Biospectroscopy Investigations

Into

Cervical Cytology

A dissertation submitted in fulfilment of the requirements for the degree of

Master of Research in Biological Sciences

By

Diane Elizabeth Halliwell

MA, BSc. Hons. Applied Biochemistry

Centre for Biophotonics

Lancaster Environment Centre

June 2016

'My biggest fear in life is to be forgotten.'

Eva Perón

First Lady of Argentina (1946 – 1952).

Died from cervical cancer, aged just 33.

Table of contents

Table of tables

Supplementary tables

Table of figures

Supplementary figures

Declaration

I, Diane Elizabeth Halliwell, confirm that the work presented in this dissertation is my own work and has not been submitted in substantially the same form for the award of a higher degree elsewhere. Those sections, which have been published with co-authors or where data has been derived from other sources, have been clearly identified within this dissertation.

All images used in this dissertation are displayed with their correct attribution and include images released into the Public Domain; those purchased under license; images used from the Microsoft Office Clip Art & Images Library and used in accordance with the Microsoft Office license agreement (http://img.labnol.org/di/microsoft-office-license.pdf); images derived from WikiMedia Commons and used in accordance with the Creative Commons licence agreements (https://wiki.creativecommons.org/wiki/4.0); images used with the permission of the owning organisation; images used under the fair use for educational and research purposes; and author-created images, where I exert my rights as creator.

For further information, please consult UK Intellectual Property Office; available at: https://www.gov.uk/guidance/exceptions-to-copyright.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Professor Francis ('Frank') L. Martin, whose continued support, advice and guidance has been invaluable throughout this research project. Throughout my time at Lancaster University, I have been indebted to my colleagues who provided vital expertise and support, especially Dr Georgios Theophilou, Dr Holly Butler and Dr Kelly Heys. I would like to extend a special thanks to Dr Maria Kyrgiou at Imperial College, London for her patience, collaboration and sharing her valuable experience. My gratitude also goes to collaborators from the Physics Department, University of Liverpool and Daresbury Laboratory, Warrington for their kind assistance during experimental time on the A.L.I.C.E accelerator (Accelerators and Lasers in Combined Experiments). Finally, I would like to thank my husband, Paul, who allowed me the space, time and finances to start my great adventure.

Abbreviations

Abbreviation	Meaning
AFM	Atomic force microscopy
A.L.I.C.E 'ALICE'	Accelerators and Lasers in Combined Experiments
AS	Age-standardised
ASC-H	Atypical squamous cells that cannot exclude high-grade intraepithelial lesions
ASCUS	Atypical squamous cells of undetermined significance
ATR-FTIR spectroscopy	Attenuated total reflectance/reflection Fourier-transform infrared spectroscopy
BSCC	British Society Clinical Cytology
CIN	Cervical intraepithelial neoplasia
CLIRSPEC	The International Society for Clinical Spectroscopy
COCP	Combined oral contraceptive pill
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor
FDA	Food and Drug Administration
FSH	Follicle Stimulating Hormone
FTIR	Fourier-transform infrared
HG-CGIN (or HGCGIN)	High-grade cervical glandular intraepithelial neoplasia
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HSIL	High-grade intraepithelial lesion
ICC	Invasive cervical cancer
IEC	Independent Ethics Committee
IR	Infrared
IRAC	International Agency for Research on Cancer
IR-FEL	Infrared free electron laser
ISD	Information Services Division

IUD	Intrauterine device
KS	Kennard-Stone
LBC	Liquid-based cytology
LCL	Lower confidence limit
LCR	Long control region
LDA	Linear Discriminant Analysis
LH	Luteinizing hormone
LG-CGIN (or LGCGIN)	Low-grade cervical glandular intraepithelial neoplasia
LLETZ	Large loop excision of the transformation zone
LSIL	Low-grade intraepithelial lesion
MIB-1	Mindbomb E3 ubiquitin protein ligase 1
NA	Not applicable
NASA	National Aeronautics and Space Administration
NCI	National Cancer Institute
NCIN	National Cancer Intelligence Network
NHS	National Health Service
NHSCSP	National Health Service Cervical Screening Programme
Pap test	Papanicolaou test
PCA	Principal Component Analysis
PCA-LDA	Principal Component Analysis coupled to Linear Discriminant Analysis
POP	Progesterone-only pill
pRb	Protein retinoblastoma
r.p.m	Revolutions per minute
SCC	Squamous cell carcinoma
SD	Standard deviation
SNOM-IR-FEL	Scanning near-field optical microscopy coupled to an infrared free electron laser
SPA	Successive Projections Algorithm

UCL	Upper confidence limit
URR	Upstream regulatory region
UK	United Kingdom
USD	United States Dollars
WHO	World Health Organisation

Abstract

Local treatment for cervical intra-epithelial neoplasia (CIN) involves the removal of the affected part of the tissue and is >95% effective in preventing re-invasive disease. However, removal of part of the cervix is linked to significant adverse sequelae, including preterm birth; with cone depth and radicality of treatment correlating to the frequency and severity of adverse events. Since pre-treatment cervix length vary amongst women, the percentage of cervix excised may correlate more accurately to risk than absolute dimensions. Attenuated total reflectance, Fourier-transform infra-red (ATR-FTIR) spectroscopy detected that treatment for CIN significantly alters the biochemical fingerprint in the cervix, compared with women who have not had treatment; this is due to the excision of cervical tissue rather than a disease controlling effect. However, the spectra did not correlate to the cone depth or proportion of cervical length excised. Post-hoc analyses of patient characteristics found that spectral absorbance was different for treated women according to whether they were current/non-smokers; nulliparous/parous; by luteal/follicular phase; and by combined oral contraceptive pill use; these patient characteristics are likely to have affected the excisional outcomes. As traditional IR techniques are limited by the effect of diffraction of ~3 μm to 30 μm, we assessed the potential of scanning near-field optical microscopy in combination with an IR free electron laser (SNOM-IR-FEL), in determining the biophysical properties of abnormal cervical cells. SNOM-IR-FEL is able to distinguish between normal and various grades of cervical abnormalities at designated wavelengths associated with DNA, amides I and II and lipids, at spatial resolutions below the diffraction limit (≥0.2 μm).

1 Introduction

1.1 The global burden of cervical cancer

In 2012, around 528,000 new cases of cervical cancer were diagnosed, ranking it the fourth most common cancer in women globally after breast, colorectal, and lung cancers, and the second most common cancer in developing countries (World Health Organization [WHO], 2015; Ferlay *et al*., 2013a). In Europe, cervical cancer is the sixth most common cancer in females with approximately 58,300 new cases in 2012 and around 24,400 deaths, with Romania having the highest World age-standardised rates of incidence and Switzerland the lowest (Cancer Research UK, 2015). Whilst immunisation and screening programs have reduced morbidity and mortality in developed regions, the global mortality rate remains high at 52%, with 85% of the 270, 000 deaths each year occurring in developing countries (WHO, 2015).

The incidence of cervical cancer varies widely by country. High risk regions (affecting more than 30 women in every 100, 000 women) include Eastern Africa (43%), Melanesia (33%), Southern and Middle Africa (32% and 31%, respectively); low risk regions include Australia and New Zealand (6%) and Western Asia (4%); (Ferlay *et al*., 2013a). Almost one fifth of all new cases are diagnosed in India (WHO, 2013). The increased incidence and high mortality in high risk regions is due to the absence of widely available, good quality screening services (Vaccarella *et al*., 2013). With little or no access to care, the death rate was estimated in 2008 to contribute to

more than 10% of the economic loss in the developing world (American Cancer Society and LIVESTRONG, 2010).

Cervical cancer is a preventable and treatable disease with an estimated 1.55 million women still alive 5 years after diagnosis in 2008 (Anolue *et al*., 2014; Ferlay *et al*., 2008). The Worldwide healthcare costs per year have been valued at 3.2 billion USD, with medication costs estimated at 865.4 million USD, and nonmedical costs at 703.6 million USD (Muka *et al*., 2015). Absenteeism and the inability to work result in a global productivity loss estimated at 1.7 million USD per year. By 2030, cervical cancer is expected to kill more than 474, 000 women per year (American Cancer Society, 2013), significantly increasing the global economic burden.

1.2 UK burden of cervical cancer

1.2.1 Incidence

In the UK, around 9 new cases per 100,000 females were diagnosed in 2012 (Table 1); (Cancer Research UK, 2015).

	England	Wales	Scotland	Northern Ireland	UK
Cases	2482	174	295	93	3044
Crude Rate	9.1	11.1	10.8	10.0	9.4
AS Rate	9.3	11.5	10.9	10.2	9.5
AS Rate - 95% LCL	8.9	9.8	9.6	8.1	9.2
AS Rate - 95% UCL	9.6	13.2	12.1	12.2	9.9

95% LCL and 95% UCL are the 95% lower and upper confidence limits around the AS Rate. AS: Age-Standardised; LCL: Lower confidence limit; UCL: Upper confidence limit.

Table 1. Number of new cases, crude and European age-standardised (AS) incidence rates per 100,000 population, females, UK (reproduced from Cancer Research UK: Incidence, 2015).

Cervical cancer is the twelfth most common cancer in the UK with the age-standardised rates being highest in Wales (Cancer Research UK, 2015). The incidence of cervical cancer varies widely by geographical region, with rates being lowest in South and East of England, and the highest rates being found in parts of Scotland, Northern Ireland and the North of England (Trent Cancer Registry; 2012; National Cancer Intelligence Network [NCIN], 2008). The differences may be linked to differences in sexual attitudes, an increased prevalence in smoking and higher parity at a younger age.

The incidence of cervical cancer is strongly associated with two, age-specific peaks: 19 new cases per 100,000 in women aged 25-29 years, and 12 new cases per 100,000 in women aged 85-89 years (Cancer Research UK, 2015). The first peak is thought to be due to women becoming sexually active and subsequently infected with human Papillomavirus (HPV), an established risk factor for cervical cancer (Foley *et al*., 2011). Between 2010 and 2012 in the UK, 78% of cervical cancers were diagnosed in 25-64 year olds, with 11% of new cases diagnosed in women aged 75 years and over (Cancer Research UK, 2015).

The reasons for the second peak in older women are not fully understood and may be a combination of factors, including a lack of knowledge of what causes cervical cancer and who can be affected, or late cytological changes that were not present earlier in life and not detected by earlier screening. Additionally, there is the perception that this cancer is a younger woman's disease which promotes a drop off in attendance for screening in women aged 60+ years (Sherman *et al*., 2015). Other factors may include an immune system compromised by advancing age and where a

latent HPV infection suddenly becomes active, or the onset of other comorbidities and the use of associated medications needed to treat and control them.

Over the last decade, the incidence of cervical cancer has decreased significantly overall due to the deployment of the UK's national cervical screening programme in 1988, with the lifetime risk of a new-born child developing cervical cancer at some point in their life being 1 in 139 (2010 estimate); (Cancer Research UK, 2015). However, although the incidence rates for women aged 25-34 initially decreased by 35% between 1985 and 1987, and again between 2002 and 2004, increased rates of HPV infection and smoking prevalence in this age group has seen incidence rates rise by 52% (Trent Cancer Registry, 2012; Foley *et al*., 2011).

Almost 76% of cervical cancers are reported as squamous cell carcinomas (SCC), 10-15% are reported as adenocarcinoma (glandular lesions) and adenosquamous cell carcinomas (mixed glandular and squamous lesions), with the remaining cases being typed as poorly specified (Vizcaino *et al*., 1998). There is mounting evidence that the incidence of SCC is falling worldwide, whilst the incidence of adenocarcinoma is increasing, as is the risk of adenocarcinoma per population (Smith *et al*., 2000; Vizcaino *et al*., 2000; Vizcaino *et al*., 1998). These results suggest current cytology screening practices that were originally developed to identify squamous lesions are inadequate in the detection of adenocarcinomas.

There is some evidence that the incidence of cervical cancer in England is associated with ethnicity, with age-standardised incidence rates for White and Black women ranging from 8.2 to 8.7 per 100,000 and 6.3 to 11.2 per 100,000, respectively; whilst the rates of cervical cancer in Asian women are significantly lower, ranging from 3.6

to 6.5 per 100,000 (NCIN and Cancer Research UK, 2009). These differences are likely to be caused by different cultural attitudes towards sex.

1.2.2 Mortality

In the UK, there were 3 deaths per 100,000 females due to cervical cancer in 2012, ranking it the 17th most common cancer death among females, with age-standardised mortality rates being highest in Scotland (Table 2); (Cancer Research UK, 2014a).

	England	Wales	Scotland	Northern Ireland	UK
Deaths	742	43	112	22	919
Crude Rate	2.7	2.7	4.1	2.4	2.8
AS Rate	2.1	2.1	3.2	2.2	2.2
AS Rate - 95% LCL	2	1.4	2.6	1.3	2.1
AS Rate - 95% UCL	2.3	2.7	3.8	3.1	2.4

95% LCL and 95% UCL are the 95% lower and upper confidence limits around the AS Rate. AS: Age-Standardised; LCL: Lower confidence limit; UCL: Upper confidence limit.

Table 2. Number of deaths, crude and European age-standardised (AS) mortality rates per 100,000 population, females, UK (reproduced from Cancer Research UK [Mortality], 2014a).

Mortality rates follow a similar pattern as for incidence and vary by health region. Rates are highest in the North of England and lowest in the South of England (NCIN, 2005). During the period 2010-2012, more than 52% of deaths from cervical cancer occurred in women aged 25-64 years, and approximately 30% of deaths occurred in women aged 75 years and over (Cancer Research UK, 2014a). Mortality rates overall have decreased significantly following improvements to the national screening programme in the 1980's which is estimated to prevent approximately 5,000 deaths each year (Peto *et al.*, 2004). Survival rates at 1 and 5 years after diagnosis are 83% and 67%, respectively (Bate and Baker, 2015).

The significant increase in mortality rates for Scottish women is thought to be due to the continuing decline in the number of women participating in the cervical screening programme in Scotland, with 70.4% of women aged between 20-60 years attending in 2014/15, continuing a 10-year fall in uptake since 2005/06, which had 78% participation (Information Services Division, Scotland, 2015). These figures suggest around 30% of Scottish women are at risk of developing cervical cancer.

1.2.3 Economic costs

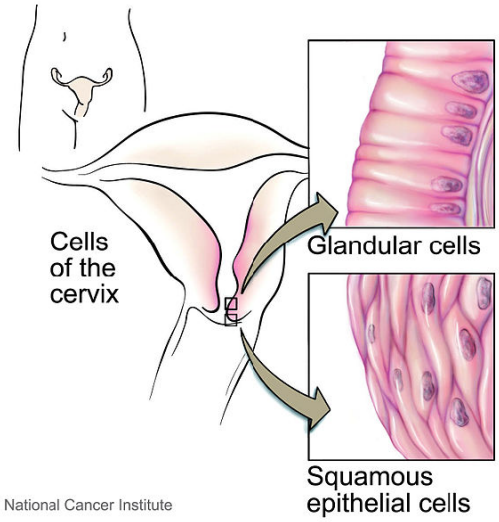
The NHS spends around £21 million treating cervical cancer per year (2011), government loses over £9 million in tax revenue as a result of absenteeism and the inability to work, and the combined financial costs to women is approximately £14 million, with a significant portion of these costs and losses associated with treating more advanced cervical cancers (Salter, 2014). Although the percentage of women aged 25-64 years in 2013/14 in England who had an adequate cervical screening test in the last 5 years was 77.8%, this was significantly lower (63.3%) for those aged between 25-29 years who are screened every 3 years (Screening and Immunisations team, Health & Social Care Information Centre, 2014), thus leaving a significant number of women with an increased risk of developing more serious cervical cancers.

The reasons behind falling screen rates are varied, often contentious, and include difficulty in being able to fit screening attendance around work and family commitments; embarrassment; experience of discomfort during the collection of cervical cells; poor awareness of what screening is for; and extending services to hard-to-reach groups such as immigrant and older women. Achieving a 100% screening uptake would reduce the incidence of more invasive cervical cancers, NHS costs

would fall by almost 50%, tax loses would fall to about a third, and total costs to women would fall by approximately 40% (Salter, 2014). Conversely, if national screening rates continue to fall to 70%, it is estimated that costs to the NHS alone would increase by another £6.5 million per year.

1.3 Structure and function of the cervix

The cervix is the tube-shaped opening to the uterus that protrudes into the vagina and plays a vital role in the overall reproductive health of women. The key functions of the cervix include permitting the passage of menstrual fluid, promoting fertility, and protecting the uterus, upper reproductive tract and a developing foetus from pathogens (Reproductive Health Technologies Project, 2000). The cervix is approximately 3-4 cm long and 2.5 cm wide, although the size and shape of the cervix differs among women according to age, hormonal status, parity, pregnancy, and surgical treatment (Martyn *et al*., 2014; Mazouni *et al*., 2005; Pardo *et al*., 2003). The cervix is composed of two parts: the ectocervix is composed squamous epithelium, and the endocervix is composed of glandular (columnar) or mucin-secreting epithelium that produces mucin glycoproteins (Martyn *et al*., 2014); (Figure 1).

