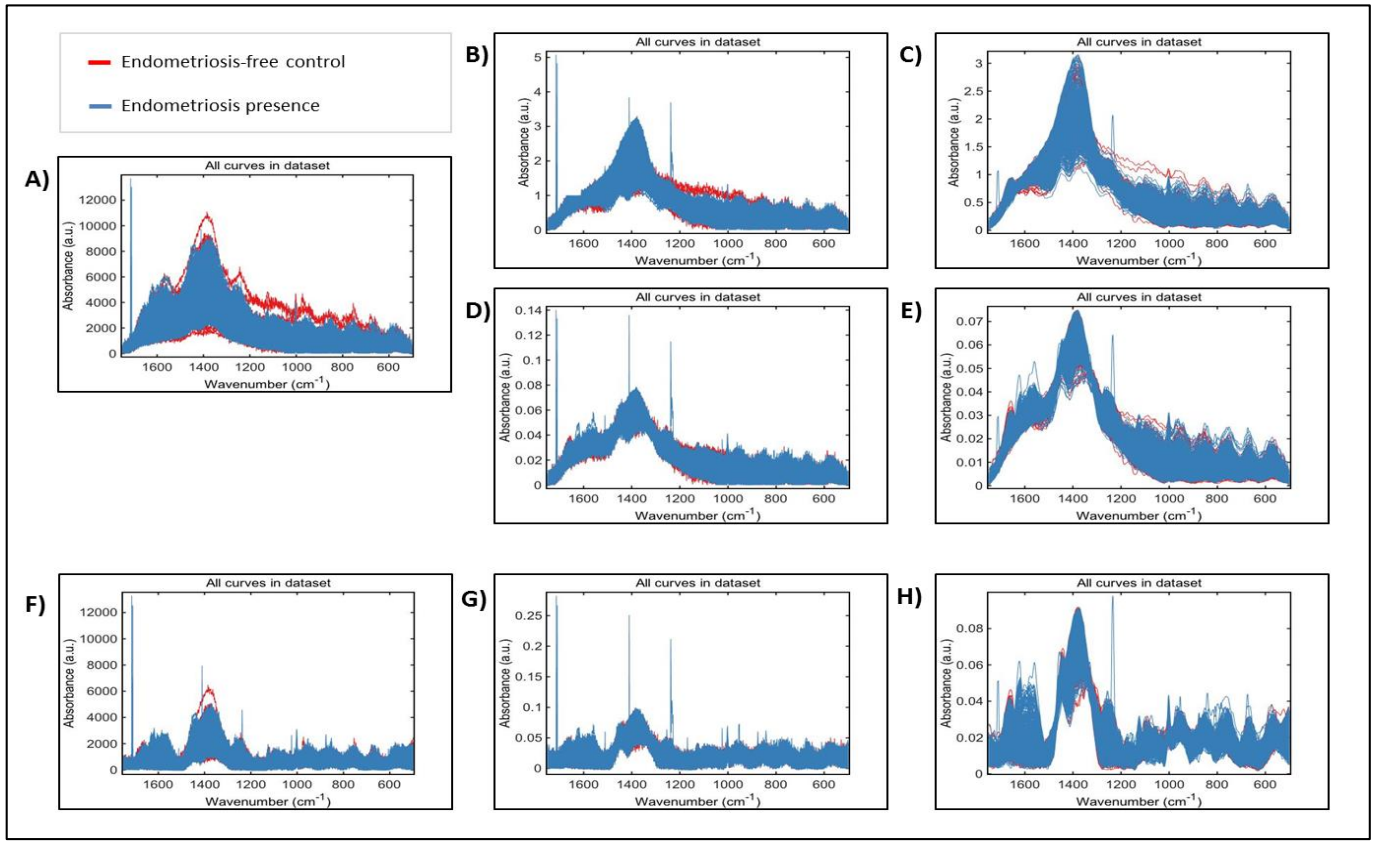


Figure 12 – Raman raw data obtained from LBC samples of endometriosis and endometriosis free patients (control) underwent several pre-processing steps in order to facilitate analysis. A combination of algorithms were applied including: A) Rubber band. B) Rubber band-Amide I. C) Rubber band-Amide I-Denoise. D) Rubber band-Vector normalisation. E) Rubber band-Vector normalisation-Denoise. F) Polynomial G) Polynomial-Vector normalisation. H) Polynomial-Vector normalisation-Denoise

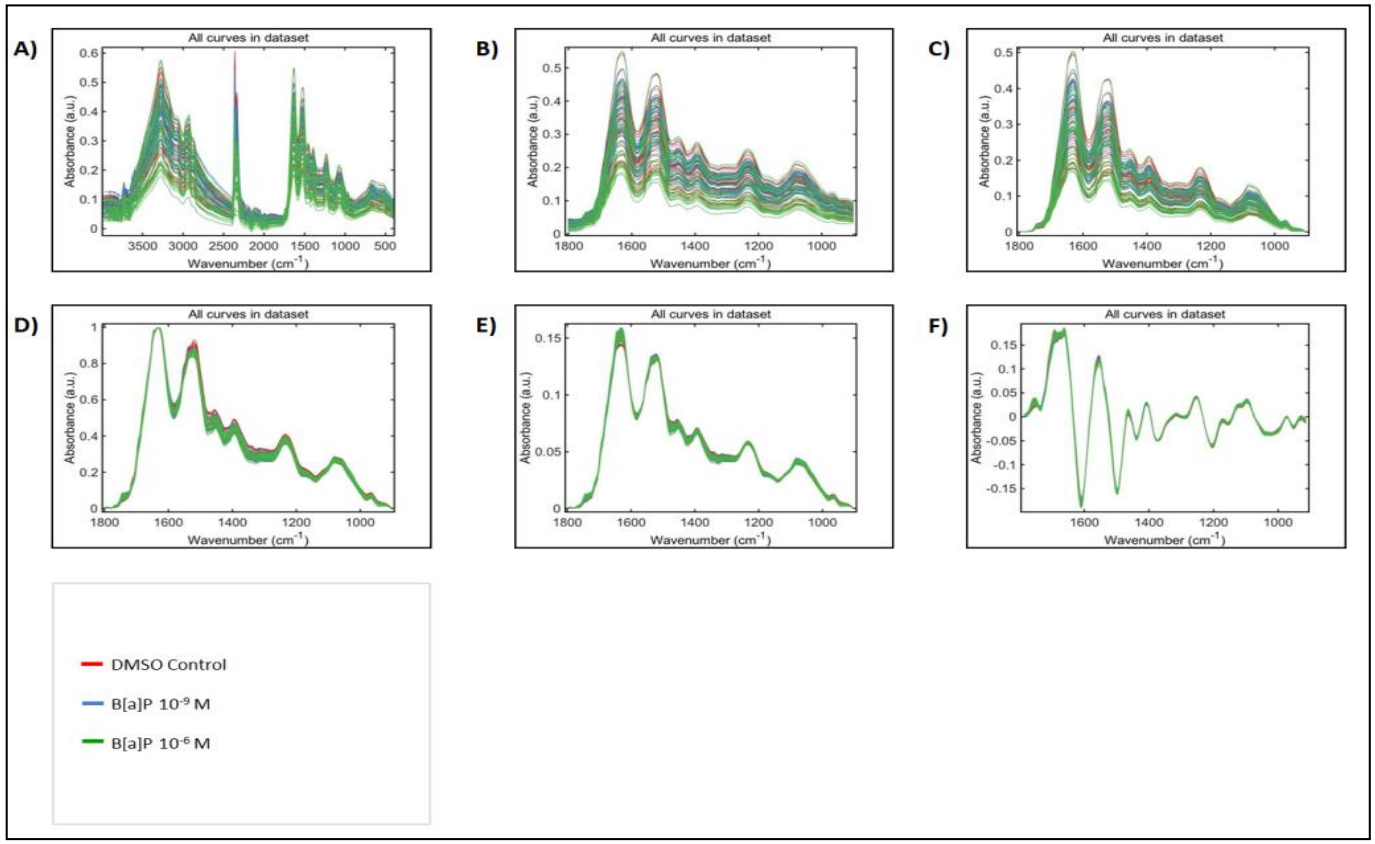


Figure 13 – FTIR-ATR raw data, obtained from MCF-7 cells exposed to B[a]P concentrations of DMSO control, B[a]P 10^{-9}M and 10^{-6}M, underwent several pre-processing steps in order to facilitate analysis. A) All raw FTIR-ATR spectra ranging from 4000 cm^{-1} and 500 cm^{-1}. B) FTIR-ATR spectra cut between 1800cm^{-1} and 900cm^{-1}. A combination of algorithms were applied including: C) Rubber band. D) Rubber band-Amide I. E) Rubber band-Vector normalisation-Trained mean centerting. F) SG differentiation-Vector normalisation-Trained mean centering

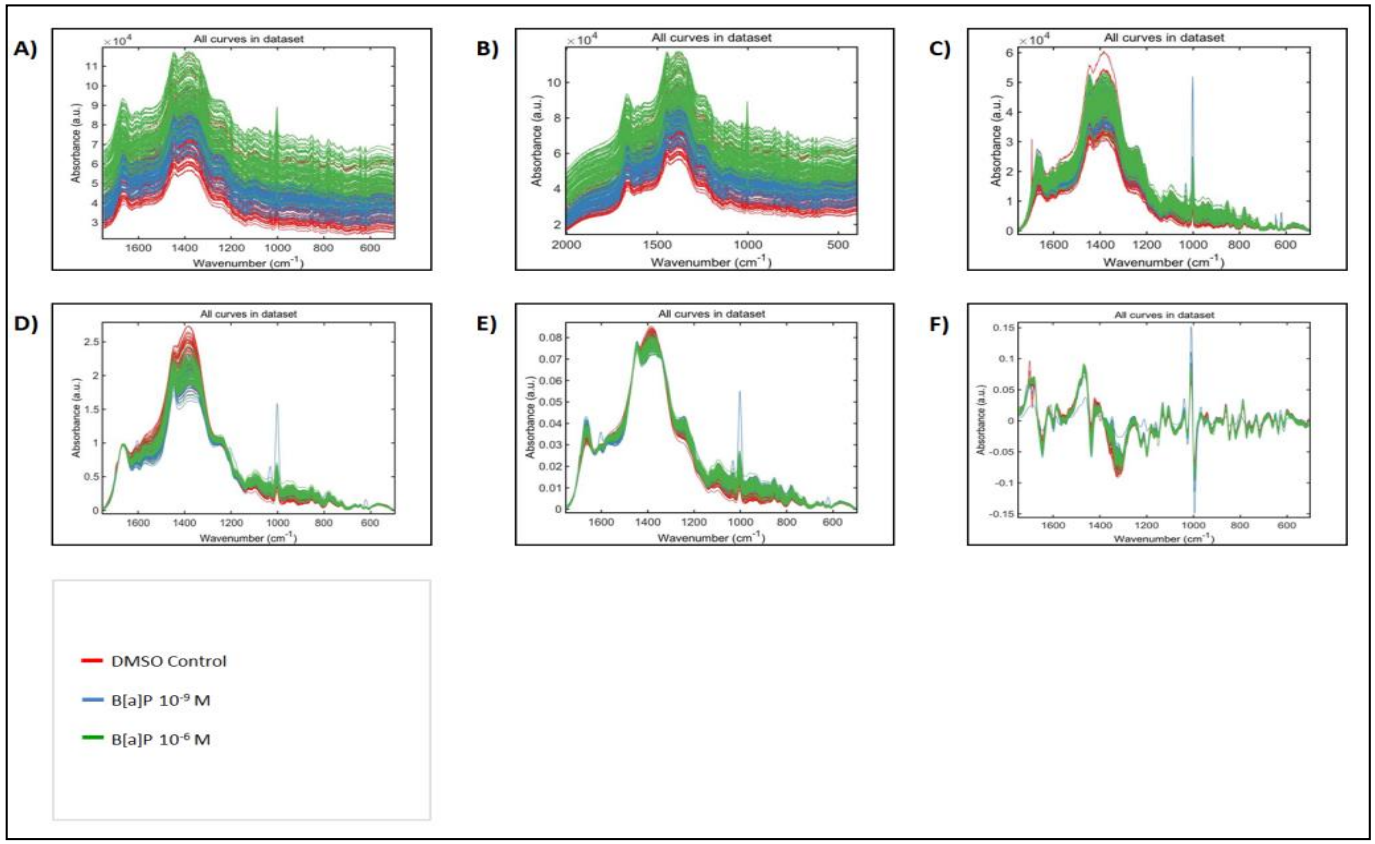


Figure 14 – Raman raw data, obtained from MCF-7 cells exposed to B[a]P concentrations of DMSO control, B[a]P 10^{-9}M and 10^{-6}M, underwent several pre-processing steps in order to facilitate analysis. A) All raw Raman spectra ranging from 4000 cm^{-1} and 500 cm^{-1}. B) Raman spectra cut between 1800cm^{-1} and 900cm^{-1}. A combination of algorithms were applied including: C) Rubber band. D) Rubber band-Amide I. E) Rubber band-Vector normalisation-Trained mean centerting. F) SG differentiation-Vector normalisation-Trained mean centering

2.5 Spectral Analysis

Pre-processed spectral data can be sorted and classified by applying a combination of both supervised and unsupervised algorithms. Supervised algorithms classify data sets according to predefined predictions based on a known data set. The supervised algorithm that was used was linear discriminant analysis (LDA). Unsupervised algorithms cluster data without previous information on the data; it based purely on the patterns that are recognized within the data set. (Trevisan et al., 2012). The unsupervised algorithm used was principal component analysis (PCA). Principal components were given a ranking score representing variance. This allowed for data point clusters and segregation to be illustrated. Principal components were chosen to capture > 95% variance held within the data set.

Large variables can be sorted when applying a combination of supervised and unsupervised algorithms. PCA minimises the variance within classes, to account for heterogeneity of similar tissue samples. Meanwhile LDA maximises the variance between classes, allowing for diagnostic differences to be noticed. Mapping PCA alongside LDA (PCA-LDA) increases the distribution of variance throughout the data set.

Processed data can be visualized in 1, 2 and 3 dimensional PCA-LDA scatter plots. The graphs show data over LD1, LD2 and LD3. This represents the PCA-LDA principal component rankings where LD1 is ranked 1st and captures > 90% of variance, LD2 is ranked 2nd then LD3 is ranked 3rd capturing the remaining variance.

The 1 dimensional graphs include a linear representation of LD1 and a linear representation of LD2. The scores plots contain a data trend line that illustrates data variance distribution. The 2 dimensional graphs show LD1 mapped alongside LD2. The 3 dimensional graph maps LD1, LD2 and LD3 together. Ellipse walls containing 95% of the data points were drawn on the 2D and 3D scatterplots to help illustrate data trends. Cluster vector loading plots illustrate the mean difference between subclasses. The differences are shown as peaks, where the height

of the peak relates to the increased dissimilarity between disease/treatment classes and the control (coefficient 0). The 8 largest peaks are labelled with their corresponding wavelengths. The scatter plots and cluster vectors graphs are shown in Section 3.

2.6 Statistical Analysis

In order to determine whether results carried statistical significance several statistical methods were applied to the data. After undergoing principal component linear discrimination analysis (PCA-LDA), LD1 values (which captures > 90% of variance) were tested for significance. P values were calculated using one-way analysis of variance to test if data means were significantly different.

Bartlett's test for equal variances analysed the significance of variance in the data. Anova tests included Dunnett's multiple comparison test and Tukey's multiple comparison test. The former analyses the significant difference between the controls and disease/treatment classes, and the latter compares each individual class to the other. An unpaired two tailed t-test and an f-test to compare variances were used for analysis of endometriosis-free control vs. no endometriosis. R^2 values were obtained through linear regression analysis to help predict the linear relationship between the data classes. All tests were performed with a confidence interval of 95% difference when $P < 0.05$. Error bars illustrating standard deviation values were added to the bar graphs and line graphs. These were calculated to measure significance of variance between treatment conditions.

2.7 Chemicals and Cell Culture Media

The chemicals used in this study, including B[a]P and DMSO, were obtained from Sigma Chemical Co. (Poole, Dorset, UK). 25µl of DMSO stock solution was used per 5ml of culture media. All consumables were obtained from Invitrogen Life Technologies (Paisley, UK).

MCF-7 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM). The media was supplemented with 10% heat inactivated foetal bovine serum (FBS) and 1% penicillin/streptomycin (penstrep). Routine culture of MCF-7 cells was carried out in T75 flasks (75cm^2 surface area), and they were incubated at 37°C in a humidified atmosphere containing 5% CO_2 in air. Every 7 days cells were passaged for maintenance using Trypsin (0.05%)/EDTA (0.02%). This is heat activated in a 37°C incubator (Figure 15) for 5 minutes to disaggregate adherent cells. Culture media and trypsin were pre-warmed in a water bath turned to 37°C for 20 minutes (Figure 16). Cell culture was performed in a class II laminar flow hood (Figure 17) which was run for 20 minutes and cleaned thoroughly with 70% EtOH before every use.

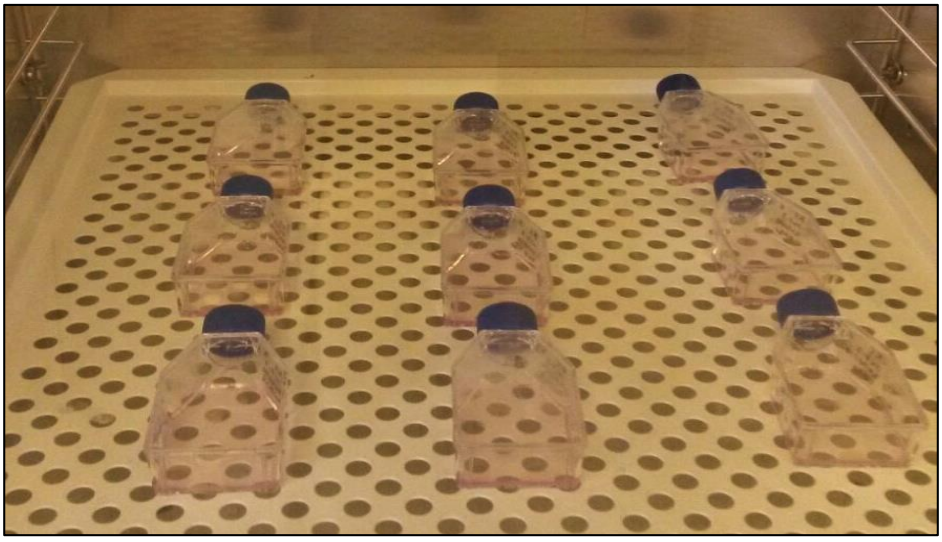


Figure 15 – Photograph showing T25 experimental flasks containing 2.5ml of trypsin being activated in a 37°C incubator

Figure 16 – Photograph showing media and trypsin being warmed in a water bath at 37°C

Figure 17 – Photograph showing the laminar flow hood used during cell culture

2.8 Clonogenic Assay

A clonogenic assay was performed to study the dose response of B[a]P on cell survival. B[a]P was diluted in DMSO to obtain concentrations of 1.0µM, 0.1µM, 0.01µM and 0.001µM. To account for DMSO presence in the dilutions, a treatment of DMSO only was used as a vehicle control. MCF-7 cells were split and then counted using a haemocytometer. 500 cells were seeded in in T25 flasks containing 5ml of DMEM complete media. Triplicate flasks were seeded for each of the five treatments (four B[a]P concentrations, one control). 25µl of treatment dose was then added to the cells on the day of seeding. The cells were incubated for 24 hours, after which media was replaced with fresh DMEM. Seven days post seeding, cells were fixed and stained for counting. Media was aspirated and cells were fixed for 30 minutes with 5ml of 70% EtOH. EtOH was aspirated and 2ml of 5% Giemsa was added to the base of the T25 flasks. The cells were left to stain for 20 minutes with occasional shaking to distribute the Giemsa evenly. Cell colonies were counted using a light box. The clonogenic assay was performed in triplicate and repeated three independent times.

Cell colony count was determined through visual observation and interpretation. Only the large colonies were counted. The same critical assessment was used throughout the counts. A visualisation of the counted colonies vs. stained cells that were not counted is shown in Figure 18.

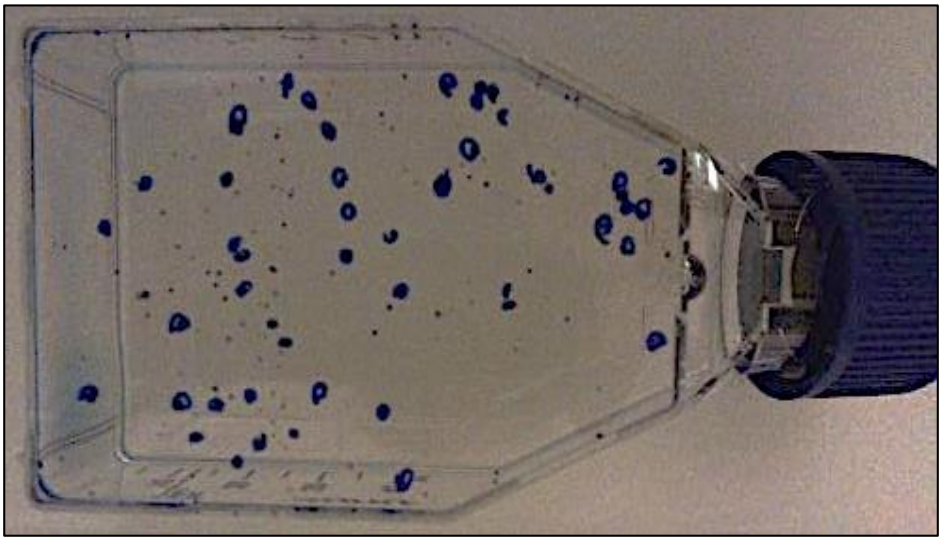


Figure 18 – Photograph showing the flasks obtained from clonogenic assay. Counted colonies were circled in blue

Cell survival and plating efficiency were calculated as percentages using the following formulae respectively:

$$\% \text{ Cell Survival} = \frac{\text{Number of colonies in experiment flask}}{\text{Number of cells seeded}} \times 100$$

$$\% \text{ Plating Efficiency} = \frac{\text{Number of colonies in control flask}}{\text{Number of cells seeded}} \times 100$$

2.9 Time Course for Cell Number

1 x 10^5 MCF-7 cells were seeded along with 5ml media in triplicate T25 flasks and were grown for 24 hours. The cells were then treated with 25µl of B[a]P, at concentrations 1 x 10^{-9}M and 1 x 10^{-6}M along with DMSO (as a vehicle control). Four separate cell counts were taken using a haemocytometer at 0 hours, 6 hours, 12 hours, 24 hours and 48 hours post treatment. Cells were trypsinised, disaggregated and re-suspended prior to cell count. The experiment was repeated on three independent occasions. Results were averaged and standard deviation was

derived. Time zero (T_0) was taken before treatment (0 hours) to normalise cell number to 100%. Relative cell number was calculated as a percentage using the following formula (T_x corresponds to the number of hours):

$$\text{Relative Cell Number (\%)} = \frac{\text{cell number at } T_x}{\text{cell number at } T_0} \times 100$$

2.10 Flow Cytometry

1 x 10^6 MCF-7 cells were seeded in triplicate flasks containing 5ml of media. The cells were exposed to nine different cell culture conditions. These included three treatment free controls and either S phase or G_0/G_1 phase concentrated cells treated with three B[a]P concentrations. Cells were grown for 24 hours to obtain S phase concentrated cells and 96 hours for G_0/G_1 phase concentrated cells. 25µl of B[a]P 1 x 10^{-9}M, 1 x 10^{-6}M and 0M (containing only DMSO as a vehicle control) was added to the cells for a 24 hour treatment. Control flasks without treatment were fixed after 0 hours, 24 hours and 96 hours incubation. These represent the cell cycle distribution at time of seeding and at time of S phase and G_0/G_1 phase treatment.

Cells were trypsinised and collected in 15ml centrifuge tubes. They were washed three times using 2ml PBS and centrifuged at 1000rpm for 5 minutes. Supernatant was aspirated and the cell pellets were subsequently fixed in 2ml ice cold 70% EtOH. This was applied drop-wise over a vortex machine to ensure the cells did not cluster together and break up into single cells. Samples were then stored at -20°C overnight. The fixed pellet was re-washed twice with PBS and centrifuged for 5 minutes at an increased speed of 1500rpm as EtOH presence impedes cell precipitation. 50µl of RNase A solution (100µg/ml) was added directly to the cell pellet to eliminate RNA presence from the sample. 400µl of Propidium Iodide (PI) (50µg/ml) was then added to the cell pellet. The sample was vortexed to mix and left to stain the DNA for 60 minutes at room temperature. PI is photosensitive so samples were protected from light during incubation. Cell pellet was re-suspended in 1.5ml PBS and run through a Becton

Dickinson FACSCaliber flow cytometer for analysis. DNA fluorescence was measured in > 10,000 events and these were processed using the manufacturers FACSDiva Version 6.1.2 software. The experiment was repeated on two independent occasions.

2.11 Western Blot

1 x 10^6 MCF-7 cells were seeded in triplicate T25 flasks and concentrated in S phase and G_0/G_1 phase prior to B[a]P treatment. 25µl of B[a]P at concentrations of 1 x 10^{-9}M, 1 x 10^{-6}M or 0M (DMSO) was added to the cells and left for 24 hours. This experiment was repeated four separate times staining for either CYP1A1, CYP1A2, P21 or BCL-2. The cells were washed within the flask with PBS twice. Cells were then lysed and collected using Radioimmunoprecipitation Assay (RIPA) lysis buffer and a cell scraper tool. The collected material was placed in a sonicator to allow DNA to break up and proteins to be released. Total protein concentration was measured through a Bradford assay with the use of a spectrophotometer. The lowest protein concentration was used to equalise of all the sample lysates so that they all carried the same amount of protein after being diluted with PBS. Electrophoresis was carried out on a 5-15% Bis-Tris gel. 20µl of lysate was added to each well along with a low molecular weight ladder (GE Healthcare Limited, Buckinghamshire, UK). The gel was run at 30 volts for 60 minutes.

Once electrophoresis was completed the gel was blotted onto a Polyvinylidene Difluoride (PVDF) membrane cut to 9 cm x 6 cm along with filter paper of the same dimensions. The membrane was equilibrated using methanol, distilled water and transfer buffer by soaking it for 10 seconds, 2 minutes and 20 minutes respectively. Care was taken to avoid drying out the membrane at all times. The gel was blotted at 115 volts for 1 hour. After transfer the PVDF membrane was washed with PBS on a rocking platform for 5 minutes. Phosphate-Buffered Saline (PBS) was replaced by 20ml of 5% dried skimmed milk dissolved in PBS with 0.1% Tween. The membrane was left to probe for one hour on a rocking platform. This step prevents any non-specific antibody from binding. Primary and secondary antibody solutions were made up

in PBS containing 0.1% Tween and 2% Bovine Serum Albumin (BSA) according to dilution guidelines provided by the manufacturers. After probing, the PVDF membrane was placed in a 50ml falcon tube together with 5ml primary antibody solution. This was incubated on a spiramix overnight at 4°C. The membrane underwent a three stop wash with PBS containing 0.1% Tween; firstly it was washed for 1 minute followed by two 15 minute washes. PVDF membrane was then incubated in 20ml of secondary antibody solution for 1 hour at room temperature. Since all primary antibodies were raised in rabbit, an anti-rabbit secondary antibody was used. A 3-step wash was again preformed to remove excess antibody presence, however pure PBS was used this time. Enhanced Chemiluminescence (ECL) reagent was used to visualise the antibodies.

The origin of the antibodies and their respective dilution concentrations are as follows: anti-p21 (BD 556431; 1:1000) from BD Science (Oxford, UK), anti-CYP1B1 (CYP1B11-A; 1:50) from Alpha Diagnostic (Hampshire, UK), anti-CYP1A1 (ab3568 1:50) was obtained from Abcam (Cambridge, MA). The loading control β-actin was visualised using (Ab6276; 1:25,000) from Abcam (Cambridge, MA). Secondary anti-rabbit antibody (CST 7074; 1:4000) was from Cell Signaling Technologies (Herts, UK).

3 Results

3.1 Endometriosis Results

The results obtained after spectral analysis, through FTIR-ATR and Raman spectroscopy, of endometriosis LBC samples fixed on low-e slides are shown in this section. Individual PCA-LDA analysis was carried out for each disease subclass using the pre-processing methods outlined in Section 2.4 (Rubber band-Amide I, Rubber band-Vector normalisation, SG differentiation-Vector normalisation, Polynomial-Vector normalisation). Statistical analysis was carried out to determine which pre-processing method best illustrated the variance between subclasses. Graphs that illustrate the best results are included in this section along with their corresponding table of statistical analysis. The remaining graphs generated through the other described pre-processing methods can be found in the appendix B. Rubber band-Amide I is shown for FTIR-ATR and Polynomial-Vector normalisation is shown for Raman spectra. Statistical analysis of LD1 results (capturing > 90% of data variance) follow a confidence interval of 95% where $P < 0.05$. None of the statistical tables indicate any statistical significance for R^2.

3.1.1 Analysis of FTIR-ATR Spectra

Graphs generated after the PCA-LDA analysis of FTIR-ATR spectra obtained from cervical LBC samples are shown in this section. Pre-processing raw spectra using Rubber band – Amide I best separated spectra of endometriosis patients from the no endometriosis control patients.

3.1.1.1 Endometriosis Presence

Spectra were classed as having either no endometriosis, or having endometriosis. In order to use spectroscopy as a diagnostic tool for detecting endometriosis, a strong separation needs to be found between the two classes.

When interpreting the 1D scatterplot (Figure 19) the data trend line shows the bulk of the data for both classes is situated along LD1 around point 0. Initially the two classes appear to contain similar variance, yet on closer inspection the trend lines move in opposite directions. Endometriosis-free negative control samples drift out to LD1 -0.4 while endometriosis positive samples drift towards LD1 0.6. The median values both classes (indicated by the boldest score) illustrated a similar shift. The control median is on the lower end of LD1 and the positive sample median is more towards higher LD1 values.

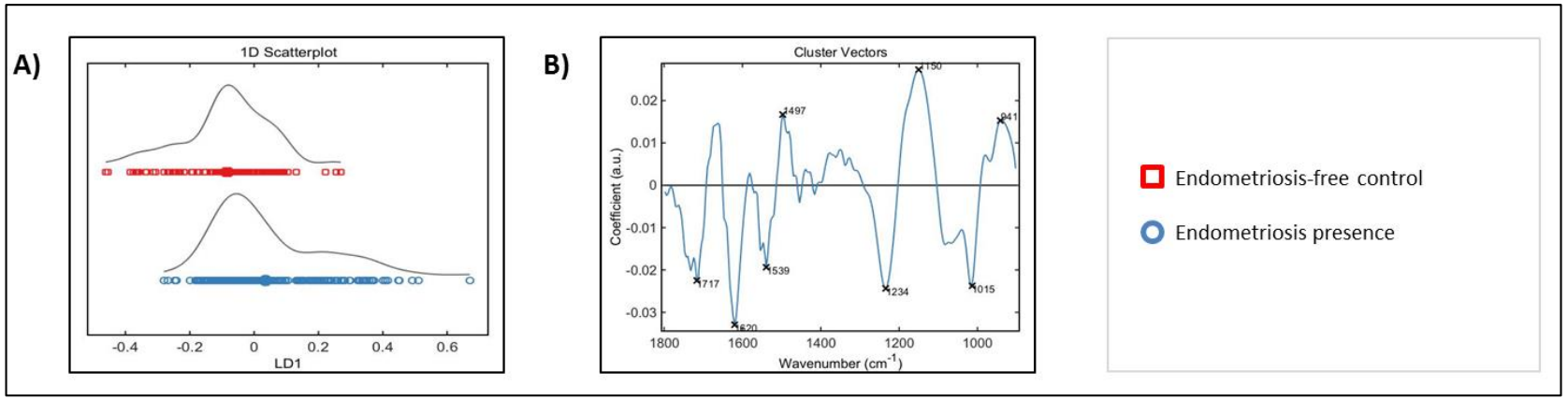


Figure 19 – FTIR-ATR spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing PCA-LDA scores along LD1. B) A cluster vector Loading plot showing distinct wavenumbers associated with endometriosis presence when compared to coefficient 0 (endometriosis-free control)

From Table 1 it is clear to see that choosing the correct pre-processing method to analyse data is of great importance. Only 1 out of 3 pre-processing methods show that there is a significant difference between endometriosis and endometriosis-free control sample spectra. It was due to this significance that Rubber Amide I pre-processed data was chosen to be illustrated in this section. For Rubber Amide I the P value is < 0.0001, meaning the initial observed shift across LD1 carries strong significance.