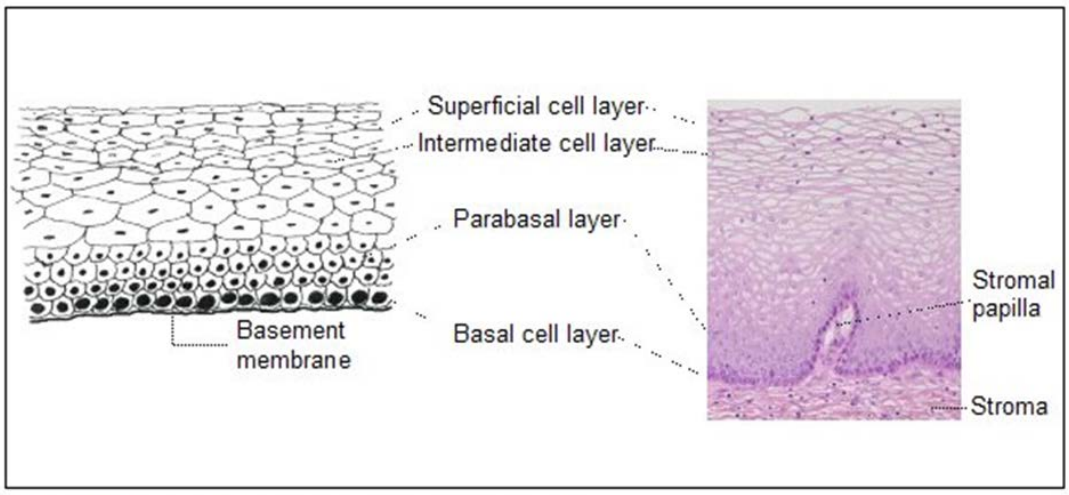


Bliss (Illustrator; source: National Cancer Institute), 2004

Figure 1. Glandular cells of the endocervix and squamous cells of the ectocervix.

1.3.1 Squamous epithelium

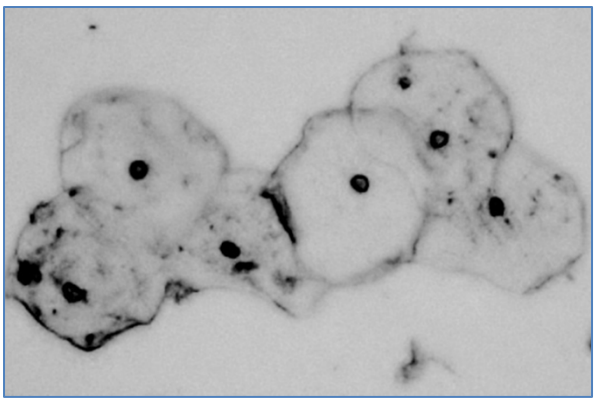
The ectocervix is covered by 15-20 layers of glycogen-rich, stratified, non-keratinizing cells which are opaque and pale pink in colour (International Agency for Research on Cancer [IRAC], 2016a). Native epithelium defines tissue that developed during embryonic life. Newly formed metaplastic tissue develops as a result of physiological influences, such as those experienced at puberty and during the early reproductive years. Age and menopausal status also affects the development of metaplastic epithelium. The histological makeup of squamous epithelium is composed of 4 cellular layers including the basal and parabasal cell layers, and the intermediate and superficial cell layers. The basal cells are attached to the basement membrane, and feature large nuclei suspended in very little cytoplasm. The membrane separates these cells from the stroma, which forms regular projections of papillae (Figure 2).



IRAC, 2016b

Figure 2. The differing layers of stratified squamous epithelium (x 20).

The basal cells divide, mature and are pushed upwards to form the parabasal layer (IRAC, 2016a). As this layer undergoes further maturation to form the intermediate layer, the cells become polygonal in shape and their nuclei smaller. Finally, these cells mature into flattened cells, with an abundance of cytoplasm containing pyknotic nuclei, where the chromatin has undergone irreversible condensation as it prepares for programmed cell death (apoptosis); (Figure 3).



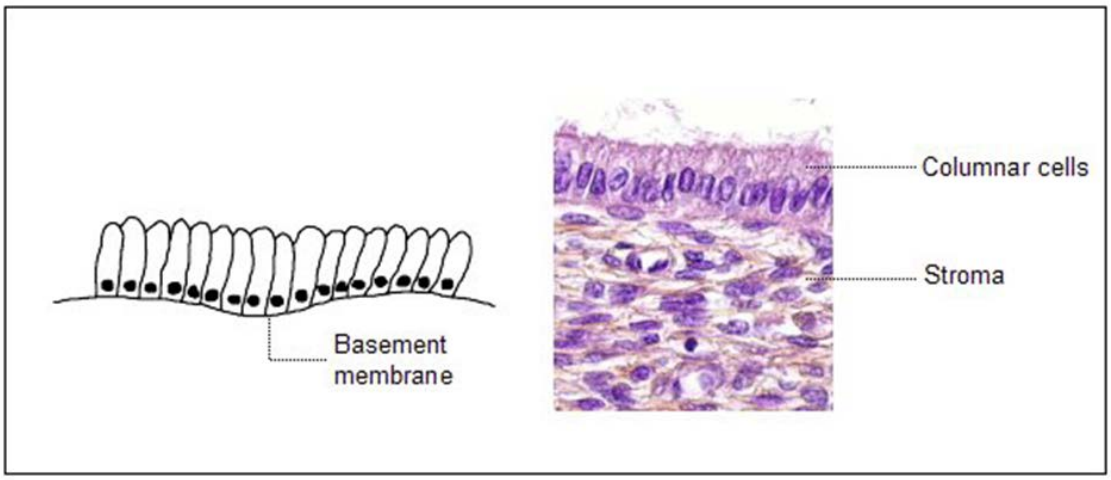
Halliwell, 2016a

Figure 3. Unstained normal squamous epithelial cells of the superficial layer of the ectocervix.

The cells of the intermediate and superficial layers have plentiful glycogen, indicative of normal maturation and development. Abnormal development is associated with an absence of glycogen production (IRAC, 2016a). Proper maturation of the squamous epithelium is controlled oestrogen. Following menopause, the cells do not mature beyond the parabasal layer or undergo glycogenation.

1.3.2 Glandular epithelium

The endocervical canal, which lies just inside the cervix, is composed of a single layer of glandular cells (IRAC, 2016a). The underlying blood vessels of the stroma are visible through the single layer, making the cells reddish in appearance. These tall 'columnar' cells are long, packed with cytoplasm, with their nuclei occupying the end of the cell that is attached to the basement membrane (Figure 4).



(IRAC, 2016c)

Figure 4. Glandular (columnar) cells of the endocervix canal (x 40).

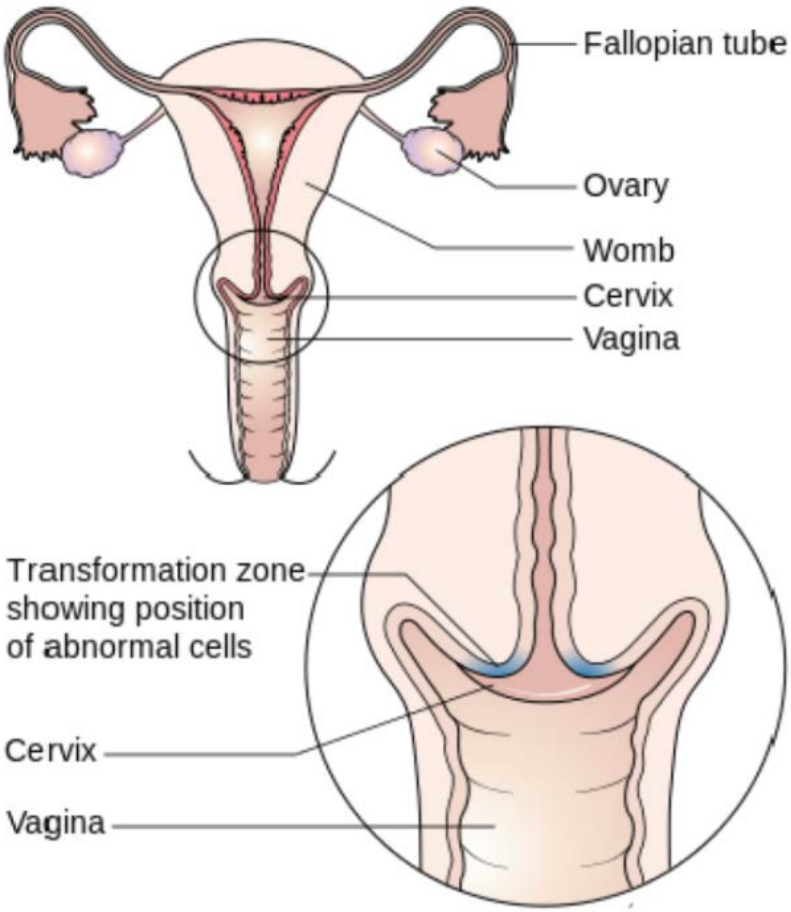
Glandular epithelium unites with the endometrial epithelium at its upper margin, and fuses with the squamous epithelium at its lower margin, called the squamo-columnar

junction. This junction may appear at different sites within the endocervix as it varies according to age, menstrual and pregnancy status, and oral contraceptive use (Eurocytology, 2015). Mitosis and glycogen production do not occur in glandular cells.

1.3.3 The transformation zone

Anaerobic conversion of vaginal glycogen to mainly acetic and lactic acids by micro-organisms, especially the *Lactobacillus* spp., maintains normal vaginal pH to between 3.8 and 4.5 (Boskey *et al*., 1999). During puberty and at first pregnancy, the influence of oestrogen triggers the cervix to increase in volume, causing the endocervix to evert onto the ectocervix (portio vaginalis), exposing the delicate glandular cells to the more acidic environment (Eurocytology, 2015). The epithelium begins to undergo dynamic changes (metaplasia); this is called the 'transformation zone' (Figure 5).

Within the transformation zone, the glandular cells are gradually replaced by the more durable squamous epithelial cells (Sun *et al*., 1992). The transformation zone is the common site of cervical neoplasia (Burghardt, 1986), whilst only 3% of cervical lesions are reported to arise in native ectocervix epithelium (Abdul-Karim, *et al*., 1982).



Cancer Research UK, 2014b

Figure 5. The transformation zone of the cervix located at the ectocervix and endocervix.

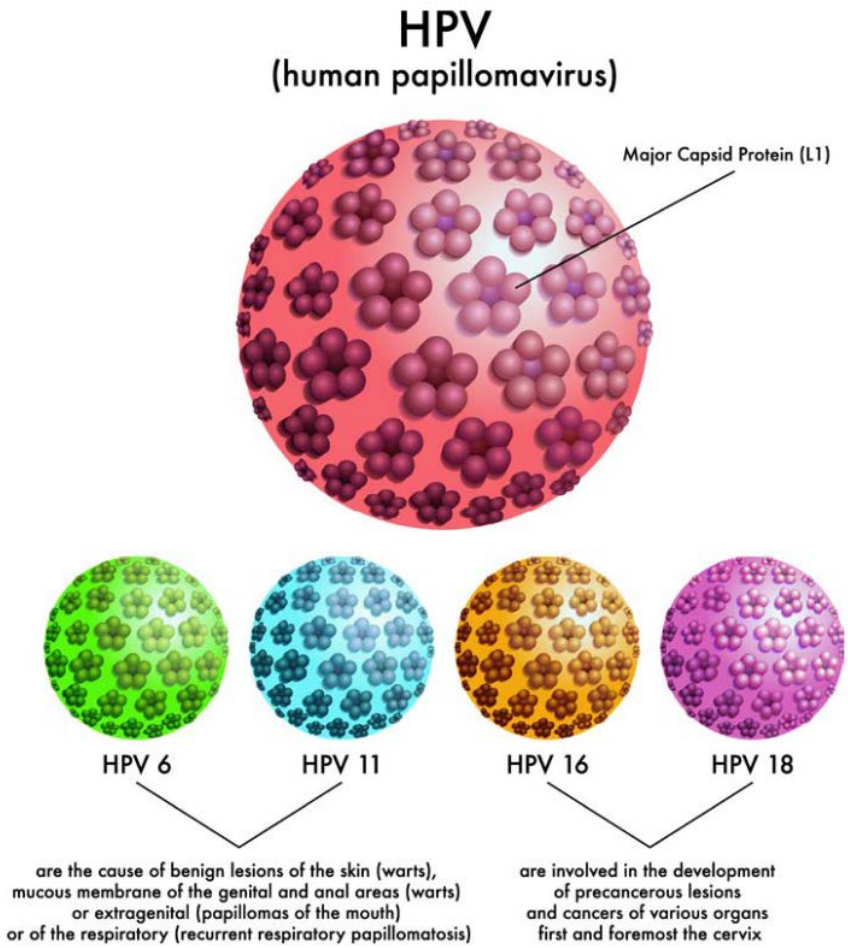
Vaginal acidity destroys the glandular epithelium, which is replaced by newly formed squamous epithelium (squamous metaplasia). The process prompts the formation of undifferentiated, sub-glandular reserve cells, from which the metaplastic squamous epithelium is eventually derived (IRAC, 2016a). Initially, there is no stratification of this immature metaplastic squamous epithelium and they are absent of glycogen. Following differentiation and stratification, the mature cells appear very similar to the original squamous epithelium. The replacement of glandular epithelium by metaplastic squamous epithelium is irreversible.

In most women, the immature metaplastic epithelium develops into mature squamous metaplastic epithelium and is functionally similar to the original squamous epithelium, including glycogen production (IRAC, 2016a). However, cervical intraepithelial neoplasia (CIN) arises following persistent infection with oncogenic types of HPV, which infect the immature basal squamous metaplastic cells, transforming them into cells with aberrant cellular functions (Bosch *et al*., 1995). These cells proliferate to form an abnormal epithelium, termed 'dyskaryosis' which can spontaneously regress to normal, remain static or progress into more invasive lesions (IRAC, 2016a). The high turnover rate and low maturation of cells in the transformation zone makes it especially vulnerable to pre-cancerous changes (Reproductive Health Technologies Project, 2000).

Both glandular and squamous epithelial cells are collected during cervical screening to detect the presence of these atypical cells within the transformation zone.

1.4 Causes of cervical cancer

HPV is a group of very common viruses that includes over 100 different types (WHO, 2015), with approximately 35 types involved in the infection of the genital tract (Bosch *et al*., 1995); (Figure 6).



Adobe Systems, 2015 (purchased under licence)

Figure 6. HPV types and associated diseases.

Cervical cancer is mainly caused through a sexually transmitted infection with HPV, with 84.1% of invasive cervical cancers associated with infection with high-risk types 16 and 18 (Ramakrishnan *et al*., 2015). Research has shown that HPV 16 is associated with SCC, and HPV 18 is more commonly associated with adenocarcinoma (Clifford *et al*., 2003). Genital HPV infection in women is largely acquired in adolescence, shortly after commencing sexual activity, although worldwide prevalence varies according to geography and is influenced by cultural attitudes to sexual behaviour (Smith *et al*., 2008). Whilst many HPV infections are transient and clear up without

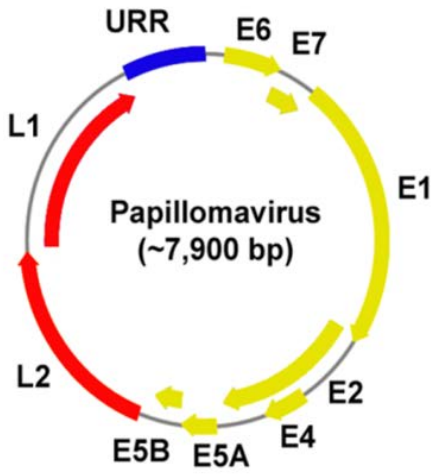
intervention within 1 to 2 years, for some women the infection persists and CIN can develop in as little as 2-4 years (Kyrgiou and Shafi, 2014).

The environmental risk factors for HPV persistence and the development of high-grade lesions and cervical cancer include high parity, long-term oral contraceptive use, and co-infection with other sexually transmitted agents, including chlamydia, Herpes simplex virus 2 (HSV-2) and Human immunodeficiency virus (HIV); (Castellsagué *et al*., 2002). Current smoking is associated with a significantly increased risk of SCC but not for adenocarcinoma (Berrington de González *et al*., 2004). Pure adenocarcinoma has a higher rate of incidence in younger patients (Sasieni and Adams, 2001), and is linked with a poorer survival outcome than SCC (Liu *et al*., 2001). Other non-environmental risk factors include the behaviour of the host's immune system, HPV genotypes, and extent of integration into the host's DNA (Castellsagué *et al*., 2002). Co-infections with taxonomically similar HPV types have been found to reduce the likelihood of progression to high-grade lesions in HIV positive women, possibly as a result of an immune cross-protection (Sobota *et al*., 2014). Research investigating a link between cervical HPV persistence and nutrition revealed that a diet rich in fruits, vegetables, Vitamins C and E, beta- and alpha-carotene, lycopene, and lutein and crytoxanthin (a natural carotenoid pigment), showed a 'possible' protective effect (García-Closas *et al*., 2005). Antioxidants may play a protective role by influencing the duration of HPV infection, since plasma concentrations of some carotenoids such as lutein and Vitamin E, were found to be inversely associated with HPV persistence (Castle and Giuliano, 2003).

1.5 The HPV genome

HPV consists of ~8000 base-pair circular DNA molecules packaged inside a protein coat Muñoz *et al*., 2006). The virus is approximately 55 nm in size, with seven important coding regions split into three functional coding groups (Faridi *et al*., 2011); (Figure 7).

'Early' genes and their associated functions
E_1/E_2: Replication & transcription (involving E6/7)
E_4: Possibly virus release
E_5: Cell immortalisation/rapid cell growth
E_6: Inhibits apoptosis (*oncogenic*)
E_7: Inhibits tumour suppression (*oncogenic*)
'Late' genes and their associated functions
L_1: Main capsid protein/L_2: Minor capsid protein
URR: promoters and enhancers
(Ramakrishnan *et al*., 2015; Faridi *et al*., 2011; Tsai and Chen, 2003)



URR: Upstream regulatory region. (Microsoft, 2016).

Figure 7. Structure of HPV.

The 'early' coding genes comprise the first functional group (E group), a long upstream regulatory region (URR) or long control region (LCR) comprises the second functional group, and separates the 'early' genes from the 'late' coding genes (L group); (Faridi *et al*., 2011; Muñoz *et al*., 2006).

The 'early' genes include E_1/E_2 which code for proteins that control the functionality of oncogenic genes E_6 and E_7; E_4 may code for a protein that governs virus release from the cell; E_5 codes for a protein that enhances cell immortalisation; E_6 codes for proteins that inhibit p53, suppressing cell apoptosis; E_7 codes for a viral protein that

binds to tumour suppressor proteins, thus overriding the control of normal cell division and keeping the cell in an unscheduled synthesis-phase (S-phase) state (Ramakrishnan *et al*., 2015; Faridi *et al*., 2011; Muñoz *et al*., 2006; Tsai and Chen, 2003). The 'late' genes, L_1/L_2, code for the structural completion of the final virus particles, and the URR (or LCR) controls the regulation of gene expression, promotes the replication of the viral genome and its eventual packaging into newly formed virus particles (Muñoz *et al*., 2006). Of these seven genes, E_5, E_6, and E_7, are considered critical in the carcinogenesis of cervical cancer (Ramakrishnan *et al*., 2015).

1.6 Mechanism of HPV carcinogenesis

1.6.1 HPV life cycle

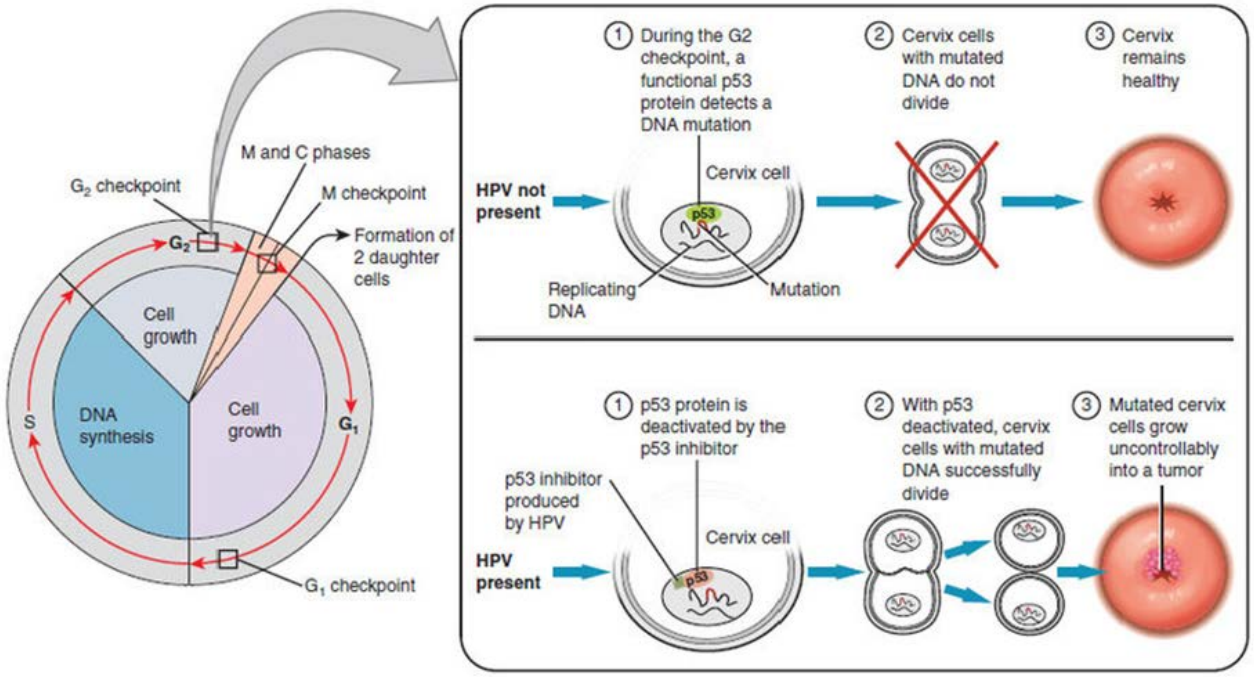
The virus is introduced to the body through a micro trauma at the site of the epithelium, possibly from small vaginal tears acquired during sexual intercourse, or friction caused by clothing. The infectious cycle begins when HPV particles reach the basal layer of the epithelium, and are thought to enter epithelial stem cells, which are abundant in hair follicles and may be an important site of entry (Doorbar, 2005). In early pre-cancerous lesions (CIN1 and CIN2), the viral DNA is present in episomal form, thus replicating independently of the host's genome (Williams *et al*, 2011). In more advanced pre-cancerous lesions (CIN3) and invasive carcinoma, the viral genome integrates with the host's genome.

The initial infection promotes a proliferative phase, increasing the number of basal cells infected with viral episomes. The viral genome is replicated to a low copy number of around 10-200 copies per cell and maintained within the infected, but still

functional cells. Proteins coded for by E_1/E_2 are considered critical for DNA replication within the basal layer (Williams *et al*., 2011). The immune system possibly keeps the infection in this state during viral persistence (Doorbar, 2005).

Normal eukaryotic cell cycle consists of four main stages. G1: the cell is metabolically active and growing; S phase: DNA replication takes place; G2: the cell prepares for division by synthesising a number of important proteins; and M or mitosis phase: the chromosomes are duplicated into sister chromatids, separate into two daughter nuclei and the cell divides into two, each new cell containing an exact copy of its parent DNA (Cooper, 2000).

G1, G2 and M phases incorporate key cellular checkpoints. These checkpoints can detect unfavourable conditions and prevent the cell from passing on the mutated DNA into daughter cells by inhibiting division. Critically in HPV infection, the protein products of E_5, E_6 and E_7, interact with a number of cellular proteins that inhibit the normal function the G2 and M checkpoints (Figure 8).



(OpenStax College, 2013).

HPV: Human papillomavirus.

Figure 8. HPV interaction of cell checkpoints, especially G2.

E_5 activates epidermal growth factor receptor (EGFR), stimulating biochemical cascades that result in an overexpression of proto-oncogenes and stimulate rapid cell growth; E_5 can also inhibit the expression of tumour suppressor gene p21 and disturb normal cell cycle checkpoints (Tsai and Chen, 2003). Protein Retinoblastoma (pRb) is a tumour suppressor protein that can prevent the cell from entering S phase (DNA replication) by binding to E2F transcription factor. However, E_7 binds to pRb and displaces E2F, triggering the expression of proteins necessary for DNA replication (Doorbar, 2005). This unscheduled S-phase would normally provoke p53, a cellular tumour antigen, to schedule the cell for apoptosis (programmed cell death). However, E_6 inhibits p53 by targeting it for degradation, thus E_6 and E_7 co-operate together to efficiently 'hijack' normal cell activity, abolish normal termination and promote cell immortalisation (Faridi *et al*., 2011).

As part of their natural turnover, aging basal cells of the epithelium are pushed to the parabasal layer and lose their ability to divide. As they breakdown, the low copy number of virus particles are released to repeat the infectious cycle (Münger *et al*., 2004). The loss of control of cell growth coupled to increasing genomic instability and oncogenic mutations, lead to the formation of CIN in the transformation zone of the cervix.

1.7 Classification of cervical intraepithelial neoplasia

1.7.1 Cervical squamous intraepithelial neoplasia (CIN)

In 2013, the NHS Cervical Screening Programme (NHSCSP) adopted the revised British Society Clinical Cytology (BSCC) terminology into their new guidelines for

the classification of cervical cytology (NHS Cervical Screening Programme, 2013). Outside of the UK, the Bethesda system is widely used (National Cancer Institute [NCI], 2001). The classification systems used both in the UK and the rest of the world is summarised in Table 3.

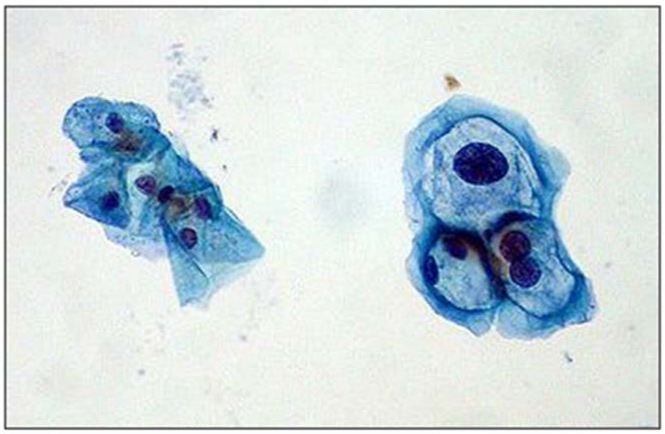
Cytology		**Histology**
NHSCSP 2013 (UK cytology system)	**Bethesda 2001 (cytology system widely used outside of the UK)**	**Used widely throughout the world**
Borderline changes in squamous/endocervical (glandular) cells	ASCUS ASC-H	HPV
Low-grade dyskaryosis	LSIL	CIN1
High-grade dyskaryosis (moderate)	HSIL	CIN2
High-grade dyskaryosis (severe)	HSIL	CIN3
High-grade dyskaryosis/?invasive SCC	HSIL SCC	SCC

ASCUS: Atypical squamous cells of undetermined significance; ASC-H: Atypical squamous cells that cannot exclude HSIL; CIN: Cervical intraepithelial neoplasia; LSIL: Low-grade intraepithelial lesion; HSIL: High-grade intraepithelial lesion; SCC: Squamous cell carcinoma.

Table 3. The BAC/NHSCSP and Bethesda systems for classification of CIN.

The histogenetic classification of pre-cancerous lesions as CIN was introduced in 1967 by Richart and now widely adopted, reflects the depth of epithelial involvement. CIN is identified through histological diagnosis and defined as the amount of squamous epithelium that is dyskaryotic or made up of undifferentiated neoplastic cells of basaloid type (Buckley *et al*., 1982). However, other factors also contribute to the classification of CIN, including stratification, the site of mitotic figures within the epithelium, and the presence of abnormal mitotic figures (Buckley *et al*., 1982). Examination of cervical cells seeks to identify atypical cells, 'koliocytes' that have

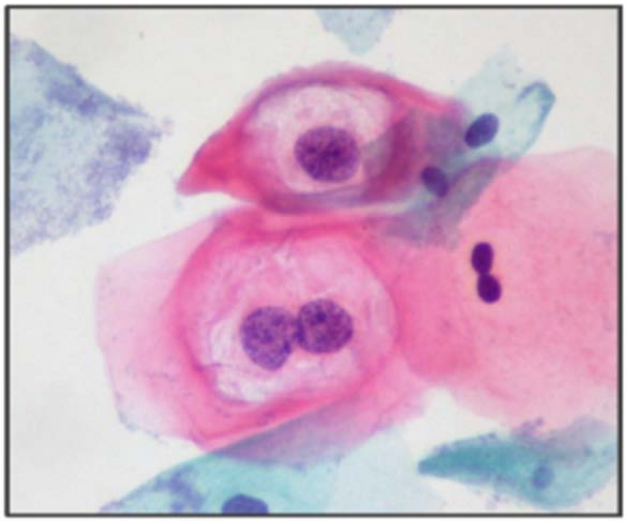
with a clear halo around the nucleus (Figure 9). The halo is thought to be a reflection of HPV infection (Edwards, 2015).



Uthman, 2006a

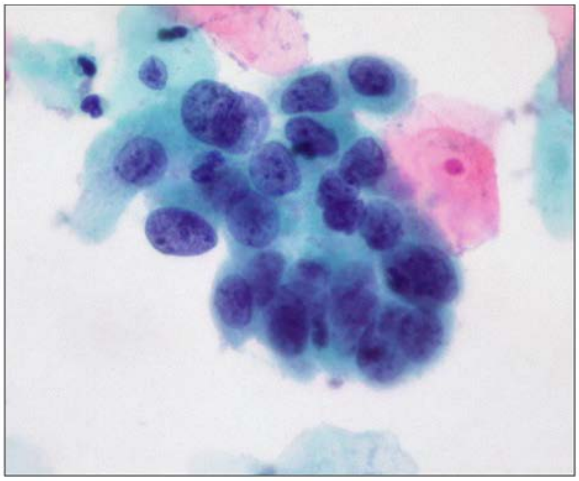
Figure 9. Normal cervical cells (left) and koliocytes (right).

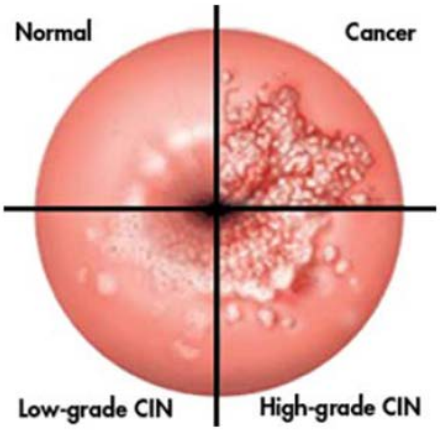
An example of low-grade dyskaryosis (Bethesda classification: LSIL) is shown in Figure 10, and high-grade dyskaryosis (Bethesda classification: HSIL), together with changes visible at the surface of the cervix, is shown in Figure 11.



Uthman, 2007a

Figure 10. Low-grade dyskaryosis (Bethesda classification: LSIL).



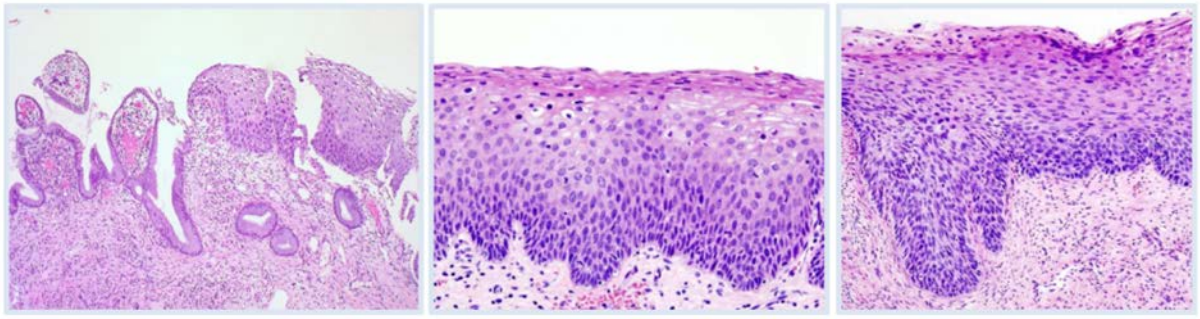


(1); Uthman, 2007b

(2); Honduras Health, no date

Figure 11. (1) High-grade dyskaryosis (Bethesda classification: HSIL); changes visible at the surface of the cervix (2).

Histological grading of CIN takes into account the level of involvement of the epithelium. For CIN1, undifferentiated cells are confined to the lower third of the epithelium. Cytological changes may be evident throughout the entire thickness of the epithelium but mitotic figures are few, and adequate cell maturation is still evident with minimal nuclear abnormalities (IRAC, 2016d). For CIN2, the abnormalities are found approximately across one half to two thirds of the thickness of the epithelium. Mitotic figures may be evident through the lower half of the tissue and the nuclear abnormalities are more prominent. For CIN3, the full thickness of the epithelium is usually involved together with abundant mitotic figures, which may have abnormal forms. Differentiation and stratification are often absent altogether (Figure 12).



CIN1; Uthman, 2013 CIN2; Uthman, 2006b CIN3: Uthman, 2006c

Figure 12. Types of cervical intraepithelial neoplasia (CIN).

1.7.2 Classification of cervical glandular intraepithelial neoplasia (CGIN)

Both the NHSCSP and the Bethesda systems (NCI, 2001) have attempted to define the range of cellular changes observed in the glandular mucosa (Table 4).

NHSCSP 2013 **(UK cytology system)**	**Bethesda 2001** **(cytology system widely used outside of the UK)**
Borderline changes in endocervical cells	Atypical glandular cells, not otherwise specified a. Endocervical b. Endometrial c. Glandular
?Glandular neoplasia a. ?Glandular neoplasia of endocervical type (divided into low-grade cervical glandular intraepithelial neoplasia [LG-CGIN], and high-grade cervical glandular intraepithelial neoplasia	Atypical glandular cells favour neoplastic a. Endocervical b. Glandular

NHSCSP 2013 (UK cytology system)	Bethesda 2001 (cytology system widely used outside of the UK)
[HG-CGIN]) b. ?Glandular neoplasia (non-cervical)	
	Endocervical adenocarcinoma *in situ* Adenocarcinoma a. Endocervical b. Endometrial c. Extra uterine d. Not otherwise specified

LG-CGIN: Low-grade cervical glandular intraepithelial neoplasia; HG-CGIN: High-grade cervical glandular intraepithelial neoplasia.