ATR Endo Statistical Tests LD1	Rubber Amide I	Rubber Vector Normalisation	Sg Differentiation Vector Normalisation
R^2	0.1109	0.0004181	0.0003018
Unpaired t-Test	P < 0.0001	P = 0.5690	P = 0.6285
F-Test	P = 0.0178	P = 0.3378	P = 0.7812

Table 1 – Table showing statistical analysis of LD1 generated by FTIR-ATR PCA-LDA analysis of endometriosis and endometriosis-free control samples. Statistical tests include R^2, Unpaired t-Test and F-test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation

The majority of the peaks shown on the cluster vector graph indicated that endometriosis samples greatly differ from the control samples mapped at coefficient 0. The higher the peak, the larger the difference. The top eight peaks were labelled and their corresponding wavenumbers were added to Table 2. The top peaks and their associations will be looked at in more detail in the discussion (Section 4).

Wavenumber (cm^{-1})	Association
941	Carotenoid
1015	$CH_{\alpha,\alpha'}$ out-of-plane bending and $C_{\alpha}==C_{\alpha'}$ torsion ν(CO), ν(CC), δ(OCH), ring (polysaccharides, pectin)

Wavenumber (cm^{-1})	Association
1150	C-O stretching vibration C-O stretching mode of the carbohydrates CH_8, CH_8'' 8 deformations v(C-O-C), ring (polysaccharides, pectin)
1234	Composed of amide III as well as phosphate vibration of nucleic acids $CH_{6,2',\alpha,\alpha'}$ rock
1497	C==C, deformation C-H
1539	Protein amide II absorption- predominately β-sheet of amide II Amide II
1620	Peak of nucleic acids due to the base carbonyl stretching and ring breathing mode
1717	C==O thymine Amide I (arises from C==O stretching vibration) C==O stretching vibration of DNA and RNA C==O stretching vibration of purine base

Table 2 – Top eight cluster vector peak wavenumbers and their associated chemical identities. Peaks were obtained from FTIR-ATR Spectra pre-processed through Rubber band-Amide I and analysed through PCA-LDA analysis. The peaks illustrate the wavenumbers that were only detected in patient samples suffering from endometriosis compared to endometriosis-free control samples

3.1.1.2 Disease Stage

Analysis was carried out on further disease subclasses to determine how extensive spectroscopy can be in identifying and separating disease characteristics. Disease stages 1 to 4 were analysed. Only one patient sample was classified with having endometriosis at stage 4, so the results for this stage are not conclusive as the sample size is too small to be specific.

In Figure 20 the 1D scores plot shows a distinct linear separation between stages 1, 2 and 3 along LD1. The median stage scores shift from LD1 0.5 down to -0.1. It would be expected that stage 0 (Endometrioses free control) would follow this trend and present itself at around LD1 0.6. Instead it appears to mirror stage 3, where its median score has an almost identical position along LD1. When classes are plotted on the 1D graph along LD2 no distinct differences can be observed. Similarly, the 2D graph shows separation across LD1 and not LD2, the ellipse walls for stage 3 and stage 0 are almost perfectly overlapping showing a high similarity between these two classes. They are better separated along LD3 on the 3D plot. However LD3 contains the third ranking of capturing most of the data variance (as LD1 captures > 90% of variance) so this is not as significant. The loading plot shows minimal peaks representing differences between stage 3 and the control class. Stages 1, 2 and 4 have similar peak distribution and contain similar wavenumber differences when compared directly to the control.

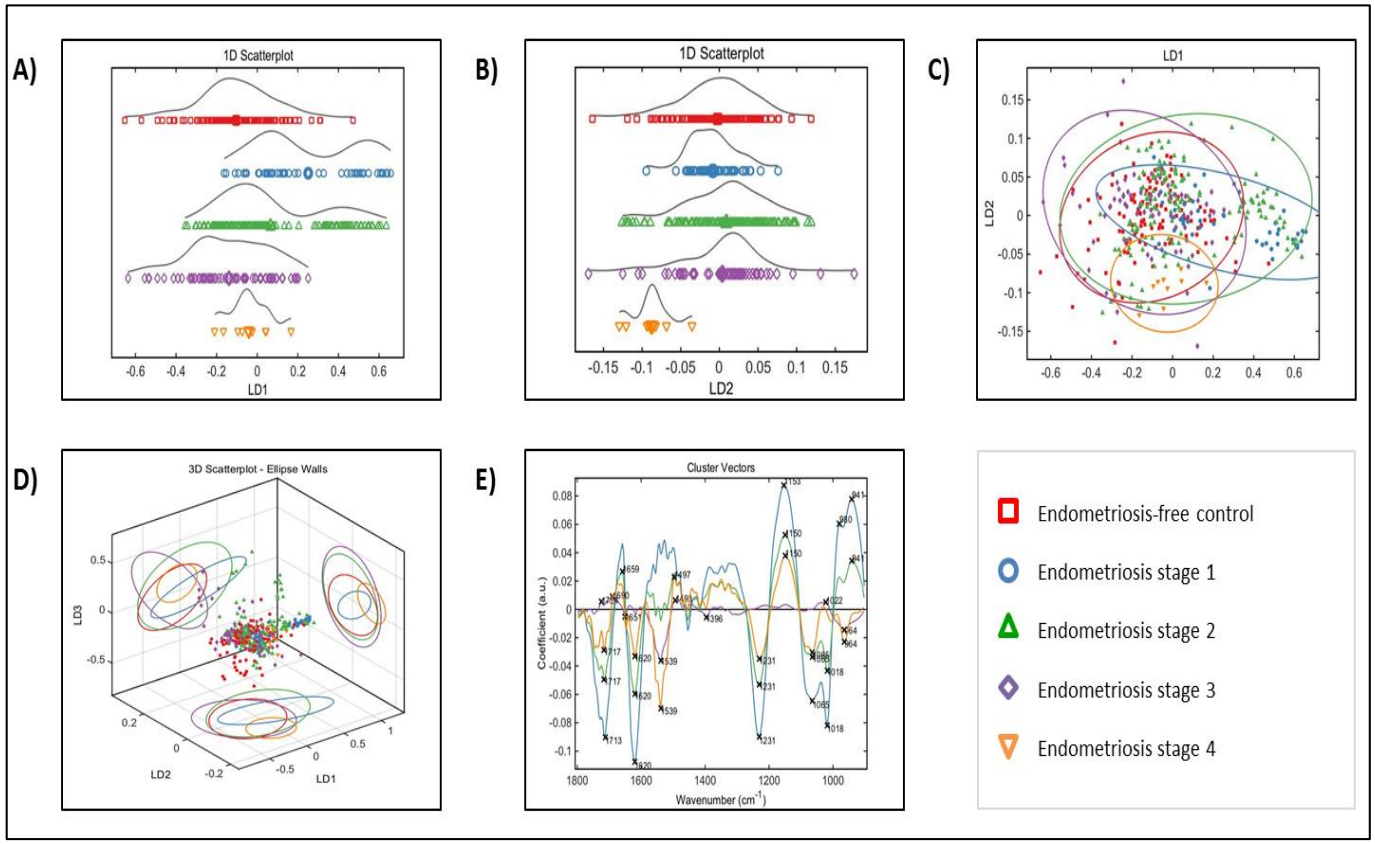


Figure 20 – FTIR-ATR spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. Spectra were separated into disease stages 1-4 A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2. D) 3D scores plot mapping LD1, LD2 and LD3 containing ellipse walls encapsulating > 95% of the data points. E) A cluster vector loading plot showing distinct wavenumbers associated with endometriosis disease stages when compared to coefficient 0 (endometriosis-free control)

When testing for statistical significance, multiple tests were used to confirm the initial observations (Table 3). Stage 4 shows very little and inconsistent significance throughout the pre-processing methods as the dataset size was insignificant. For all of the statistical tests applied, in all of the pre-processing methods, a significant difference was determined between the all of the different stages. The only exception is when stage 0 is compared to stage 3.

ATR Stage Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Sg Differentiation Vector Normalisation	
R^2	0.2449		0.2469		0.3642	
ANOVA	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	0 vs 1	P < 0.001	0 vs 1	P < 0.001	0 vs 1	P < 0.001
	0 vs 2	P < 0.001	0 vs 2	P < 0.001	0 vs 2	P < 0.001
	0 vs 3	P > 0.05	0 vs 3	P > 0.05	0 vs 3	P > 0.05
	0 vs 4	P > 0.05	0 vs 4	P > 0.05	0 vs 4	P < 0.001
	1 vs 2	P < 0.001	1 vs 2	P < 0.001	1 vs 2	P < 0.001
	1 vs 3	P < 0.001	1 vs 3	P < 0.001	1 vs 3	P < 0.001
	1 vs 4	P < 0.01	1 vs 4	P < 0.01	1 vs 4	P > 0.05
	2 vs 3	P < 0.001	2 vs 3	P < 0.001	2 vs 3	P < 0.001
	2 vs 4	P > 0.05	2 vs 4	P > 0.05	2 vs 4	P > 0.05
	3 vs 4	P > 0.05	3 vs 4	P > 0.05	3 vs 4	P < 0.001
Dunnett's Multiple Comparison Test	0 vs 1	P < 0.01	0 vs 1	P < 0.01	0 vs 1	P < 0.01
	0 vs 2	P < 0.01	0 vs 2	P < 0.01	0 vs 2	P < 0.01
	0 vs 3	P > 0.05	0 vs 3	P > 0.05	0 vs 3	P < 0.05
	0 vs 4	P > 0.05	0 vs 4	P > 0.05	0 vs 4	P < 0.01

Table 3 –Table showing statistical analysis of LD1 generated by FTIR-ATR PCA-LDA analysis of endometriosis stage subclasses. Statistical tests include R^2, ANOVA, Tukey's multiple comparison test and Dunnett's multiple comparison test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. Numbers 0-4 indicate disease stage, with 0 being no endometriosis presence

3.1.1.3 Disease Morphology

The manifestation of lesions during endometriosis can cause great discomfort in patients. Red lesions occur in early endometriosis stages and can be removed through ablation. These are often overlooked during examination and when given the time to fester they will progress into

black lesions as disease stage progresses. Being able to detect red lesions early on could prevent the disease from growing into an untreatable state.

Scores plots across LD1 show that the endometriosis-free control and black + red lesions are highly similar (Figure 21). Only red lesion presence is also similar, however, it displays a different trend from the control. Black lesions scores are clearly shifted towards a lower LD1 when compared to the other classes. The medians plotted on the 1D scores plot across LD2 all stay within LD2 0 region. The same can be seen in the 2D and 3D plots, where the majority of the separation is observed solely for black lesions along LD1. The cluster vector loading show the largest peaks were present for black lesions with respect to the endometriosis-free control at coefficient 0.

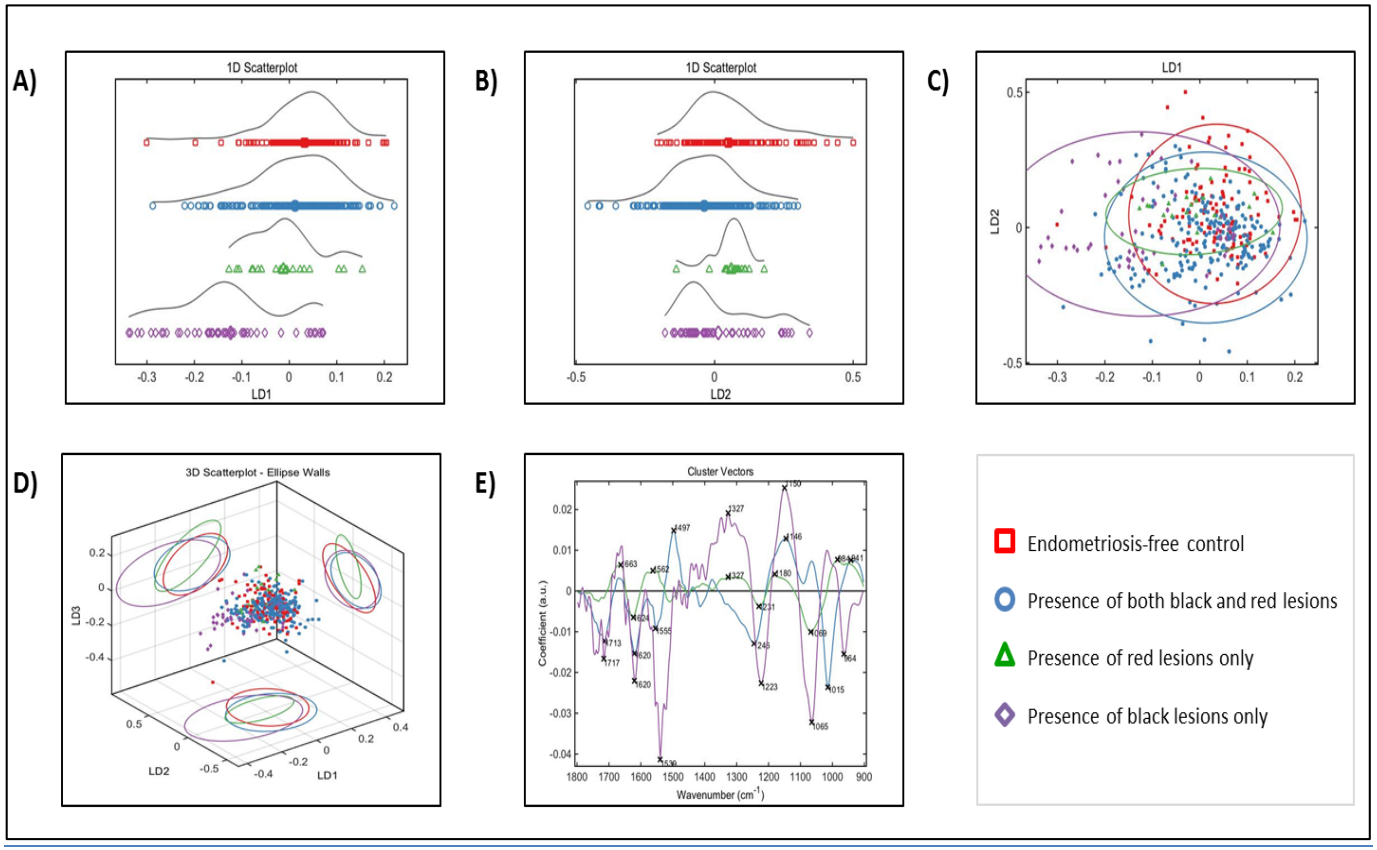


Figure 21 – FTIR-ATR spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. Spectra were separated into disease morphologies of black and red lesion presence. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) 3D scores plot mapping LD1, LD2 and LD3 containing ellipse walls. E) A cluster vector loading plot showing distinct wavenumbers associated with endometriosis morphologies when compared to coefficient 0 (endometriosis-free control)

After observing a shift in LD1, statistical analysis in confirmed that black lesions were significantly different to the control, the red lesions, and the black and red lesions combined (Table 4). The latter is surprising as some similarity was excepted since black lesions are present in the samples therefore some spectral readings should match up. The dissimilarity of the red lesions compared to the black ones might have obscured underlying similarities. No other lesions were successfully separated from the control.

ATR Black Red Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Sg Differentiation Vector Normalisation	
R^2	0.2327		0.2226		0.2708	
ANOVA	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	0 vs R	P > 0.05	0 vs R	P > 0.05	0 vs R	P > 0.05
	0 vs BR	P > 0.05	0 vs BR	P > 0.05	0 vs BR	P < 0.05
	0 vs B	P < 0.001	0 vs B	P < 0.001	0 vs B	P < 0.001
	R vs BR	P > 0.05	R vs BR	P > 0.05	R vs BR	P > 0.05
	R vs B	P < 0.001	R vs B	P < 0.001	R vs B	P < 0.001
	BR vs B	P < 0.001	BR vs B	P < 0.001	BR vs B	P < 0.001
Dunnett's Multiple Comparison Test	0 vs R	P > 0.05	0 vs R	P > 0.05	0 vs R	P > 0.05
	0 vs BR	P > 0.05	0 vs BR	P > 0.05	0 vs BR	P < 0.05
	0 vs B	P < 0.01	0 vs B	P < 0.01	0 vs B	P < 0.01

Table 4 – Table showing statistical analysis of LD1 generated by FTIR-ATR PCA-LDA analysis of endometriosis morphology subclasses. Statistical tests include R^2, ANOVA, Tukey's multiple comparison test and Dunnett's multiple comparison test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. Subclass morphology is indicated through letters, with Black lesions (B), Red lesions (R), both black and red lesions (BR) and endometriosis-free (0)

3.1.1.4 Ovarian Endometriosis

Ovarian endometriosis is one of the more severe manifestations of endometriosis. It is the strongest cause of sub fertility induced by endometriosis. The presence of ovarian endometriosis scores at stage 3 of disease severity.

Figure 22 showing analysed spectra from patients suffering from endometriosis with the presence of ovarian endometriosis shows very interesting results. The 1D graphs illustrate that the shift between classes only occurs along LD1. The control and the ovarian endometriosis free scores show distinctly different trend lines and the medians are widely separated. The positive ovarian endometriosis median is very close though the control median along LD1. Ovarian endometriosis scores have a compact score range except for one clear outlier on LD1. This outlier appears to skews the data trend line and considered to be removed as an anomaly. However this same outlier can be seen in the 2D map, on the edge of the no ovarian endometriosis ellipse. Here, only a total of 5 ovarian endometriosis positive scores are not contained within the red control ellipse. The insignificance of the shift caused by the outlier prevented it from being removed to maintain the integrity of the data set. The cluster vectors show large peak differences for both classes compared to the control.

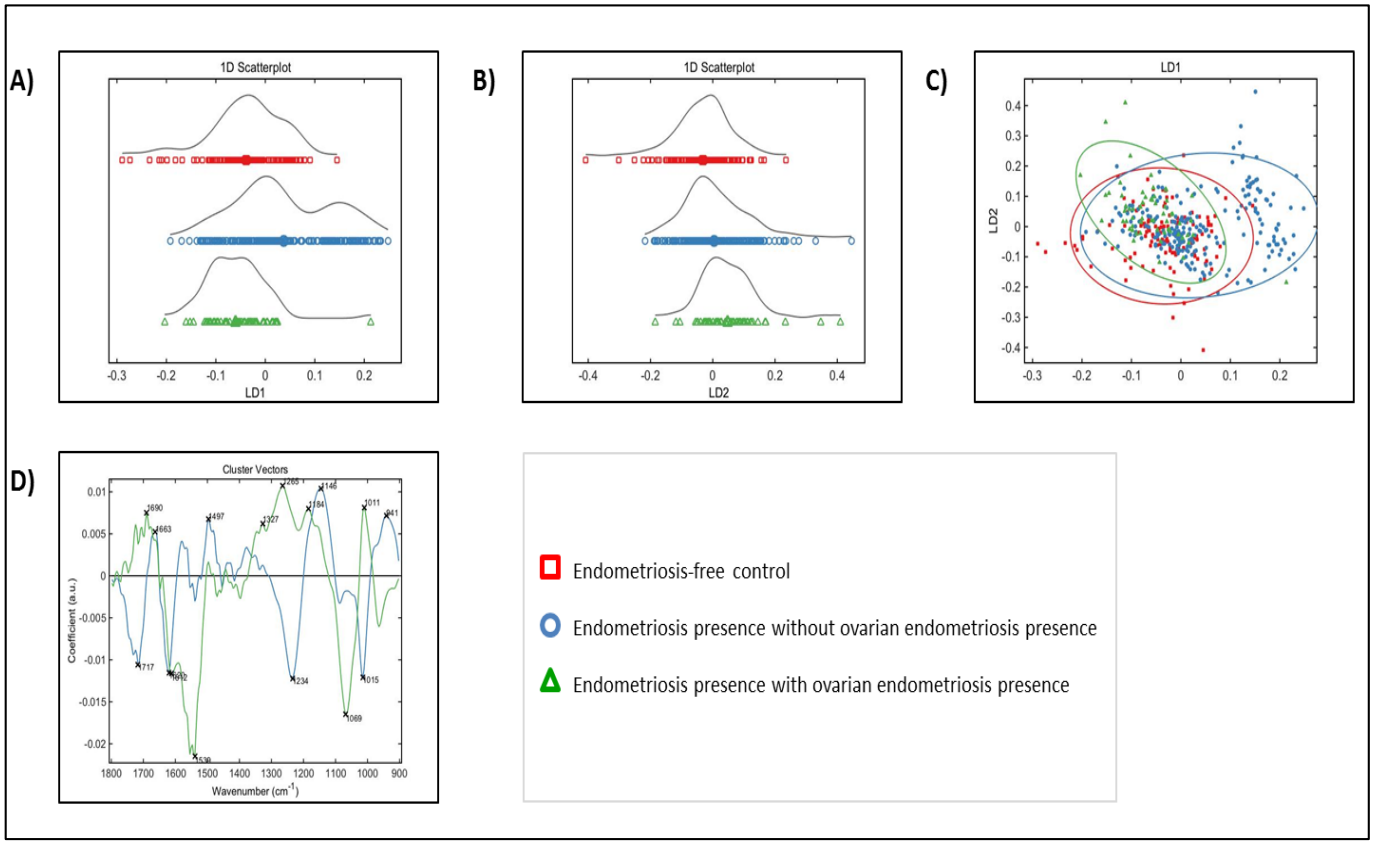


Figure 22 – FTIR-ATR spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. Spectra were separated into disease morphologies of black and red lesion presence. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data. D) A cluster vector loading plot showing distinct wavenumbers associated with endometriosis with and without ovarian endometriosis presence when compared to coefficient 0 (endometriosis-free control)

The same statistics results were obtained throughout all of the pre-processing methods (Table 5). These confirm the observations of LD1 from the graphs. The control is successfully separated from samples without ovarian endometriosis, whilst the ovarian endometriosis positive samples did not show any significant difference to the control. This shows that spectroscopy cannot readily detect ovarian endometriosis. It does reinforce the initial findings that endometriosis positive (in this case with no ovarian endometriosis) can be identified from no endometriosis control.

ATR Ovarian Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Sg Differentiation Vector Normalisation	
R^2	0.1893		0.1825		0.2935	
ANOVA	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	0 vs N	P < 0.001	0 vs N	P < 0.001	0 vs N	P < 0.001
	0 vs Y	P > 0.05	0 vs Y	P > 0.05	0 vs Y	P < 0.05
	N vs Y	P < 0.001	N vs Y	P < 0.001	N vs Y	P < 0.001
Dunnett's Multiple Comparison Test	0 vs N	P < 0.01	0 vs N	P < 0.01	0 vs N	P < 0.01
	0 vs Y	P > 0.05	0 vs Y	P > 0.05	0 vs Y	P < 0.01

Table 5 – Table showing statistical analysis of LD1 generated by FTIR-ATR PCA-LDA analysis of ovarian endometriosis. Statistical tests include R^2, ANOVA, Tukey's multiple comparison test and Dunnett's multiple comparison test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. Subclass morphology is indicated through letters, ovarian endometriosis presence being yes (Y), endometriosis patients without ovarian endometriosis are No (N) and endometriosis-free is 0

3.1.2 Analysis of Raman Spectra

Graphs generated after the PCA-LDA analysis of Raman spectra obtained from cervical LBC samples are shown in this section. Raw data that underwent pre-processing through Polynomial-Vector normalisation best separated endometriosis and no endometriosis control. Data mirrors results obtained for the FTIR-ATR spectral analysis, this illustrates that spectroscopy is a robust characterisation tool.

3.1.2.1 Endometriosis Presence

Endometriosis was successfully identified from the no endometriosis control through IR spectroscopy. Vibrational spectroscopy in the form of Raman spectroscopy applied the same analysis techniques in order to replicate that success.

In Figure 23 a similarity can be seen in the data trend line, situated above the scores on the graph. The two classes appear similar in shape and distribution, however a definitive shift can be seen for the median scores. Endometriosis is shifted towards LD1 -0.2; more so than the control.

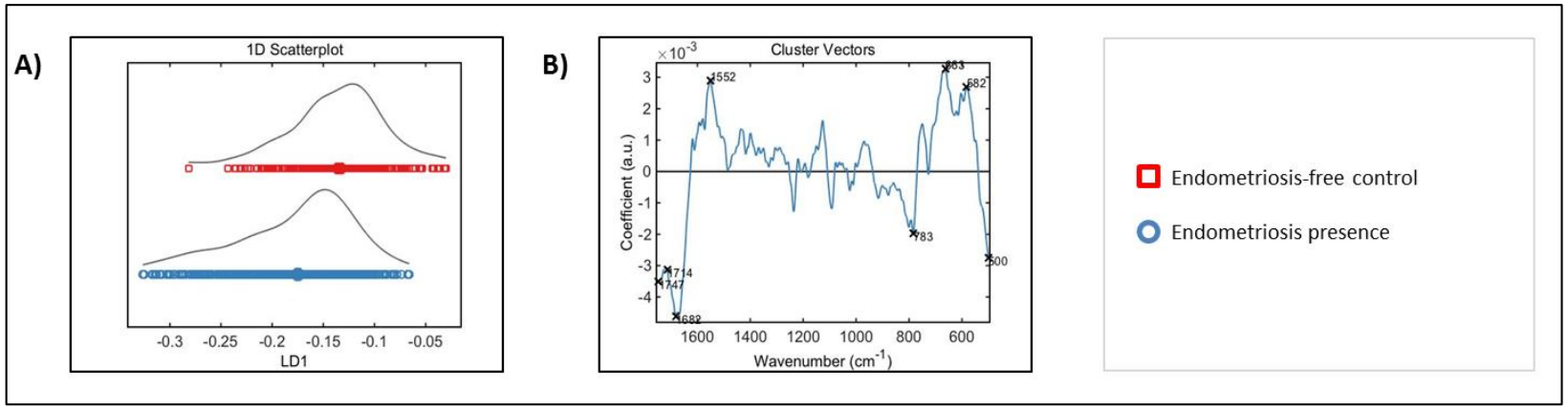


Figure 23 – Raman spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Polynomial-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing PCA-LDA scores along LD1. B) A cluster vector Loading plot showing distinct wavenumbers associated with endometriosis presence when compared to coefficient 0 (endometriosis-free control).

The statistical analysis from the Raman results (Table 6) is comparable to the FTIR-ATR results. Again only one pre-processing method shows statistical significant separation between the no endometriosis control and the endometriosis samples. Instead of Amide I-Vector normalisation the significance is found in Polynomial-Vector normalisation where $P < 0.0001$.

Raman Endo Statistical Tests LD1	Rubber Amide I	Rubber Vector Normalisation	Polynomial Vector Normalisation
R^2	0.00005051	0.00001223	0.1234
Unpaired t-Test	P = 0.7695	P = 0.8853	P < 0.0001
F-Test	P = 0.7703	P = 0.7639	P < 0.0001

Table 6 – Table showing statistical analysis of LD1 generated by Raman PCA-LDA analysis of endometriosis and endometriosis-free control samples. Statistical tests include R^2, Unpaired t-Test and F-test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and Polynomial-Vector normalisation

The peaks representing the differences between the classes are very sharp in the loading plot. The labeled wavenumbers and their known association have been included in Table 7. These will be looked at in more detail in Section 4.

Wavenumber (cm^{-1})	Association
500	C-OH3 torsion of methoxy group(1)
582	OH out of plane bending (free)
663	C-S stretching mode of cystine (collagen type I)
783	DNA Thymine, cytosine, uracil RNA U, T, C (ring breathing modes in the DNA/RNA bases)

Wavenumber (cm^{-1})	Association
1552	Tryptophan v(C==C), tryptophan (protein assignment) v(C==C), porphyrin
1682	One of absorption positions for the C==O stretching vibrations of cortisone
1714	One of absorption positions for the C==O stretching vibrations of cortisone C==O
1747	C==O, lipids

Table 7 – Top eight cluster vector peak wavenumbers and their associated chemical identities. Peaks were obtained from Raman Spectra pre-processed through Polynomial-Vector normalisations and analysed through PCA-LDA analysis. The peaks illustrate the wavenumbers that were only detected in patient samples suffering from endometriosis compared to endometriosis-free control samples

3.1.2.2 Disease Stage

Disease stage represents the quantification of the severity of the disease from 1-4. In this data set only one patient had endometriosis stage 4 so the data for this subclass if for illustration purposes only as the small sample size does not carry significance.

In Figure 24, the 1D scores plot a linear shift can be seen for stages 1, 2 and 3 where the scores medians are separated steadily along LD1. The endometriosis control can be seen right above stage 3 where their medians share the same position along LD1. The ellipse walls for the

control contain all the scores plots for stage 3. This same observation has been made for the FTIR-ATR analysed spectra. LD1 shift was observed and stage 3 similarities with the control were noted. However, no shift was seen in LD2 or LD3. In contrast, the Raman spectra show a large shift for stage 1 and 2 along LD2. The loading plot peaks show relatively low peaks for stage 3 again.

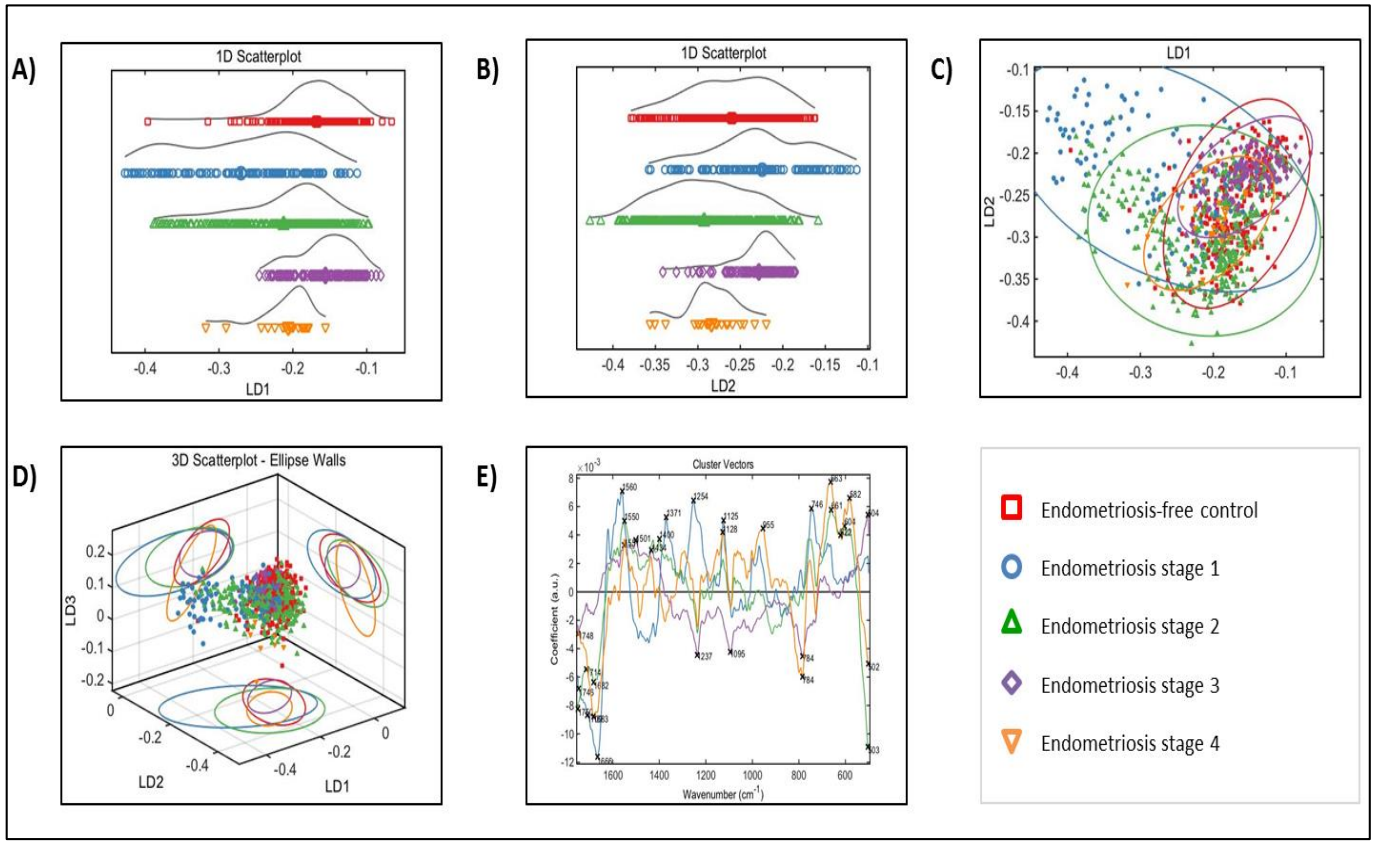


Figure 24 – Raman spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Polynomial-Vector normalisation and analysed through PCA-LDA multivariate analysis. Spectra were separated into disease stages 1-4 A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) 3D scores plot mapping LD1, LD2 and LD3 containing ellipse walls. E) A cluster vector loading plot showing distinct wavenumbers associated with endometriosis disease stages when compared to coefficient 0 (endometriosis-free control)

Endometriosis stage 3 appears to have a similar variance distribution to the endometriosis control. When looking at the statistical results for this data (Table 8) it can be seen that different methods obtain different results. Stage 4 is extremely varied throughout due to its small data set. Overall it appears that disease stages 1 and 2 are significantly different from the control. Stage 3 is only significantly different from the control for Rubber vector normalisation. However Polynomial Vector Normalisation best separates endometriosis from no endometriosis so this is the preferred method.

Raman Stage Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Polynomial Vector Normalisation	
R^2	0.2863		0.2958		0.2841	
ANOVA	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	0 vs 1	P < 0.001	0 vs 1	P < 0.001	0 vs 1	P < 0.001
	0 vs 2	P < 0.001	0 vs 2	P < 0.001	0 vs 2	P < 0.001
	0 vs 3	P > 0.05	0 vs 3	P < 0.01	0 vs 3	P > 0.05
	0 vs 4	P > 0.05	0 vs 4	P < 0.05	0 vs 4	P < 0.05
	1 vs 2	P < 0.001	1 vs 2	P < 0.001	1 vs 2	P < 0.001
	1 vs 3	P < 0.001	1 vs 3	P < 0.001	1 vs 3	P < 0.001
	1 vs 4	P < 0.001	1 vs 4	P < 0.001	1 vs 4	P < 0.001
	2 vs 3	P < 0.001	2 vs 3	P < 0.001	2 vs 3	P < 0.001
	2 vs 4	P > 0.05	2 vs 4	P > 0.05	2 vs 4	P > 0.05
	3 vs 4	P > 0.05	3 vs 4	P < 0.001	3 vs 4	P < 0.01
Dunnett's Multiple Comparison Test	0 vs 1	P < 0.01	0 vs 1	P < 0.01	0 vs 1	P < 0.01
	0 vs 2	P < 0.01	0 vs 2	P < 0.01	0 vs 2	P < 0.01
	0 vs 3	P > 0.05	0 vs 3	P < 0.01	0 vs 3	P > 0.05
	0 vs 4	P > 0.05	0 vs 4	P < 0.01	0 vs 4	P < 0.01

Table 8 – Table showing statistical analysis of LD1 generated by Raman PCA-LDA analysis of endometriosis stage subclasses. Statistical tests include R^2, ANOVA, Tukey's multiple comparison test and Dunnett's multiple comparison test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and Polynomial-Vector normalisation. Numbers 0-4 indicate disease stage, with 0 being no endometriosis presence

3.1.2.3 Disease Morphology

Lesions progress from red, more mild and hidden lesions into hard rooted black lesions that are difficult to treat. Analysis on this disease subclass was performed and compared to the results obtained through FTIT-ATR.

It appears that for this data set (Figure 25) there is a slight shift along LD1 as the scores medians do not overlap. In LD2 however there is a much greater similarity in the graph plot. Red as well as black lesions have shifted towards LD1 -0.45. A distinct cluster of scores plots can be seen for Red scores, especially along LD1.

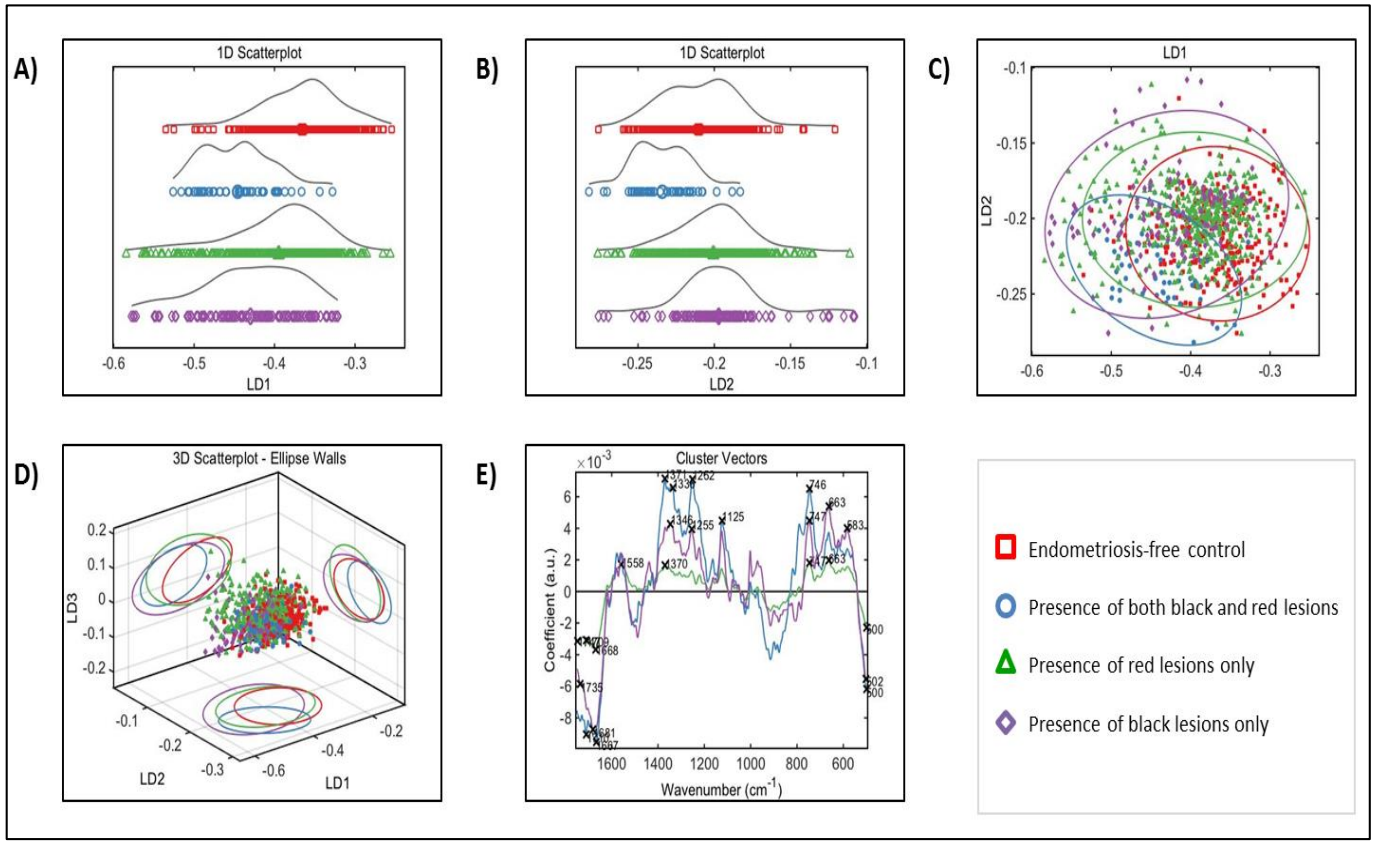


Figure 25 – Raman spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Polynomial-Vector normalisation and analysed through PCA-LDA multivariate analysis. Spectra were separated into disease morphologies of black and red lesion presence. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) 3D scores plot mapping LD1, LD2 and LD3 containing ellipse walls. E) A cluster vector loading plot showing distinct wavenumbers associated with endometriosis morphologies when compared to coefficient 0 (endometriosis-free control)

Statistical tests in Table 9 show that all of the disease morphology classes are significantly different to one another. This is shown for all of the pre-processing methods. The exception is for red lesions compared to black lesions, where no difference was observed. This is a different result than was obtained by FTIR-ATR, which showed a difference between these two classes.

Raman Black Red Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Polynomial Vector Normalisation	
R^2	0.1441		0.1443		0.1647	
ANOVA	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	0 vs R	P < 0.001	0 vs R	P < 0.001	0 vs R	P < 0.001
	0 vs BR	P < 0.001	0 vs BR	P < 0.001	0 vs BR	P < 0.001
	0 vs B	P < 0.001	0 vs B	P < 0.001	0 vs B	P < 0.001
	R vs BR	P < 0.001	R vs BR	P < 0.001	R vs BR	P < 0.001
	R vs B	P > 0.05	R vs B	P > 0.05	R vs B	P > 0.05
	BR vs B	P < 0.001	BR vs B	P < 0.001	BR vs B	P < 0.001
Dunnett's Multiple Comparison Test	0 vs R	P < 0.01	0 vs R	P < 0.01	0 vs R	P < 0.01
	0 vs BR	P < 0.01	0 vs BR	P < 0.01	0 vs BR	P < 0.01
	0 vs B	P < 0.01	0 vs B	P < 0.01	0 vs B	P < 0.01

Table 9 – Table showing statistical analysis of LD1 generated by Raman PCA-LDA analysis of endometriosis morphology subclasses. Statistical tests include R^2, ANOVA, Tukey's multiple comparison test and Dunnett's multiple comparison test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and Polynomial-Vector normalisation. Subclass morphology is indicated through letters, with Black lesions (B), Red lesions (R), both black and red lesions (BR) and endometriosis-free (0)

3.1.2.4 Ovarian Endometriosis

Patients are often diagnosed with ovarian endometriosis when they experience fertility issues. This disease characteristic is very severe and will be classified into the endometriosis a stage 3 or 4.

The 1D scores plot in Figure 26 shows a true difference between the three classes along LD1. When looking at LD2 no separation can be seen at all. The 2D scores plot shows that there is a large difference between the control and the endometriosis no ovarian sample. The ovarian

endometriosis positive scores are all held within the 95% ellipse for the no endometriosis control. While the endometriosis samples without ovarian endometriosis are much better separated. Loading plots illustrate a similar difference between endometriosis samples with and without ovarian endometriosis presence.

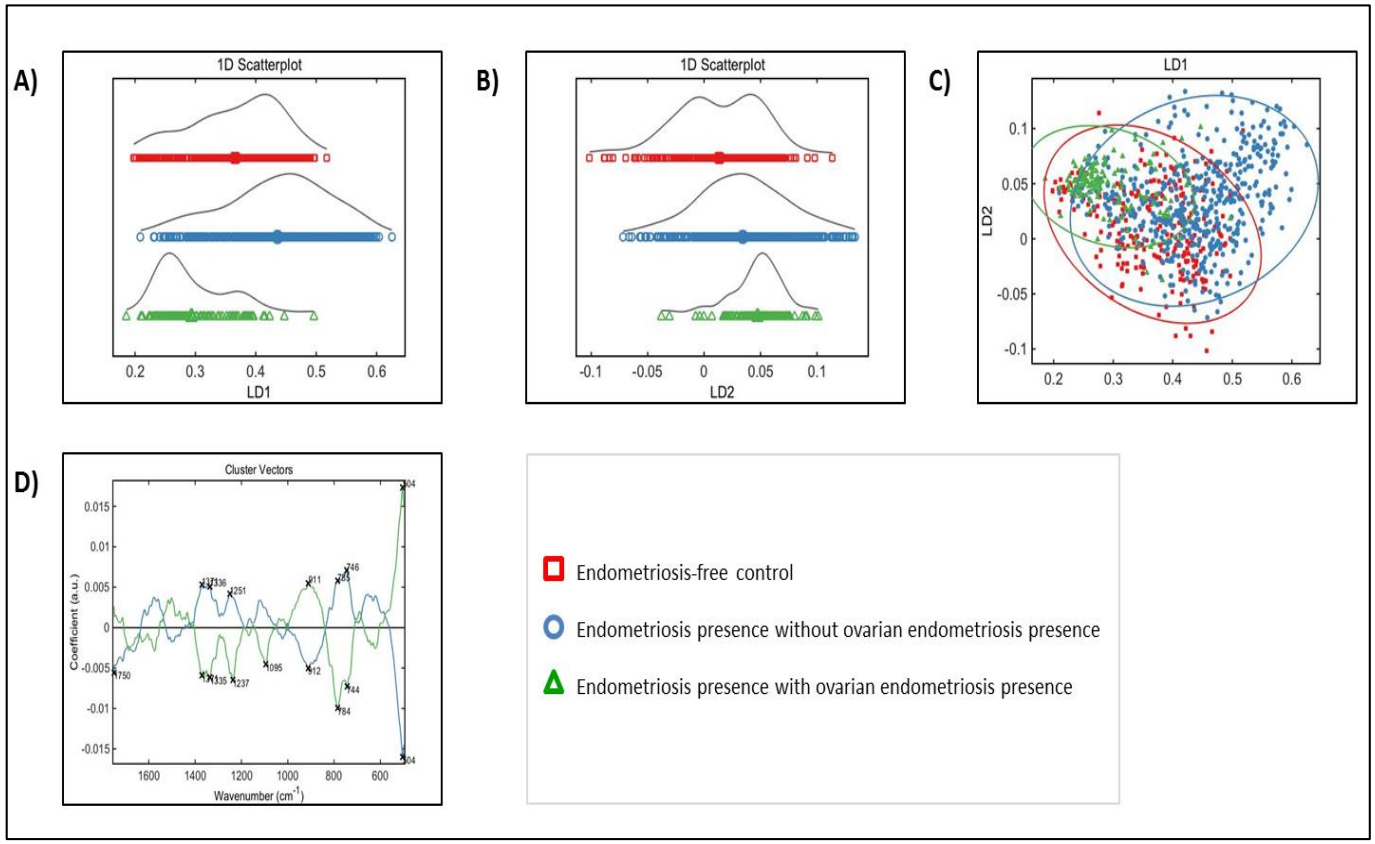


Figure 26 – Raman spectra obtained from LBC samples of endometriosis and endometriosis free patients (control) were pre-processed with Polynomial-Vector normalisation and analysed through PCA-LDA multivariate analysis. Spectra were separated into disease morphologies of black and red lesion presence. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data. D) A cluster vector loading plot showing distinct wavenumbers associated with endometriosis with and without ovarian endometriosis presence when compared to coefficient 0 (endometriosis-free control)

Statistical analysis of LD1 shown in Table 10 indicated P value < 0.01 or < 0.0001 when comparing samples. This determines that there is strong statistical significant difference between all of the samples, despite the similarity between the endometriosis-free control and the endometriosis with ovarian presence observed in the PCA-LDA scores plots. This again confirms that samples containing endometriosis can be successfully identified from endometriosis free samples, this shows great promise for the diagnostic potential of Raman spectroscopy

Raman Ovarian Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Polynomial Vector Normalisation	
R^2	0.2893		0.3036		0.3011	
ANOVA	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	0 vs N	P < 0.001	0 vs N	P < 0.001	0 vs N	P < 0.001
	0 vs Y	P < 0.001	0 vs Y	P < 0.001	0 vs Y	P < 0.001
	N vs Y	P < 0.001	N vs Y	P < 0.001	N vs Y	P < 0.001
Dunnett's Multiple Comparison Test	0 vs N	P < 0.01	0 vs N	P < 0.01	0 vs N	P < 0.01
	0 vs Y	P < 0.01	0 vs Y	P < 0.01	0 vs Y	P < 0.01

Table 10 – Table showing statistical analysis of LD1 generated by Raman PCA-LDA analysis of ovarian endometriosis. Statistical tests include R^2, ANOVA, Tukey's multiple comparison test and Dunnett's multiple comparison test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and Polynomial-Vector normalisation. Subclass morphology is indicated through letters, ovarian endometriosis presence being yes (Y), endometriosis patients without ovarian endometriosis are No (N) and endometriosis-free is 0

3.2 MCF-7 Cell Results

The effect of B[a]P exposure of MCF-7 cells was tested using several experimental methods. Different doses of B[a]P were tested for toxicity in the a clonogenic assay in Section 3.2.1. A time course assay analysed the proliferation kinetics of MCF-7 cells after B[a]P exposure in Section 3.2.3. The flow cytometry data demonstrating the cell cycle position of the MCF-7 cells during different times of B[a]P exposure is shown in Section 3.2.3. Protein expression was

quantified by western blot analysis in section 3.2.4. Error bars depicting standard deviation values have been included on all graphs.

The results obtained after analysis of MCF-7 samples through FTIR-ATR and Raman spectroscopy are shown in Section 3.2.5 and 3.2.6. PCA-LDA analysis was carried out to generate scores plots included in this section along with their corresponding table of statistical analysis on LD1. Statistical tests follow a confidence interval of 95% where $P < 0.05$.

Spectral analysis underwent three combinations of pre-processing methods outlined in Section 2.4 (Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation). As all three pre-processing methods illustrated the same statistical significance and cohesive trends for the data set, all of the generated graphs are shown to illustrate the robustness of the data. The statistical tests used follow a confidence interval of 95% where $P < 0.05$

3.2.1 B[a]P Toxicity

The dose response curve obtained after a 7 day treatment of increasing B[a]P concentrations shows that B[a]P has an immediate and strong effect on MCF-7 cell survival (Figure 27). A distinct relationship can be seen between B[a]P concentration and decreasing cell survival percentage. Not a single cell survived the highest treatment dose of 1µM.

Plating efficiency was calculated to be an average of 10.8%. A total of 500 cells were seeded for treatment and approximately 54 colonies were counted at day 7. Whilst colonies account for a large number of single cells, the low plating efficiency indicates that DMSO in itself may have a toxic effect on cell survival.

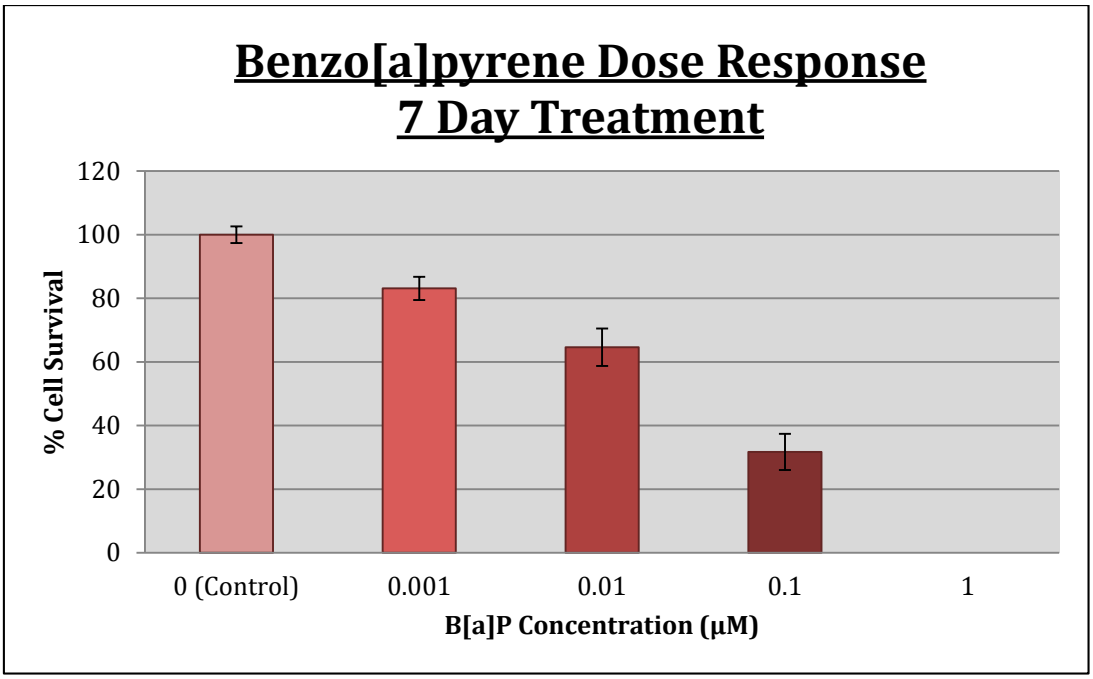


Figure 27 – Graph illustrating the dose response of MCF-7 cells following a 7 day B[a]P treatment. DMSO is used as a vehicle control for B[a]P concentration 0µM, where cell survival is normalised to 100%. Percentage Cell Survival = (no. colonies in experiment flask/ no. cells seeded) x 100. Error bars depict the standard deviation

3.2.2 Time Course Analysis of Cell Number

Quantification of relative cell number progression over a time period of 48 hours (Figure 28) allowed for a more in depth view of the proliferation response of MCF-7 cells exposed to B[a]P. There is a similar trend followed both by the control and the low B[a]P dose. Cells steadily increase in cell number over time and this accumulation speeds up as cells reach their exponential growth phase at the 24 hour time point. The low B[a]P dose mimics the clonogenic assay results in showing an overall slightly reduced cell number compared to the DMSO control.

There is however a unmistakable contrast presented by the high B[a]P dose. It shows an increased relative cell number percentage when compared to the control and the low dose. This is apparent for the initial 24 hours. After this the cell number does not just plateau but appears to drop. It is interesting to note that 6 hours post treatment both of the B[a]P dose

concentrations have a greater cell number than the control. These results outline the importance of the cell cycle for this study.

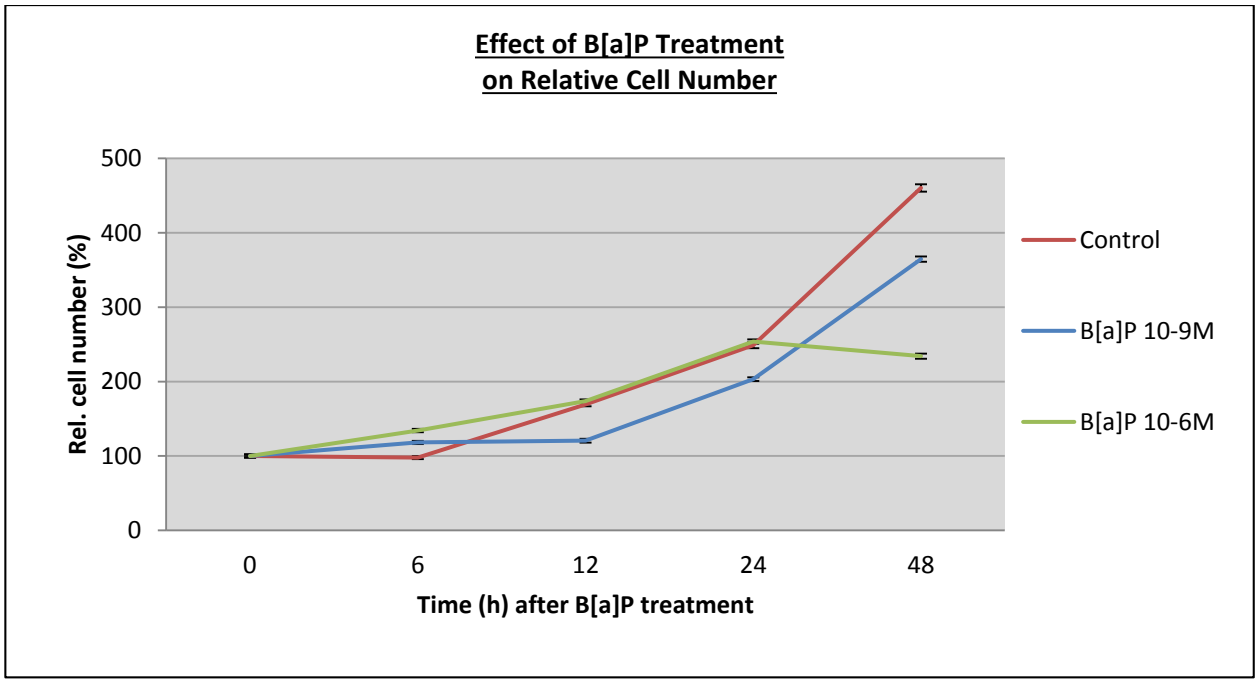


Figure 28 – Graph showing the effect of B[a]P treatment on MCF-7 cells over a period of 48 hours. Cells were treated with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Relative cell number was calculated as a percentage [%; i.e., ratio of the cell number at indicated time point relative to that determined at (normalised to 100%) x100]. Standard deviation can be visualised in the error bars

3.2.3 Identification of Cell Cycle Phase

Cells were run through a flow cytometer to determine MCF-7 cell cycle distribution. The results are presented in bar graphs in Figures 29, 30, and 31. The distribution of G_0/G_1 phase, S phase and G_2/M phase is visualised as a percentage. Culture conditions refer to the amount of time cells were grown post seeding. Treatment conditions outline the B[a]P concentrations. Cell cycle distribution of cells at time of seeding can be seen in the 0 hour control. Standard deviation was calculated and can be visualised through the error bars.

Control measurements were taken to check that 24 and 96 hours were the appropriate time points to concentrate cells in S phase and G_0/G_1 phase (Figure 29). At time of seeding (0 hour)

most cells were found to be in early G phase (G_0/G_1). Cells grown for 96 hours were successfully shown to contain the majority of cells in G_0/G_1 phase. Growing cells for 24 hours greatly increased the percentage of cells located in S phase. Even though the cell cycle distribution was equal for all the phases, this is the highest observed S phase distribution and lowest G phase distribution.

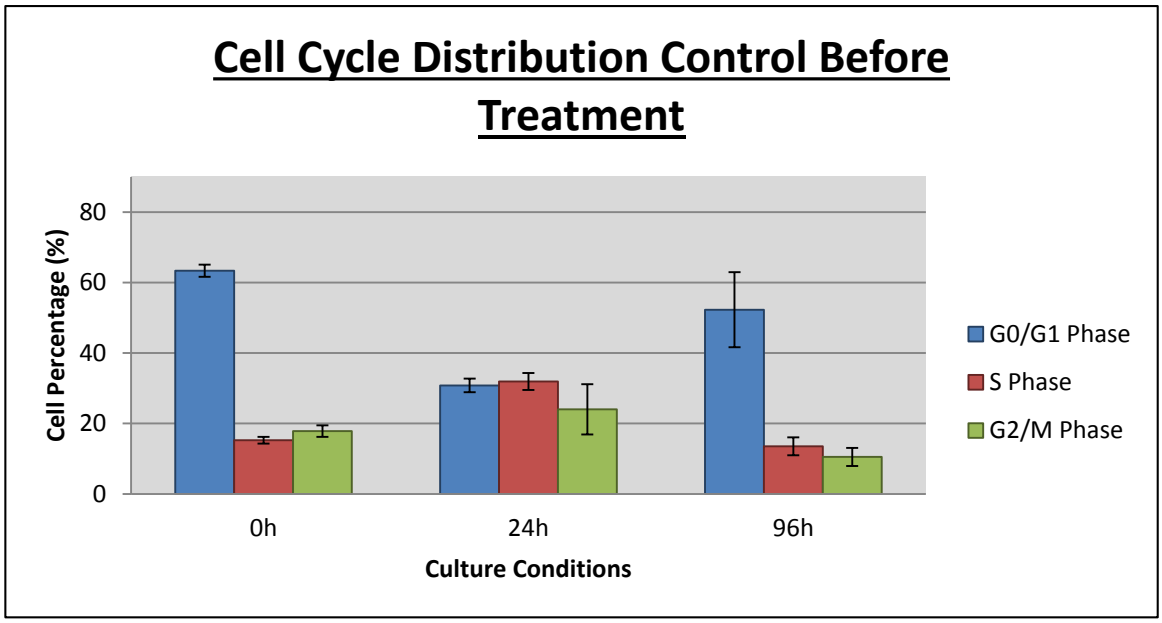


Figure 29 – Graph showing the cell cycle distribution of MCF-7 cells before treatment of B[a]P. Cells were grown for 0 hours, 24 hours and 96 hours before being analysed by flow cytometry. Cells are distributed in G_0/G_1, S or G_2/M phases. The experiment was repeated two independent times with triplicate flasks. Data was averaged and error bars were added to depict the standard deviation

Using the knowledge that 24 hour growth results in S phase concentrated cells, cells were treated with a 24 hour B[a]P treatment during S phase (Figure 30). Varying concentrations of B[a]P resulted in a different cell cycle distribution. DMSO and a low treatment dose showed similar results of a high G_0/G_1 presence and a low even distribution of S and G_2/M phase. A high concentration of B[a]P showed a very different result. It echoes the findings of the time course experiment, where the low B[a]P dose and the control are similar, whilst only the high dose elicits a change. The percentage of cells located in S phase greatly increases for B[a]P 10^{-6}M. It appears that when cells are treated with a high concentration of B[a]P after a 24 hour

incubation (when a third of the cell population is in S phase) cells gain the ability to surpass G_0/G_1 more easily and therefore show an increased presence in S phase.

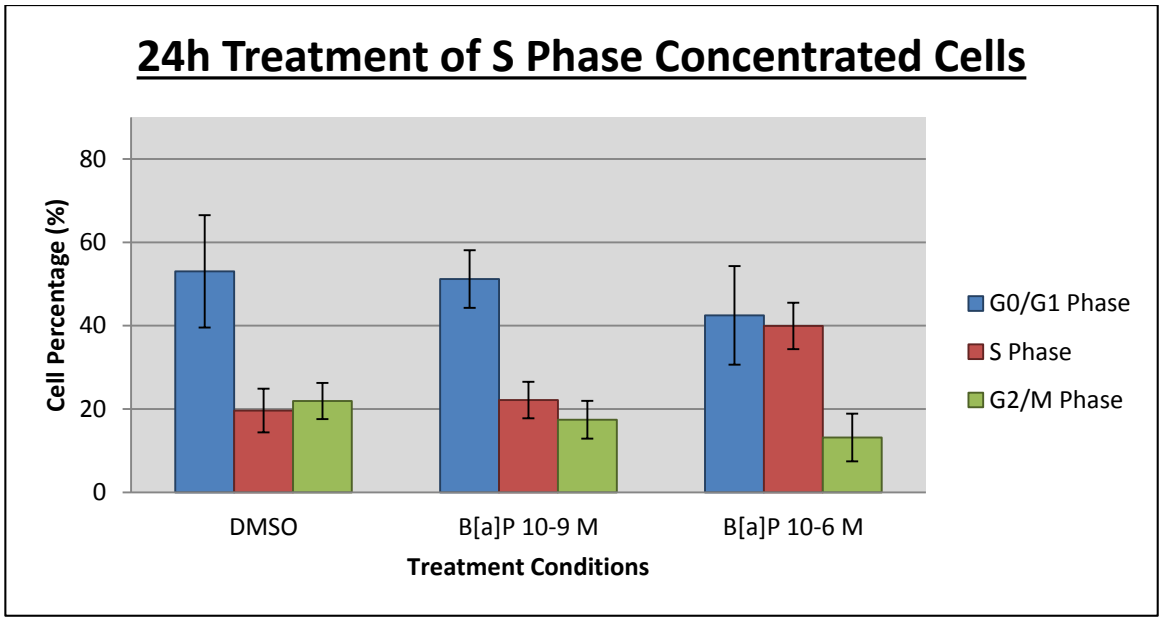


Figure 30 – Graph showing the cell cycle distribution of MCF-7 cells after a 24 hour treatment of B[a]P. Cells were grown for 24 hours (concentrated in S phase) then treated with concentrations of 10^{-9}M B[a]P, 10^{-6}M B[a]P or a DMSO only control and subsequently analysed by flow cytometry. Cells are distributed in G_0/G_1, S or G_2/M phases. The experiment was repeated two independent times with triplicate flasks. Data was averaged and error bars were added to depict the standard deviation

When G_0/G_1 phase concentrated cells were treated with B[a]P very minimal difference was observed between the varying treatment concentrations (Figure 31). All three conditions showed an overwhelming G_0/G_1 presence. A slight shift of cell percentage distribution can be seen from G_0/G_1 into S phase when cells treated with a high B[a]P dose. Yet the large error bars representing standard deviation make this shift insignificant.

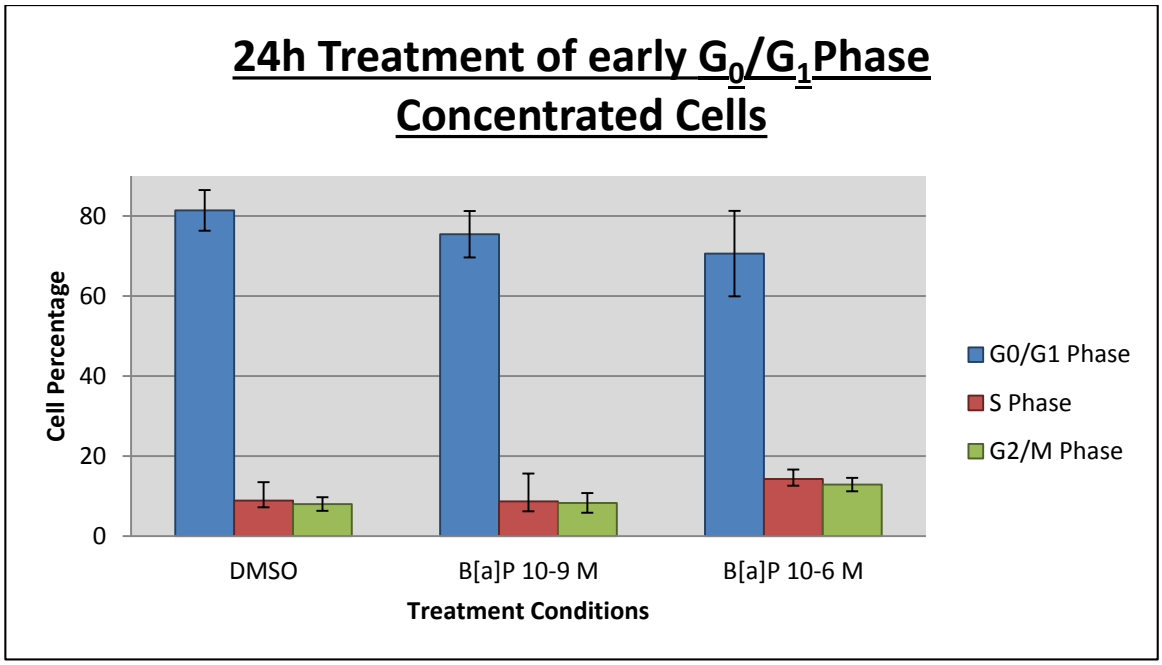


Figure 31 – Graph showing the cell cycle distribution of MCF-7 cells after a 24 hour treatment of B[a]P. Cells were grown for 96 hours (concentrated in early G_0/G_1 phase) prior to the treatment with concentrations of 10^{-9}M B[a]P, 10^{-6}M B[a]P or a DMSO only control and then analysed by flow cytometry. Cells are distributed in G_0/G_1, S or G_2/M phases. The experiment was repeated two independent times with triplicate flasks. Data was averaged and error bars were added to depict the standard deviation

3.2.4 Protein Expression

A western blot was preformed to examine how different concentrations of B[a]P was affecting protein expression when applied to cells in S phase and G_0/G_1 phase (Figure 32). This was done in the hope to uncover a better understanding of the molecular mechanisms occurring during the observed S phase proliferation phenomenon induced by a high dose of B[a]P.