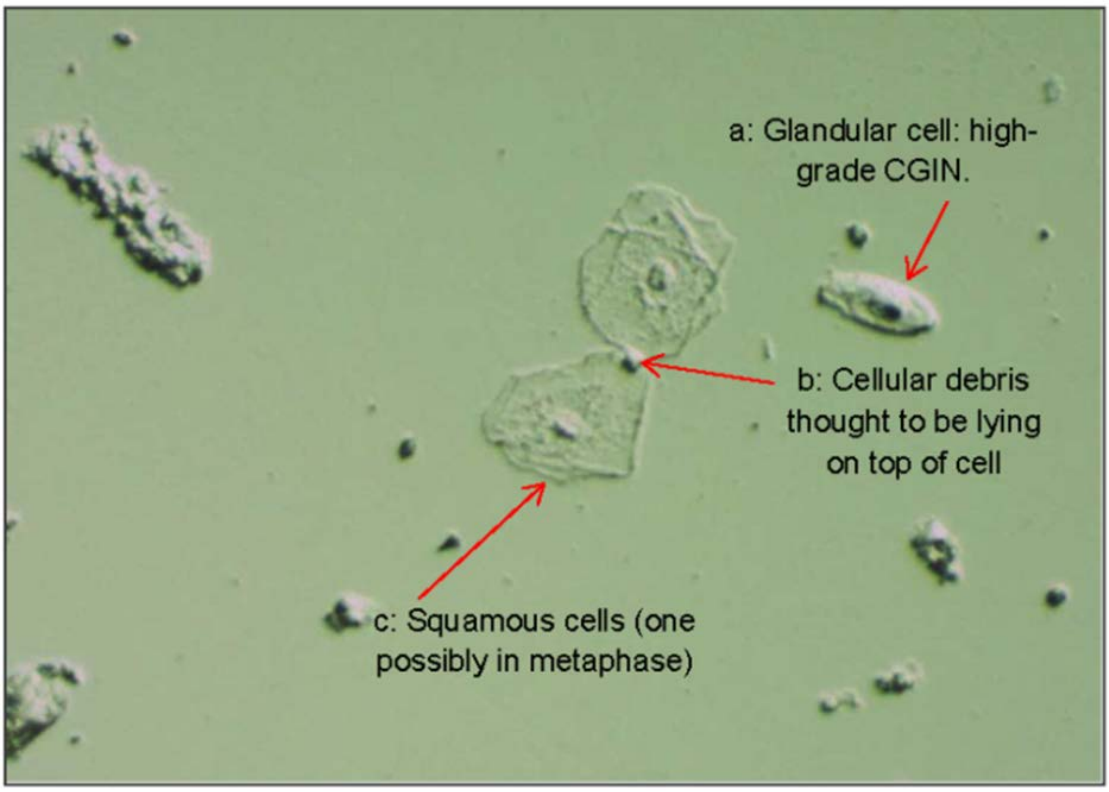
Table 4. The BAC/NHSCSP and Bethesda systems for classification of glandular neoplasia

The biology of glandular epithelial lesions is less well understood than squamous epithelial lesions. Around 60% of adenocarcinomas are pure and involve only the glandular cells, with the remaining 40% composed of both squamous and glandular cells (adenosquamous carcinoma); (Kyrgiou, 2013). Adenocarcinomas present with more histological sub types than squamous carcinomas.

Low-grade cervical glandular intraepithelial neoplasia (CGIN) affects the basal two thirds of the epithelium, may present as altered glandular profiles, and is characterised by hyperchromasia with small changes to nuclear size, increased mitosis and loss of polarity (Brown, no date). High-grade CGIN, which includes adenocarcinoma in situ, affects the whole depth of the epithelium, where the glands are budded or branched, and characterised by hyperchromasia, atypical nuclei and increased and abnormal

mitosis. These changes are accompanied by an increased bioavailability of the Ki-67 protein, a marker for cell proliferation, and the enzyme, mindbomb E3 ubiquitin protein ligase 1 (MIB-1) which is involved in the control of apoptosis.

Since cervical screening collects both glandular and squamous cells, they are often present together (Figure 13).



Halliwell, 2016b

Sampled from adenocarcinoma Stage 1B1 (unstained) a: High-grade cervical glandular intraepithelial cell with atypical nuclei and increased hyperchromasia. b: Cellular debris, possibly a stray nucleus on top of cell. c: Low-grade squamous cell; elongated nucleus may be indicative of metaphase with chromosomes lined up along the metaphase plate.

Figure 13. Adenocarcinoma Stage 1B1: mixed grade are often present.

1.8 Conventional screening for cervical cancer

1.8.1 National policy

In the UK, the national policy on cervical screening is divided into two age groups. Women aged 25-49 years are invited for routine screening every 3 years, and women aged 50-64 are invited for screening every 5 years (Screening and Immunisations Team, Health & Social Care Information Centre, 2014). Invasive disease in women under 25 years is rare, which prompted a revision in 2003 of the screening starting age from 20 years to 25 years.

For women aged 30 years and under, the prevalence of HPV infection is high, whilst the risk of cervical cancer is low; these patients often present with clinically insignificant lesions that will regress without intervention (Women's Health and Education Centre, 2010).

For women aged 65 and over, only women who have never been screened, or those who have recently had abnormal tests on three occasions are invited for screening (Screening and Immunisations Team, Health & Social Care Information Centre, 2014), despite evidence that approximately 20% of new cases of cervical cancer occur in women aged 65 years or over (The Guardian, 2015), and about half of all deaths occur in women in this age group (Sherman *et al*., 2015).

The diagnosis of celebrity Jade Goody in 2008 and her subsequent death in 2009 at the age of 27 from cervical cancer, prompted an extra 478, 000 screening attendances in England between 2008/09; with 28% of these attendances occurring in the 25-49 age group and whose screening was overdue (Lancucki *et al*., 2012). The high media

attention surrounding Goody's diagnosis focused much attention on lowering the age of first screening, and gave rise to the perception of cervical cancer is a disease affecting only younger women (Sherman *et al*., 2015). As the proportion of the UK's population aged 65 and over is predicted to rise to 24.3% by 2037 (Office for National Statistics, 2015), and a significant number will be women, the number of women in this age group affected by the disease will also increase. UK policy on cervical screening in this age group needs to reflect this pattern of predicted change.

1.8.2 The Papanicolaou smear

Originally described by Dr George N Papanicolaou and Dr Herbert Traut (1941), the Papanicolaou cervical smear test (Pap smear) was the first method of screening for cervical cancer in many countries, and has contributed significantly to reducing cervical cancer incidence and mortality by 75% (Mehta *et al*., 2009). The Pap smear collects exfoliated cells scraped from the cervix which are spread directly onto a glass slide and viewed microscopically to identify atypical cells (dyskaryosis). However, the test has a number of important limitations. Inadequate sampling, which accounts for around 8% of samples, occurs when the transformation zone is not sampled correctly, and only squamous cells are collected (Mehta *et al*., 2009). Conventional cytology screening is subjective, leading to misinterpretation (Nanda *et al*., 2000). The false-negative rate for the test has been reported as high as 36% (Spence *et al*., 2007), with a reported sensitivity of 30% to 87%, and a specificity of 86% to 100% (Nanda *et al*., 2000). The Pap smear can be compromised by contaminants such as blood, bacteria and yeast (Mehta *et al*., 2009).

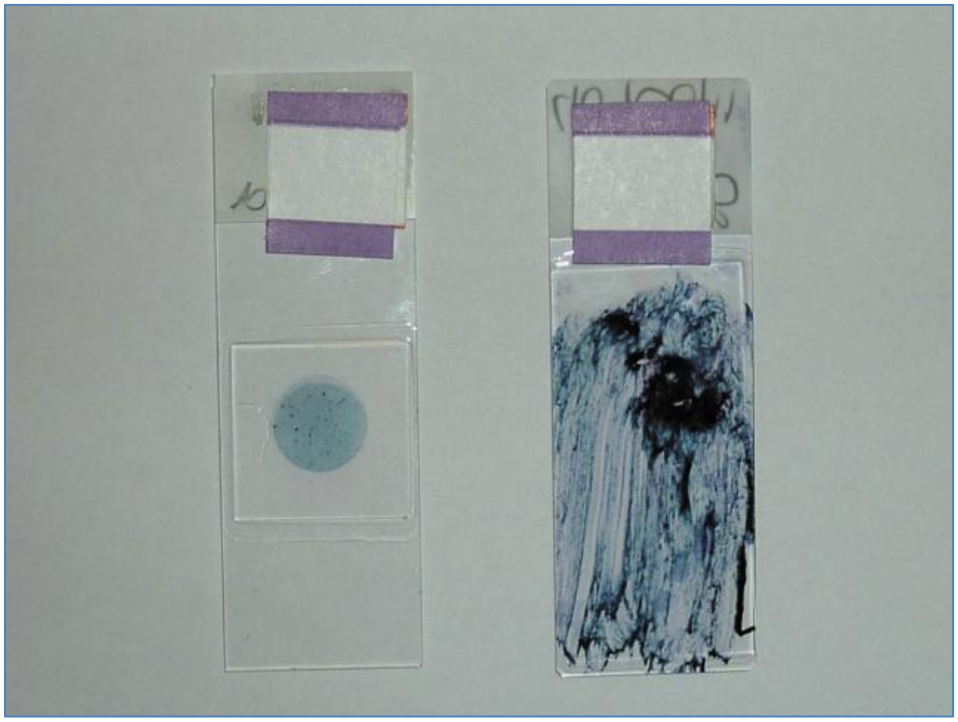
The Pap smear was originally developed to detect squamous cell carcinomas, and has a lower sensitivity for adenocarcinoma (Scheiden *et al*., 2004). Since a quarter of new cases of cervical cancer are adenocarcinomas, the Pap smear will become an increasingly less useful diagnostic tool. The most important limitation is human error, either through inaccurate observer interpretation or bias, and it has a low sensitivity in detecting CIN2 and CIN3 (Boone *et al*., 2012). The sensitivity becomes more variable with age, increasing in women aged 50 years and over (Cuzick *et al*., 2006). Since a typical Pap smear slide contains 50,000-150,000 cells with the possibility of only a few being abnormal, it is not surprising that underlying atypia is missed (Mehta *et al*., 2009).

1.8.3 Liquid-based cytology

In many countries, including the UK, the Pap smear has been replaced by liquid-based cytology (LBC) which has greatly reduced the number of inadequate samples (Safaeian and Solomon, 2007). This technique uses a cervical speculum that brushes cells from the cervix and the speculum is then plunged directly into an ethanol-based preservative/fixative and vigorously stirred, producing a suspension of filtered cells ready for mounting onto a slide. This sampling technique offers some advantages over the conventional Pap smear, including providing a higher number of cells in the media and the removal of contaminants such as blood.

LBC is also semi-automated, provides an even layer of cells that are easier to interpret, and can be used for testing for HPV DNA and other biomarkers (Safaeian *et al*., 2007), which removes the necessity for the patient to return for an additional visit. LBC has been shown to have comparable sensitivity in the detection of CIN2 or worse

lesions compared to the conventional Pap smear but can also produce more positive findings leading to a lower predictive value (Ronco *et al*, 2007). This higher rate of satisfactory LBC smears reduces both the patient's anxiety and financial costs associated with repeat Pap smears (Figure 14).



Continuing Medical Education, 2014

Figure 14. An example of a LBC slide (left) and a conventional Pap cytology slide (right) after staining and mounting.

1.8.4 HPV testing

An investigation of HPV testing showed that it was not cost-effective nor did it add any significant benefit when used in combination with LBC; however, it was shown to be highly effective as sole primary screening method reported to provide 60-70% better protection against invasive cervical cancer when compared with LBC

(Kitchener *et al*., 2009). HPV testing has other advantages over LBC in terms of a high negative predictive value and is based on fully automated platforms enabling high throughput. In the UK, it is likely that HPV-based screening will eventually replace LBC screening and may optimise the screening interval to 5 years for women aged 30 years and above (Kyrgiou and Shafi, 2014).

1.9 Alternative screening approaches

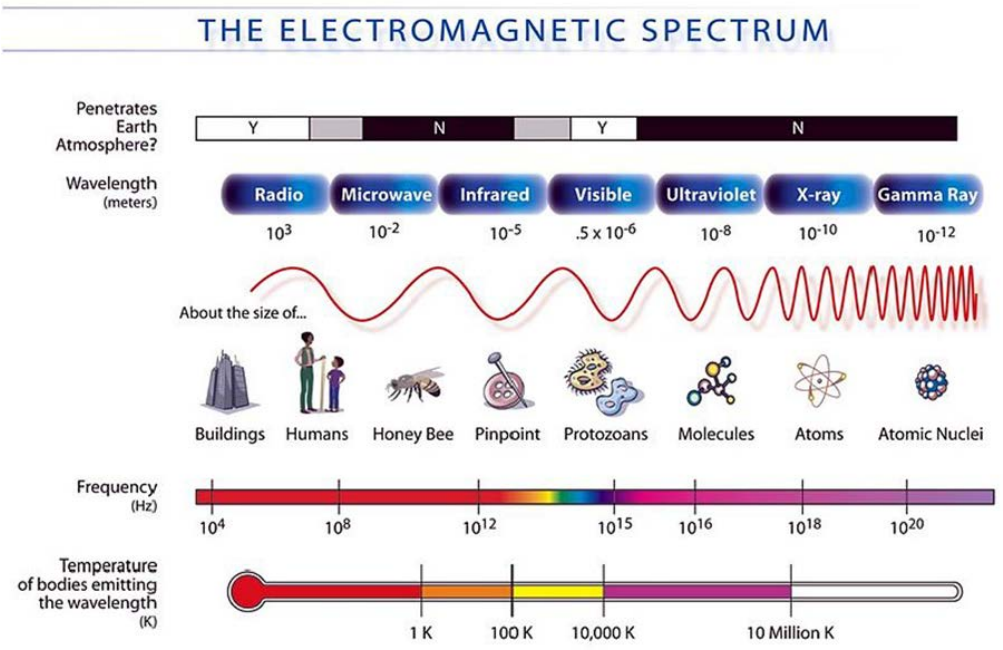
1.9.1 Infrared biospectroscopy

Infrared (IR) radiation was discovered in 1800 by Sir William Hershel and most IR instruments were at that time dependent on prisms or grating monochromators. Albert Michelson developed the interferometer in 1881 and by 1949; Peter Fellgett obtained the first IR spectrum by using a Fourier-transform IR (FTIR) spectrometer. '*Fellgett's advantage*' as it became known, improved the signal-to-noise ratio by means of the simultaneous measurement at different wavelengths (Fellgett, 1949). By the 1960's, FTIR spectrometers became commercially available, and in 1965, an algorithm was developed to quickly compute a Fourier transform (Cooley and Tukey, 1965). Since then, FTIR spectroscopy has greatly improved the quality of IR spectra and reduced the operational time in gathering spectral data. Today, depending upon the machine's settings, a single Fourier-transform spectrum can be obtained in as little as ~30 seconds.

1.9.1.1 Infrared-active and infrared-inactive molecules

Infrared spectroscopy exploits the light properties of a specific part of the electromagnetic spectrum (Figure 15), which is composed of 7 main types of

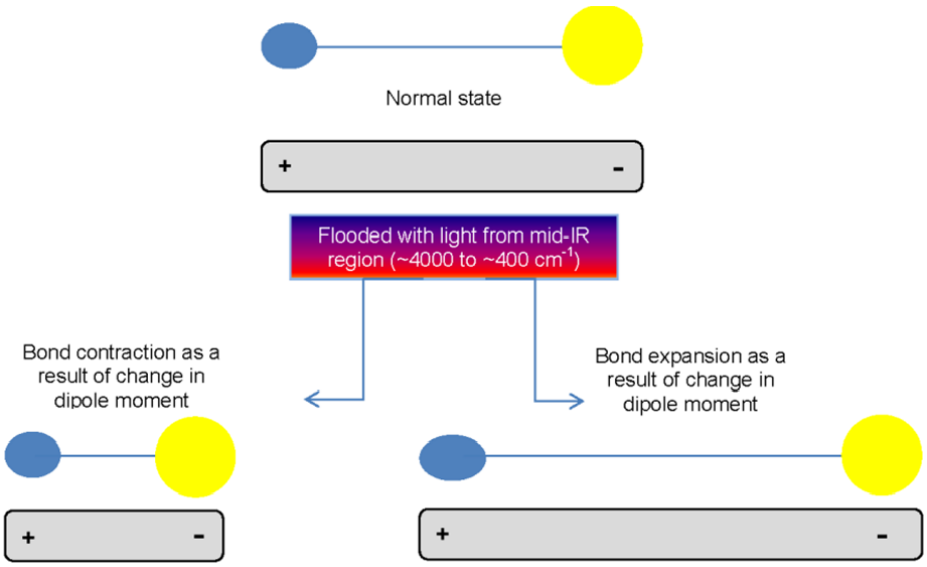
radiation: gamma rays, X-rays, ultraviolet, visible light, infrared, microwaves and radio waves. The IR part of the spectrum is longer in wavelength and lower in frequency than visible light and is divided into 4 main regions: the near IR region (~14286 to ~7143 cm^{-1}); the shortwave IR region (~7143 to 4000 cm^{-1}); the mid IR region (~4000 to 400 cm-1); and the far IR region (< ~400 cm^{-1}).



National Aeronautics and Space Administration (NASA), 2007

Figure 15. The electromagnetic spectrum.

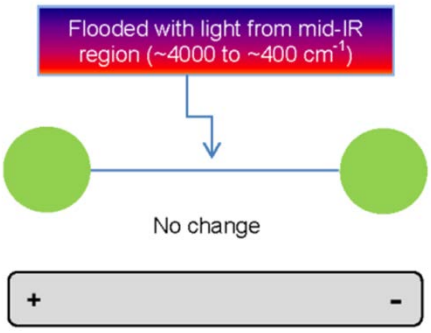
When exposed to frequencies within the mid-IR spectrum (4000-400 cm^{-1}), biological materials vibrate with distinct frequencies that can be detected. For a molecule to be IR-active, a change must occur in the permanent electric dipole moment of the molecule that is non-zero as the bond expands and contracts as result of being flooded with IR light. This is called the 'selection rule' for infrared spectroscopy (Stuart, 2004). Figure 16 illustrates an example of an IR-active heteronuclear molecule.



Halliwell, 2016c

Figure 16. Potential changes in the dipole moments of a heteronuclear molecule.

By contrast, an example of an IR-inactive molecule is a mononuclear diatomic molecule, where molecules are composed of only one type of element such as O_2, and where the bond is symmetrical and IR light has no effect on the dipole moment (Figure 17).