Protein expression can be seen for CYP1A1 during S phase and G_0/G_1 phase. Its presence is observed when B[a]P is exposed to G_0/G_1 phase concentrated cells. It is also seen in S phase at both concentrations. However, in the replicates, either of the concentrations expressed CYP1A1. This, coupled with the poor β-actin blot, brings into question the validity of this data.

It can be observed that CYP1B1 is clearly expressed in G_0/G_1 phase cells. This is true for all G_0/G_1 phase concentrations, although the control expressed the lower amounts of protein.

The high dose of B[a]P expressed the most. During S phase, a reduced protein expression was observed. However, it is faintly present throughout the treatment concentrations with the darkest band observed at the DMSO control. It seems S phase and G_0/G_1 phase show conflicting results.

P21 expression is extremely faint. However, a slight band can be seen in the high dose treatment for both S phase and G_0/G_1 phase concentrated cells. Again, a poor β-actin blot disputes the validity of this data and explains the weak band intensity for P21.

For BCL-2, the low molecular weight ladder showing a band at 25kDA was chosen to be included in the western blot results. Whilst no bands can be seen at 26kDA, some can be observed just below the 25kDA band. According to the β-actin staining the blot was very successful. Therefore it can be determined that no BCL-2 presence was expressed in any of the treatment and culture conditions.

BCL-2 expression is coupled with P21 expression and the expectation is to observe down regulation of P21 when BCL-2 is overexpressed. Since no BCL-2 was detected, it can be concluded that the P21 western blot would require repeating, as it was not suppressed by BCL-2 presence.

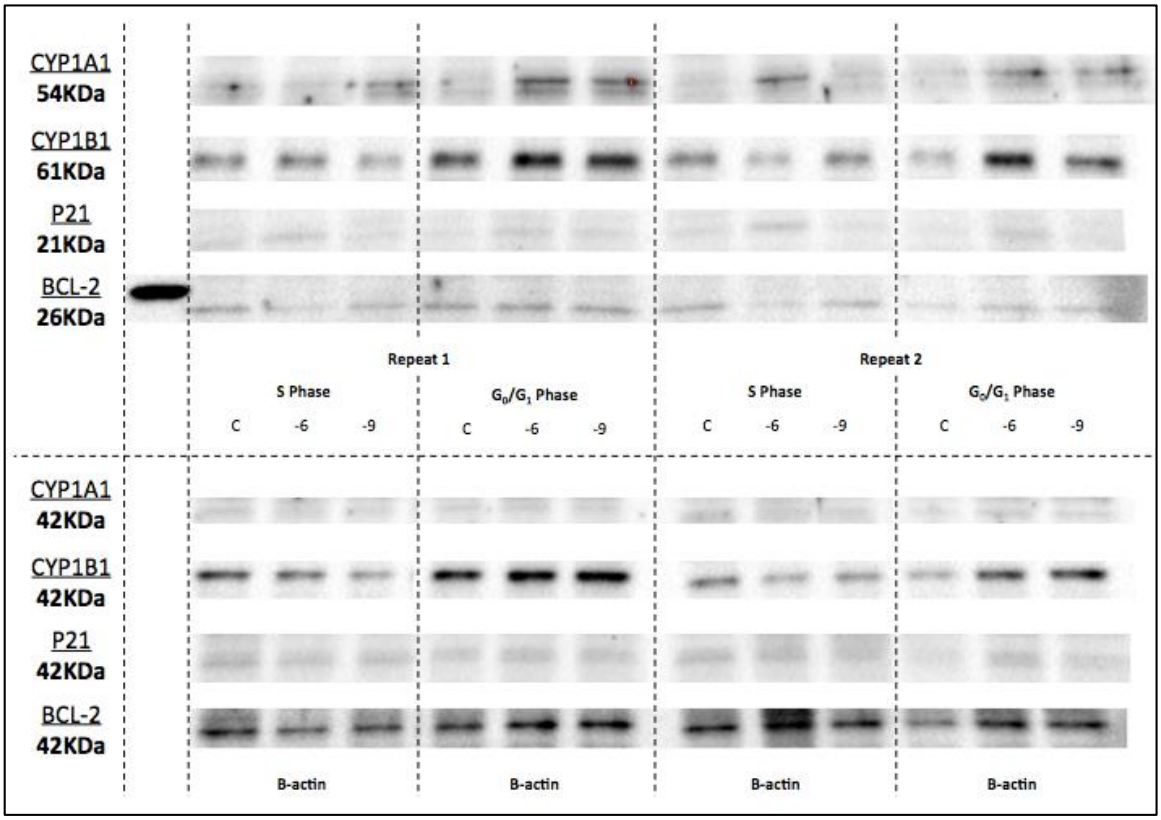


Figure 32 – Western Blot showing CYP1A1, CYP1B1, P21 and BCL-2 protein expression. The 25kDa band can be seen in the first well of BCL-2. Culture conditions include C= DMSO control, -6 = B[a]P 10^{-6}M and -9 = B[a]P 10^{-9}M. Duplicate blots can be seen as repeat 1 and repeat 2. B-actin expression was measured at 42kDa and can be seen on the lower half of the figure

3.2.5 Analysis of FTIR-ATR Spectra

Graphs generated after the PCA-LDA analysis of spectra obtained through FTIR-ATR spectroscopy on MCF-7 cells fixed on low-e slides are shown in this section. MCF-7 cells were concentrated in either S phase or G_0/G_1 prior to a 24h treatment with either DMSO (control), B[a]P 10^{-9}M or B[a]P 10^{-6}M. Graphs obtained after PCA-LDA analysis are shown below along tables showing statistical evaluation of LD1 results (capturing > 90% of data variance). Statistical tests follow a confidence interval of 95% where $P < 0.05$. Three different pre-processing methods were applied to the data set before PCA-LDA analysis, these include Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. All three pre-processing methods resulted in near identical results.

3.2.5.1 S Phase Concentrated Cells

S phase concentrated MCF-7 cells treated with varying B[a]P concentrations underwent spectral analysis to determine if B[a]P induces a changes in the biochemical fingerprint of sample when treated at a specific point in the cell cycle. Previous results indicate a significance of the S phase when a high B[a]P treatment is applied.

In Figure 33 (Rubber band-Amide I), Figure 34 (Rubber band-Vector normalisation) and Figure 35 (SG differentiation-Vector normalisation) a distinct separation can be seen for the highest dose of B[a]P compared to the DMSO control and B[a]P 10^{-9}M across LD1. A shift is observed towards LD1 -0.1 whilst the lower dose and the control stay by LD1 0.1. In the 2D scores plot only two scores connect the 95% ellipse of B[a]P 10^{-6}M to the control and in SG differentiation-Vector normalisation these two conditions are separated entirely across LD1 and LD2.

When comparing the DMSO control to the low Dose of B[a]P, a definite overlap can be seen across LD1 on both the 1D and 2D scores plots. This is the lease for SG differentiation-Vector normalisation best separated these treatment conditions, interesting it is the only pre-processing method that illustrated a shift across LD2. The B[a]P treatments both shift towards LD2 0.01 compared to the control at LD2 -0.01.

The cluster vector illustrates very large peaks for the B[a]P 10^{-6}M when set against DMSO control (coefficient 0), smaller but seemingly significant peaks can be seen for B[a]P 10^{-9}M.

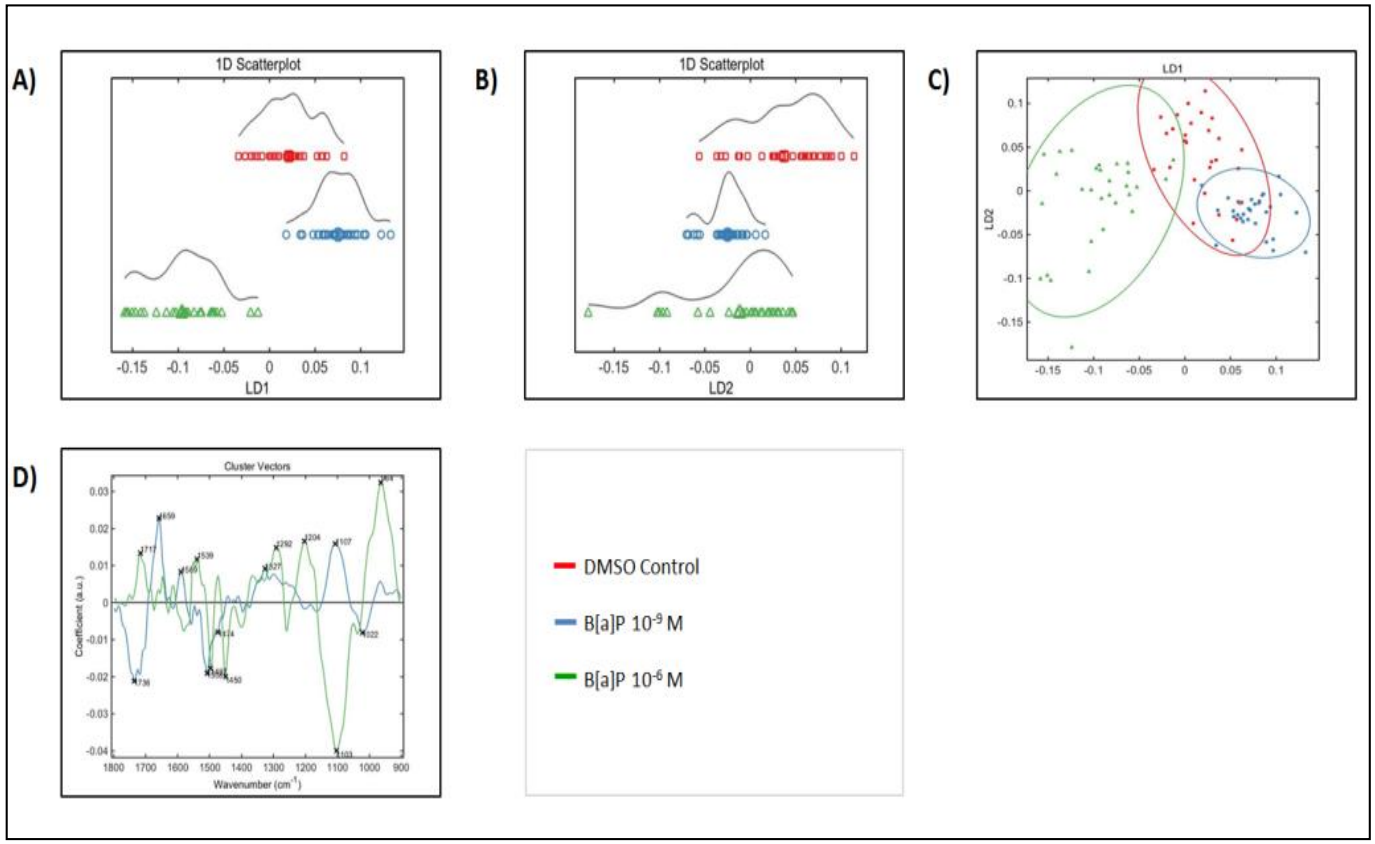


Figure 33 – FTIR-ATR spectra obtained from S phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

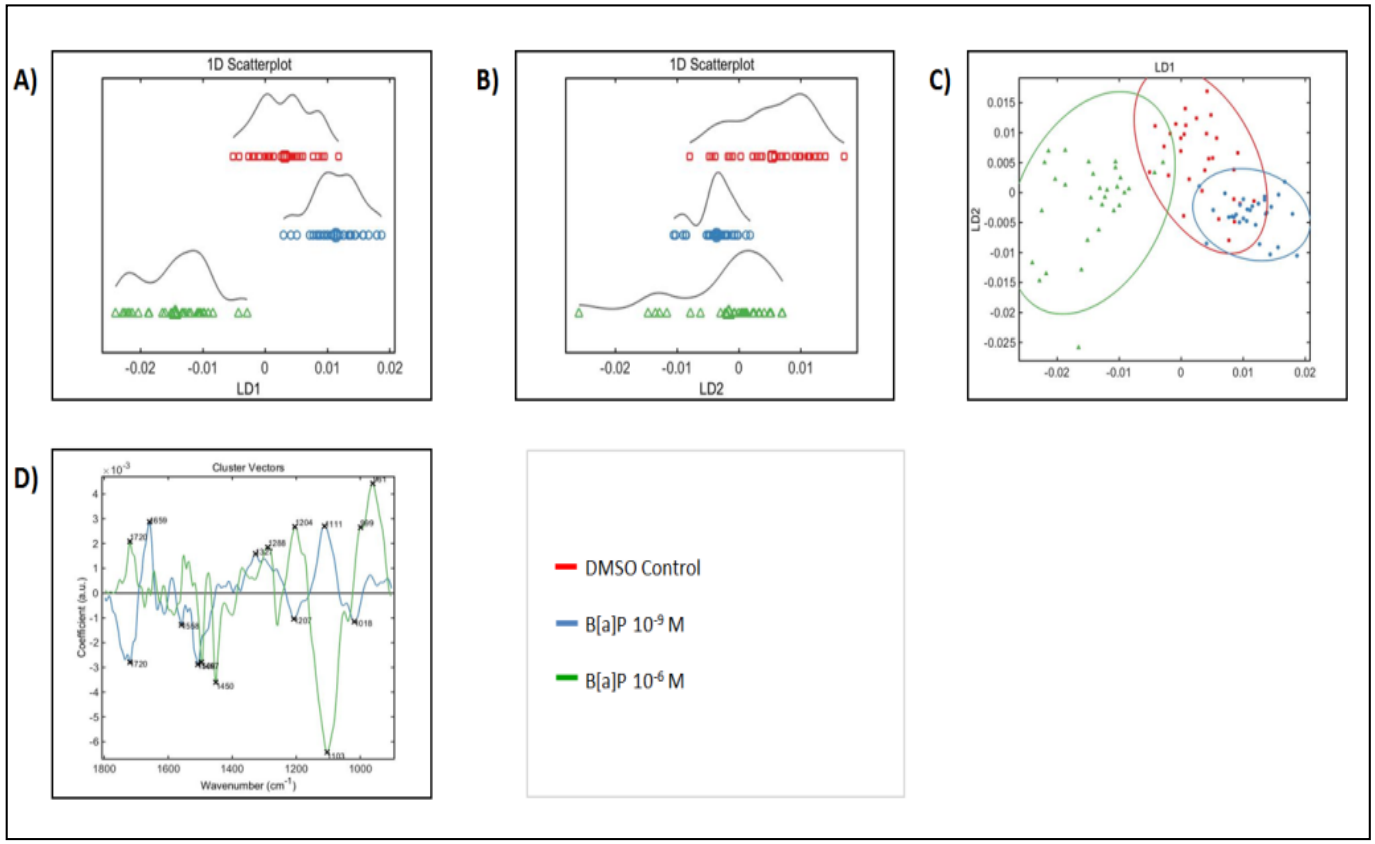


Figure 34 – FTIR-ATR spectra obtained from S phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

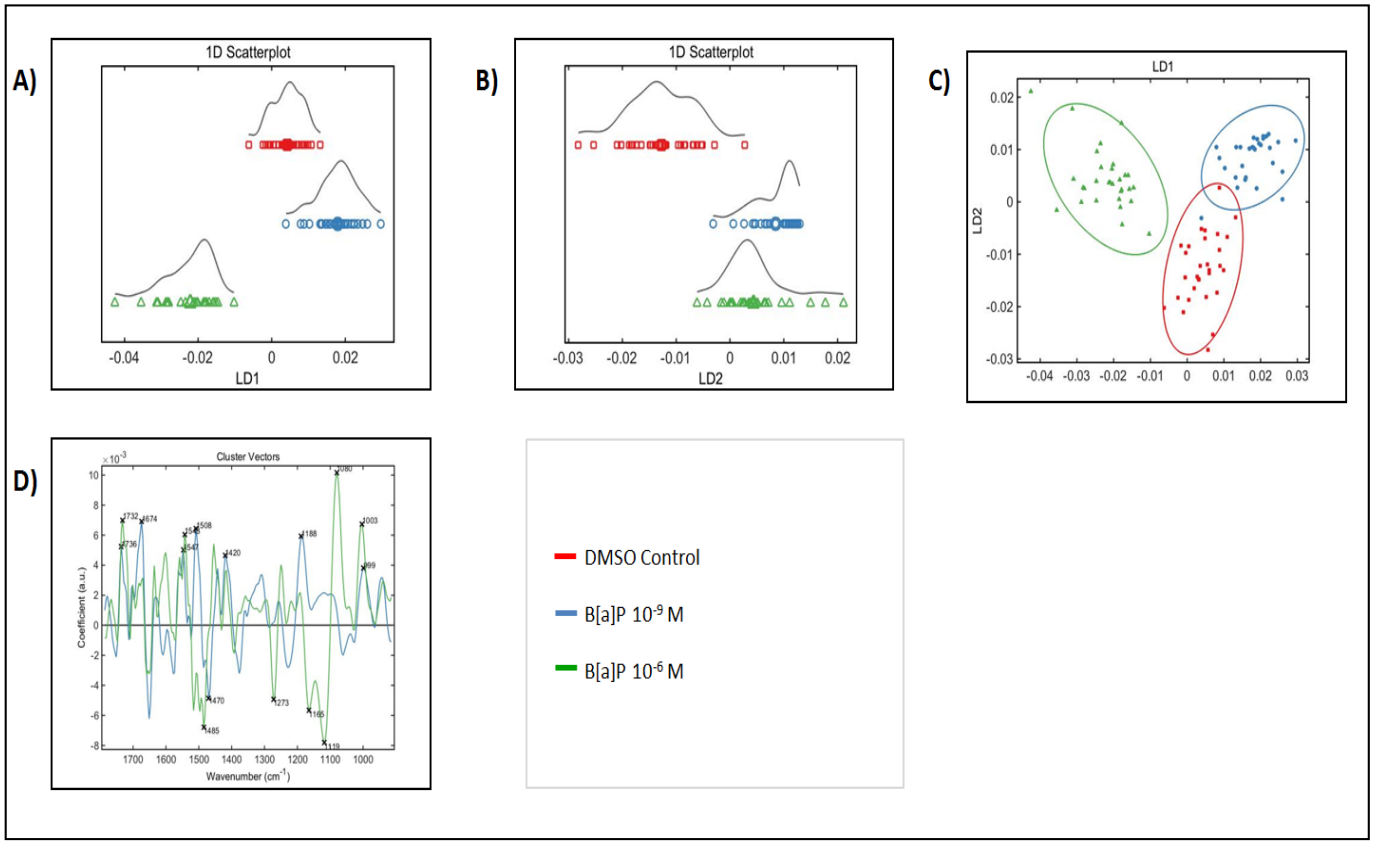


Figure 35 – FTIR-ATR spectra obtained from S phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with SG differentiation-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

The initial observed separation along LD1 of B[a]P 10^{-6}M treated cells from the control, DMSO only treated cells, is validated through statistical analysis in Table 11. It confirms that the high dose of B[a]P is significantly different to the control and the low dose B[a]P, with a P value < 0.001. It also shows that the DMSO control and the low B[a]P dose can be significantly be separated from each other despite their scores plots overlap.

ATR S Phase Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Sg Differentiation Vector Normalisation	
R^2	0.8397		0.8479		0.8956	
Unpaired t-Test	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	C vs -9	P < 0.001	C vs -9	P < 0.001	C vs -9	P < 0.001
	C vs -6	P < 0.001	C vs -6	P < 0.001	C vs -6	P < 0.001
	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001

Table 11 - Table showing statistical analysis of LD1 generated by FTIR-ATR PCA-LDA analysis on S phase concentrated MCF-7 cells treated with B[a]P. Statistical tests include R^2, Unpaired t-Test and F-test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. Treatment conditions are indicated through letters, where C = DMSO control. -9 = B[a]P 10^{-9}M. -6 = B[a]P 10^{-6}M

3.2.5.2 G_0/G_1 Phase Concentrated Cells

MCF-7 cells concentrated in G_0/G_1 phase and treated with B[a]P at concentrations of 10^{-9}M and 10^{-6}M underwent spectral analysis to determine if B[a]P susceptibility varies when cells are treated at different stages in the cell cycle.

Figures 36 (Rubber band-Amide I) Figure 37 (Rubber band-Vector normalisation) and Figure 38 (SG differentiation-Vector normalisation) show a clear separation between all three treatment conditions. There is a gradual shift along LD1 as the concentrations are becoming increasingly different, this is apparent in the 1D and the 2D scores plots. Interestingly the ellipse walls (with a 95% confidence interval) show that the B[a]P dose of 10^{-9}M is entirely contained within the region of overlap between the control and B[a]P 10^{-6}M. No separation of

treatment conditions can be observed along LD2. The cluster vector displays a similar trend where the higher the B[a]P dose, the larger the peak.

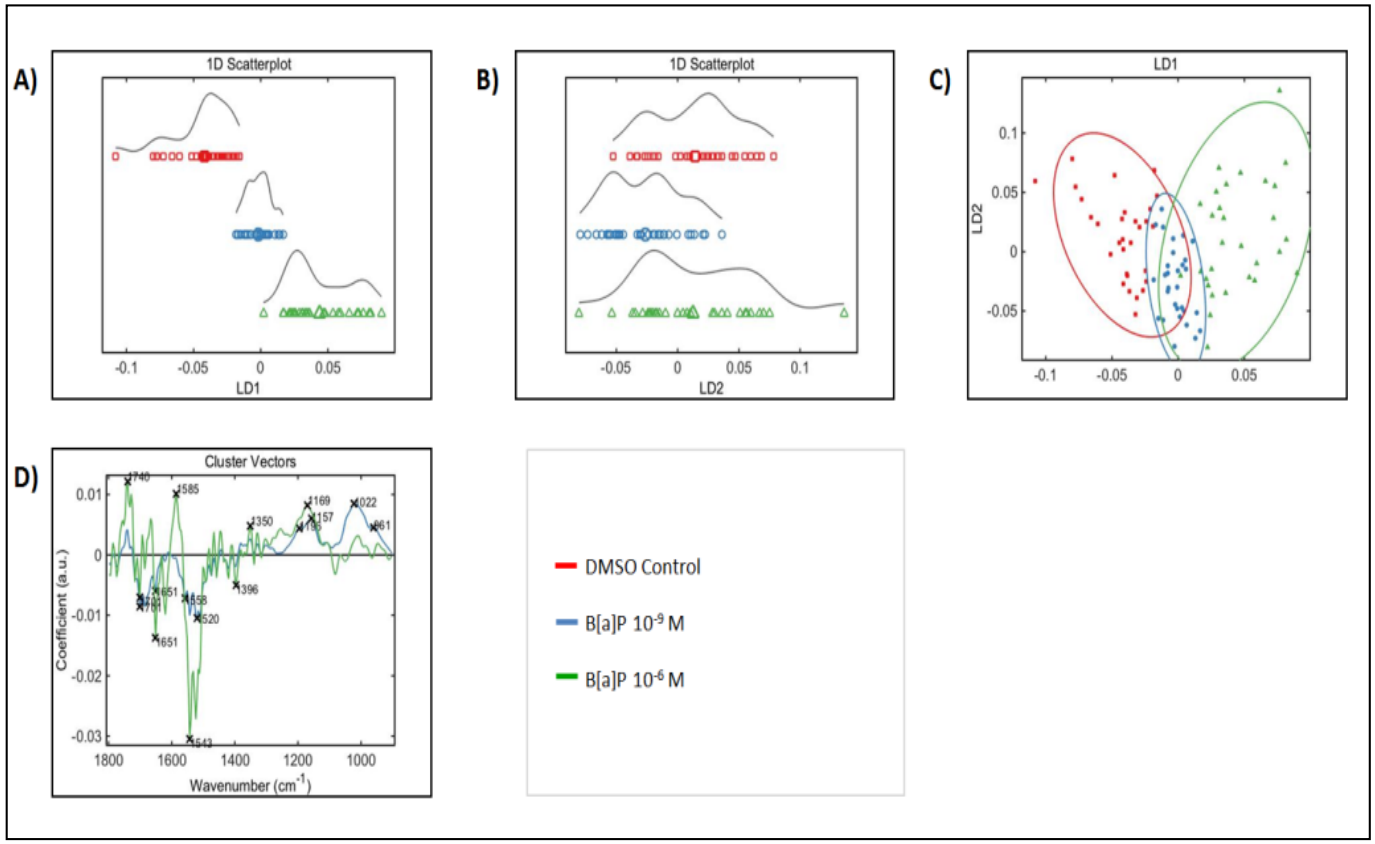


Figure 36 – FTIR-ATR spectra obtained from G_0/G_1 phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

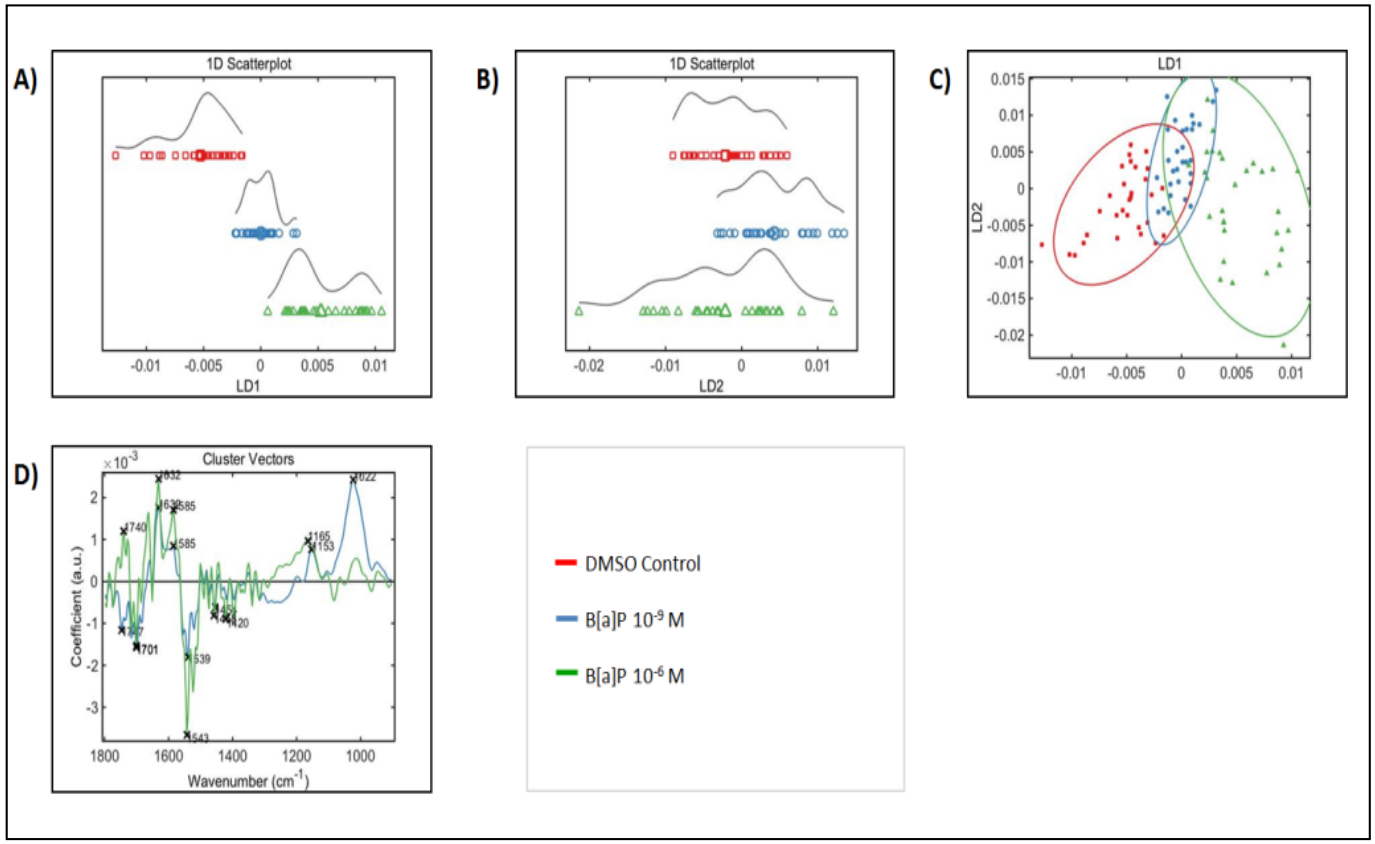


Figure 37 – FTIR-ATR spectra obtained from G_0/G_1 phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

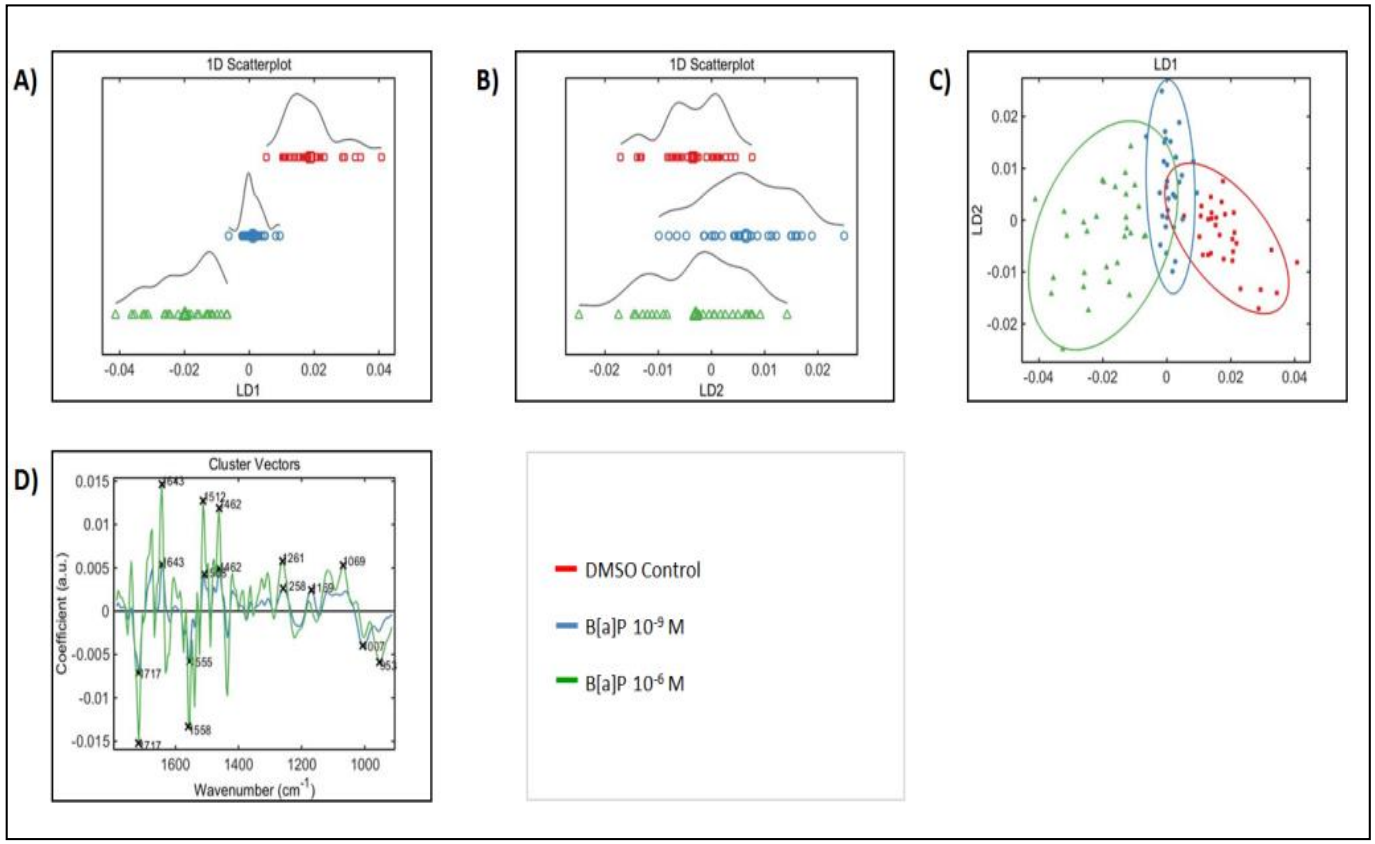


Figure 38 – FTIR-ATR spectra obtained from G_0/G_1 phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with SG differentiation-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

Table 12 shows that all treatment conditions carry significant statistical difference and can be separated successfully with a P value < 0.001. The P value results for this data is identical to the data obtained for S phase treated cells. The only difference between cells treated in S phase and G_0/G_1 phase is the visual trend observed in the scores plots where S phase shows to be susceptible to only the high, B[a]P 10^{-6}M, dose. This trend is not statistically significant and can therefore be disputed.