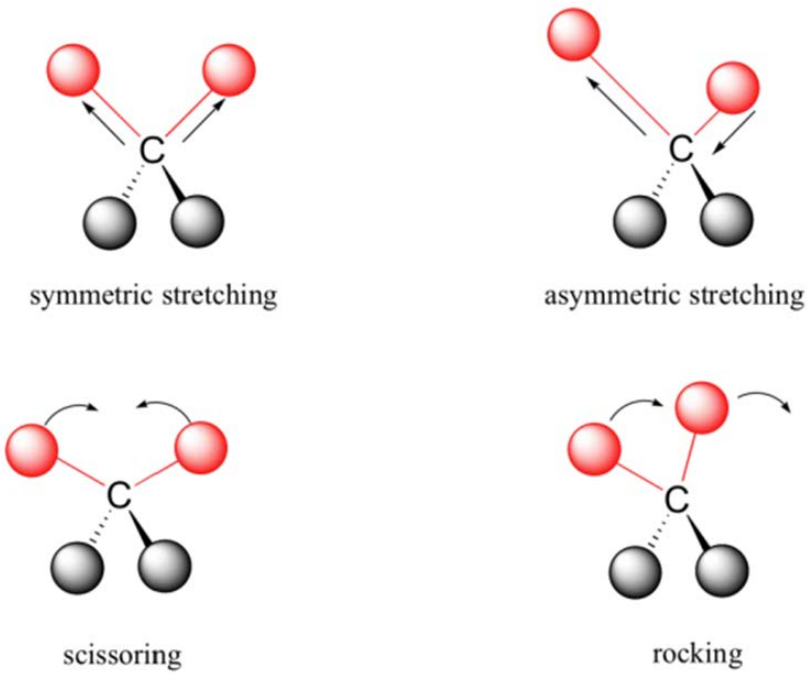
Halliwell, 2016d

Figure 17. An example of an IR-inactive molecule.

This molecule has no permanent dipole, has only a single motion and therefore its dipole moment will always remain at zero no matter how long the bond.

1.9.1.2 Vibrational modes of biological materials

The atoms in a polyatomic molecule are in constant motion causing the molecule to undergo translational and rotational motions, known as ‘normal modes’. Atoms in molecules can also move relative to each other, promoting changes in bond length or move out of their current plane. These stretching and bending movements of molecules irradiated with IR light are collectively referred to as ‘vibrations’ (Stuart, 2004). Some vibrations are due to bonds stretching in-phase (i.e., symmetrical stretching) or out-of-phase (i.e., asymmetrical stretching); other bending vibrations can also occur that add to IR spectra and including rocking and scissoring (Figure 18).



UC Davis ChemWiki, no date

Figure 18. Potential vibrations that can contribute to infrared spectra.

The degrees of freedom for a polyatomic molecule containing many atoms (*N*) are governed by whether it is a linear or non-linear molecule. For both types of molecule, the degrees of freedom describe the translational, rotational, and vibrational motions of the molecule around the three Cartesian axes: *x*, *y*, and *z* (i.e., 3*N* degrees of freedom). A linear molecule can only rotate around 2 of the Cartesian axes whilst a non-linear can rotate around all three. Obtaining the vibrational degrees of freedom for molecules is performed by subtracting the number of translational and rotational degrees of freedom from the number of atoms in the molecules (Table 5).

Types of degrees of freedom	**Linear (Example: hydrogen cyanide; HCN)**	**Non-linear (Example: chloroform; $CHCl_3$)**
Number of atoms	3	5
Translational	3	3
Rotational	2	3
Vibrational	3N - 5	3N - 6
Formula	3N = (3 x 3 – 5)	3N = (3 x 5 – 6)
Total number of vibrational modes	4	9

Table 5. Calculating the degrees of freedom for polyatomic molecules.

A common unit used to present data collected at spectral regions include the 'wavelength' expressed as metres or microns (e.g., 1 μm = 10^{-6} m) and are typically used by physicists. For spectroscopists, the 'wavenumber' is the preferred unit of choice and is the number of waves in a length of one centimetre:

$$\tilde{\nu} = 1/\lambda = \nu/c$$

The wavelength unit (cm^{-1}) is used extensively in the following chapters.

1.9.1.3 Biospectroscopy and cancer studies

The application of Fourier-transform biospectroscopy to cancer cytology and histology is a rapidly expanding area of research (Bellisola and Sorio, 2012; Diem *et al*., 2004). Since the rapid synthesis of proteins, lipids and nucleic acids is a condition for the proliferation of cancer cells (Baenke *et al*, 2013), changes in these biomarkers can be detected by infrared spectroscopy.

The most important spectral regions of biological interest are found within the biochemical 'fingerprint region' (1800 cm^{-1} to 900 cm^{-1}); (Purandare *et al*., 2013), and contain features specific lipids, DNA, polypeptides (as suggested by bonding types amide I and amide II), glycomaterials and collagen (Table 6); (Movasaghi *et al*., 2008). As the bioavailability of these biomarkers increases or decreases, the area under the curve of a peak associated with the relevant wavenumber within a spectrum will increase or decrease accordingly, and can be quantified by IR spectroscopy to reveal important patterns in intracellular change (Griffiths and De Haseth, 2007).

Tentative assignment of biomarkers to wavenumber[a]	Wavenumber (cm-1)
Lipids, fatty acids	~1750
Amide I – predominantly in α helix formation	~1651
Protein Amide II – predominantly β sheet	~1550
Methylene chains in lipids	~1470
Phosphate I - asymmetric (DNA)	~1223
C-O bands from glycomaterials and proteins	~1169
Phosphate I - asymmetric (DNA)	~1072
Glycogen & collagen	~1030

[a] Movasaghi *et al*, 2008; N.B.: other biomarkers may also be associated with wavenumbers.

Table 6. Tentative assignment of important biomarkers.

Of the three main forms of FTIR-spectroscopic sampling modes (transmission, transflection and attenuated total reflectance/reflection [ATR]); (Baker *et al*., 2014), this dissertation focuses on the proven advantages of the application of attenuated total reflectance, Fourier-transform infra-red (ATR-FTIR) spectroscopy to cervical cytology. ATR-FTIR spectroscopy has demonstrated advantages over conventional cervical cytology screening, proving to be an inexpensive and robust method. Previous research has shown this technique is able to segregate grades of cervical cytology (Purandare *et al*., 2014), classify cervical cytology based on HPV infection with low- or high-risk types and according to the patient's age (Lima *et al*., 2014), and diagnose underlying disease more accurately that conventional cytology screening (Gajjar *et al*., 2014).

Other applications of ATR-FTIR include being used to differentiate the different stages of the cell cycle (Stuart, 2004), complement histological classification of skin cancer (Lima *et al*., 2015), detect the recurrence of bladder from bladder wash (Gok *et al*., 2016), and differentiate brain tumours, including severity and organ of metastatic disease, from human serum (Hands *et al*., 2016).

1.9.1.4 Spectral acquisition

Detailed instructions for spectral acquisition are given in Project One, and the laboratory protocol for the preparation of liquid-based cervical cytology samples in provided in Appendix C.

1.9.1.5 Pre-processing of raw spectra

Pre-processing aims to correct problems associated with spectral data acquisition, thereby improving the robustness and accuracy of any subsequent

multivariate analysis by data mining relevant features and removing redundant data (Lasch, 2012). A typical spectrum collected by ATR-FTIR spectroscopy contains 32 scans collecting some 235 data points. The result is a very large array of absorbance which is difficult to work with and extract meaningful outcomes. Following initial visual inspection of the raw spectra to identify errors (e.g., mounting the sample onto the wrong slide of the slide, which will produce spectra specific to glass and show as a distinctive 'quiff' to the far right), feature selection is applied and the spectra are cut to the 'fingerprint region.' This process discards irrelevant data that lies outside the region of interest. A baseline correction is then applied to the cut spectra, followed by a normalisation technique (Table 7).

Baseline correction techniques	Normalisation techniques
Rubber-band correction	Normalise to Amide I/II or other (including Vector)
1st Order Differentiation correction	Normalise to Vector
2nd Order Differentiation correction	Normalise to Vector

Table 7. The options for pre-processing of spectra and application of PCA-LDA following pre-processing.

Rubber-band baseline correction is based on the idea that a spectrum is divided into n ranges (n = number of baseline points) of equal size. The minima y-value of each baseline is determined. Connecting the minima with straight lines creates a baseline. Starting from below the baseline, a rubber-band is stretched over this curve, making it the baseline. Any points that do not lie on the rubber-band are discarded.

The application of differentiation as a baseline correction serves to segregate the gradients of spectral bands, making them more prominent (Trevisan *et al*., 2012). First-order differentiation is the most widely used, although second-order differentiation can also be used. The limitation of this approach is that with each successive order of differentiation, the method augments any spectral noise by one order of magnitude (Griffiths and Haseth, 2007).

Differences in sample thickness or concentration can contribute a significant amount of spectral variation between samples. Normalizing to Amide I, Amide II, etc, forces all spectra to have the same absorbance intensity at the Amide I/II peaks (Baker *et al*., 2014). The disadvantage of normalising this way is that any changes at peaks associated with Amide I/II may be disguised. Vector normalisation (formally referred to as Euclidean), is the preferred choice as it will not disguise changes in biologically important peaks. It is the only available normalisation technique application following 1^{st} and 2^{nd} Order differentiation.

Normalising is critical to allow us to make data comparable. Without it, the data will remain distorted. The technique assists in compensating for differences in sample quality. The choice of baseline and normalisation approaches should be based upon the characteristics of each data set (Trevisan *et al*., 2012) and for this reason, the spectra should be initially explored comparing several techniques to determine the method that best suits the data.

1.9.1.6 Multivariate analysis

Multivariate analysis is the statistical technique used to analyse data that arises from more than one variable. Once the spectral datasets have been baseline corrected and normalised, multivariate analysis techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), can be applied to reduce a large number of data into small data sets or clusters, making them easier to compare.

1.9.1.7 Principal Component Analysis and Linear Discriminant Analysis

PCA is an unsupervised technique commonly used as the first step in analysing large, multivariate data sets. Unsupervised techniques require no information from the user but rely instead on an internal criterion to guide learning. In unsupervised learning, the system forms clusters (groupings, regions of data space). In general terms, PCA reduces the dimensionality of large data sets and using mathematical projection, the original data set which may have involved many variables, can often be interpreted in just a few variables (the Principal Components; PCs). This reduced dimensional data set will allow the user to spot trends, patterns and outliers in the data, far more easily than would have been possible without performing the PCA. When applied to spectra, PCA identifies common sources of variance across spectra and collates them into a small number of dimensions. PCA is often not enough to segregate out data classes or clusters sufficiently, since the heterogeneity within the data is often due to within-class variability rather than between-class variability. Since PCA is unsupervised and has had no instruction from the user, it cannot distinguish between these sorts of variability (Martin *et al*., 2010). By applying a supervised technique such as LDA to the PCA output, it promotes inter-class variation to be identified whilst preventing over-fitting of the data.

LDA is supervised technique because it takes into account the nomination of classes determined by the user. It is a generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition and machine learning, to find a linear combination of features that separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification. LDA works when the measurements made on independent variables for each observation are continuous quantities.

1.9.2 Alternative infrared techniques

1.9.2.1 Scanning near field optical microscopy coupled to free electron laser

Although ATR-FTIR spectroscopy has proven advantages over conventional cervical cytology screening, it is limited in spatial resolution by the effect of diffraction. Diffraction describes the phenomenon of the bending of light waves around obstacles in its path causing it to spread out. This effect restricts the spatial resolution of ATR-FTIR spectroscopy to about half the wavelength of light or ~3 μm to 30 μm (Harrison *et al*., 2013). Resolution is a measure of how closely the lines of an image can be resolved (i.e., the number of independent pixels per value per unit length).

The nanoscopic technique of scanning near field optical microscopy (SNOM) has shown promise in providing detailed information on cell topography and cytoplasmic structures. SNOM has a clear advantage over conventional infrared IR microscopy in terms of spatial resolution because it is able to overcome the diffraction limit, by employing an apertured fibre optic scanning tip. However, the SNOM technique requires relatively high photon intensities such as those provided by an IR free

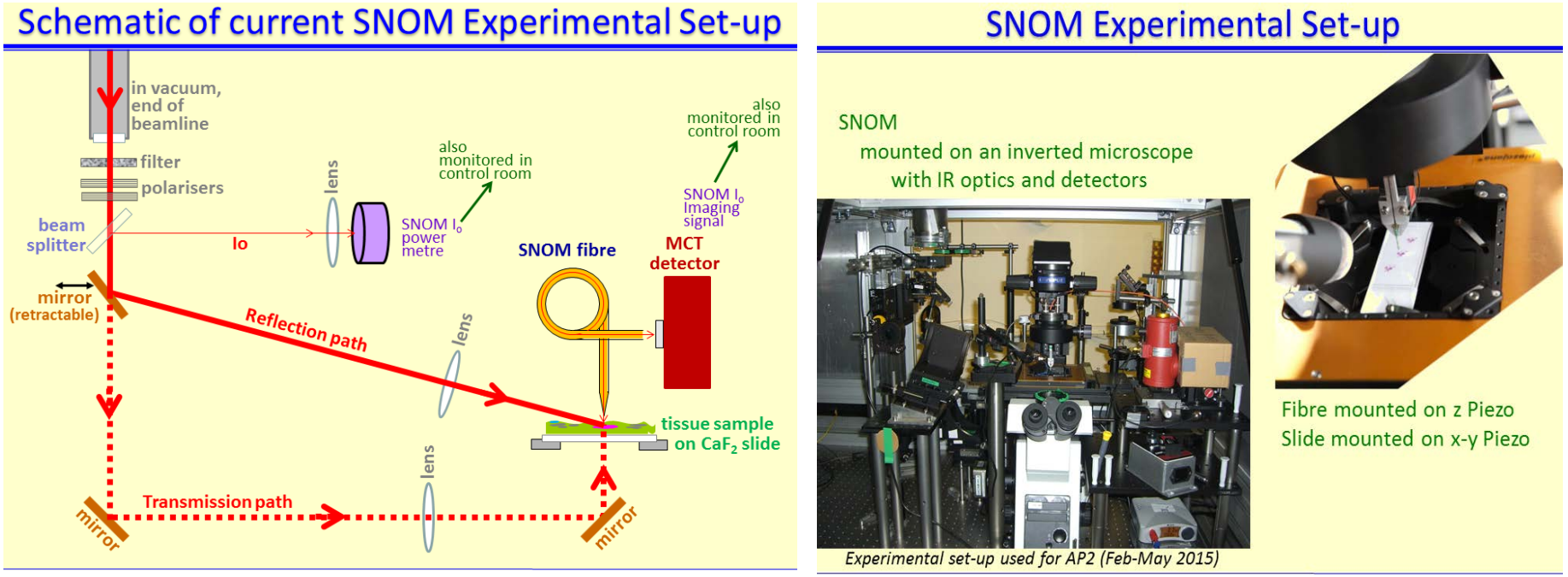
electron laser (IR-FEL) which allows the simultaneous collection of topography and optical features at scales not normally achieved with conventional IR techniques to produce high quality, chemically-rich images at designated wavelengths with a spatial resolution of ≥0.2 μm (Smith *et al*, 2013; Cricenti *et al*., 2002).

The IR-FEL on the ALICE accelerator at Daresbury Laboratory (Warrington, UK) is tuneable over the range of 5.5 μm to 8.8 μm (~1818 cm^{-1} to ~1136 cm^{-1}), which includes a number of biologically important biomarkers as previously mentioned (see Table 6), at designated wavenumbers (Movasaghi *et al*., 2008). Many of these biomarkers have previously been used to separate normal, low- and high-grade dyskaryosis and cancer cells from each other (Gajjar *et al*., 2014).

SNOM has been used to investigate the localisation of molecules within cell membranes of prostate cancer cells (Walker *et al*., 2012). Further research demonstrated SNOM can accurately define both the cell surface and internal structures in both healthy and anomalous sperm, including the acrosome, nucleus and the organisation of mitochondria in the mid-piece region of the tail (Andolfi *et al*., 2015). Furthermore, it has demonstrated the potential for single molecule imaging (Zhong *et al*., 2008).

The application of SNOM to oesophageal cancer tissue studies provided evidence of its potentiality for cancer diagnosis (Smith *et al*, 2013). The increased spatial resolution of SNOM has the potential to reveal and quantify highly localised cancer-related changes in cervical cells at the sub-cellular level (1 - 0.1 μm), and more accurately and precisely than conventional IR techniques. The majority of the above described IR-SNOM-FEL studies were all carried out in reflection mode.

This dissertation reports the collection of data obtained using a SNOM-IR-FEL in transmission mode at the ALICE facility in Daresbury, Warrington, to image whole cervical cells. The experimental period ran from February to May 2015 (Project Two). The set-up of the SNOM-IR-FEL used to collect data for this project is presented in Figure 19.



IR: Infrared; SNOM: Scanning near-field optical microscopy. CaF_2: Calcium fluoride: MCT: Mercury Cadmium Telluride (HgCdTe); SNOM: Scanning near-field optical
microscopy.

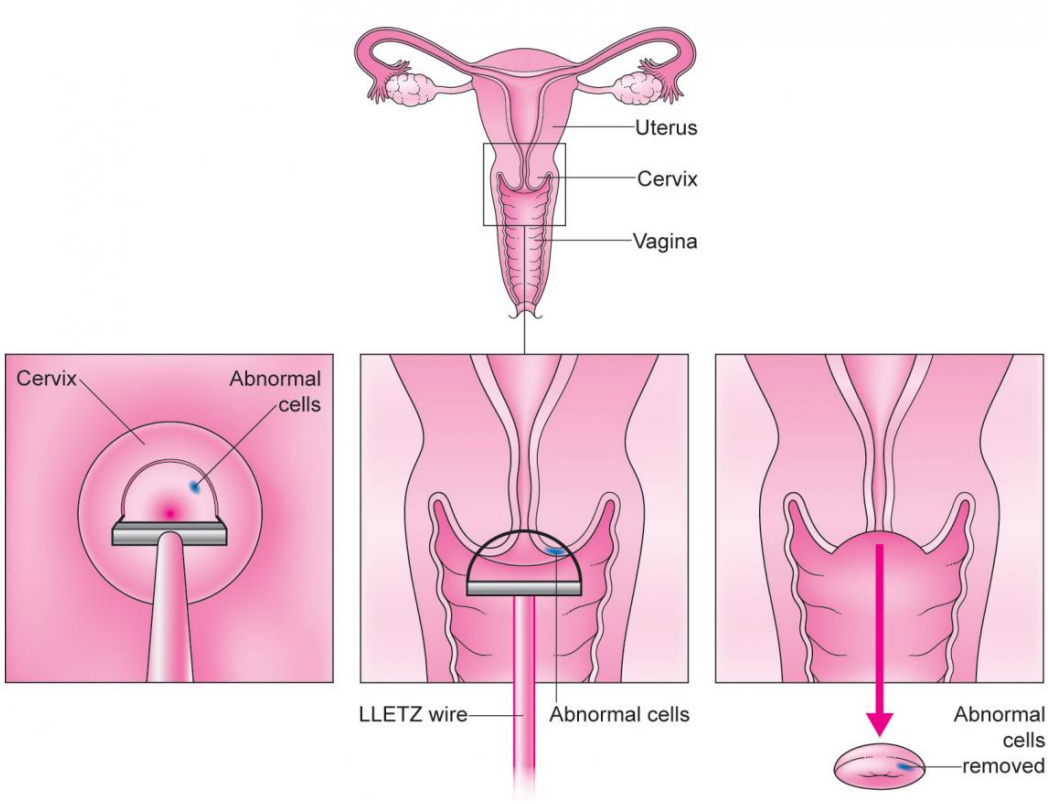
Used with kind permission from Dr M.R.F Siggel-King (Department of Physics, Liverpool University), 2016

Figure 19. Schematic of SNOM-IR-FEL set-up used to conduct Project Two.

1.10 Treatment approaches for pre-malignant CIN

In the UK, the choice of treatment for pre-malignant CIN takes into account both the level of dyskaryosis and the histological classification of CIN (Cancer Research UK, 2014c). Ablative techniques, such as laser therapy and cryotherapy, destroy abnormal cells at the surface of the cervix, allowing normal cells to re-populate the treated area. Excisional treatments to remove the area of the transformation zone that contains abnormal cells that could develop into cervical cancer, include cold knife conisation, laser conisation and large loop excision of the transformation zone (LLETZ). Since LLETZ offers many advantages including low cost, high success rate and ease of use (Carcopino *et al*., 2013), it is the preferred treatment in most countries. Local conservative treatments for pre-invasive CIN are highly effective in preventing recurrent cervical disease and future invasion (95% or better, with the exception of cryotherapy which has a lower clearance rate of 85%); (Wai and Patil, 2008).

The LLETZ technique involves using a wire loop, available in different sizes, through which an electrical current is passed. The loop is applied to the transformation zone of the cervix to a depth of around 7-8 mm and extending 4-5 mm beyond the lesion itself; the transformation zone can vary in size and shape, but for most patients, the entire zone can be removed by a loop 2 cm in diameter and 1.5 cm in depth (Prendiville *et al*., 1989); (Figure 20).



LLETZ: large loop excision of the transformation zone.

Used with kind permission from Jo's Cervical Cancer Trust, 2016

Figure 20. The LLETZ procedure.

However, there is a substantial amount of evidence to suggest that excisional treatment is coupled to a significant risk to future pregnancies, including preterm birth and mid-trimester loss (Kyrgiou *et al*., 2015a; Kyrgiou *et al*., 2015b; Kyrgiou *et al*., 2014; Arbyn *et al*, 2008; Kyrgiou *et al*., 2006). Further research also found that women with a shorter time interval between excision treatment and subsequent pregnancy, have an increased risk of spontaneous abortion (Connor *et al*., 2013).

Removal of part of the cervix may undermine the mechanical integrity of the cervix as a result of collagen disruption (Stout *et al*., 2014; Phadnis *et al*., 2011). Additionally, a function of the endocervical glands is to produce the mucus plug during pregnancy,

which has antimicrobial properties against Gram positive and negative bacteria; thus their removal may damage the host defence mechanism against ascending infections when pregnant (Hein *et al*., 2001). The length of cone removed is directly correlated to the frequency and severity of the adverse events (Kyrgiou *et al*., 2015a).

Pre-treatment cervical length and dimensions vary amongst women (Kyrgiou *et al*., 2015a), as does the does the volume of cervical tissue excised. The assessment of the proportion (percentage) of excised cervical tissue may correlate more accurately to the chances of adverse sequelae than the absolute dimensions (i.e., cone depth and volume), and therefore provide a clinical cut-off that signifies women at risk (Arbyn *et al*., 2014; Khalid *et al*., 2012; Kyrgiou *et al*, 2012., Founta *et al*., 2010).

Given that ATR-FTIR spectroscopy is able to detect underlying disease more accurately than conventional cytology (Lima *et al*., 2014), the 'biochemical fingerprint' produced for normal or dysplastic cells, and cells before and after treatment, may differ based on changes in availability of lipids, proteins, carbohydrates and so on. This may arise from altered biochemical processes as a result of excision or treatment of the disease.

1.11 Aims and objectives

This research project was conducted via two studies using samples from the same cohort of patients. Both studies involved using whole cervical cells collected from women attending colposcopy. The objectives of each study are given below.

1.11.1 Project One objective

The aim of this study was to assess the impact that local treatment may have on the cervix by comparing the spectral absorbance within the 'fingerprint region' of cervical cells before and after treatment, and to evaluate how these changes may be affected by the absolute cone dimensions or the proportion (percentage) of cervix excised.

1.11.2 Project Two objective

The aim of this pilot study was to assess the potential of SNOM in combination with an IR-FEL in the detection of the biophysical properties of cervical cell abnormalities. Spectra were also collected using traditional ATR-FTIR biospectroscopy to investigate the differences between the techniques.

2 Project One

This paper is to be submitted for publication to *Nature Communications* as:

"Tracking the Impact of Excisional Cervical Treatment on the Cervix using Biospectroscopy."

Diane E. Halliwell, Maria Kyrgiou, Anita Mitra, Ilkka Kalliala, Evangelos Paraskevaidis, Georgios Theophilou, Pierre L. Martin-Hirsch and Francis L. Martin.

Contribution:

I collaborated with the co-authors on this project by preparing the samples for spectral acquisition; obtaining the spectra; pre-processing of the spectra and applying multivariate analysis; performing the statistical comparisons of the spectra; extracting the demography of the patient population; and helping to prepare the manuscript for publication.

…… ……

Diane E. Halliwell Professor Francis L. Martin

Nature Communications

Tracking the Impact of Excisional Cervical Treatment on the Cervix using Biospectroscopy

Diane E. Halliwell,[a] *Maria Kyrgiou,[b,c] Anita Mitra,[b,c] Ilkka Kalliala,[b,c] Evangelos Paraskevaidis,[d] Georgios Theophilou,[e] Pierre L. Martin-Hirsch,[f] and *Francis L. Martin[a]

[a] Centre for Biophotonics, LEC, Lancaster University, Lancaster, UK; [b] Institute of Reproductive and Developmental Biology, Department of Surgery & Cancer, Faculty of Medicine, Imperial College, London, UK; [c] West London Gynaecological Cancer Centre, Imperial College NHS Healthcare, London, UK; [d] Department of Obstetrics and Gynaecology, University of Ioannina, Ioannina, Greece; [e] St James Hospital, Leeds, West Yorkshire, UK; [f] Department of Obstetrics and Gynaecology, Lancashire Teaching Hospitals NHS Trust Foundation, Preston, UK.

Correspondence and request for materials should be addressed to:

Dr Maria Kyrgiou, MSc, PhD, MRCOG, 3rd Floor, Institute of Reproductive and Developmental Biology, Department of Surgery and Cancer, Imperial College, Hammersmith Campus, Du Cane Road, W12 0NN, London, UK. Tel: +44 (0)20 7594 2177; Email: m.kyrgiou@imperial.ac.uk.

Professor Francis L Martin, Centre for Biophotonics, Lancaster University, Lancaster, UK. Tel: +44 1524 510206; Email: f.martin@lancaster.ac.uk.

Abstract

Local excisional treatment for cervical intra-epithelial neoplasia (CIN) is linked to significant adverse sequelae including preterm birth, with cone depth and radicality of treatment correlating to the frequency and severity of adverse events. Attenuated total reflectance Fourier-transform infra-red (ATR-FTIR) spectroscopy can detect underlying cervical disease more accurately than conventional cytology. The biochemical 'fingerprint' produced for cells before and after treatment may differ as a result of altered biochemical processes due to excision, or treatment of the disease. Since pre-treatment cervix length vary amongst women, the percentage of cervix excised may correlate more accurately to risk than absolute dimensions. We show that treatment for CIN significantly alters the biochemical fingerprint in the cervix, compared with women who have not had treatment; this is due to the excision of cervical tissue rather than a disease controlling effect. However, the spectra do seem to correlate to the cone depth or proportion of cervical length excised. Future research should aim to explore the impact of treatment on the biochemical 'fingerprint' in a larger cohort.

Keywords: cervical intraepithelial neoplasia, CIN; preterm birth; excisional treatment; LLETZ; biospectroscopy; biochemical fingerprint

The introduction of systematic call and recall screening programs in the UK over the past 20 years has resulted in a profound decrease in the incidence and mortality from invasive cervical cancer, since pre-invasive lesions (cervical intra-epithelial neoplasia; CIN) can be detected by the screening program and treated appropriately[1,2]. The choice of treatment technique varies across countries, although outpatient large loop excision of the transformation zone (LLETZ) is the preferred treatment in most. This preference is because LLETZ offers many advantages including low cost, high success rate and ease of use.[3]

The mean age of women undergoing treatment for pre-invasive cervical disease is similar to the age of women having their first child. Although local conservative treatments for pre-invasive cervical lesions are highly effective in preventing recurrent cervical disease and future invasion[2], there is mounting evidence that excisional treatment is associated with a significant risk to future pregnancies, including preterm birth and mid-trimester loss[4,5,6]. It has been postulated that removal of part of the cervix possibly leads to acquired mechanical weakness as a result of collagen disruption[7,8]. Furthermore, endocervical glands produce the mucus plug during pregnancy which has antimicrobial properties against Gram positive and negative bacteria; thus their removal may damage the host defence mechanism against ascending infections when pregnant[9].

The cone depth and radicality of treatment have been found to be directly correlated to the frequency and severity of the adverse events[10]. Given that the pre-treatment cervix length and dimensions vary amongst women[11], it is

biologically plausible that the proportion (percentage) of the cervix excised may more accurately correlate to the chances of adverse sequelae than the absolute dimensions, and that there may be a cut-off for the proportion of excision that signifies women at risk[12,13,14,15].

Attenuated total reflectance Fourier-transform infra-red (ATR-FTIR) spectroscopy has shown potential in the field of cancer screening, being an inexpensive but robust technique. Previous research demonstrated that the technique is able to segregate grades of cervical cytology[16,17], classify cervical cytology based on HPV infection with low- or high-risk types[18], and diagnose underlying disease more accurately that conventional cytology screening[19]. Infra-red (IR) spectroscopy exploits the molecular vibrations of biologically active molecules that are created by dipole moments as a result of being flooded with IR light. The 'biochemical fingerprint' produced for normal or dysplastic cells, and cells before and after treatment may differ based on changes in availability of lipids, proteins, carbohydrates and so on. This may arise from altered biochemical processes as a result of excision or treatment of the disease.

The aim of this study was to assess the impact that local treatment may have on the cervix by comparing the spectral absorbance within the 'fingerprint region' of cervical cells before and after treatment, and to evaluate how these changes may be affected by the absolute cone dimensions or the proportion (percentage) of cervix excised.

Results

We enrolled a total of 58 women planned to undergo cervical treatment into the study and 27 healthy controls (cytology negative); 20 of these were also HPV test negative. Out of the 58 treated women, 34 women had paired samples before and after treatment and 58 had at least post-treatment samples; 39 out of these 58 women had normal results post-treatment (cytology and HPV DNA test negative). The depth of the excised cone was available for all 58 women who underwent treatment and the proportion of cervical length removed available for 53 of these (91%). The different clinical groups are described in Figure 1.

Out of the 34 women with paired samples before and after treatment (Comparison 1), 3 samples pre- and 1 sample post-treatment did not provide reliable spectra, allowing 29 and 33 samples for analysis, respectively. There were no major differences between the pre- and post-treatment samples. The rate of women who had sexual intercourse less than 48 hours from the sample collection was 90% and 85%, at pre- and post-treatment sampling, respectively ($p=0.71$). None of the patients reported the use of vaginal douching.

Thirty-nine out of 58 women that had at least a post-treatment sample were normal with negative cytology and a negative HPV test post-treatment. The characteristics of this subgroup (n=39) were largely similar to the normal healthy controls (cytology/HPV test negative) that have had no local cervical treatment (n=20); (Comparison 2). The mean age (SD) for treated and control groups was 30.8 years (4.5), and 30.6 years (4.2), respectively. The rate of

women who were current smokers was significantly greater in the treated group (31%), compared with controls (5%); (p=0.04), as was the proportion of women taking combined oral contraceptives (51% vs 15% respectively, p=0.01). Fewer controls were of Caucasian race (55% vs 77%), and a greater proportion of women in the treated group had been sampled in the luteal phase (54%) compared with controls (30%), although neither of these differences however were statistically significant. The proportion of women reporting recent intercourse (75% vs 87%), with bacterial vaginosis diagnoses on high-vaginal swabs (10% vs 8%) and vaginal pH was also comparable between the control and treated group.

A total of 58 women had data on the cone depth and proportion of depth excised; of these 5 were lacking data on the percentage of excision but had data on the depth. Mean depth (SD) of excision for the 58 women was 10.41 mm (3.8 mm); (range 2 – 17 mm), and the mean proportion of excision (SD) in the 53 women for whom this data was available was 29.13% (9.9%); (range 2% – 56%). Seven of the 58 treated women (12%) had involvement of the cone margins at histopathological analysis; 6/58 (10%) with HSIL and 1/58 (2%) with LSIL, however, only two of these 7 patients (3%) had abnormal cytology at 6 month follow-up. The characteristics of this group were overall similar to healthy controls with normal cytology irrespective of HPV status (n=27); (Comparison 3); (Table 1 and Supplementary Table 2).

Once again the proportion of women using combined oral contraceptive pills was significantly greater in treated women (48% vs 15%, p=0.003), and a

higher percentage of women in the treated group had been sampled in the luteal phase (50%) compared with controls (26%), which did not reach statistical significance. The groups were otherwise similar with respect to age, ethnicity, smoking, parity, recent intercourse, vaginal pH, bacterial vaginosis and HPV DNA status.

Treatment for CIN significantly changes the biochemical fingerprint in the cervix, compared with women who have not had treatment. When we compared the absorbance spectra before and after local excisional cervical treatment, we detected a statistically significant difference between the pre- and post-treatment paired samples ($p<0.0001$; 95 CI = -0.17 to -0.08; Fig. 2 [**a**]). A significant positive rate of change was found for absorbance associated with lipids ($p=0.0015$), and for glycomaterials/proteins ($p=0.0006$) for the pre-treatment samples as compared to the post-treatment samples, indicating higher bioavailability in the former group (Fig. 2 [**b**]). A significant positive rate of change was found for absorbance associated with proteins featuring amide I ($p=0.034$) and amide II ($p=0.0004$) type bonding for post-treatment samples as compared the pre-treatment ones, indicating higher bioavailability of polypeptides in the post-treatment group. A significant negative rate of change was detected for absorbance associated with glycogen/collagen ($p=0.0008$) and symmetric phosphate of DNA ($p=0.0001$) in pre- as opposed to the post-treatment group, signifying lower bioavailability in the pre-treatment group. Similarly, absorbance associated asymmetric phosphate for DNA was shown to have a significant negative rate of change for post- as opposed to pre-treatment

samples (p=0.0001); (Fig. 1[**b**]). The steps for pre-processing, multivariate analysis and extraction of wavenumbers are summarized in Supplementary Figures 1, 2 and 3.

Changes in the biochemical fingerprint are due to the excision of cervical tissue rather than the removal of the disease. In order to explore whether the observed differences before and after treatment were attributed to the treatment rather than the removal of a disease, we performed a subgroup analysis comparing all women that had at least one post-treatment sample with negative cytology and HPV DNA test (n=39) versus healthy controls negative for cytology and HPV (n=20). We found that the difference in the spectra post-treatment remained significant, evidencing that the difference observed before and after treatment was due to treatment intervention rather than the treatment of the disease ($p<0.0001$; 95% CI = -0.18 to -0.07; Fig 3 [**a**]). A significant positive rate of change was found for absorbance associated with proteins featuring amide II bonding (p=0.001) in the treated group compared with controls; no other significant changes were detected (Fig. 3 [**b**]).

Spectra do not seem to correlate to the cone depth or proportion of cervical length excised. We further assessed whether the observed difference in the absorbance spectra correlated to the depth of the cone and the proportion of the depth excised in the pre-specified treated groups and compared these to healthy controls that were cytology negative irrespective of HPV status (n=27). We found overall that the spectra of treated women were different to the spectra of healthy controls but this did not seem to correlate to the cone depth and

proportion of cervical length excised. More specifically, we detected statistically significant differences in the spectra when the healthy samples was compared to samples from treated patients with a cone depth of <10 mm (p=0.0008; 95% CI = 0.03 to 0.12), a cone depth of ≥ 15 mm (p=0.001; 95% CI = 0.03 to 0.15), but not for a cone depth between 10-14 mm (Fig. 4 [**a**]). A significant positive rate of change was observed for absorbance associated with proteins featuring amide II bonding for both the <10 mm and ≥ 15 mm groups compared to healthy samples (p=0.004, p=0.0004, respectively). Cone depth ≥ 15 mm also had a significant positive rate of change for absorbance associated with polypeptides featuring amide I bonding (p=0.008), and a significantly negative rate of change for asymmetric phosphate of DNA (p=0.0009) as compared to healthy controls (Fig. 4 [**b**]).