ATR G Phase Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Sg Differentiation Vector Normalisation	
R^2	0.7755		0.7760		0.8260	
Unpaired t-Test	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	C vs -9	P < 0.001	C vs -9	P < 0.001	C vs -9	P < 0.001
	C vs -6	P < 0.001	C vs -6	P < 0.001	C vs -6	P < 0.001
	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001

Table 12 – Table showing statistical analysis of LD1 generated by FTIR-ATR PCA-LDA analysis on G_0/G_1 phase concentrated MCF-7 cells treated with B[a]P. Statistical tests include R^2, Unpaired t-Test and F-test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. Treatment conditions are indicated through letters, where C = DMSO control. -9 = B[a]P 10^{-9}M. -6 = B[a]P 10^{-6}M

3.2.6 Analysis of Raman Spectra

Graphs generated after the PCA-LDA analysis of spectra obtained through Raman spectroscopy on MCF-7 cells fixed on low-e slides are shown in this section. MCF-7 cells concentrated in S phase or G_0/G_1 phase underwent a 24hour treatment with either a DMSO (control), B[a]P 10^{-9}M or B[a]P 10^{-6}M. Graphs generated after PCA-LDA analysis are shown below together with tables showing statistical evaluation of LD1 results. Statistical tests follow a confidence interval of 95% where P < 0.05.

3.2.6.1 S Phase Concentrated Cells

Raman spectroscopy was applied to MCF-7 cells concentrated in S phase and subsequently treated with varying concentrations of B[a]P. Spectral analysis through PCA-LDA resulted in the graphs shown in Figures 39 (Rubber band-Amide I) Figure 40 (Rubber band-Vector normalisation) and Figure 41 (SG differentiation-Vector normalisation). S phase concentrated cells have been shown to have a higher susceptibility to B[a]P's genotoxic effects.

Only a slight shift across LD1 can be seen between the different treatment conditions in the 1D and 2D scores plots. This is mainly apparent for the high dose B[a]P which is partially separated from the low dose B[a]P and the control (DMSO only). No shift is visible across LD2 as all three conditions remain parallel. On the 2D scores plot, the ellipse walls overlap significantly, with the entire control ellipse being contained within the B[a]P 10^{-9}M ellipse walls. It is clear that only the high dose condition can be separated from the other conditions through Raman spectroscopy whilst the control and low dose B[a]P contain very similar spectral properties.

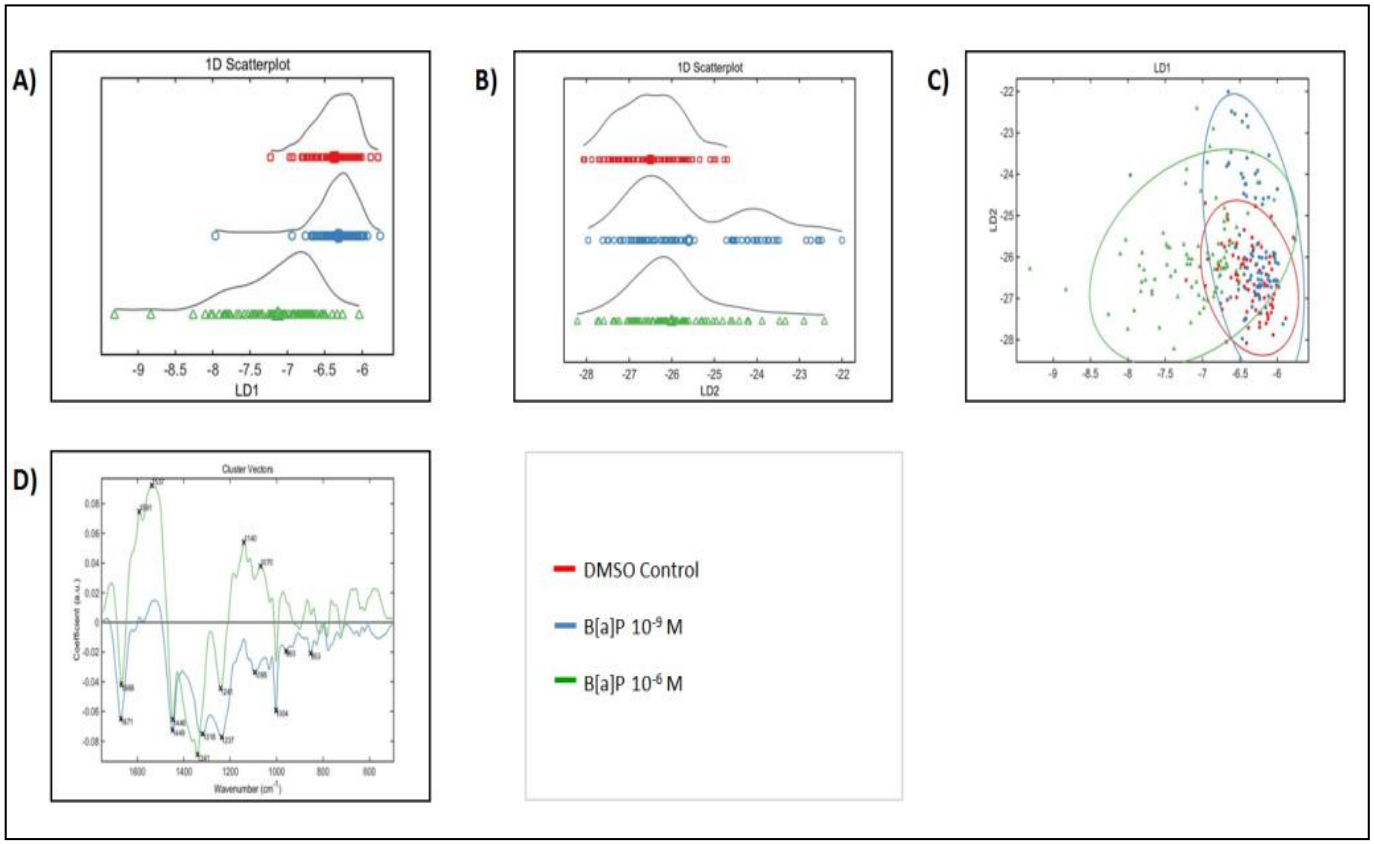


Figure 39 – Raman spectra obtained from S phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

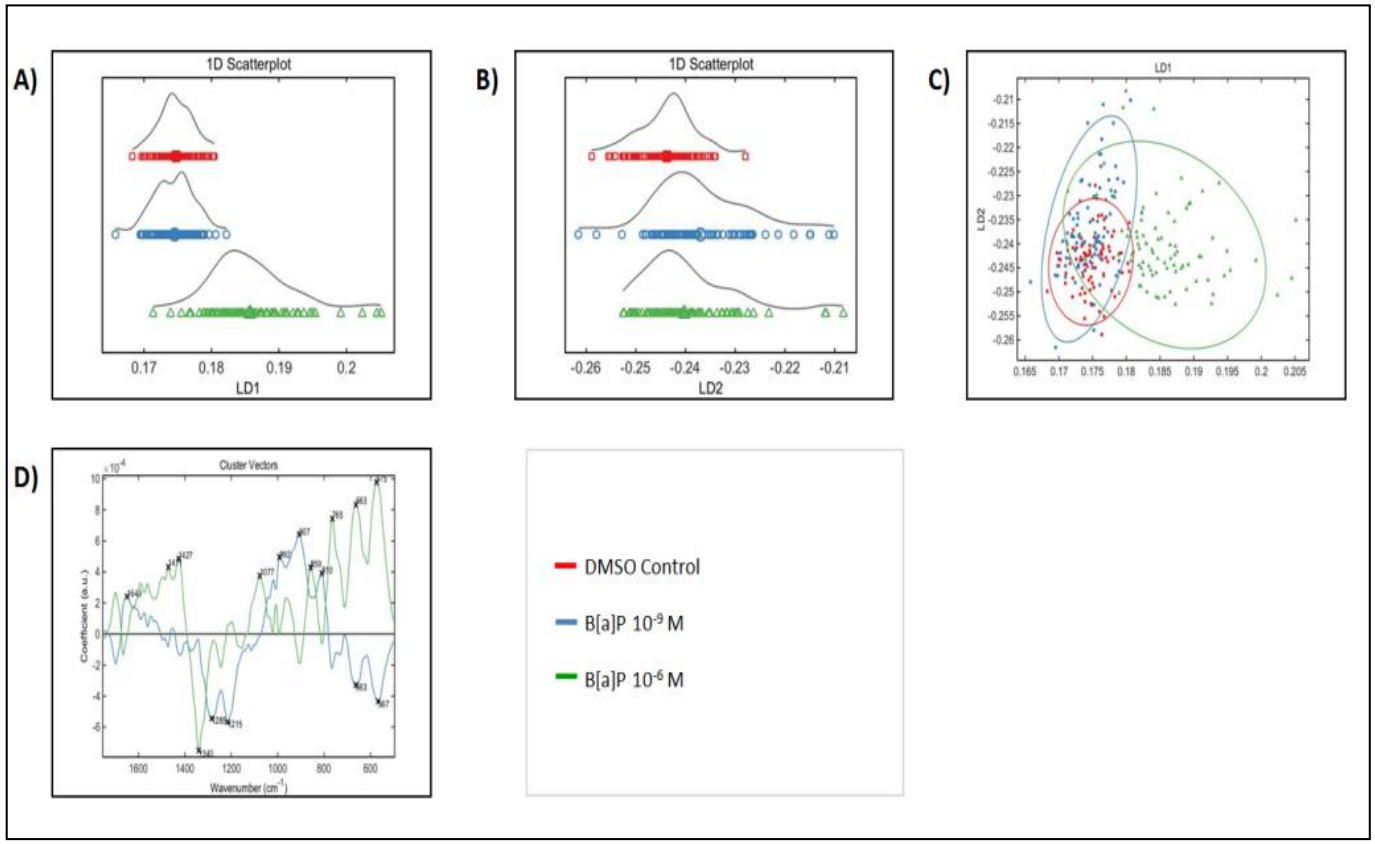


Figure 40 – Raman spectra obtained from S phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

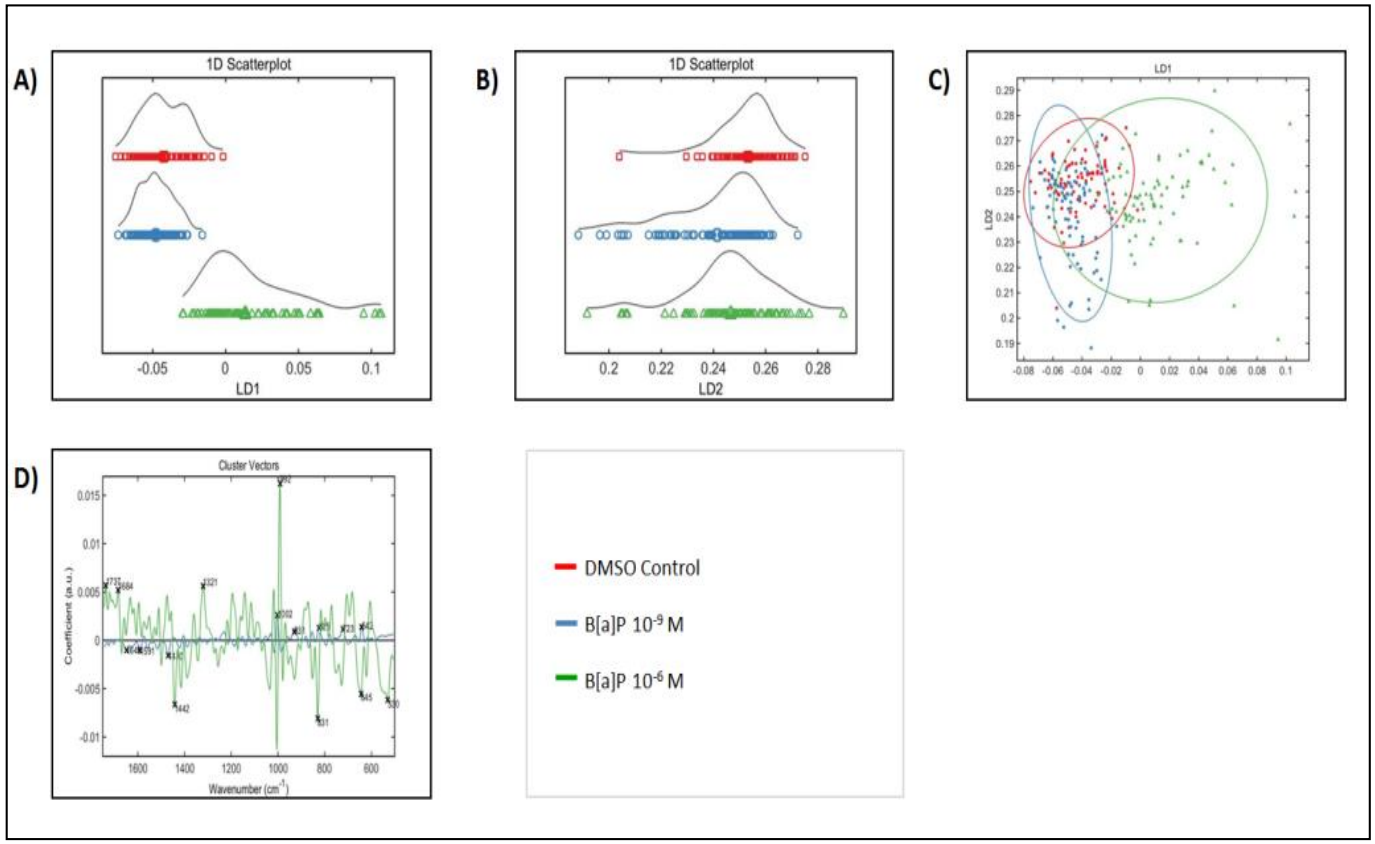


Figure 41 – Raman spectra obtained from S phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with SG differentiation-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

The observations obtained from the scores plots are reinforced through the statistical analysis shown in Table 13. The DMSO control (containing no B[a]P) is not found to be significantly different than the low dose B[a]P of 10^{-6}M as P > 0.05. The analysed spectra acquired from the high B[a]P concentration varies greatly from the control and the low dose, where P < 0.001. This is true for all three of the pre-processing methods, as they show coherent results.

Raman S Phase Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Sg Differentiation Vector Normalisation	
R^2	0.4651		0.6101		0.6474	
Unpaired t-Test	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	C vs -9	P > 0.05	C vs -9	P > 0.05	C vs -9	P > 0.05
	C vs -6	P < 0.001	C vs -6	P < 0.001	C vs -6	P < 0.001
	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001

Table 13 – Table showing statistical analysis of LD1 generated by Raman PCA-LDA analysis on S phase concentrated MCF-7 cells treated with B[a]P. Statistical tests include R^2, Unpaired t-Test and F-test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. Treatment conditions are indicated through letters, where C = DMSO control. -9 = B[a]P 10^{-9}M. -6 = B[a]P 10^{-6}M

3.2.6.2 G_0/G_1 Phase Concentrated Cells

When G_0/G_1 concentrated MCF-7 cells were treated with B[a]P and analysed through Raman spectroscopy the results echoed the findings from the S phase concentrated cells. Figures 42 (Rubber band-Amide I), Figure 43 (Rubber band-Vector normalisation) and Figure 44 (SG differentiation-Vector normalisation) contain scores plots and cluster vectors after PCA-LDA analysis.

A very clear shift across LD1 can be observed in the 1D and 2D scores plots. As consistently stated, B[a]P 10^{-6}M is strongly separated from the control and 10^{-9}M. The DMSO control and low treatment dose are not seemingly different to each other. On the 2D scores plot, the

ellipse walls of the DMSO control encapsulate the majority of the scores from the low B[a]P dose. The ellipse walls of the high B[a]P dose barely overlap at all with the other scores and hereby show a strong separation along LD1.

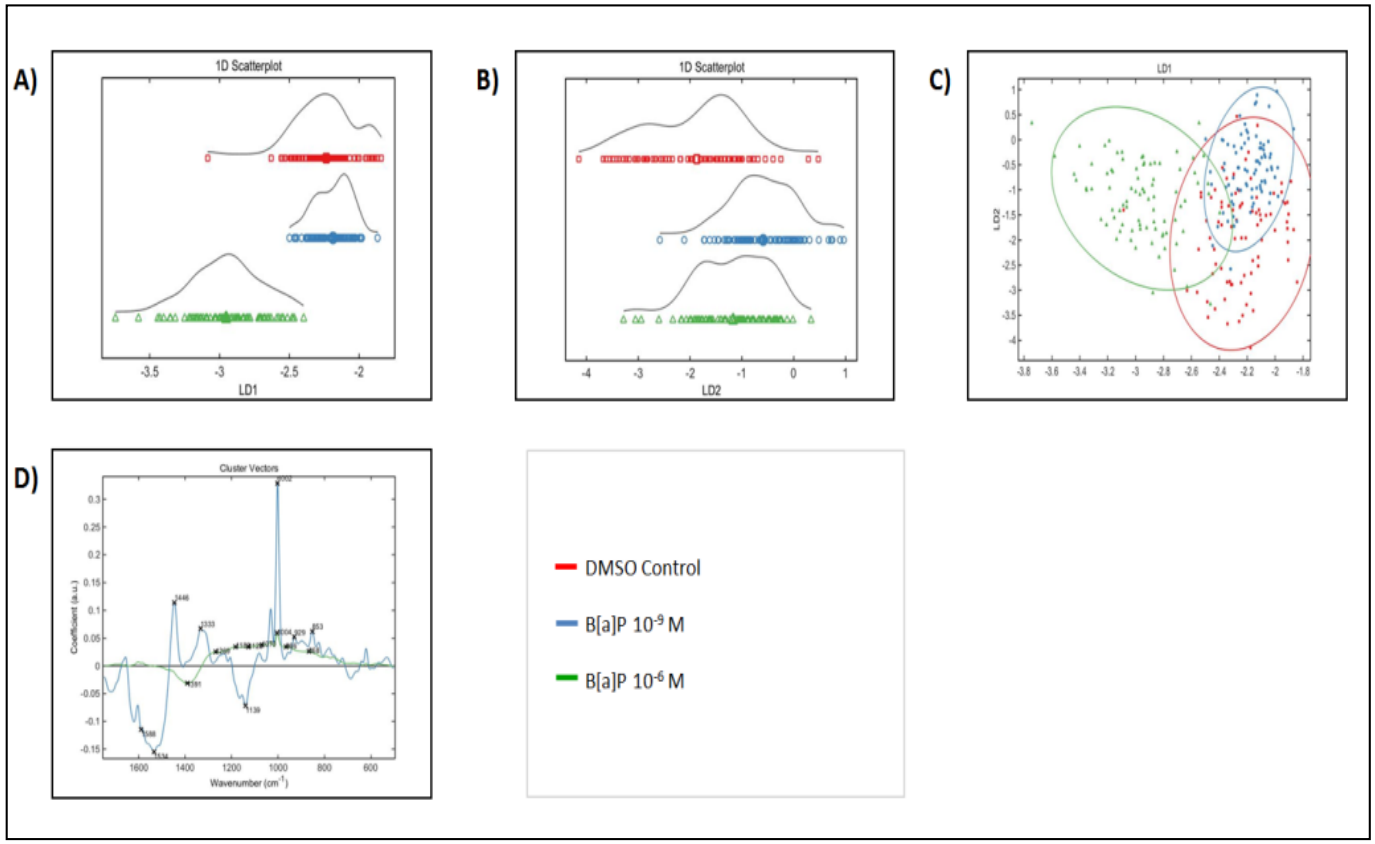


Figure 42 – Raman spectra obtained from G_0/G_1 phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Amide I and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

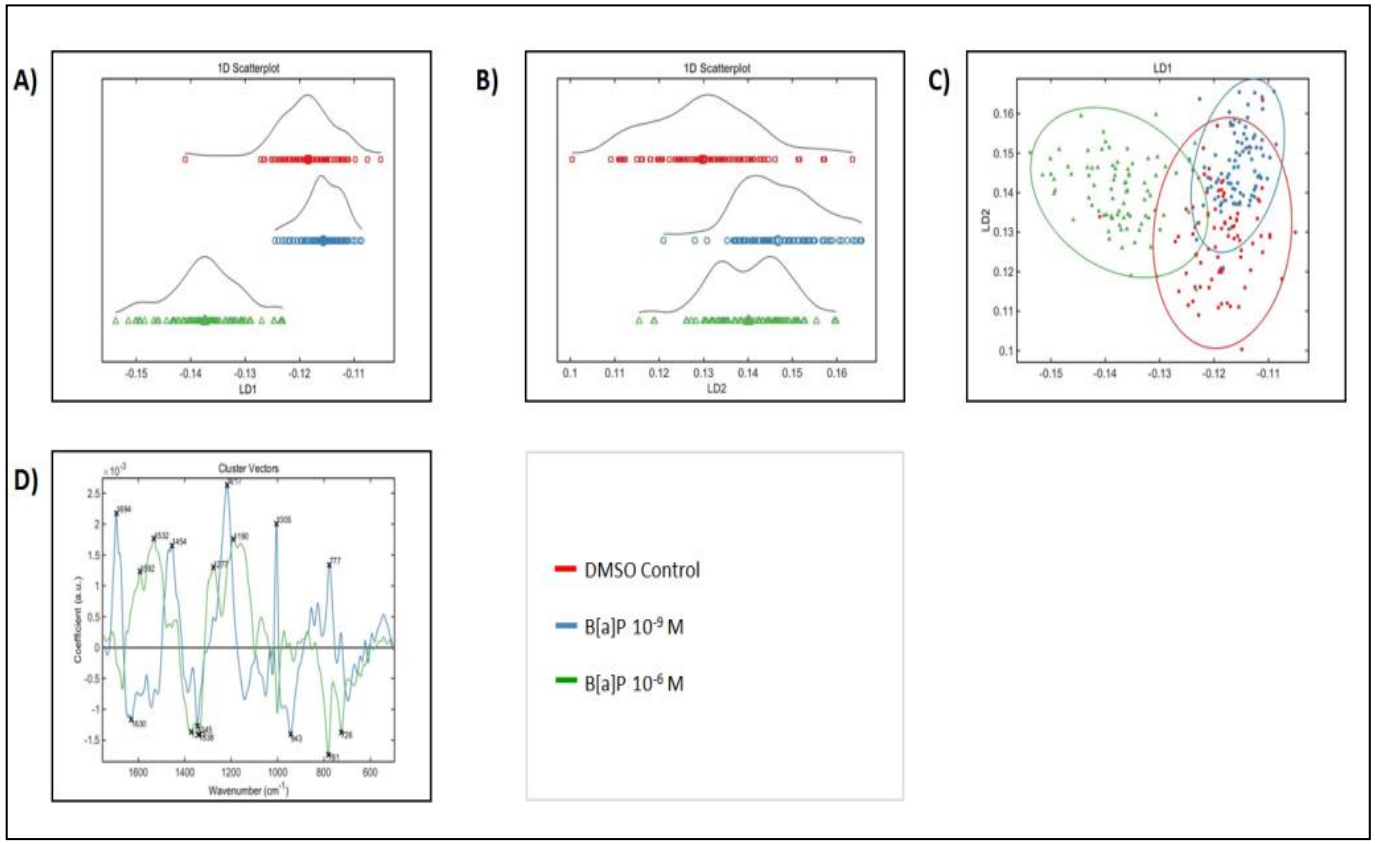


Figure 43 – Raman spectra obtained from G_0/G_1 phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with Rubber band-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

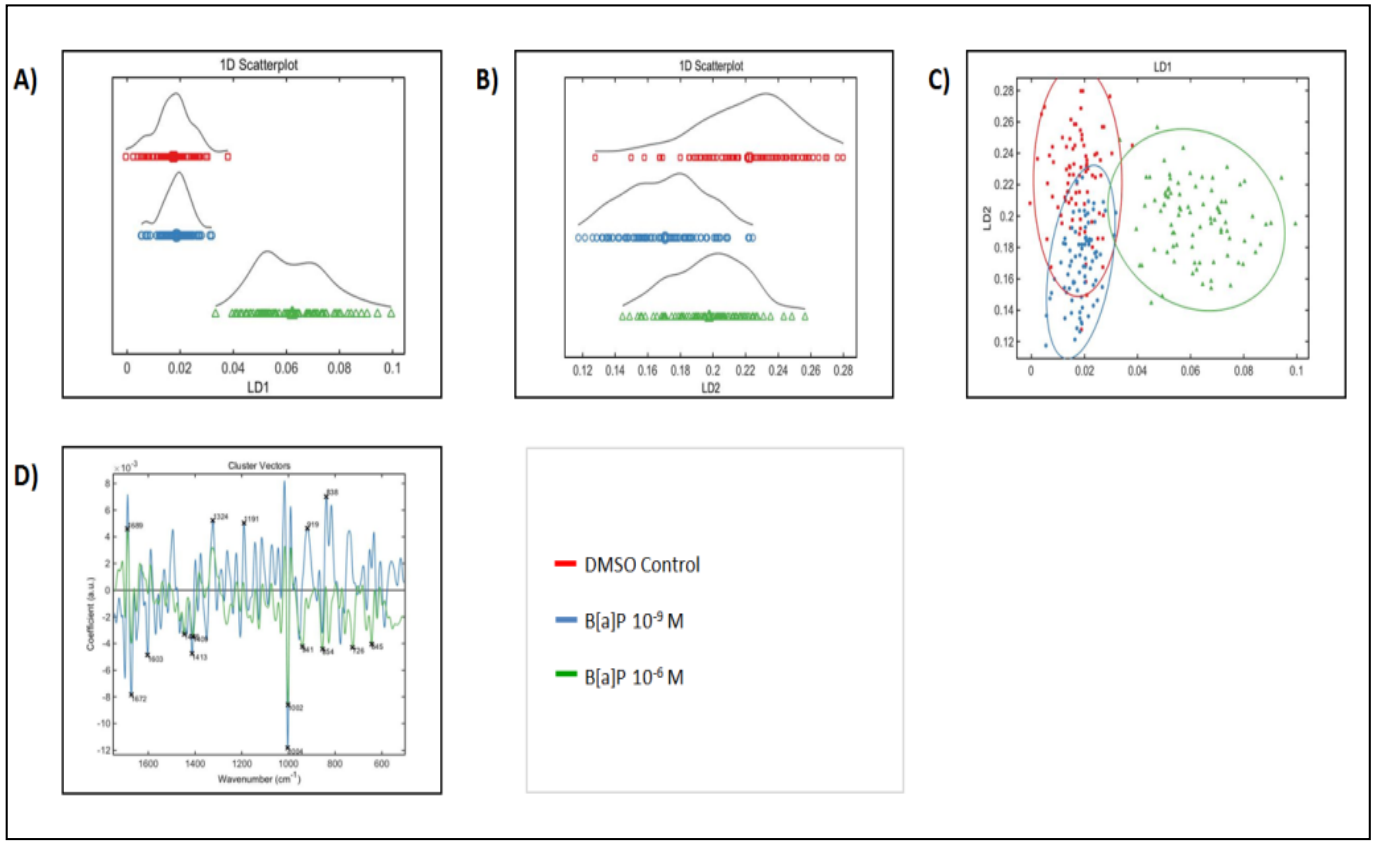


Figure 44 – Raman spectra obtained from G_0/G_1 phase concentrated MCF-7 cells treated for 24h with concentrations of 10^{-9}M B[a]P and 10^{-6}M B[a]P alongside a DMSO only control. Spectra were pre-processed with SG differentiation-Vector normalisation and analysed through PCA-LDA multivariate analysis. A) 1D Scores plot showing scores along LD1. B) 1D scores plot showing scores along LD2. C) 2D scores plot mapping LD1 along LD2 containing ellipse walls encapsulating > 95% of the data points. D) A cluster vector loading plot showing distinct wavenumbers associated with B[a]P treatment when to coefficient 0 (DMSO control)

Table 14 shows that P is smaller than 0.0001, therefore a significant difference is observed between the control and the B[a]P low dose of 10^{-9}M and also between the two B[a]P treatment concentrations. However, the control DMSO and the low B[a]P concentrations show distinct similarities. This is again suggested after statistical analysis. Variance found between the control and B[a]P 10^{-9}M is only statistically different for the Rubber Band and Vector normalisation pre-processed data where is has a P value < 0.01. This carries less significance than the P values of < 0.001, obtained when cross comparing control to B[a]P 10^{-9}M or B[a]P 10^{-9}M to B[a]P 10^{-6}M.

Raman G Phase Statistical Tests LD1	Rubber Amide I		Rubber Vector Normalisation		Sg Differentiation Vector Normalisation	
R^2	0.7376		0.7726		0.8349	
Unpaired t-Test	P < 0.0001		P < 0.0001		P < 0.0001	
Tukey's Multiple Comparison Test	C vs -9	P > 0.05	C vs -9	P < 0.01	C vs -9	P > 0.05
	C vs -6	P < 0.001	C vs -6	P < 0.001	C vs -6	P < 0.001
	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001	-9 vs -6	P < 0.001

Table 14 – Table showing statistical analysis of LD1 generated by Raman PCA-LDA analysis on G_0/G_1 phase concentrated MCF-7 cells treated with B[a]P. Statistical tests include R^2, Unpaired t-Test and F-test. Analysis was done on FTIR-ATR data pre-processing through Rubber band-Amide I, Rubber band-Vector normalisation and SG differentiation-Vector normalisation. Treatment conditions are indicated through letters, where C = DMSO control. -9 = B[a]P 10^{-9}M. -6 = B[a]P 10^{-6}M

4 Discussion

4.1 Endometriosis

4.1.1 Detecting Endometriosis

Whilst separation of endometriosis presence is not immediately apparent from the PCA-LDA scores plot, after testing the data for statistical significance, variance was proven to be highly significant. Both FTIR-ATR and Raman spectroscopy have shown to be capable of distinguishing endometriosis presence within an LBC sample. For both of the applied spectroscopy techniques, the obtained P values were < 0.0001 for the unpaired t-test. Raman had slightly more sensitivity in detecting endometriosis as its P value was < 0.0001 for the F-test. This was only 0.0178 for FTIR-ATR. It is important to note that one out of three pre-processing methods was successful in detecting endometriosis for both techniques. These pre-processing methods differed for each machine as Raman was successful with Polynomial-Vector normalisation whilst FTIR-ATR was successful with Rubber band-Amide I. The ability to distinguish endometriosis presence confirms the notion that cells and tissue from endometriosis are histologically different to normal endometrial cells.

Main peak differences found in FTIR- ATR spectra were found to be associated with nucleic acids and DNA/RNA vibration frequencies at 1717cm^{-1} and 1234cm^{-1}. This suggests that endometriosis has a genetic underlying cause. It is possible that the cells have retained some stem cell DNA or RNA, allowing them to have an increased plasticity (Laganà et al., 2013) This also reinforces the theory of the hereditary link to endometriosis. A difference for Carotenoid vibration at 951cm^{-1} was one of the largest peaks observed. This is very interesting as the dietary carotenoid lycopene has been linked to inhibit endometrial cancer (Nahum et al., 2001). Therefore carotenoid can be considered capable of eliciting an effect in endometrial tissue. The role of carotenoid β-carotene in cancer has been long disputed. Some studies show that it contains anti-cancerous properties (Peto et al. 1981) however other studies indicate

that it plays a major role in the development of cancer (Omenn et al., 1996). It is clear that carotenoid is involved in abnormal proliferation of tissue, whether the effect is more preventative or causative is yet to be determined.

Raman spectra uncovered RNA/DNA (783cm^{-1}) and more importantly the 1682cm^{-1} and 1714cm^{-1} regions that are associated with cortisone. Increased levels of cortisone and cortisol presence have been previously suggestive of an increased susceptibility to endometriosis (Miki et al,. 2014). It is theorised that women with high cortisol levels have a lower immune response and are not as capable of removing cell debris deposited by retrograde menstruation (Miki et al,2014). However it is unclear whether the increased cortisol levels are a result of the disease causing discomfort and stress, resulting in the release of cortisol from the adrenal gland.