The results were similar when we correlated the spectra to the percentage of excision. The group of women that had <10% of their cervix excised included only 2 patients and was therefore excluded from the analysis. For all the remaining groups, we detected significant differences as compared with healthy cervix [11-20% (p=0.002, mean rank difference = 31.11); 21-30% (p=0.03, mean rank difference = 17.85); 31-40% (p=0.007, mean rank difference = 23.33); and for >40% (p=0.023, mean rank difference = 27.24)]; (Fig. 5 [**a**]). A significant positive rate of change in absorbance associated with biomarkers was found only for 11-20% compared with healthy cervix (lipids: p=0.008) and polypeptides featuring amide II bonding (p=0.003); (Fig. 5 [**b**]).

The in-between group comparisons of the absorbance spectra demonstrated significant differences for most comparisons, apart from, rather surprisingly, the comparison of the most extreme values (<10 mm with >=15 mm). Most of the comparisons between the clinical groups of the proportional cervical length excised did not show a significant difference. (Supplementary Figures 4 and 5).

The high variability of polypeptides featuring amide I bonding appeared to be consistent across untreated, treated and healthy controls, suggesting it is unlikely to be due to human sampling error (either through collection of liquid-based cytological samples or acquisition of spectra), and may suggest that synthesis of these molecules is independent of disease status.

Discussion

Local conservative treatment for CIN has been associated with significant adverse sequelae in subsequent pregnancies[4,5,6]. The frequency and severity of these effects seem to correlate directly to the radicality and depth of the treatment[10].

The analysis of the biochemical fingerprints obtained from samples collected from the cervix before and 6 months post-treatment revealed that excision of cervical tissue and endocervical glands impact on the absorbance spectra that appeared to be significantly different after treatment. The significant increase in polypeptides in the post-treatment group, as evidenced by amide I/II bonding, may be a response to localized injury caused by the excisional treatment. Normal wound healing is typified by three overlapping phases; the inflammatory phase, the proliferation phase and the maturation phase[20,21]. At

around 30 days following the initial injury and when the wound is closed, the maturation phase begins and cellular activity diminishes, the number of blood vessels regress and collagen is remodelled from type III to type 1. Although previous work has shown that cervical regeneration is almost complete 6 months after excisional treatment[22], our findings suggests that cellular function remains elevated at this time point. The presence of scar tissue at the site of injury may account for the increased polypeptides, since collagen is composed of three separate polypeptide chains[23], and production is increased 2-3 times more in fibroblasts isolated from scar tissue than from normal tissue[24]. Despite this, a healed wound will only achieve a maximum of 80% of the tensile strength of normal epithelium. Further follow up may have shown the differences resolve with time or, the distinction is a true reflection of the newly formed epithelium capping the original wound. Additionally, the number of endocervical cells collected in post-LLETZ liquid-based cytology samples has been shown to be significantly decreased[25]. Therefore, any changes in biochemical function following treatment are likely to be a reflection of the cellular function largely associated with squamous cells.

When the spectra in healthy controls post-treatment were compared to healthy untreated controls, the differences remained significant, suggesting that it is the treatment rather than the excision of the disease that alters the biochemical balance within the cervix. Cervical regeneration is dependent upon the depth of excision, the percentage of cervix excised and/or the remaining cervical tissue immediately after treatment[26]. Our findings show that patients

who were classified as cytologically free of cervical abnormality and HPV negative, remained biochemically distinct from healthy controls, with an increase in polypeptides (amide II bonding) remaining elevated in the treated group, which may be due to the previously hypothesized causes.

When the absorbance spectra were assessed for different treatment cone depths and proportions of cervical lengths excised, we found no direct influence of the different clinical groups on the spectra, although the number of samples was small in each group.

The mechanism behind the high variability associated with the production of polypeptides that appears to be independent of disease status is unknown, and may be a reflection genetic variation or other patient-specific characteristics.

Several reports have assessed the impact of cervical excision on subsequent clinical reproductive outcomes; others suggested that it is the presence of the pre-cancer itself that also contributes to the adverse sequelae[27]. It is likely that excision of cervical tissue causes a disruption of the immune defence mechanisms, the natural production of antimicrobials and the mucus secretion from endocervical glands. These changes together with alterations caused by the disease are likely to interact with genetic, viral and microbial factors in a complex interplay within the vagina[28,29].

This is the first report that assessed the impact that local excisional treatment for cervical pre-invasive disease has on the biochemical fingerprint and molecular processes within the cervix. We were able to show that excision causes major alterations in the cervico-vaginal environment and further research

should establish the pathophysiological processes that may be correlated to adverse obstetric sequelae. We did not identify direct correlation of the changes to the radicality of the excision measured by cone depth and cervical proportion excised, but the numbers in each group were small and the results should be interpreted with caution. Further research should explore in more detail the impact of the severity/grade of CIN and/or presence of HPV infection on the biochemical spectra and should further describe the biochemical alterations in treated individuals with or without positive cytology and/or HPV infection post-treatment in larger cohort. This will allow a more comprehensive exploration of the impact of treatment; this was not feasible in this analysis as the number of samples in these subgroups was small for any valid comparisons.

The mechanism that accounts for the increased risk of second trimester loss and preterm birth associated with CIN and its treatment is not yet clarified. While acquired mechanical weakness of the cervix secondary to surgery might seem a logical assumption, more subtle mechanisms may be involved. Histological changes in the healed cervix affecting the tensile strength or changes in the innate immune system and vaginal microenvironment may also be involved. Removing part of the cervix or simply being infected with human papillomavirus (HPV) may impair the host's defence mechanisms, change the chemical microenvironment and prevent a pregnancy being maintained to full term. Conversely, it may be that women at risk have intrinsic compromised defences that promote the persistence of oncogenic HPV infections and the development of ascending infections during pregnancy. All these interactions

may be best described in the absorbance spectra of the metabolic changes in the vagina as described in this study.

Ascending infection from the vagina into the fetoplacental unit and associated inflammation are presumed to be causative in preterm labour. The uterus in pregnancy is protected by the cervix via its mucous plug, the local synthesis of antimicrobial peptides and proteins and by a 'benign' *Lactobacillus*-dominated vaginal microbiota. *Lactobacilli* spp. inhibits pathogen growth by maintaining a hostile pH and secreting species-specific metabolites and bacteriocins that limit the growth of other organisms[30]. By causing scarring and a change in the histological structure of the cervix, it is likely that the protein expression and metabolites of the cervix are also altered after treatment, resulting in an environment, which is more or less hospitable to particular bacterial species. It remains unknown whether these changes underlie the increased risk of adverse obstetric sequelae, but are a priority area of research. FTIR spectroscopy is capable of detecting alterations in the microbiota composition[31], and thus spectral information should be correlated with 16s ribosomal RNA gene sequencing data, as well as proteomic data in future prospective studies to facilitate a better understanding of the structural, biochemical, metabolic and microbiological changes that result from excisional treatment for CIN.

In conclusion, this study clearly demonstrates that local treatment for cervical pre-invasive disease has a direct impact on the biological and biochemical processes within the cervix, and this may correlate to the adverse sequelae

described in future pregnancies that include preterm birth and premature rupture of the membranes, possibly as a results of ascending infections and disruption of the immune defence mechanisms. Correlation of the specific biochemical markers in the produced spectra with the outcomes of subsequent pregnancies in the future may allow the detection of women at high risk of preterm birth and enable the selection of women at high risk that would benefit from intensive antenatal surveillance when pregnant[32]. Furthermore, further exploration of the mechanistic aspects leading to these changes in the metabolic spectra may permit the use of more targeted cause-directed preventative treatment in the future.

Materials and methods

Study population – Inclusion and Exclusion criteria

Ethical approval was obtained from the National Research Ethics Service Committee London – Fulham (Approval number 13/LO/0126). This study was conducted according to the principles of the Declaration of Helsinki and all other applicable national or local laws and regulations. All patients gave written informed consent before any protocol-specific procedure was performed.

We included pre-menopausal, non-pregnant women of reproductive age (18-45 years of age) who attended the colposcopy and were planned to undergo local cervical treatment at Imperial College NHS Healthcare Trust. We collected samples before treatment, and a repeat sample 6 months after treatment. We also recruited a population of women with normal cytology (+/- negative HPV DNA test) attending the colposcopy or general gynaecology

clinics that would serve as healthy controls. The samples for the healthy group were collected at one time point. The recruitment commenced in May 2013 and was completed in May 2015.

Women were included irrespective of their ethnicity, parity, smoking habits, phase in their cycle and use of contraception. The type of contraception and the time of their cycle (follicular or luteal) were documented. Women who were HIV or hepatitis B/C positive, women with autoimmune disorders, and women that received antibiotics or pessaries within 14 days of sampling were excluded. Detailed medical and gynaecological history was collected for each patient including time since last sexual intercourse. Ethnicity was self-reported as Caucasian, Asian or Black.

Patients were anonymized and assigned a unique identifier. For each patient and visit, we collected data on the cytology, HPV DNA test and typing and histology, if available. The cytology result was classified as normal, borderline or mild dyskaryosis (low-grade squamous intraepithelial lesion [LSIL]), moderate or severe dyskaryosis (high-grade squamous intraepithelial lesions [HSIL]) and invasive cervical cancer (ICC). The histology was defined as normal, CIN1, CIN2, CIN3 or invasive cervical cancer.

Transvaginal ultrasound (Voluson E6 with a 5-9 MHz(RIC5-9-RS series) transvaginal probe (GE Healthcare, Zipf, Austria) was used to measure the cervical length and volume immediately prior to excision. The dimensions of the cervical cone were measured using electronic callipers. The volume was also determined by water displacement, using a 50ml syringe. These measurements were taken prior to fixation

in formaldehyde, which may result in a degree of sample shrinkage, and thus underestimate the size of the excised specimen. These data were used to calculate the proportion (percentage) of the length or volume excised. Patients included in the analysis of cone depth were categorized according to 1 of 4 categories (untreated healthy cervix; treated/cone depth: 1 = <10 mm, 2 = 10-14 mm, 3 = ≥15 mm). Patients included in the analysis of percentage excision were categorized according to 1 of 6 categories (untreated healthy cervix; treated/percentage excision: 1 = 0-10%, 2 = 11-20%, 3 = 21-30%, 4 = 31-40%, 5 = >40%).

Sample collection and processing

A sterile, disposable speculum was inserted, without lubricant, and a cervical sample of ThinPrep, liquid-based cytology (LBC) was taken from the cervix (ThinPrep, HOLOGIC Inc., Bedford, USA). This was analysed for cytological diagnosis and HPV DNA test and typing. HPV DNA test and 16/18 genotyping was carried out according to manufacturer's guidelines using the Abbott RealTime High Risk (HR) HPV assay on Abbott M2000 platform; a clinically validated in vitro polymerase chain reaction (PCR) assay with identification of HPV-16, -18 or any other of 12 HR HPV subtypes (31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68)[33]. From the remaining methanol-based fluid, 1 ml was stored at 4° centigrade and was used for biospectroscopy analysis at the Centre for Biophotonics, Lancaster University, England. Routine high vaginal microbiology swabs were taken and sent for microscopy, culture and sensitivity. These swabs were also used to diagnose bacterial vaginosis based on the Hay/Ison criteria[34]. Vaginal pH was measured using CarePlan VpH gloves (Inverness Medical, Unipath Ltd., Bedford, UK).

Biospectroscopy: slide preparation

Each sample was agitated to disperse the cell pellet, then a 500µl aliquot collected into a clean micro tube. Samples were centrifuged at 2000 rpm for 5 minutes and the ThinPrep supernatant was aspirated to remove its spectral signature (i.e., the methanol fixative). Each sample was then immersed in 500 µl of distilled H2O, agitated and centrifuged again. The supernatant was removed again and the wash step repeated once more. The final pellet was immersed in 100 µl of distilled H2O, agitated and dispensed onto IR-reflective glass slides (Low-E; Kevley Technologies Inc., Chesterland, OH, USA) and allowed to bench dry for a minimum of 24 hours. Samples were then stored in a desiccator for a minimum of 48 hours to remove any residual water before spectral analysis. For those samples with a resultant poor spectral signal considered to be due to fewer cells on the slide, repeat samples were prepared using the remaining 500 µl of the original sample and prepared as described above. However, the final 100 µl solution was dispensed as 2 x 50 µl aliquots, with a 24-hour drying period in between each dispensing, and each aliquot being dispensed directly on top of the previous one to achieve a uniform spread and thickness of cells. These samples were then desiccated as described above.

ATR-FTIR spectroscopy, computational and statistical analyses

Spectral acquisition

Spectra were acquired using a Tensor 27 FTIR spectrometer with a Helios ATR attachment (Bruker Optik GmbH). The instrument was set to take 32 scans at 8 cm^{-1} wavenumber spacing with 2 x interferogram zero-filling. Before the

spectra were taken, the crystal was cleaned with distilled H_2O and inspected by video camera to be free of any contaminants. A background spectrum was acquired before the sample slide was mounted and stage moved to bring the cervical cells in contact with the diamond. Spectra were collected from ten random sites on the slide. Spectra were converted to absorbance by Bruker OPUS software (Bruker Inc., Billerica, MA, USA).

Pre-processing and multivariate analysis of spectra

Each spectrum was classed, cut to the 'fingerprint region' (1800-900 cm^{-1}), Savitsky Golay differentiated to 1st order and vector normalized. Combining vector normalization with differentiation is a typical pre-processing approach[35]. This was followed by Principal Component Analysis coupled to Linear Discriminant Analysis (PCA-LDA) cascade using MATLAB R2014a software (Mathworks Inc., Natick, MA, USA) together with the toolbox 'IRootLab' (http://trevisanj.github.io/irootlab/). The number of principal components (n_{PCS}) for PCA was calculated to guarantee a ratio of 20 between the number of spectra and the number of variables inputted into LDA, to avoid potential arbitrary class separation. Typically, the number of PCs was either 9 or 10. The PCA-LDA output provided a 'feature' for each spectrum for each patient (i.e., 10 for each patient). An average of these 10 features was taken and used to conduct the statistical analyses.

Peaks found within the 'biochemical fingerprint' region contain features specific to biomarkers as shown in Supplementary Table 1[36]. Changes in the availability of these biomarkers will result in changes in peaks (absorbance),

which can reveal patterns of intracellular change. We therefore extracted 7 wavenumbers associated with biologically important biomarkers for each analysis, using means, SD and multiple *t* tests (corrected for multiple *t* testing using the Holm-Sidak method) to produce a rate of change for each wavenumber. These were extracted from pre-processed data (i.e., cut, 1st order differentiated, vector normalized) at specified wavelengths within MATLAB.

Impact of treatment

Demographic data were summarized using RStudio Version 3.2.1 (RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, USA). Differences in the patient characteristics between the compared groups were assessed using Fishers exact test (GraphPad Prism 6, [GraphPad Software Inc., La Jolla, CA, USA]). P-value <0.05 was considered to be statistically significant.

We assessed the impact that local treatment may have on the cervix by comparing the cervical cell spectral absorbance of women with paired samples before and after treatment using means, standard deviations (SD) and a Student's *t* test (Comparison 1). In order to control whether the observed differences were a result of the treatment itself or due to the removal of the disease (CIN), we performed an additional subgroup analysis to compare normal women post-treatment (negative cytology and negative HPV DNA test), to untreated normal controls (negative cytology and negative HPV DNA test). This was also done using means, SD and a Student's *t* test (Comparison 2). We further assessed whether the changes in the biochemical fingerprint of the cervix correlated to the absolute cone depth (mm) and the proportion

(percentage) of cervical length excised. We compared the different cone depth/proportions groups and also compared these with healthy controls with negative cytology irrespective of HPV status by means, SD and one-way ANOVA (Comparison 3). All statistical analyses of spectra were conducted using GraphPad Prism 6 (GraphPad Software Inc., La Jolla, CA, USA).

Acknowledgements

We thank all the participants of the study. This work was supported by the British Society of Colposcopy Cervical Pathology Jordan/Singer Award (P47773)(MK); the Imperial College Healthcare Charity (P47907) (AM, MK); Genesis Research Trust (P55549)(MK); Sigrid Juselius Fellowship (P52483) (IK, MK) and the Imperial Healthcare NHS Trust NIHR Biomedical Research Centre (P45272) (MK). None of the funders have had any influence on the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Author Contributions

The study was conceived and designed by MK and FM. The data was acquired and collated by MK, AM, IK, GT, DH and analysed by DH, MK and AM. The manuscript was drafted and revised critically for important intellectual content by all authors (DH, MK, AM, IK, EP, PMH, FM). DH and MK are joint first authors. MK and FM are joint corresponding authors. All authors gave final approval of the version to be published and have contributed to the manuscript.

Additional information

Supplementary Information accompanies this paper.

Declaration of competing interests

The authors declare no competing financial interests.

References

1. Peto, J., Gilham, C., Fletcher, O., & Matthews FE. The cervical epidemic that screening has prevented in the UK. *Lancet.* **364**, 249-256 (2004).
2. Wai, T.T., & Patil, D. Modern management of abnormal cervical smear. *BJMP*. **1**, 18-22 (2008).
3. Carcopino, X. *et al*. Direct colposcopic vision used with the LLETZ procedure for optimal treatment of CIN: results of joint cohort studies. *Arch. Gynecol. Obstet.* **288,** 1087-1094 (2013).
4. Kyrgiou, M., Koliopoulos, G., Martin-Hirsch, P., Arbyn, M., Prendiville, W., & Paraskevaidis, E. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. *Lancet.* **367,** 489-498 (2006).
5. Kyrgiou, M. *et al*. Fertility and early pregnancy outcomes after treatment for cervical intraepithelial neoplasia: systemic review and meta-analysis. *BMJ.* **349,** g6192 (2014). doi:10.1136/bmj.g6192.
6. Kyrgiou, M. *et al*. Fertility and early pregnancy outcomes after treatment for cervical intraepithelial neoplasia. *Cochrane Database Syst Rev*. **9,** CD008478 (2015). doi:10.1002/14651858.CD008478.pub2.
7. Stout, M.J. *et al*. Loop electrosurgical excision procedure and risk of vaginal infections during pregnancy: an observational study. *BJOG*. **122,** 545-551 (2014). doi:10.1111/1471-0528.13252.