4.1.2 Similarities between Severe Endometriosis and Control

Black lesions are a trait linked to endometriosis progression, whilst red lesions appear at early stages. The former are more pronounced and easier to detect by eye as well as with FTIR-ATR spectroscopy. Black endometriosis was significantly separated from disease free patients, whilst red lesions were not shown to be significantly dissimilar to the control. It is not surprising that early symptoms of endometriosis have more spectral similarities to patients without endometriosis, as smaller regions will be affected and the lesions are not as established.

When red and black lesions were analysed using Raman spectroscopy, no significance was found between the control and any type of lesion. However, Raman spectroscopy showed to have the delicate ability to actually differentiate between a black and a red region. Both spectral techniques have shown an incredible cohesion between results obtained. It is only in lesion detection that a strong difference was found between the techniques.

An interesting finding is that no dissimilarity was observed when control samples were directly compared to samples suffering from some of the more severe characteristics of

endometriosis. This was found for patients in disease stage 3 and for those with ovarian endometriosis presence (in FTIR-ATR). This is curious, as it is expected that disease progression would become increasingly apparent and more easily detectable. The presence of black lesions and ovarian endometriosis are often found during the later disease stages 3 or 4. It coincides with the finding that ovarian endometriosis presence was much harder to distinguish from the endometriosis free samples, the same way stage 3 was harder to distinguish. However, in contrast to this black lesions were the only lesions that were successfully separated from the control (for FTIR-ATR only). A possible explanation for this is that endometriosis symptoms do not directly correlate with disease stage. The stage categorisation is a subjective guideline for doctors, and black lesions can also be found in earlier stages where patients have few other symptoms. Detecting ovarian endometriosis, however, is very severe and even if this is the only disease phenotype the patient will be will be classified with stage 3. In this particular sample group all of the ovarian endometriosis positive patients were classified at stage 3.

The similarity between separation difficulty for stage 3 and ovarian endometriosis presence can be explained through the fact that endometriosis severity does not correlate to specific phenotypes. When separating stage 3 spectra from the control and when separating ovarian endometriosis presence from the control, the same spectral variance will be observed because the two classes fall into the same category.

However, the observed similarity between severe endometriosis samples (disease stage 3 and ovarian endometriosis presence) and endometriosis free samples brings into question the underlying cause of the disease. One explanation could be down to endometriosis density. Whilst endometriosis tissue is histologically different from endometrium tissue, it is possible that as this tissue becomes more established it will more closely resemble endometrial tissue as its structures are becoming more pronounced.

It could be seen as detrimental that more severe endometriosis proves to be more difficult to distinguish. However, this incidentally means that even very primitive endometriosis at stage 1 is more easily detected. This shows promise for spectroscopy to be used as an early screening tool, allowing women to be diagnosed at earlier stages, when the disease is most likely to be overlooked. Catching endometriosis early allows for much more successful treatment options. It is more valuable and efficient to detect the disease at its early stages rather than later when it is more established.

4.1.3 Challenges

This study applied several different pre-processing methods to spectral acquisitions in order to separate class variance as a tool for detecting endometriosis. A significant difference was detected between subclasses. However, pre-processing methods that best separated the classes differed for each endometriosis characteristic. Whilst the potential to use spectroscopy as a high throughput and low cost diagnostic tool is apparent, its application needs to be further refined. To allow for widespread use of this technique in hospital laboratories, consistency is required to select a suitable pre-processing method for data analysis.

When a patient sample is screened for endometriosis, this would have to be compared to a spectral database with information on other endometriosis sufferers. Using a shared open access database carries a challenge, as patient data is regarded as confidential and ethical regulations need to be taken into consideration. Different spectroscopy machines used during sample screening will require their own specific databases.

Both Raman and FTIR-ATR spectroscopy showed specificity when separating endometriosis presence. While both machines can be used for screening, data needs to be compared to the database built specifically to each machine. A simpler solution would be if only one database was created. This would, however, limit the number of laboratories able to participate in

screening, as only the ones with the correct machine would have the ability to compare their spectra to the database.

43 patient samples were compared in this study. In order for the shared database to carry a high specificity and sensitivity rate, a much larger data set is required. Having a higher number of control samples would contribute to reduced false positive diagnoses. This notion comes hand-in-hand with the challenge faced with diagnosing endometriosis in the first place, as there is a possibility that many of the control patients actually have endometriosis. This presence would reduce the reliability of the dataset, as control spectra obtained would be contaminated.

As an increasing number of cervical cancer screening is being performed through spectral analysis of LBC samples. If the spectra acquired from these samples were to be run though the endometriosis database, women would have access to conclusive endometriosis screening and diagnoses at an almost negligible cost. Whilst this is a great opportunity, implementing a nationwide screening through spectroscopy has already proven to be challenging (Gajjar et al., 2014) as not all laboratories are properly equipped. Governmental action would be required in order to establish new screening laws; a process that could be extremely time consuming.

Due to endometriosis being prevalent throughout the entire body, future challenges involve being able to detect endometriosis in areas outside the pelvic cavity. Surface-Enhanced Raman Spectroscopy (SERS) could be applied to the samples to enhance spectral readings, although this can is very costly. SERS is a technique that uses metal surfaces, most commonly silver and gold, to increase the signal to noise ratio by enhancing the amount of Raman scattering detected (Butler et al., 2015).

4.2 MCF-7 Cells

4.2.1 B[a]P Dose Response

The clonogenic assay was conclusive and showed that DMSO in itself has an effect on MCF-7 cell survival since the plating efficiency was very low. It can be said that counting stained colonies by eye carries a large bias. However from the error bars contained in Figure 27 it is clear that despite choosing to count only large colonies, consistency ensured equal measures were taken for all of the experimental flasks. MCF-7 cells show to have a strong linear dose response when treated with B[a]P as percentage cell survival drastically decreases as B[a]P concentration increases. Cells exhibit a slight reduction in plating efficiency which accounts for the toxic properties of the vehicle control DMSO itself.

4.2.2 Cell Number Over Time

The observed increase in cell number for a high dose B[a]P compared to the control is a contradiction, as B[a]P should damage the cells and decrease proliferation. B[a]P induced CYP1 isoforms cause increased proliferation as a way of encouraging damaged and mutated cells to finish the cell cycle and become abundant (Harper et al., 2010). This provides and explanation for increased proliferation after B[a]P treatment.

4.2.3 Cell Cycle Progression

The 0 hour control indicated that at the time of seeding 63.4% of the cells were in G_0/G_1 phase. The remaining 36.6% of the population was distributed evenly through S and late G2 phase. This greatly resembles the population distribution of cells found in a 96 hour growth culture. G_0/G_1 phase is a clear outlier and over half the population resides in this phase. MCF-7 cells cultured for 24 hours are found to be very evenly distributed within the cell cycle phases. It is surprising that cells in S phase are twice as present.

These findings illustrate the reasoning behind choosing to culture cells for 24 hours and 96 hours respectively in order to concentrate cells in either S or G2/M phases. When referring to

an S phase concentrated culture, this refers to the 24 hour point in which the highest presence of cells in S phase is observed. Whilst this only accounts for 31.9% of the population, it relates to a third of the population.

Cells that were grown for 24 hours prior to B[a]P treatment were concentrated in S phase. Following a 24 hour treatment, the cell cycle distribution appears similar to 0 hour and 96 hour cultures, where G_0/G_1 phase dominates over half of the population and S and G2/M account for the remaining 40% evenly. However, cells treated with a high dose of 10^{-6}M B[a]P showed a steep incline in S phase concentrated cells. S phase concentrated cells treated with DMSO control do not show an even distribution of all phases. Instead, approximately 40% of the cells are in G_0/G_1 or S phase, whilst only 13% are found in G2/M. This same effect does not occur with a lower B[a]P concentration.

The clonogenic assay indicated that B[a]P elicits a reduction in cell survival when applied to MCF-7 cells, it appears that only a higher concentration is able to induce cell cycle alterations. This is both observed through flow cytometry as well as the time course assay. Both experiments showed a strong response in the S phase/ 24 hour window.

When a 24 hour B[a]P treatment was applied to G_0/G_1 phase concentrated cells (grown for 96 hours), very little difference between treatment conditions was observed. There was only a slight increase of S phase and G_2/M phase when the cells were exposed to 10^{-6}M B[a]P compared to the DMSO control and lower B[a]P dose. This was within the range of the standard deviation error bars, meaning it does not carry much significance. Around 70 to 80% of the cells were present in G_0/G_1 phase throughout all treatment conditions.

4.2.4 S Phase Accumulation

When the accumulation is observed in S phase, this is due to a high number of cells having passed the restriction point and avoiding G_1 arrest/ G_0 quiescence. Therefore, more cells are passed on and get detected in S phase. Cells are continuously able to re-enter the cell cycle

despite damage, causing a high number of faulty cells to complete the cell cycle. This in turn gives them the ability to continuously re-enter the cell cycle, resulting in unregulated growth and proliferation capacity commonly seen in tumours. It is also possible that the cells have successfully surpassed the checkpoint and have passed into S phase where they are accumulating as there are further checkpoints employed within S phase (Ye et al., 1996).

4.2.5 Complementary Spectra

Results obtained through PCA-LDA and statistical analyses were very similar for FTIR-ATR and Raman spectra throughout the three pre-processing methods used. This indicates that FTIR-ATR and Raman spectroscopy are complementary to one another and can be used in unison. Raman analysis concluded that there was no difference between low dose B[a]P and the DMSO control. FTIR-ATR did identify a difference between all of the treatment classes. The separation of classes reinforces the accuracy of subcellular detection in both vibrational and infrared spectroscopy. Consistent results were obtained through Raman and FTIR-ATR spectroscopy when analysing MCF-7 cells. Both spectroscopy techniques were able to differentiate between cells treated with B[a]P in S phase and G_0/G_1 phase. No cell cycle phase difference was observed through biospectroscopy in this study, even though the flow cytometry and time course assay reinforce the notion that cell cycle phase affects B[a]P effects. Loading plots identified wavenumber regions associated with lipids and specifically with DNA vibrations when cells were treated with B[a]P. These tables are shown in the appendix D. Here is apparent that more DNA vibrations were observed when cells were treated in S phase.

5 Conclusion

Spectroscopy is a very sensitive tool which carries the ability to analyse both in vitro and in vivo cells. Raman and FTIR-ATR spectroscopy are compatible as they showed very similar results, even after undergoing different pre-processing methods. This was true for both endometriosis detection and MCF-7 examination.

Endometriosis can successfully be detected from a no endometriosis control. This is very promising for the medical field as it can provide a non-invasive, high throughput and high specificity diagnostic tool. The Tao Brush offers a comfortable way of obtaining an endometrium sample that can be used for multiple screenings. The implementation of LBC would open up many more screening opportunities in the UK and offers a conclusive diagnostic method without the need for surgery. Raman spectroscopy proved to be the most successful in detecting endometriosis. It was able to recognise endometriosis presence with very high significance. It could distinguish between different stages, expect for late stage 3 (same as FTIR-ATR). It was able to detect spectral differences between red and black lesions, and even found variance for endometriosis in the ovaries. Progressive endometriosis appears to carry a biochemical fingerprint similar to control samples without endometriosis for ATR spectra. It is unclear if this is a trait of the disease pathology, or if IR spectroscopy lacked specificity.

As for the MCF-7 cell cycle study of toxicity, research project demonstrates that B[a]P is capable of inducing alterations in exposed cell populations even at very low doses. B[a]P gradually reduces cell survival after 7 days exposure. However, when exposed for 6 to 24 hours, when the cells are completing S phase, a high dose of 10^{-6} M B[a]P induces short term increased proliferation and accumulation in S phase. After 24 hours the long term inhibition of cell growth caused by B[a]P is re-established. Flow cytometry showed evidence that the cells are in the presumed cell cycle phases. Biospectroscopy analysis (FTIR-ATR and Raman)

clearly distinguished between cells that have been concentrated in G_0/G_1 and S phase and exposed to B[a]P concentrations of 10^{-9}M and 10^{-6}M. Despite being affected by B[a]P, the cells are able to survive. The surviving cells suffered from altered protein expression, in particular CYP1B1 involved in xenotic metabolism of B[a]P.

Overall this study showed that spectroscopy is a novel tool that has multiple biological, chemical and physical applications. Its benefits are the non-destructive and non-invasive nature of the acquisition procedure, and it is a technique which needs to be exposed to further research in order to uncover its full potential.

6 References

American Society for Reproductive Medicine. (1997). Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertility and Sterility. 67 (5), 817-21.

Baker M J, Trevisan J, Bassan P, Bhargava R, Butler H J, Dorling K M, Fielden P R, Fogarty S W, Fullwood N J, Heys K A, Hughes C, Lasch P, Martin-Hirsch P L, Obinaju B, Sockalingum G D, Sulé-Suso J, Strong R J, Walsh M J, Wood B R, Gardner P, Martin F L. (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols*. 9 (8), 1771-91.

Berka K, Hendrychová T, Anzenbacher P, Otyepka M. (2011). Membrane Position of Ibuprofen Agrees with Suggested Access Path Entrance to Cytochrome P450 2C9 Active Site. *Journal of Physical Chemistry A*. 115 (41), 11248–11255.

Bersten D C, Sullivan A E, Peet D J, Whitelaw M L. (2013). bHLH–PAS proteins in cancer. *Nature Reviews Cancer*. 13, 827–841.

Bérubé S, Marcoux S, Maheux R. (1998). Characteristics related to the prevalence of minimal or mild endometriosis in infertile women. Canadian Collaborative Group on Endometriosis. *Epidemology*. 9 (5), 504-10.

Brosens I, Benagiano G. (2011). Endometriosis, a modern syndrome. *Indian J Med Res.* 133 (6), 581-93.

Bulletti C, Coccia M E, Battistoni S, Borini A. (2010). Endometriosis and infertility. *Journal of Assisted Reproduction and Genetics*. 27, 441-447.

Butler H J, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, Esmonde-White K, Fullwood N J, Gardner B, Martin-Hirsch P L, Walsh M J, McAinsh M R, Stone N, Martin F L. (2016). Using Raman spectroscopy to characterize biological materials. *Nature Protocols*. 11 (4), 664-87.

Butler H J, Fogarty S W, Kerns J G, Martin-Hirsch P L, Fullwood N J, Martin F L. (2015). Gold nanoparticles as a substrate in bio-analytical near-infrared surface-enhanced Raman spectroscopy. *Analyst*. 140 (9), 3090–3097.

Carter E A, Tam K K, Armstrong R S, Lay A. (2009). Vibrational spectroscopic mapping and imaging of tissues and cells. *Biophysical Reviews*. 1, 95-103.

Chen F P, Soong Y K, Lee N, Lo S K. (1998). The use of serum CA-125 as a marker for endometriosis in patients with dysmenorrhea for monitoring therapy and for recurrence of endometriosis. *Acta Obstetricia et Gynecologica Scandinavica*. 77 (6), 665-70.

Cheung A N Y, Szeto E F, Leung B S Y, Khoo U-S, Ng A W Y. (2003). Liquid-Based Cytology and Conventional Cervical Smears. *Cancer Cytopathology*. 99 (6), 331-335.

Cheung K T, Trevisan J, Kelly J G, Ashton K M, Stringfellow H F, Taylor S E, Singh M N, Martin-Hirsch P L, Martin F L. (2011). Fourier-transform infrared spectroscopy discriminates a spectral signature of endometriosis independent of inter-individual variation. *Analyst*. 136, 2047-2055.

Chiriboga L, Xie P, Yee H, Vigorita V, Zarou D, Zakim D, Diem M. (1998). Infrared spectroscopy of human tissue. I. Differentiation and maturation of epithelial cells in the human cervix. *Biospectroscopy*. 4 (1), 47-59.

Cox J T. (2004). Liquid-based cytology: evaluation of effectiveness, cost-effectiveness, and application to present practice. *Journal of the National Comprehensive Cancer Network*. 2 (6), 597-611.

Dai W, Sun C, Huang S, Zhou Q.(2016) Carvacrol suppresses proliferation and invasion in human oral squamous cell carcinoma. *OncoTargets and Therapy*. 9, 2297-2304.

Davidson J M, Duronio R J. (2012). S Phase–Coupled E2f1 Destruction Ensures Homeostasis in Proliferating Tissues. *PLOS Genetics*. 8 (8).

Davis C, Bhana S, Shorrocks A J, Martin F L. (2002). Oestrogens induce G1 arrest in benzo[a]pyrene-treated MCF-7 breast cells whilst enhancing genotoxicity and clonogenic survival. *Mutagenesis*. 17 (5), 431-438.

Dawson A E. (2004). Can We Change the Way We Screen?: The ThinPrep Imaging System. *Cancer Cytopathology*. 106 (6), 340-344.

Del Priore G, Williams R, Harbatkin C B, Wan L S, Mittal K, Yang G C. (2001). Endometrial brush biopsy for the diagnosis of endometrial cancer. *Journal of Reproductive Medicine*. 46 (5), 439-43.

Fogarty S W, Patel I I, Trevisan J, Nakamura T, Hirschmugl C J, Fullwood N J, Martin F L. (2013). Sub-cellular spectrochemical imaging of isolated human corneal cells employing synchrotron radiation-based Fourier-transform infrared microspectroscopy. *Analyst*. 138, 240-248.

Foster D A, Yellen P, Xu L, Saqcena M. (2011). Regulation of G1 Cell Cycle Progression: Distinguishing the Restriction Point from a Nutrient-Sensing Cell Growth Checkpoint(s). *Genes & Cancer*. 1 (11), 1124-1131.

Fung M, Senterman M, Eid P, Faught W, Mikhael N Z, Wong P T. (1997). Comparison of Fourier-transform infrared spectroscopic screening of exfoliated cervical cells with standard Papanicolaou screening. *Gynecologic Oncology*. 66 (1), 10-5.

Gajjar K, Martin-Hirsch P L, Martin F L. (2012). CYP1B1 and hormone-induced cancer. *Cancer Letters*. 324 (1), 13-30.

Gajjar K, Trevisan J, Owens G, Keating P J, Wood N J, Stringfellow H F, Martin-Hirsch P L, Martin F L. (2013). Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer. *Analyst*. 138, 3917-3926.

Gajjar K, Ahmadzai A A, Valasoulis G, Trevisan J, Founta C, Nasioutziki M, Loufopoulos A, Kyrgiou M, Stasinou S M, Karakitsos P, Paraskevaidis E, Da Gama-Rose B, Martin-Hirsch P L, Martin F L. (2014). Histology Verification Demonstrates That Biospectroscopy Analysis of Cervical Cytology Identifies Underlying Disease More Accurately than Conventional Screening: Removing the Confounder of Discordance. *PLos ONE*. 9 (1), 1-16.

Genies C, Jullien A, Lefebvre E, Revol M, Maitre A, Douki T. (2016). Inhibition of the formation of benzo[a]pyrene adducts to DNA in A549 lung cells exposed to mixtures of polycyclic aromatic hydrocarbons.*Toxicology in Vitro*. S0887-2333 (16), 30090-X.

Gruppo italiano per lo studio dell'endometriosi. (1994). Endometriosis: Prevalence and anatomical distribution of endometriosis in women with selected gynaecological conditions: results from a multicentric Italian study. *Human Reproduction*. 9 (6), 1158-1162.

Guerreiro C B B, Horálek J, De Leeuw F, Couvidat F . (2016). Benzo(a)pyrene in Europe: Ambient air concentrations, population exposure and health effects. *Environmental Pollution*. 214, 657–667.

Hamouchene H, Arlt V M, Giddings I, Phillips D H. (2011). Influence of cell cycle on responses of MCF-7 cells to benzo[a]pyrene. *BMC Genomics*. 12 (333), 1-16.

Harper L J, Costea D E, Gammon L, Fazil B, Biddle A, Mackenzie I C. (2010). Normal and malignant epithelial cells with stem-like properties have an extended G2 cell cycle phase that is associated with apoptotic resistance. *BMC Cancer*. 10 (166), 1-16.

Harris A T, Lungari A, Needham C J, Smith S L, Lones M A, Fisher S E, Yang X B, Cooper N, Kirkham J, Smith D A, Martin-Hirsch D P, High A S. (2009). Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. *Head & Neck Oncology*. 1 (34).

Her-Juing Wu H, Harshbarger K E, Berner H W, Elsheikh T M. (2000). Histologic Diagnosis of 200 Cases With Complementary Cytology: An Accurate Sampling Technique for the Detection of Endometrial Abnormalities. *American Journal of Clinical Pathology*. 114, 412-418.

Husain O A N, Butler E B, Evans M D, Macgregor J E, Yule R. (1974). Quality control in cervical cytology. *Journal of Clinical Pathology*. 27, 935-944.

Jiao H, Allinson S L, Walsh M J, Hewitt R, Cole K J, Phillips D H, Martin F L. (2007). Growth kinetics in MCF-7 cells modulate benzo[a]pyrene-induced CYP1A1 up-regulation. *Mutagenesis*. 22 (2), 111-116.

Kelly, J.G., Martin-Hirsch, P.L. & Martin, F.L. (2009). Discrimination of base differences in oligonucleotides using mid-infrared spectroscopy and multivariate analysis. Anal. Chem. 81, 5314–5319.

Kennedy S, Bergqvist A, Chapron C, D'Hooge T, Dunselman G, Greb R, Hummelshoj L, Prentice A, Saridogan E, on behalf of the ESHRE Special Interest Group for Endometriosis and Endometrium Guideline Development Group. (2005). ESHRE guideline for the diagnosis and treatment of endometriosis. *Human Reproduction*. 20 (10), 2698-2704.

Krafft C, Steiner G, Beleites C, Salzer R. (2009). Disease recognition by infrared and Raman spectroscopy. *Journal of Biophotonics*. 2 (1-2), 13-28.

Laganà A S, Sturlese E, Retto G, Sofo V, Triolo O. (2013). Interplay between Misplaced Müllerian-Derived Stem Cells and Peritoneal Immune Dysregulation in the Pathogenesis of Endometriosis. *Obstetrics and Gynecology International*. 527041.

Latham Macer M, Taylor H S. (2012). Endometriosis and Infertility: A review of the pathogenesis and treatment of endometriosis-associated infertility. *Obstetrics and Gynecology Clinics of North America*. 39 (4), 535-49.

Leyendecker G. (2000). Redefining endometriosis: Endometriosis is an entity with extreme pleiomorphism. *Human Reproduction*. 15 (1), 4-7.

Leyendecker G, Kunz G, Noe M, Herbertz M, Mall G. (1998). Endometriosis: a dysfunction and disease of the archimetra. *Human Reproduction Update*. 4 (5), 752-62.

Lin W, Arthur G. (2007). Phospholipids are synthesized in the G2/M phase of the cell cycle. *The International Journal of Biochemistry & Cell Biology*. 39, 597-605.

Lykkesfeldt A E, Briand P. (1986). Indirect mechanism of oestradiol stimulation of cell proliferation of human breast cancer cell lines. *British Journal of Cancer*. 53 (1), 29-35.

Machairiotis N, Stylianaki A, Dryllis G, Zarogoulidis P, Kouroutou P, Tsiamis N, Katsikogiannis N, Sarika E, Courcoutsakis N, Tsiouda T, Gschwendtner A, Zarogoulidis K, Sakkas L, Baliaka A, Machairiotis C. (2013). Extrapelvic endometriosis: a rare entity or an under diagnosed condition?. *Diagnostic Pathology.* 8, 1-12.

Martin FL, Kelly JG, Llabjani V, Martin-Hirsch PL, Patel II, Trevisan J. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. Nature protocols 5 (11), 1748-1760

Matthews J, Gustafsson J-Å. (2006). Estrogen receptor and aryl hydrocarbon receptor signaling pathways. *Nuclear Receptor Signaling*. 4:e016.

Miki Y, Takagi K, Doe Z, Tanaka S, Suzuki T, Sasano H., Ito K. (2014). SAT-0288: Intratumoral Concentration of Stress Hormone, Cortisol in Endometrial Carcinoma. *Hormone Dependent Tumors - Meeting Abstracts*.

Movasaghi Z, Rehman S, Rehman I U. (2007). Raman Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews.* 42, 493–541.

Movasaghi Z, Rehman S, Rehman I U. (2008). Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews*. 43 (2), 134-179.

Muñoz-Hernando L, Muñoz-Gonzalez J L, Marqueta-Marques L, Alvarez-Conejo C, Tejerizo-García Á, Lopez-Gonzalez G, Villegas-Muñoz E, Martin-Jimenez A, Jiménez-López J S. (2015). Endometriosis: alternative methods of medical treatment. *International Journal of Women's Health*. 11 (7), 595-603.

Nahum A, et al., *Lycopene inhibition* of *cell cycle progression* in *breast* and *endometrial cancer cells* is *associated* with *reduction* in *cyclin D levels* and *retention* of *p27 (Kip1*) in the *cyclin E-cdk2 complexes*, Oncogene 20 (26) (2001) 3428-3436

Nasse M J, Walsh M J, Mattson E C, Reininger R, Kajdacsy-Balla A, Macias V, Bhargava R, Hirschmugl C J. (2011). High-resolution Fourier-transform infrared chemical imaging with multiple synchrotron beams. *Nature Methods*. 8 (5), 413-418.

Omenn et al., 1996. Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease, N. Engl. J. Med 334:1150-1155

Overton K W, Spencer S L, Noderer W L, Meyer T, Wang C L. (2014). Basal p21 controls population heterogeneity in cycling and quiescent cell cycle states. *Proceedings of the National Academy of Sciences of the United States of America*. 111 (41): E4386-E4393.

Pang W, Li J, Ahmadzai A A, Heppenstall L D, Llabjani V, Trevisan J, Qui X, Martin F L. (2012). Identification of benzo[a]pyrene-induced cell cycle-associated alterations in MCF-7 cells using infrared spectroscopy with computational analysis. *Toxicology*. 298, 24-29.

Peto et al. 1981. Carotenoids and cancer: an update with emphasis on human intervention studies, Nature 290:201-208).

Purandare N C, Patel I I, Lima K M G, Trevisan J, Ma'Ayeh M, McHugh A, Von Bünau G, Martin-Hirsch P L, Prendiville W J, Martin F L. (2014). Infrared spectroscopy with multivariate analysis segregates low-grade cervical cytology based on likelihood to regress, remain static or progress. *Analytical Methods*. 6 (13), 4576-4584.

Rullo S, Severi Silvestrini I, Sorrenti G, Rossitto C, Pietropolli A, Piccione E. Endometrial biopsy in the diagnosis of abnormal uterine bleeding with a new intra-sheath device during hysteroscopy: preliminary study. Available: http://www.seedos.co.uk/SeeDOS%20Ltd%20-%20VABRA%20Clinical%20Literature.pdf. Last accessed 20 Sep 2015.

Sampson J. (1927). Metastatic or Embolic Endometriosis, due to the Menstrual Dissemination of Endometrial Tissue into the Venous Circulation. *American Journal of Pathology*. 3 (2), 93-110.

Sandaka A, Sandaka J, Zborowskab M, Prądzyński W. (2010). Near infrared spectroscopy as a tool for archaeological wood characterization. *Journal of Archaeological Science*. 37 (9), 2093-2101.

Schledermann D, Ejersbo D, Hoelund B. (2006). Improvement of Diagnostic Accuracy and Screening Conditions With Liquid-Based Cytology. *Diagnostic Cytopathology*. 34 (11), 780-785.

Schledermann D, Hyldebrandt T, Ejersbo D, Hoelund B. (2007). Automated Screening Versus Manual Screening: A Comparison of the ThinPrep Imaging System and Manual Screening in a Time Study. *Diagnostic Cytopathology*. 35 (6), 348-352.

Soule H D, Vazguez J, Long A, Albert S, Brennan M. (1973). A human cell line from a pleural effusion derived from a breast carcinoma. *Journal of the National Cancer Institute*. 51 (5), 1409-16.

Sourial S, Tempest N, Hapangama D K. (2014). Theories on the Pathogenesis of Endometriosis. *International Journal of Reproductive Medicine*. 2014, 1-9.

Speert H. (1958). John A. Sampson and Pelvic Endometriosis, In: Obstetric and Gynaecologic Milestones. Essays in Eponymy, p 397 – 399, The Macmillan Company, New York, 1958.

Taylor S E, Cheung K T, Patel I I, Trevisan J, Stringfellow H F, Ashton K M, Wood N J, Keating P J, Martin-Hirsch P L, Martin F L. (2011). Infrared spectroscopy with multivariate analysis to interrogate endometrial tissue: a novel and objective diagnostic approach. *British Journal of Cancer*. 104 (5), 790–797.

Theophilou G, Lima K M, Briggs M, Martin-Hirsch P L, Stringfellow H F, Martin F L. (2015). A biospectroscopic analysis of human prostate tissue obtained from different time periods points to a trans-generational alteration in spectral phenotype. *Scientific Reports*. 5, Article number: 13465.

Theophilou G, Lima KM, Martin-Hirsch PL, Stringfellow HF, Martin FL. (2016) ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer. *The Analyst*. PMID 26090781

Trevisan J, Angelov P P, Carmichael P L, Scott A D, Martin F L. (2012). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst*. 137, 3202-3215.

Trevisan J, Angelov P P, Scott A D, Carmichael P L, Martin F L. (2013). IRootLab: a free and open-source MATLAB toolbox for vibrational biospectroscopy data analysis. *Bioinformatics*. 29 (8), 1095-1097.

Tsuchikawa S, (2007). A review of recent near infrared research for wood and paper, Appl. Spectrosc. Rev. 42(1), 43–71 (2007).

Van Der Sneppen L S, Ritchie G, Hancock G, Ariese F, Gooijer C, Ubachs W, Haselberg R, Somsen G W, De Jong G J. (2010). Evanescent-Wave Cavity Enhanced Spectroscopy as a Tool in Label-Free Biosensing. *Conference on Lasers and Electro-Optics 2010*, OSA Technical Digest (CD) (Optical Society of America, 2010), paper AMC2.

Walsh M J, Singh M N, Pollock H M, Cooper L J, German M J, Stringfellow H F, Fullwood N J, Paraskevaidis E, Martin-Hirsch P L, Martin F L. (2007). ATR microspectroscopy with multivariate analysis segregates grades of exfoliative cervical cytology. *Biochemical and Biophysical Research Communications*. 352, 213-219.

Walsh M J, German M J, Singh M, Pollock H M, Hammiche A, Kyrgiou M, Stringfellow H F, Paraskevaidis E, Martin-Hirsch P L, Martin F L. (2007). IR microspectroscopy: potential applications in cervical cancer screening. *Cancer Letters.* 246 (1-2), 1-11.

Williams A R W, Brechin S, Porter A J L, Warner P, Critchley H O D. (2008). Factors affecting adequacy of Pipelle and Tao Brush endometrial sampling. *BJOG*. 115, 1028-1036.

Yang G C H, Wan L S. (2000). Endometrial Biopsy Using the Tao Brush Method. *The Journal of Reproductive Medicine*. 45 (2), 109-114.

Ye X S, Fincher R R, Tang A, O'Donnell K, Osmani S A. (1996). Two S-phase checkpoint systems, one involving the function of both BIME and Tyr15 phosphorylation of p34cdc2, inhibit NIMA and prevent premature mitosis. *The EMBO Journal*. 15 (14), 3599-3610.

Zhu J, Norman I, Elfgren K, Gaberi V, Hagmar B, Hjerpe A, Andersson S. (2007). A comparison of liquid-based cytology and Pap smear as a screening method for cervical cancer. *Oncology Reports*. 18 (1), 157-60.

Zölzer F, Mußfeldt T, Streffer C. (2014). Differential S-phase progression after irradiation of p53 functional versus non-functional tumour cells. *Radiology and Oncology*. 48 (4), 354-360.

Zondervan K T, Treloar S A, Lin J, Weeks D E, Nyholt D R, Mangion J, MacKay I J, Cardon L R, Martin N G, Kennedy S H, Montgomery G W. (2007). Significant evidence of one or more susceptibility loci for endometriosis with near-Mendelian inheritance on chromosome 7p13-15. *Human Reproduction*. 22 (3), 717-28.

7 Appendix

Additional graphs, tables and other supplementary information can be found here.

Appendix A – Poster presented at the European Environmental Mutagenesis Society (EEMS) annual conference hosted by UKEMS, July 2014

Cell Cycle Modifies Susceptibility of MCF-7 Cells to B[a]P

K.Snijders[1] and F.L.Martin[1]

Centre of Biophotonics, LEC, Lancaster University, Lancaster, LA1 4YQ

Attendance at the meeting was supported by a bursary from UKEMS

Aims

- To produce an integrated response profile of the environmental effects of Benzo[a]Pyrene
- To show that B[a]P plays a crucial role in evading G_0 arrest
- To increase our understanding of the risks and underlying mechanisms of pro-mutagens and pro-carcinogens

Introduction

Benzo[a]Pyrene is an environmental contaminant mainly found in tobacco smoke, car exhaust fumes and industrial wastes. While it has been found to have pro-mutagenic and pro-carcinogenic properties it is not toxic until it enters the plasma membrane and binds to the Aryl hydrocarbon receptor (AhR). Once bound it activates AhR causing it to translocate into the nucleus where it dissociates. Active AhR is involved in controlling the transcription and expression of many genes.

MCF-7 cells are a breast cancer cell line established in 1970. They are great for research use because of their resilience: they are easy to grow and hard to kill. They also incorporate estrogen receptors; these are thought to interact with AhR during gene regulation.

The Cell cycle: The G_1 phase is where cell growth occurs while the S phase controls replication and DNA synthesis. Cells have to pass the restriction point, between G_1 and S phase, in order to undergo replication and later division. If the restriction point is not passed then cells remain in G_0 arrest (quiescence) and are unable to enter the subsequent phase.

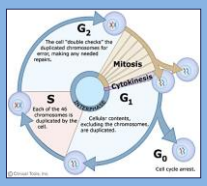


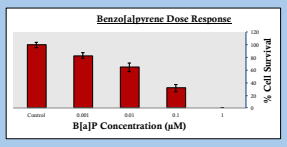
Figure 1: The Cell Cycle (Obtained from Clinical Tools, Inc)

Failure to regulate the restriction point results in the avoidance of arrest by cells. This characteristic has been linked to the uncontrolled growth of tumours, as the cells are able to continuously re-enter the cell cycle allowing for continuous proliferation.

Methods

How different concentrations of B[a]P affect MCF-7 cells

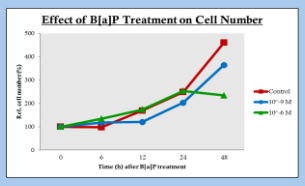
- MCF-7 cells were cultured in complete DMEM medium
- 500 cells seeded per T25 flask
- Added 25 µl of B[a]P concentrations
- Used DMSO as a vehicle control
- Treated for **7 days**
- Stained with 5% Giemsa and counted



As B[a]P concentration increases the % cell survival decreases. This effect is seen for B[a]P concentrations of 10^{-3} → 1 µ M

How B[a]P exposure time affects proliferation of MCF-7 cells

- MCF-7 cells were cultured in complete DMEM medium
- 10000 cells seeded per T25 flask, grown for 24h
- Added 25 µl of B[a]P/DMSO concentrations
- Treated for **0h**, **6h**, **12h**, **24h** and **48h**
- After treatment cells were disaggregated with trypsin, centrifuged and re-suspended in complete DMEM
- Cell counts taken at each time point using a haemocytometer



Results are expressed as relative cell number [%; i.e., ratio of the cell number at indicated time point relative to that determined at 0h (normalised to 100%)

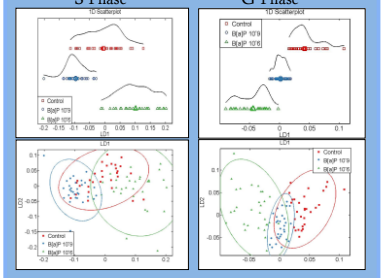
10^{-9} M concentration shows similar trend to the control but with a reduced cell number.
10^{-6} M shows an increased cell number at 6h and a drop in cell number at 48h when compared to 10^{-9} M and the control.

Spectral Analysis

How cell cycle affects MCF-7 cells treated with B[a]P

- MCF-7 cells were cultured in complete DMEM medium
- 10^7 cells seeded in T25 flask
- Grown for **24h** (S phase) or **96h** (G_1/G_0 phase)
- Treated with 25 µl of B[a]P/DMSO for 24h
- Fixed onto Low-E reflective glass slides in 70% ethanol
- **ATR-FTIR** Spectroscopy used to take wavenumber measurements (Attenuated Total Reflectance Fourier Transform Infrared)
- Data was normalised and multivariate analysis was performed
- PCA-LDA scatter plots are shown below (Principal Component Analysis and Linear Discriminant Analysis)

S Phase G Phase

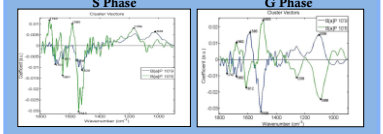


Statistical Tests	S Phase	G Phase
R^2	0.4126	0.7755
ANOVA	P<0.0001	P<0.0001
Dunnett's Multiple Comparison Test	C vs -9 P < 0.05	C vs -9 P < 0.01
	C vs -6 P < 0.01	C vs -6 P < 0.01
	-9 vs -6 P < 0.01	-9 vs -6 P < 0.01

Distinguishing regions of IR spectra from cluster vector plots

- Spectral regions that separate treated cells from the vehicle control
- Finding the **wavenumbers** of bio molecular entities most altered by identifying the 5 largest coefficient peaks

S Phase G Phase



S Phase	Wave number (cm-1)	Association
B[a]P 10^{-9}	1747	Lipids, Triglycerides
	1652	C=O Stretching
	1585	Ring Base
	1508	C-H Phenyl Ring
	1099	Phosphate II (PO_2^-)
B[a]P 10^{-6}	1682	Amide II
	1612	DNA Vibrational
	1493	C-H Bending
	1258	Phosphate I (PO_2^-)
	1085	Phosphate I (PO_2^-)

G Phase	Wave number (cm-1)	Association
B[a]P 10^{-9}	1703	C=O Stretching
	1653	Amide II
	1550	Ring Base
	1520	Amide III
	1026	Glycogen
B[a]P 10^{-6}	1740	Phospholipids
	1651	Amide II
	1580	C-C Phenyl Ring
	1543	Amide III
	1169	ν_{as} C-O-C

Discussion

Thus far this study demonstrates that B[a]P is capable of inducing alterations in exposed cell populations even at very low doses. B[a]P gradually reduces cell survival after 7days exposure. However when exposed for 6h→24h, when the cells are completing S phase, a high dose of 10^{-6} M B[a]P induces short term increased proliferation. After 24h the long term inhibition of cell growth caused by B[a]P is re-established. Despite being affected by B[a]P the cells are able to survive. The hypothesis that the surviving cells suffer from altered gene expression, in particular cytochrome P450 genes, will be tested in future work.

Spectral analysis through PCA-LDA showed nearness similarity in S phase. Statistical tests showed r^2 was low and Dunnett's test indicates that the control and 10^{-9} are not significantly different. However B[a]P 10^{-6} vs control is. Segregation of category clusters in PCA-LDA represents a high degree of dissimilarity in G phase. All concentrations where found to be significantly different. Cluster vectors indicated high peaks associated with alterations in DNA/RNA (ν_{as} + PO_2^-), suggesting genotoxic mechanisms of B[a]P. This was particularly prevalent in S phase cells with B[a]P 10^{-6}. This further fuels hypothesis that B[a]P induces an increased proliferation in S phase to ensure cells that have undergone DNA damage can successfully complete the cell cycle.

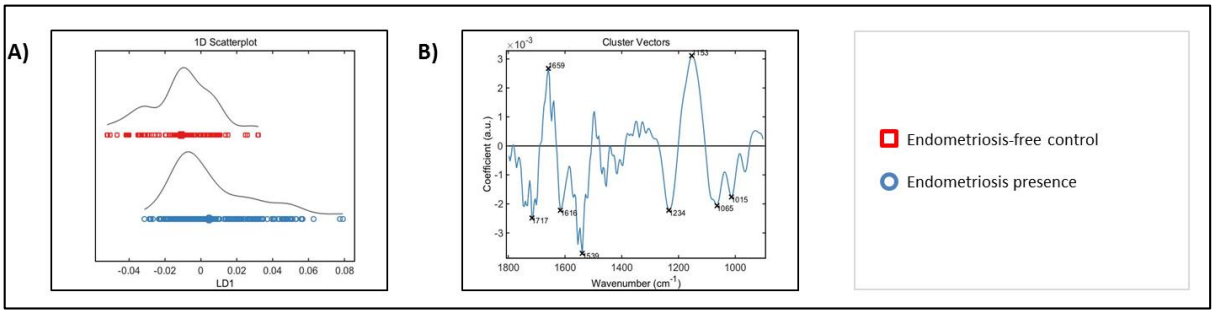
Future work includes

- Flow Cytometry to gain evidence that the cells are in the presumed cell cycle phases
- Western Blots and RT-PCR to identify if B[a]P alters gene expression of CYP1A1, CYP1B1, BCL-2 and P21 during cell cycle

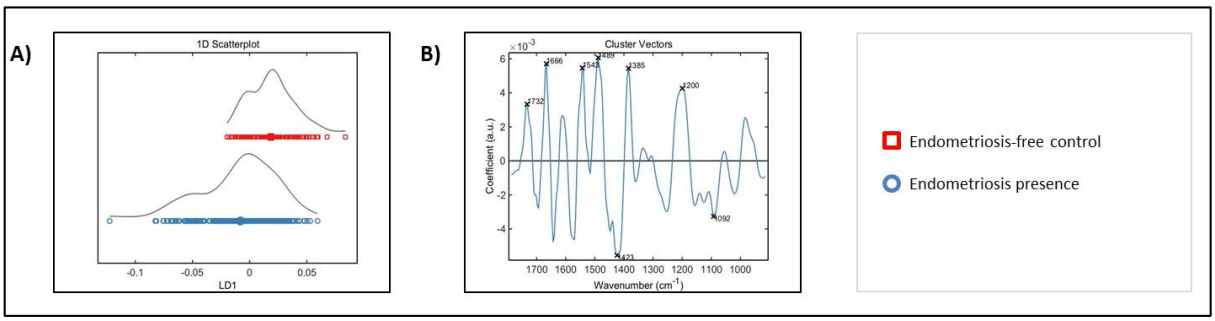
References-
Hamouchene, H. (2011). Influence of cell cycle on responses of MCF-7 cells to benzo[a]pyrene. *BMC Genomics*. 12
Jiao, HY. (2007). Growth kinetics in MCF-7 cells modulate benzo[a]pyrene-induced CYP1A1 up-regulation. *Mutagenesis*. 22
Pang, W. (2012). Identification of benzo[a]pyrene-induced cell cycle-associated alterations in MCF-7 cells using infrared spectroscopy with computational analysis *Toxicology*. 298

Appendix B – The graphs in this section illustrate 1D, 2D, 3D scores plots and cluster vectors for spectral data obtained by FTIR-ATR and Raman spectroscopy. Graphs show analysis of endometriosis presence, disease stage, disease morphology and ovarian endometriosis presence. Pre-processing methods Rubber band-Vector normalisation and SG differentiation – Vector normalisation are shown for the FTIR-ATR spectra. The Raman spectra were pre-processed using Rubber band-Vector normalisation and Rubber band-Amide I.

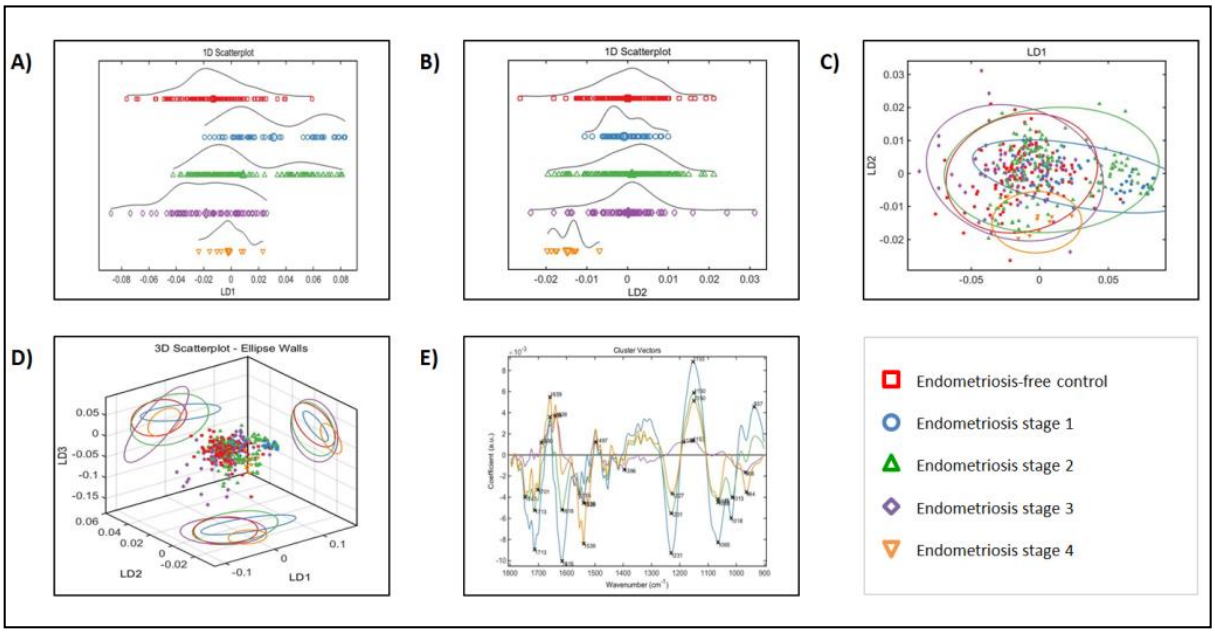
PCA-LDA results for endometriosis presence obtained from FTIR-ATR spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:



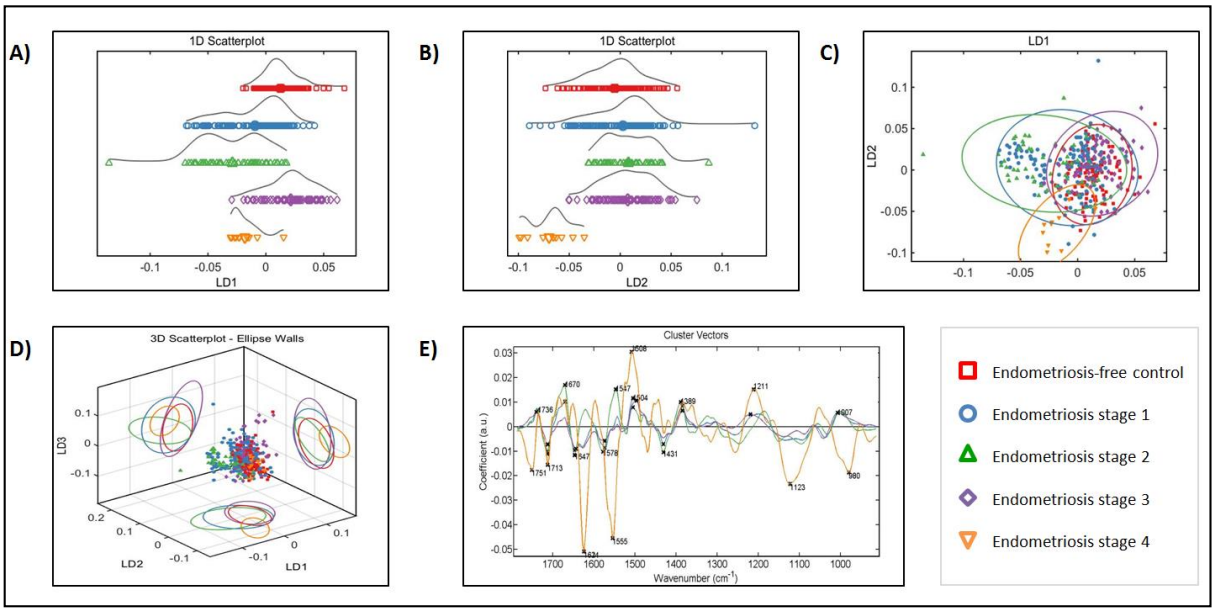
PCA-LDA results for endometriosis presence obtained from FTIR-ATR spectroscopy spectra that underwent SG differentiation-Vector normalisation pre-processing:



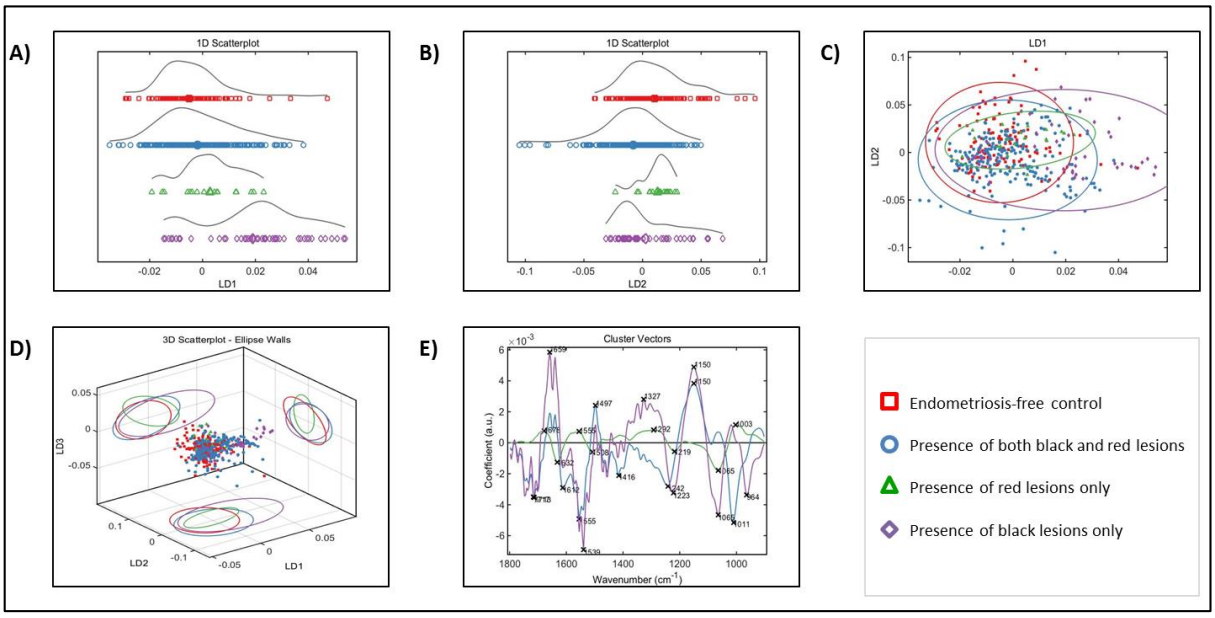
PCA-LDA results for disease stage obtained from FTIR-ATR spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:



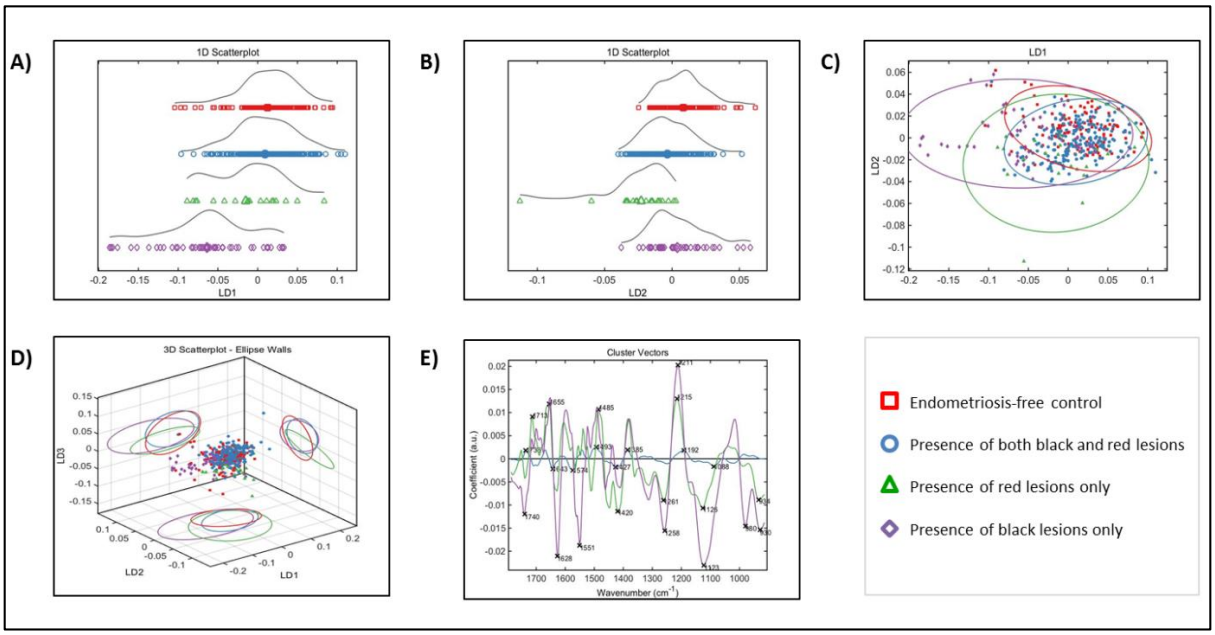
PCA-LDA results for disease stage obtained from FTIR-ATR spectroscopy spectra that underwent SG differentiation-Vector normalisation pre-processing:



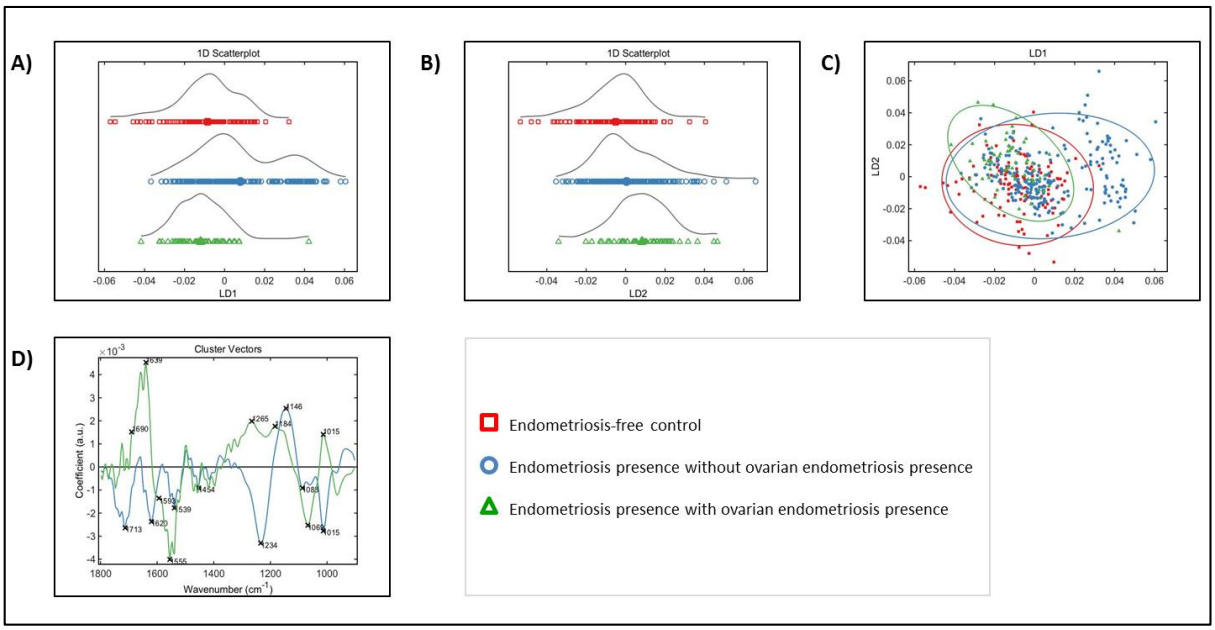
PCA-LDA results for disease morphology obtained from FTIR-ATR spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:



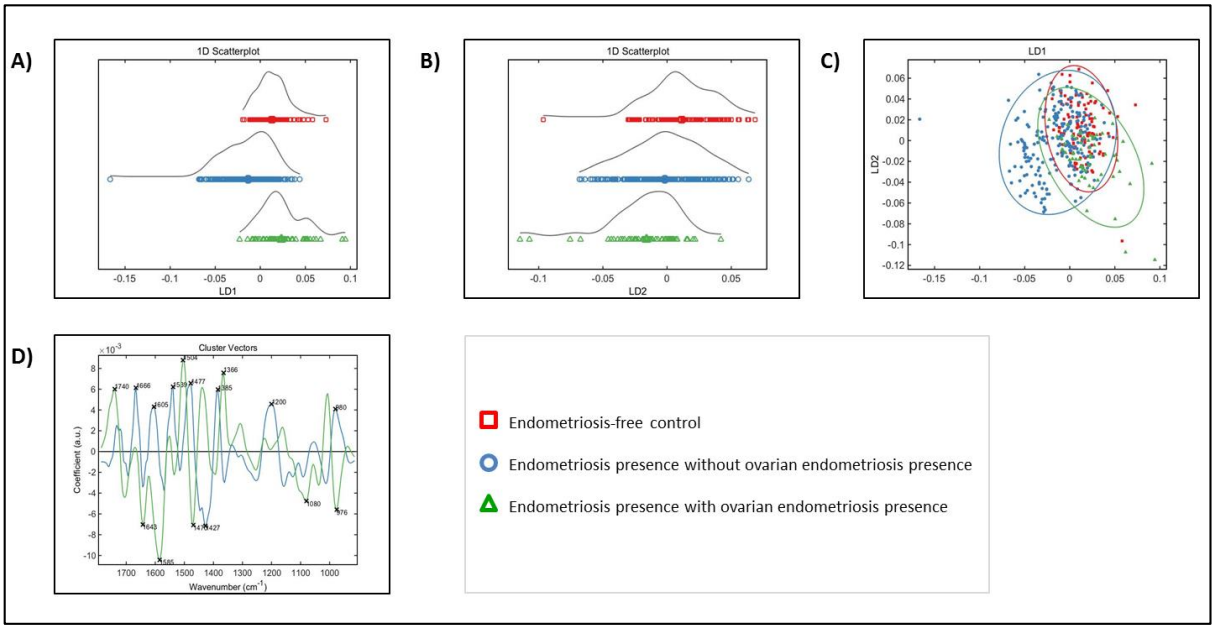
PCA-LDA results for disease morphology obtained from FTIR-ATR spectroscopy spectra that underwent SG differentiation-Vector normalisation pre-processing:



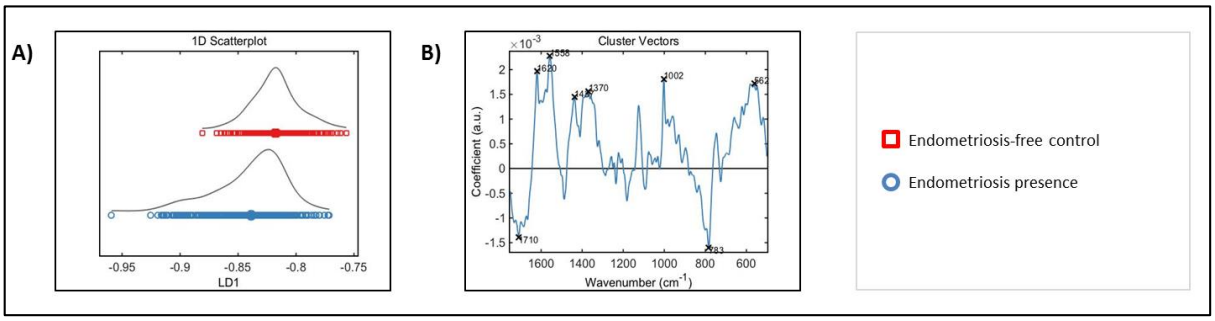
PCA-LDA results for ovarian endometriosis presence obtained from FTIR-ATR spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:



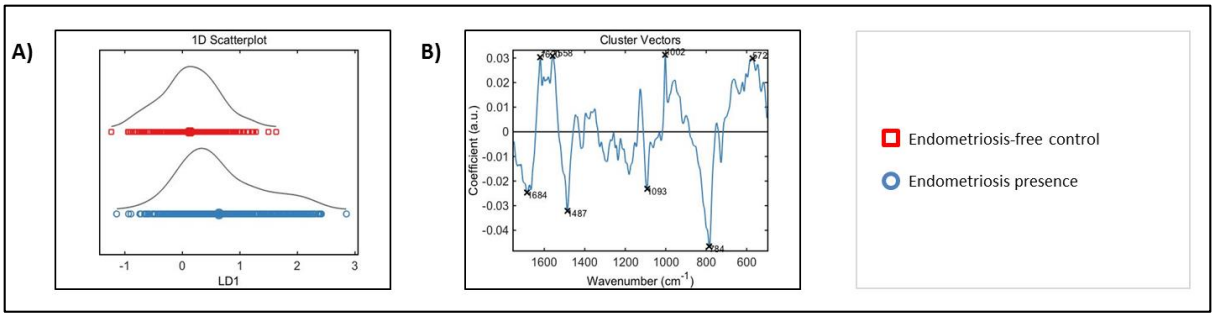
PCA-LDA results for ovarian endometriosis presence obtained from FTIR-ATR spectroscopy spectra that underwent SG differentiation-Vector normalisation pre-processing:



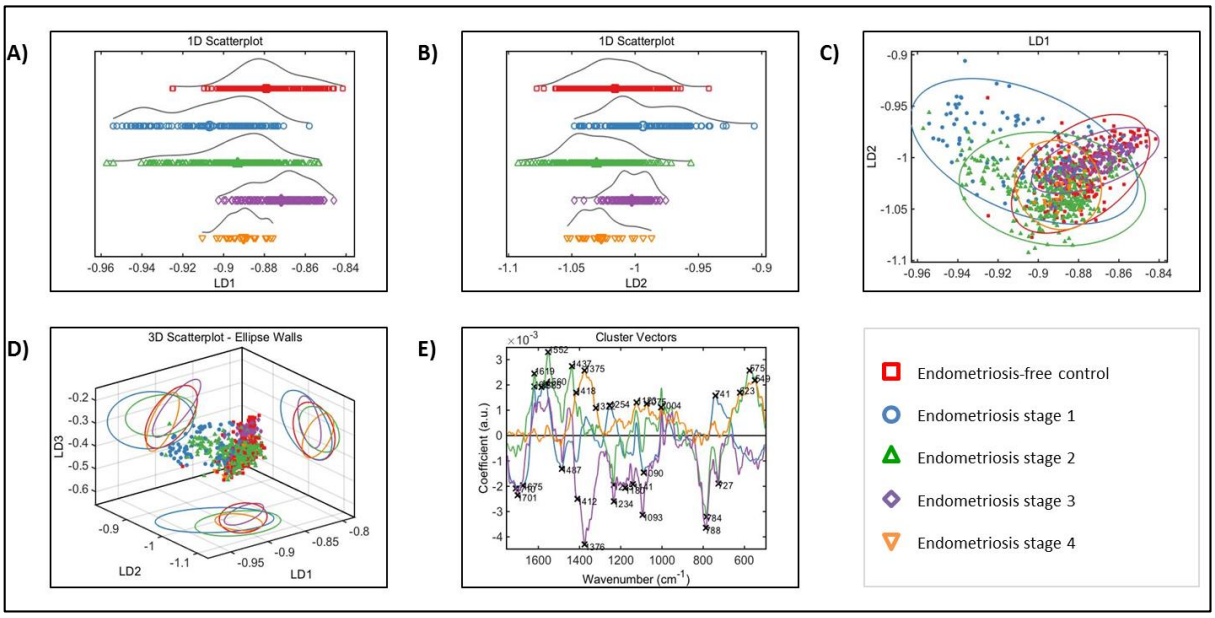
PCA-LDA results for endometriosis presence obtained from Raman spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:



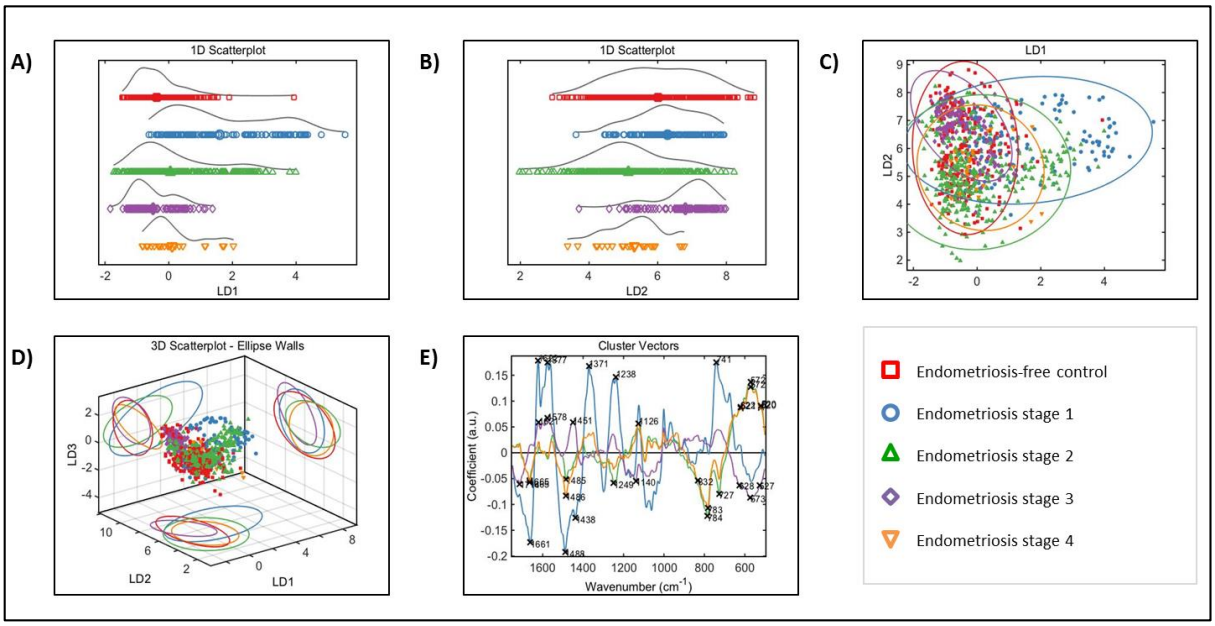
PCA-LDA results for endometriosis presence obtained from Raman spectroscopy spectra that underwent Rubber band-Amide I pre-processing:



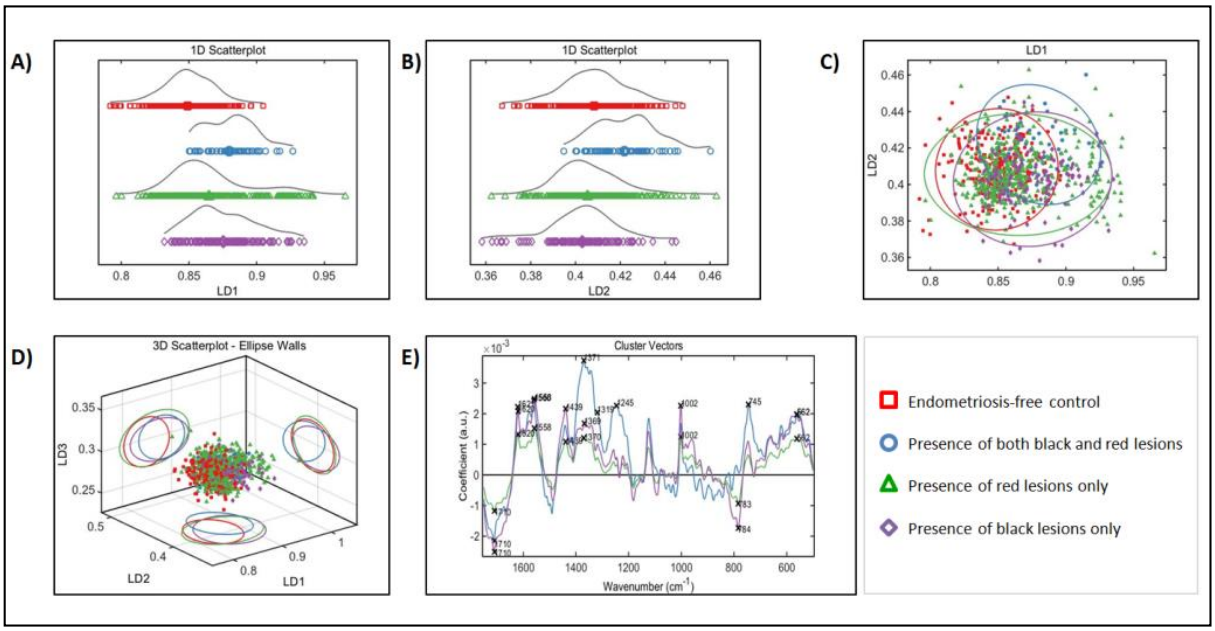
PCA-LDA results for disease stage obtained from Raman spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:



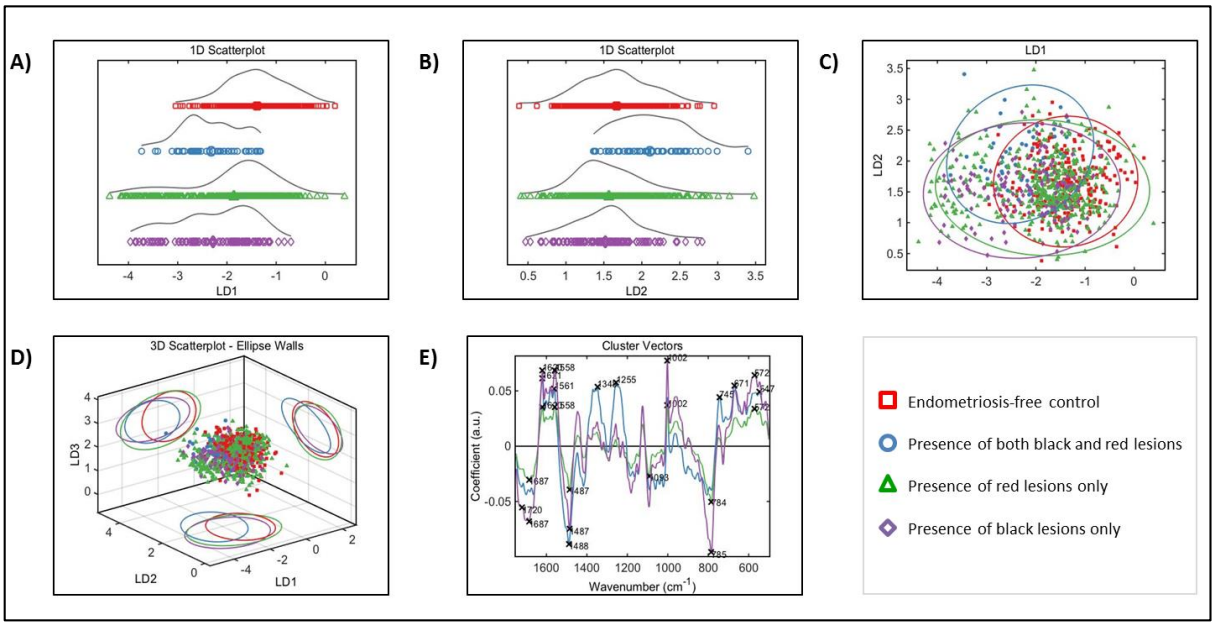
PCA-LDA results for disease stage obtained from Raman spectroscopy spectra that underwent Rubber band-Amide I pre-processing:



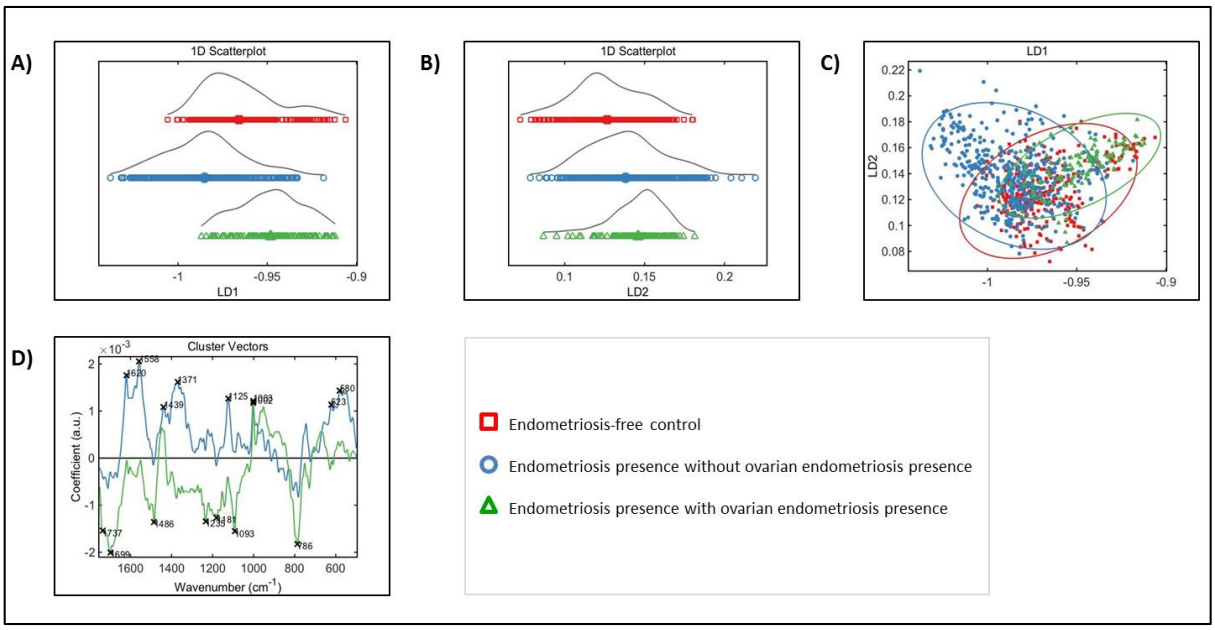
PCA-LDA results for disease morphology obtained from Raman spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:



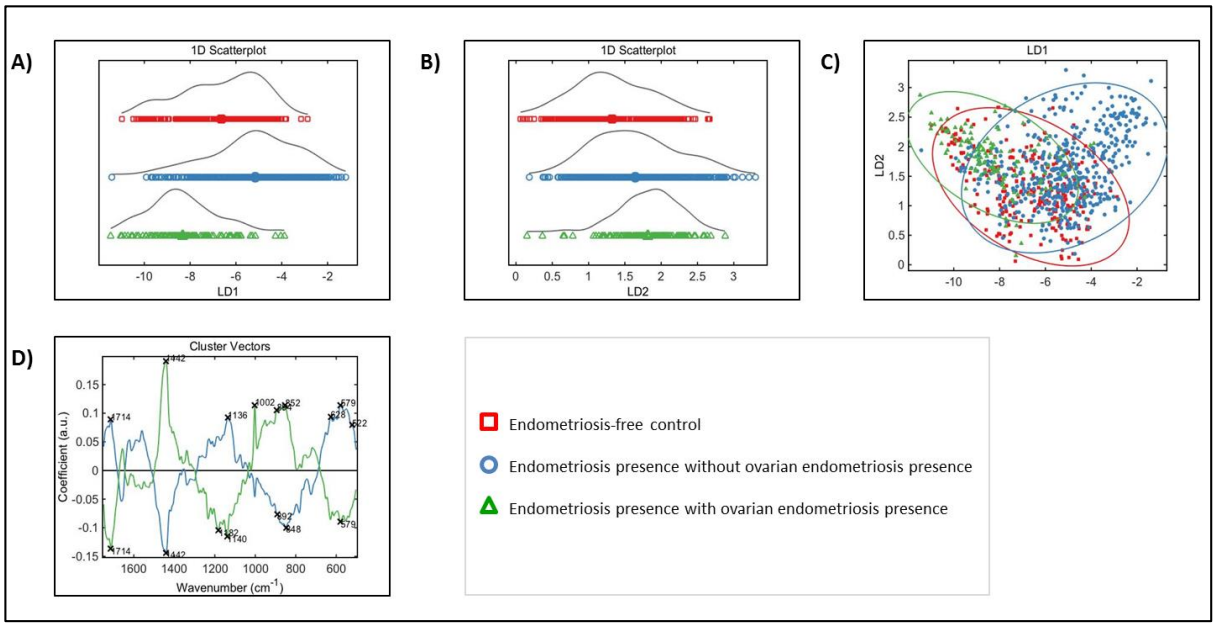
PCA-LDA results for disease morphology obtained from Raman spectroscopy spectra that underwent Rubber band-Amide I pre-processing:



PCA-LDA results for ovarian endometriosis presence obtained from Raman spectroscopy spectra that underwent Rubber band-Vector normalisation pre-processing:

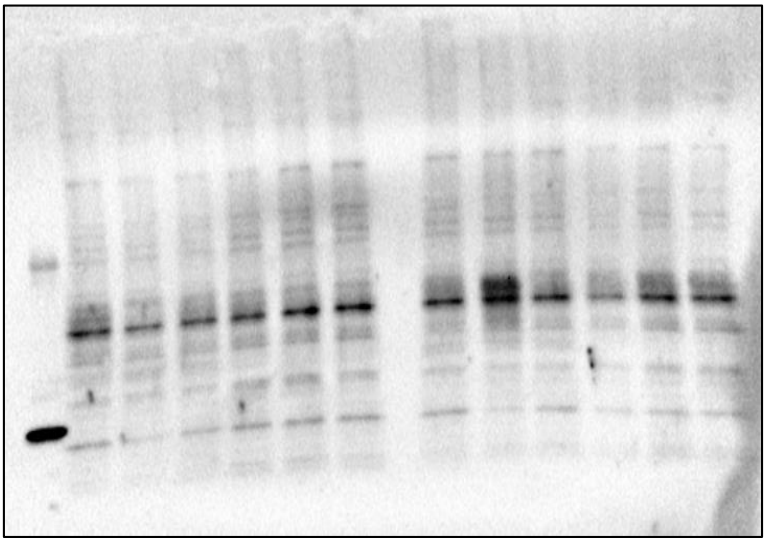


PCA-LDA results ovarian endometriosis presence obtained from Raman spectroscopy spectra that underwent Rubber band-Amide I pre-processing:

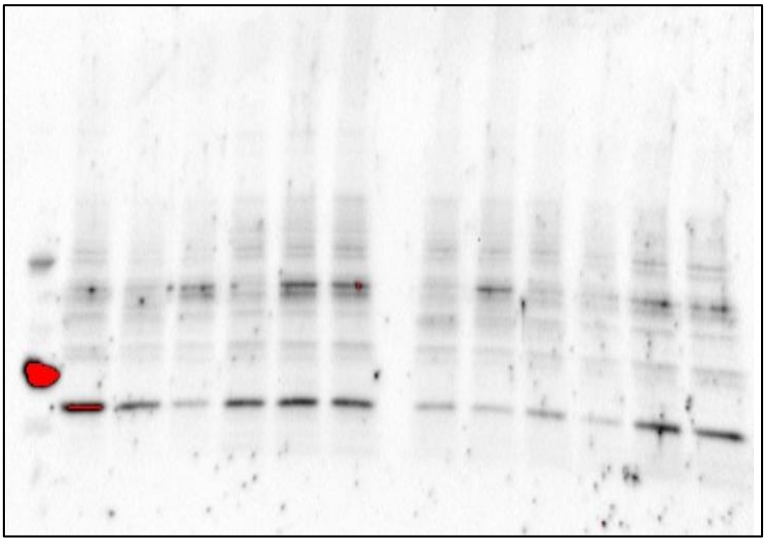


Appendix C – The figures in this section show raw data from the western blots. Low molecular weight ladder is shown in the first well on the left. The darkest band on the ladder is 25kDa whilst the top band indicates 75kDA.

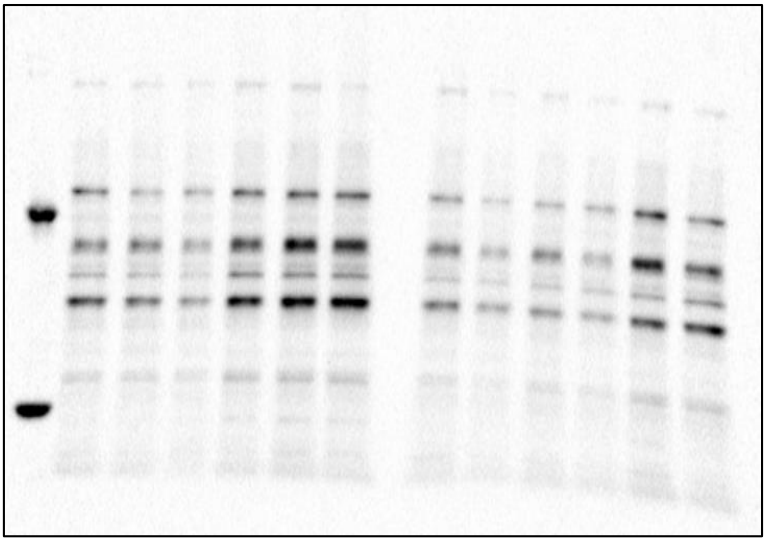
Raw data for western blot of BLC-2 protein expression. Band should be seen at 26kDA:



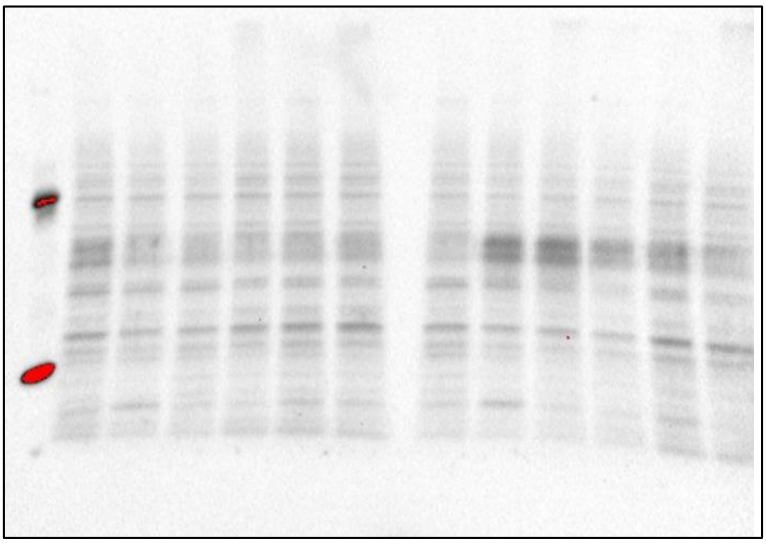
Raw data for western blot of CYP1A1 protein expression. Band should be seen at 54kDA:



Raw data for western blot of CYP1B1 protein expression. Band should be seen at 61 kDA:



Raw data for western blot of P21 protein expression. Band should be seen at 54kDA:



Appendix D – Tables illustrating the peak wavenumbers and the associated chemical identities for S phase and G phase treated MCF-7 cells. Peak wavenumbers were generated from PCA-LDA cluster vector loading plots from spectral data obtained from FTIR-ATR or Raman spectroscopy.

FTIR-ATR results for MCF-7 cells treated in S phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1736	C==O stretching (lipids)
1659	Amide I
1589	Ring C-C stretch of phenyl
1482	Not specified
1474	CH2 bending of the methylene chains in lipids
1327	Stretching C-N thymine, adenine
1107	v(CO), v(CC), ring (polysaccharides, pectin)
1022	Glycogen

FTIR-ATR results for MCF-7 cells treated in S phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1717	C==O thymine Amide I (arises from C==O stretching vibration) C==O stretching vibration of DNA and RNA C==O stretching vibration of purine base
1539	Not specified
1497	C==C, deformation C-H
1450	Methylene deformation in biomolecules Polyethylene methylene deformation modes
1292	N-H thymine
1204	Vibrational modes of collagen proteins-amide III C-O-C, C-O dominated by the ring vibrations of polysaccharides C-O-P, P-O-P Collagen
1103	Symmetric stretching P-O-C
964	C-C and C-O in deoxyribose of DNA of tumor cells C-O deoxyribose, C-C C-C, C-O deoxyribose

FTIR-ATR results for MCF-7 cells treated in S phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1720	C==O
1659	Amide I
1558	Ring base
1452	Not specified
1327	Stretching C-N thymine, adenine
1207	PO_2^- asymmetric (phosphate I) Collagen
1111	ν(CO), ν(CC), ring (polysaccharides, cellulose)
1018	ν(CO), ν(CC), δ(OCH), ring (polysaccharides, pectin)

FTIR-ATR results for MCF-7 cells treated in S phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1720	C==O
1842	Not specified
1450	Methylene deformation in biomolecules Polyethylene methylene deformation modes
1288	N-H thymine
1204	Vibrational modes of collagen proteins-amide III C-O-C, C-O dominated by the ring vibrations of polysaccharides C-O-P, P-O-P Collagen
1103	Symmetric stretching P-O-C
999	Ring stretching vibrations mixed strongly with CH in-plane bending C-O ribose, C-C
961	C-O deoxyribose, C-C

FTIR-ATR results for MCF-7 cells treated in S phase with a B[a]P 10^{-9}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1736	C==O stretching (lipids)
1674	Amide I (anti-parallel β-sheet) ν(C==C) trans, lipids, fatty acids Stretching C==O vibrations that are H-bonded (changes in the C==O stretching vibrations could be connected with destruction of old H-bonds and creation of the new ones)
1547	Protein band Amide II (δN-H, νC-N) Peptide amide II Amide II Amide II of proteins
1508	In-plane CH bending vibration from the phenyl rings CH in-plane bend
1470	CH_2 bending of the methylene chains in lipids
1420	$\nu_s(COO^-)$ (polysaccharides, pectin) Deformation C-H
1188	Deoxyribose
999	Ring stretching vibrations mixed strongly with CH in-plane bending C-O ribose, C-C

FTIR-ATR results for MCF-7 cells treated in S phase with a B[a]P 10^{-6}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1732	Absorption band of fatty acid ester Fatty acid ester band
1543	Amide II
1485	C_8-H coupled with a ring vibration of guanine Deformation C-H
1273	$CH_{\alpha'}$ rocking
1165	Mainly from the C-O stretching mode of C-OH groups of serine, threosine, & tyrosine of proteins) ν(CC), δ(COH), ν(CO) stretching C-O stretching (in normal tissue) Hydrogen-bonded stretching mode of C-OH groups C-O stretching band of collagen (type I)
1119	Symmetric stretching P-O-C C-O stretching mode
1080	ν PO_2^- Phosphate vibration Symmetric phosphate [PO_2^- (sym)] stretching Collagen & phosphodiester groups of nucleic acids
1003	Not specified

FTIR-ATR results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1701	Not specified
1651	Amide I absorption (predominantly the C==O stretching vibration of the amide C==O) Protein amide I absorption C==O, stretching C==C uracyl, NH_2 guanine Peptide amide I Amide I
1558	Ring base
1520	Amide II
1196	Collagen
1157	C-O stretching vibration v (C-C)-diagnostic for the presence of a carotenoid structure, most likely a cellular pigment vC-O of proteins and carbohydrates
1022	Glycogen
961	C-O deoxyribose, C-C

FTIR-ATR results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1740	C==O C==O stretching (lipids) Ester C==O stretching vibration (phospholipids)
1701	Not specified
1651	Amide I absorption (predominantly the C==O stretching vibration of the amide C==O) Protein amide I absorption C==O, stretching C==C uracyl, NH_2 guanine Peptide amide I Amide I
1585	Ring C-C stretch of phenyl
1543	Amide II
1396	Symmetric CH3 bending of the methyl groups of proteins
1350	Stretching C-O, deformation C-H, deformation N-H
1169	v_{as} CO-O-C

FTIR-ATR results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1747	Ester group (C==O) vibration of triglycerides v(C==O) (polysaccharides, pectin)
1701	Not specified
1632	Ring C-C stretch of phenyl
1604	C==C uracyl, C==O
1585	Ring C-C stretch of phenyl
1539	Protein amide II absorption- predominately β-sheet of amide II
1345	Not specified
1153	Stretching vibrations of hydrogen-bonding C-OH groups

FTIR-ATR results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1740	C==O C==O stretching (lipids) Ester C==O stretching vibration (phospholipids)
1701	Not specified
1632	Ring C-C stretch of phenyl
1604	C==C uracyl, C==O
1585	Ring C-C stretch of phenyl
1543	Amide II
1345	Not specified
1274	Not specified

FTIR-ATR results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-9}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1717	C==O thymine Amide I (arises from C==O stretching vibration) C==O stretching vibration of DNA and RNA C==O stretching vibration of purine base
1643	Amide I band (arises from C==O stretching vibrations)
1555	Ring base
1508	In-plane CH bending vibration from the phenyl rings
1462	Paraffin
1258	PO_2^- asymmetric (phosphate I)
1169	ν_{as} CO-O-C C-O bands from glycomaterials and proteins
1107	ν(CO), ν(CC), ring (polysaccharides, pectin)

FTIR-ATR results for MCF-7 cells treated in G phase with a B[a]P 10^{-6}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1717	C==O thymine Amide I (arises from C==O stretching vibration) C==O stretching vibration of DNA and RNA C==O stretching vibration of purine base
1643	Amide I band (arises from C==O stretching vibrations)
1558	Ring base
1512	In-plane CH bending vibration from the phenyl rings CH in-plane bend
1512	ν(C==C)-diagnostic for the presence of a carotenoid structure, most likely a cellular pigment
1462	Paraffin
1261	PO_2^- asymmetric (phosphate I)
1069	Stretching C-O ribose

Raman results for MCF-7 cells treated in S phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1671	Amide I C==C stretching vibrations Cholesterol & its esters C==C stretching vibration mode of steroid ring Amide I (anti-parallel β-sheet) ν(C==C) trans, lipids, fatty acids C==C stretch Amide I band (C==O stretch coupled to a N-H bending) Ceramide
1449	C-H vibration (proteins) C-H vibration (lipids) Lipids
1318	G (ring breathing modes of the DNA/RNA bases)-C-H deformation (protein) Amide III (α-helix)
1237	Amide III & CH2 wagging vibrations from glycine backbone & proline side chains
1095	Lipid ν(C-N) Phosphodioxy group (PO2- in nucleic acids)

Wave number [cm^{-1}]	Association
1004 1004	Phenylalanine (of collagen) vs(C-C), symmetric ring breathing, phenylalanine (protein assignment) Phenylalanine (collagen assignment) Phenyl breathing mode v(C-C) phenylalanine
960	Symmetric stretching vibration of v1PO43- (phosphate of HA) Calcium-phosphate stretch band (high quantities of cholesterol) Quinoid ring in-plane deformation
853	Ring breathing mode of tyrosine & C-C stretch of proline ring Glycogen

Raman results for MCF-7 cells treated in S phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1668	Protein band C==C stretching band α-helical structure of amide I Structural protein modes of tumors Carbonyl stretch (C==O) Carbonyl stretch (C==O) Cholesterol ester
1591	Carbon particles
1537	Amide carbonyl group vibrations and aromatic hydrogens
1446	CH2 bending mode of proteins & lipids CH2 deformation
1341	Nucleic acid mode Differences in collagen content Nucleic acid modes indicating the nucleic acid content in tissues G (DNA/RNA) CH deformation (proteins and carbohydrates)
1241	Asymmetric phosphate [PO2- (asym.)] stretching modes (phosphate stretching modes originate from the phosphodiester groups of nucleic acids and suggest an increase in the nucleic acids in the

Wave number [cm^{-1}]	Association
1241	malignant tissues). (Generally, the PO2- groups of phospholipids do not contribute to these bands)
1140	Palmitic acid Fatty acid Carbohydrates peak for solids
1070	Triglycerides (fatty acids) Symmetric PO2- stretching of DNA (represents more DNA in cell)

Raman results for MCF-7 cells treated in S phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1649	(C==C) Amide I Protein amide I absorption Amide I
1285	Differences in collagen content Cytosine
1215	Stretching of C-N
992	C-O ribose, C-C
907	Formalin contamination peak on fixed tissues
810	Phosphodiester (Z-marker)
663	C-S stretching mode of cystine (collagen type I)
657	C-S stretching mode of cystine (collagen type I) C-C twisting mode of tyrosine

Raman results for MCF-7 cells treated in S phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1471	C==N stretching Paraffin
1427	Deoxyribose, (B,Z-marker)
1340	Nucleic acid mode Differences in collagen content Nucleic acid modes indicating the nucleic acid content in tissues
1077	C-C (lipid in normal tissues) Symmetric stretching vibration of v3PO43- v(C-C) or v(C-O), phospholipids (lipid assignment) Phospholipids C-C or C-O stretching mode of phospholipids Carbohydrate peak for solids C-C or C-O stretch (lipid), C-C or PO2 stretch (nucleic acid) v(CC) skeletal v(CC) vitors (PO2-) nucleic acid
859	Tyrosine, collagen
765	Pyrimidine ring breathing mode
663	C-S stretching mode of cystine (collagen type I)
575	Phosphatidylinositol

Raman results for MCF-7 cells treated in S phase with a B[a]P 10^{-9}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1649	(C==C) Amide I Protein amide I absorption Amide I
1591	Carbon particles
1470	C==N stretching
1002	C-C aromatic ring stretching Phenylalanine Phenylalanine (collagen assignment)
931	Carbohydrates peak for solutions and solids
825	Phosphodiester O-P-O stretch DNA
723	DNA
642	C-C twisting mode of tyrosine

Raman results for MCF-7 cells treated in S phase with a B[a]P 10^{-6}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1737	C==O ester (lipids) Lipids
1684	Amide I (disordered structure; non hydrogen bonded)
1442	Fatty acids CH2 bending mode Due to changes in chemical environment of the CH2 bending mode CH3, CH2 deformation (collagen assignment) Triglycerides (fatty acids)
1321	Amide III (α-helix)
992	C-O ribose, C-C
831	Asymmetric O-P-O stretching, tyrosine
645	C-C twisting mode of phenylalanine (proteins)
530	S-S disulfide stretching in proteins Phosphatidylserine ν(S-S) gauche-gauche-trans (aminoacid cysteine)

Raman results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1588	Phenylalanine, hydroxyproline
1534	Amide carbonyl group vibrations and aromatic hydrogens
1446	CH2 bending mode of proteins & lipids CH2 deformation
1333	Guanine
1139	Palmitic acid Fatty acid
1002	C-C aromatic ring stretching Phenylalanine Phenylalanine (collagen assignment)
929	v(C-C), stretching-probably in amino acids proline & valine (protein band)
853	Ring breathing mode of tyrosine & C-C stretch of proline ring Glycogen

Raman results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Amide I:

Wave number [cm^{-1}]	Association
1391	CH rocking
1269	Amide III (collagen assignment)
1183	Cytosine, guanine, adenine
1128	C-N stretching (proteins) C-O stretching (carbohydrates) v(C-C) skeletal of acyl backbone in lipid (transconformation)
1070	Triglycerides (fatty acids)
1004	Phenylalanine (of collagen) vs(C-C), symmetric ring breathing, phenylalanine (protein assignment) Phenylalanine (collagen assignment) Phenyl breathing mode v(C-C) phenylalanine
966	Hydroxyapatite
868	C-C stretching, hypro (collagen assignment) Monosaccharides (β-fructose), (C-O-C) skeletal mode Disaccharide (sucrose), (C-O-C) skeletal mode Polysaccharides, amylose Polysaccharides, amylopectin

Raman results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-9}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1694	Amide I (turns and bands)
1630	Amide C==O stretching absorption for the β-form polypeptide films
1454	CH2 stretching/CH3 assymetric deformation Overlapping asymmetric CH3 bending & CH2 scissoring (is associated with elastin, collagen, & phospholipids) Collagen & phospholipids
1345	δ(CH), residual vibrations
1217	Stretching of C-N
1005	Phenylalanine (proteins) Proteins Symmetric ring breathing of phenylalanine
943	Skeletal modes (polysaccharides, amylose) Skeletal modes (polysaccharides, amylopectin)
777	Phosphatidylinositol

Raman results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-6}M and pre-processing using Rubber band-Vector normalisation:

Wave number [cm^{-1}]	Association
1592	C==N and C==C stretching in quinoid ring
1532	Amide carbonyl group vibrations and aromatic hydrogens
1372	T, A, G (ring breathing modes of the DNA/RNA bases)
1338	Tryptophan CH2/CH3 wagging, twisting &/or bending mode of collagens & lipids CH2/CH3 wagging & twisting mode in collagen, nucleic acid & tryptophan
1277	Amide III
1190	Nucleic acids and phosphates Electronic structure of nucleotides Amide III
781	Cytosine/uracil ring breathing (nucleotide)
726	C-S (protein), CH2 rocking, adenine

Raman results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-9}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1672	C==C stretch Amide I band (C55O stretch coupled to a N-H bending) Ceramide
1603	C==C in-plane bending mode of phenylalanine & tyrosine Ring C-C stretch of phenyl
1413	vs COO- (IgG)
1324	CH3CH2 wagging mode present in collagen & purine bases of DNA
1191	Nucleic acids and phosphates Electronic structure of nucleotides Amide III
1004	Phenylalanine (of collagen) vs(C-C), symmetric ring breathing, phenylalanine Phenylalanine (collagen assignment) Phenyl breathing mode v(C-C) phenylalanine
919	C-C stretch of proline ring/glucose/lactic acid C-C, praline ring (collagen assignment)
838	Deformative vibrations od amine groups

Raman results for MCF-7 cells treated in G_0/G_1 phase with a B[a]P 10^{-6}M and pre-processing using SG differentiation-Vector normalisation:

Wave number [cm^{-1}]	Association
1689	Amide I (disordered structure; non hydrogen bonded)
1446	CH2 bending mode of proteins & lipids CH2 deformation
1409	vs COO- (IgG)
1002	C-C aromatic ring stretching Phenylalanine Phenylalanine (collagen assignment)
941	Skeletal modes (polysaccharides, amylose) Skeletal modes (polysaccharides, amylopectin)
854	(C-O-C) skeletal mode of α-anomers (polysaccharides, pectin) Ring breathing tyrosine (proteins)
726	C-S (protein), CH2 rocking, adenine
645	C-C twisting mode of phenylalanine (proteins)