8. Phadnis, S.V. *et al*. Regeneration of cervix after excisional treatment for cervical intraepithelial neoplasia: a study of collagen distribution. *BJOG*. **13,** 1585-1591 (2011). doi:10.1111/j.1471-0528.2011.03085.x
9. Hein, M., Helmig, R.B., Schønheyder, H.C., Ganz, T., & Uldbjerg, N. An in vitro study of antimicrobial properties of the cervical mucus plug in pregnancy. *Am J Obstet Gynecol*. **185,** 586-592 (2001).
10. Arbyn, M. *et al*. Perinatal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: meta-analysis. *BMJ*. **337,** a1284. (2008). doi:10.1136/bmj.a1284.
11. Kyrgiou, M. *et al*. Proportion of cervical excision for cervical intraepithelial neoplasia as a predictor of pregnancy outcomes. *Int J Gynaecol Obstet*. 128, 141-147 (2015). doi:10.1016/j.ijgo.2014.07.038.
12. Founta, C. *et al*. Proportion of excision and cervical healing after large loop excision of the transformation zone for cervical intraepithelial neoplasia. *BJOG*. **117,** 1468-1474 (2010). doi:10.1111/j.1471-0528.2010.02709.x.
13. Khalid, S. *et al*. The thickness and volume of LLETZ specimens can predict the relative risk of pregnancy-related morbidity. *BJOG*. **119,** 685-681 (2012). doi:10.1111/j.1471-0528.2011.03252.x.
14. Kyrgiou, M., Arbyn, M., Martin-Hirsch, P., & Paraskevaidis, E. Increased risk of preterm birth after treatment for CIN. *BMJ*. **345,** e5847 (2012). doi:10.1136/bmj.e5847.

15. Arbyn, M., Kyrgiou, M., Gondry, J., Petry, K.U., & Paraskevaidis, E. Long term outcomes for women treated for cervical precancer. *BMJ.* **348,** f7700 (2014). doi:10.1136/bmj.f7700.
16. Purandare, N.C. *et al*. Biospectroscopy insights into the multi-stage process of cervical cancer development: probing for spectral biomarkers in cytology to distinguish grades. *Analyst.* **138,** 3909-3916 (2013). doi:10.1039/c3an36527a.
17. Purandare, N.C. *et al*. Infrared spectroscopy with multivariate analysis segregates low-grade cervical cytology based on likelihood to regress, remain static or progress. *Anal. Methods*. **6,** 4576-4584 (2014). doi:10.1039/c3ay42224k.
18. Lima, K.M.G. *et al*. Classification of cervical cytology for human papilloma virus (HPV) infection using biospectroscopy and variable selection techniques. *Anal. Methods.* **6,** 9643-9652 (2014). doi:10.1039/c4ay01736f.
19. Gajjar, K. *et al*. 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,** e82416:1 (2014) doi:10.1371/journal.pone.0082416.
20. Leaper, D.J., & Harding, K.G. Wounds: Biology and Management. Oxford (UK): Oxford University Press (1998).
21. Hutchinson, J. The Wound Programme. Dundee (UK): Centre for Medical Education (1992).
22. Paraskevaidis, E. *et al*. Cervical regeneration after diathermy excision of cervical neoplasia as assessed by transvaginal sonography. *Eur. J. Obstet. Gynecol Reprod. Biol.* **102,** 88-91 (2002).

23. Flynn, M.E., & Rovee, D.T. Wound Healing Mechanisms. *The American Journal of Nursing.* **82,** 1544-1549 (1982).
24. Diegelmann, R.F., & Evans, M.C. Wound healing: an overview of acute, fibrotic and delayed healing. *Frontiers in Bioscience*. **9,** 283-289 (2004).
25. Maguire, A., Turner, L., Magee, D., & Gibbons, D. Decrease in numbers of glandular cell groups in post-LLETZ liquid-based cytology samples. *Cytopathology*. **19,** 44-47 (2008).
26. Papoutsis, D., Rodolakis, A., Mesogitis, S., Sotiropoulou, M., & Antsalis, A. Regeneration of uterine cervix at 6 months after large loop excision of the transformation zone for cervical intraepithelial neoplasia. *BJOG.* **119,** 678-684 (2012). doi:10.1111/j.1471-0528.2012.03275.x.
27. Bruinsma, F., Lumley, J., Tan, J., & Quinn, M. Precancerous changes in the cervix and risk of subsequent preterm birth. *BJOG.* **114,** 70-80 (2007).
28. Mitra, A. *et al*. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci Rep*. **5,** 16865 (2015). doi: 10.1038/srep16865.
29. Kyrgiou, M., Mitra, A., & Moscicki, A-B. Does the vaginal microbiota play a role in the development of cervical cancer? *Tran Res (in press)*. (2016)
30. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature.* **486,** 207-214 (2012). doi:10.1038/nature11234.
31. Daniel, H., *et al*. High-fat diet alters gut microbiota physiology in mice. *ISME J*. **8,** 295-308 (2014). doi: 10.1038/ismej.2013.155.

32. Kindinger, L.M. *et al*. Preterm birth prevention post-conization: a model of cervical length screening with targeted cerclage. *Plos ONE (in press).* (2016).
33. Coutlée, F., Rouleau, D., Ferenczy, A., & Franco, E. The laboratory diagnosis of genital human papillomavirus infections. *Can J Infect Dis Med Microbiol*. **16,** 83-91 (2005).
34. Ison, C.A., & Hay, P.E. Validation of a simplified grading of Gram stained vaginal smears for use in genitourinary medicine clinics. *Sex Transm Infect*. **78,** 413-415 (2002).
35. Baker, M.J. *et al*. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat Protoc.* **9,** 1771-1791 (2014). doi:10.1038/nprot.2014.110.
36. Movasaghi, Z., Rehman, S., & ur Rehman, I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Appl. Spectosc. Rev*. **43,** 134-179 (2008). doi:10.1080/05704920701829043.q.

Figure legends

Figure 1: Flowchart of the included population and different comparison groups.

Figure 2. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1: Pre- vs Post-treatment (a) together with absorbance per wavenumber (b). The paired samples pre- and post-treatment were significantly different along LD1 (Mean/SD (**a**): 0.004/0.10 for 'A'; -0.12/0.08 for 'B'; p<0.0001, 95% CI = -0.17 to -0.08). Absorbance associated with lipids, glycomaterials and proteins was shown to have a significant positive rate of change for the pre-treatment group compared with the post-treatment group, indicating their higher bioavailability. Similarly, absorbance associated amide I and amide II was shown to have a significant positive rate of change for the post-treatment group compared within the pre-treatment group. Absorbance associated with glycogen, collagen and symmetric phosphate of DNA was shown to have a significant negative rate of change for the pre-treatment group compared with the post-treatment group, suggesting lower bioavailability. Similarly, absorbance associated asymmetric phosphate for DNA was shown to have a significant negative rate of change for the post-treatment group compared with the pre-treatment group (**b**). ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared; CI: Confidence interval; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis coupled to Linear Discriminant Analysis; SD: Standard deviation.

Figure 3. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1: controls vs treated (normal cytology and HPV –ve); (a) together with absorbance per wavenumber (b). The 2 groups were significantly different along LD1 (Mean/SD (**a**): 0.32/0.10 for Controls; 0.20/0.11 for Treated; p<0.0001, 95% CI = -0.18 to -0.07). These results evidence that the difference in LD1 was due to the impact of treatment. Absorbance associated with amide II was shown to have a significant positive rate of change for the treated group compared with controls, indicating higher bioavailability (**b**). No other significant changes were detected. ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared; CI: Confidence interval; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis coupled to Linear Discriminant Analysis; SD: Standard deviation.

Figure 4. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1: Healthy Cervix vs Cone Depth (a) together with absorbance per wavenumber (b). Mean/SD (**a**) for each group was: 0.22/0.07 for healthy cervix; 0.15/0.07 for <10 mm; 0.19/0.07 for 10-14 mm; 0.13/0.04 for ≥15 mm. A significant difference along LD1 was detected for healthy cervix vs <10 mm (p=0.0008; 95 CI = 0.03 to 0.12); and for healthy cervix *vs* ≥15 mm (p=0.001; 95 CI = 0.03 to 0.15). No significant difference along LD1 was detected for healthy cervix vs 10-14 mm (p=0.13; 95 CI = -0.01 to 0.08). Absorbance associated with amide II was shown to have a significant positive rate

of change for <10 mm group compared with healthy cervix, indicating higher bioavailability (**b**). Similarly, absorbance associated with amides I and II were shown to have a significant positive rate of change for ≥15 mm compared with healthy cervix, whilst absorbance associated with DNA was shown to have a negative rate of change, indicating lower bioavailability. ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared; CI: Confidence interval; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis coupled to Linear Discriminant Analysis; SD: Standard deviation.

Figure 5. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1: Healthy Cervix vs Percentage Excision (a) together with absorbance per wavenumber (b). Mean/SD (**a**) for each group was: 0.30/0.07 for healthy cervix; 0.16/0.12 for 11-20%; 0.24/0.06 for 21-30%; 0.22/0.09 for 31-40%; 0.21/0.07 for >40%. A significant difference along LD1 was detected for healthy cervix vs 11-20% (p=0.002; mean rank difference: 31.11); for healthy cervix vs 21-30% (p=0.03; mean rank difference: 17.85); for healthy cervix vs 31-40% (p=0.007; mean rank difference 23.33; and for healthy cervix vs >40% (p=0.023; mean rank difference: 27.24). Absorbance associated with lipids and amide II was shown to have a significant positive rate of change for 11-20% compared with healthy cervix, indicating higher bioavailability (**b**). No other significant differences were detected for healthy cervix vs all other groups. ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared; CI: Confidence interval; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis coupled to Linear Discriminant Analysis; SD: Standard deviation.

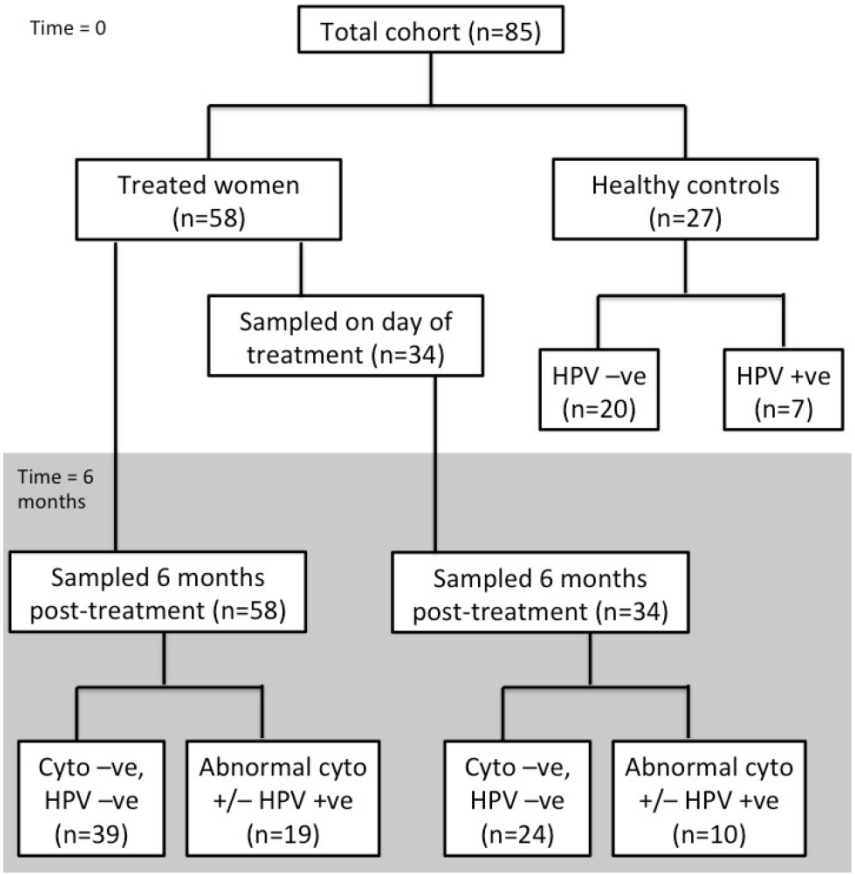
Table 1. Patient characteristics

	Comparison 1: Treated women with paired samples			Comparison 2 : Normal post-treatment vs. normal untreated controls			Comparison 3: Treated by cone depth/proportion vs. normal untreated controls		
Characteristics	**Pre-treatment (n= 29)**	**Post treatment (n= 33)**	***P-value***	**Normal post-treatment, (Cytology negative, HPV negative) (n = 39)**	**Normal controls (Cytology negative, HPV negative) (n=20)**	***P-value***	**Post treatment (n=58)**	**Controls (Cytology negative; HPV status ignored) (n=27)**	***P-value***
Age, years			0.69			0.87			0.50
Mean (SD, range)	30.3 (4.9, 25-43)	30.8 (5.0, 25-43)		30.8 (4.5, 25-43)	30.6 (4.2, 24-37)		30.7 (4.5, 25-42)	30.0 (4.4, 22-37)	
Ethnicity, n/N (%)			1.00			0.12			0.24
Caucasian	22/29 (76)	25/33 (76)		30/39 (77)	11/20 (55)		46/58 (79)	18/27 (67)	
Asian	5/29 (17)	6/33 (18)		6/39 (15)	4/20 (20)		8/58 (14)	4/27 (15)	
Black	2/29 (7)	2/33 (6)		3/39 (8)	5/20 (25)		4/58 (7)	5/27 (18)	
Smoking status, n/N (%)			0.77			**0.04***			0.18
Non-smoker	23/29 (79)	25/33 (76)		27/39 (69)	19/20 (95)		41/58 (71)	23/27 (85)	
Current smoker	6/29 (21)	8/33 (24)		12/39 (31)	1/20 (5)		17/58 (29)	4/27 (15)	
Contraception, n/N (%)			0.98			0.05			0.06
Nil	11/29 (37)	11/33 (33)		10/39 (25)	13/20 (65)		16/58 (28)	16/27 (59)	
Condoms	1/29 (4)	2/33 (6)		6/39 (15)	2/20 (10)		8/58 (14)	4/27 (15)	
COCP	14/29 (48)	17/33 (52)		20/39 (51)	3/20 (15)		28/58 (48)	4/27 (15)	
POP	2/29 (7)	2/33 (6)		1/39 (3)	1/20 (5)		2/58 (3)	1/27 (4)	
Implant	0/20 (0)	0/33 (0)		0/39 (0)	0/20 (0)		0/58 (0)	1/27 (4)	
Mirena IUS	1/29 (4)	1/33 (3)		1/39 (3)	1/20 (5)		2/58 (3)	1/27 (4)	
Copper IUD	0/29 (0)	0/33 (0)		0/39 (0)	0/20 (0)		1/58 (2)	0/27 (0)	
Vaginal ring	0/29 (0)	0/33 (0)		1/39 (3)	0/20 (0)		1/58 (2)	0/27 (0)	
Parity, n/N (%)			1.00			0.52			0.51
Nulliparous	21/29 (72)	24/33 (73)		32/39 (82)	15/20 (75)		45/58 (78)	22/27 (81)	
Parous	8/29 (28)	9/33 (27)		7/39 (18)	5/20 (25)		13/58 (22)	5/27 (19)	
Time since last intercourse, n/N (%)			0.71			0.28			0.31
>48 hours	26/29 (90)	28/33 (85)		34/39 (87)	15/20 (75)		52/58 (90)	22/27 (81)	
<48 hours	3/29 (10)	5/33 (15)		5/39 (13)	5/20 (25)		6/58 (10)	5/27 (19)	
Phase of menstrual cycle, n/N (%)			0.40			0.14			0.08
Luteal	15/29 (52)	18/33 (55)		21/39 (54)	6/20 (30)		29/58 (50)	7/27 (26)	

	Comparison 1: Treated women with paired samples			Comparison 2 : Normal post-treatment vs. normal untreated controls			Comparison 3: Treated by cone depth/proportion vs. normal untreated controls		
Follicular	13/29 (45)	11/33 (33)		14/39 (36)	9/20 (45)		23/58 (40)	14/27 (52)	
Unknown	1/29 (3)	4/33 (12)		4/39 (10)	5/20 (25)		6/58 (10)	6/27 (22)	
Vaginal pH			0.61			0.92			0.37
<4.5	12/29 (41.5)	13/33 (39)		16/39 (41)	9/20 (45)		24/58 (41)	15/27 (56)	
≥4.5	16/29 (55)	20/33 (61)		21/39 (54)	9/20 (45)		31/58 (53)	10/27 (37)	
Unknown	1/29 (3.5)	0/33 (0)		2/39 (5)	2/20 (10)		3/58 (6)	2/27 (7)	
Bacterial vaginosis, n/N (%)			1.00			0.13			0.16
No	27/29 (93)	31/33 (94)		35/39 (89)	15/20 (75)		52/58 (89)	21/27 (78)	
Yes	2/29 (7)	2/33 (6)		3/39 (8)	2/20 (10)		5/58 (9)	3/27 (11)	
Unknown	0/29 (0)	0/33 (0)		1/39 (3)	3/20 (15)		1/58 (2)	3/27 (11)	
HPV DNA test, n/N (%)			**0.0001***			1.00			0.59
Negative	1/29 (3)	27/33 (82)		39/39 (100)	20/20 (100)		46/58 (79)	20/27 (74)	
Positive	28/29 (97)	6/33 (18)		0/39 (100)	0/20 (0)		12/58 (21)	7/27 (26)	

COCP: Combined oral contraceptive pill; HPV: Human Papillomavirus; HSIL: High-grade intraepithelial lesion; IUD: Intrauterine device; IUS: Intrauterine system; POP: Progesterone-only pill; SD: Standard deviation.

Figure 1



Cyto: cytology; HPV: human papilloma virus; -ve: negative; +ve: positive.

Figure 2

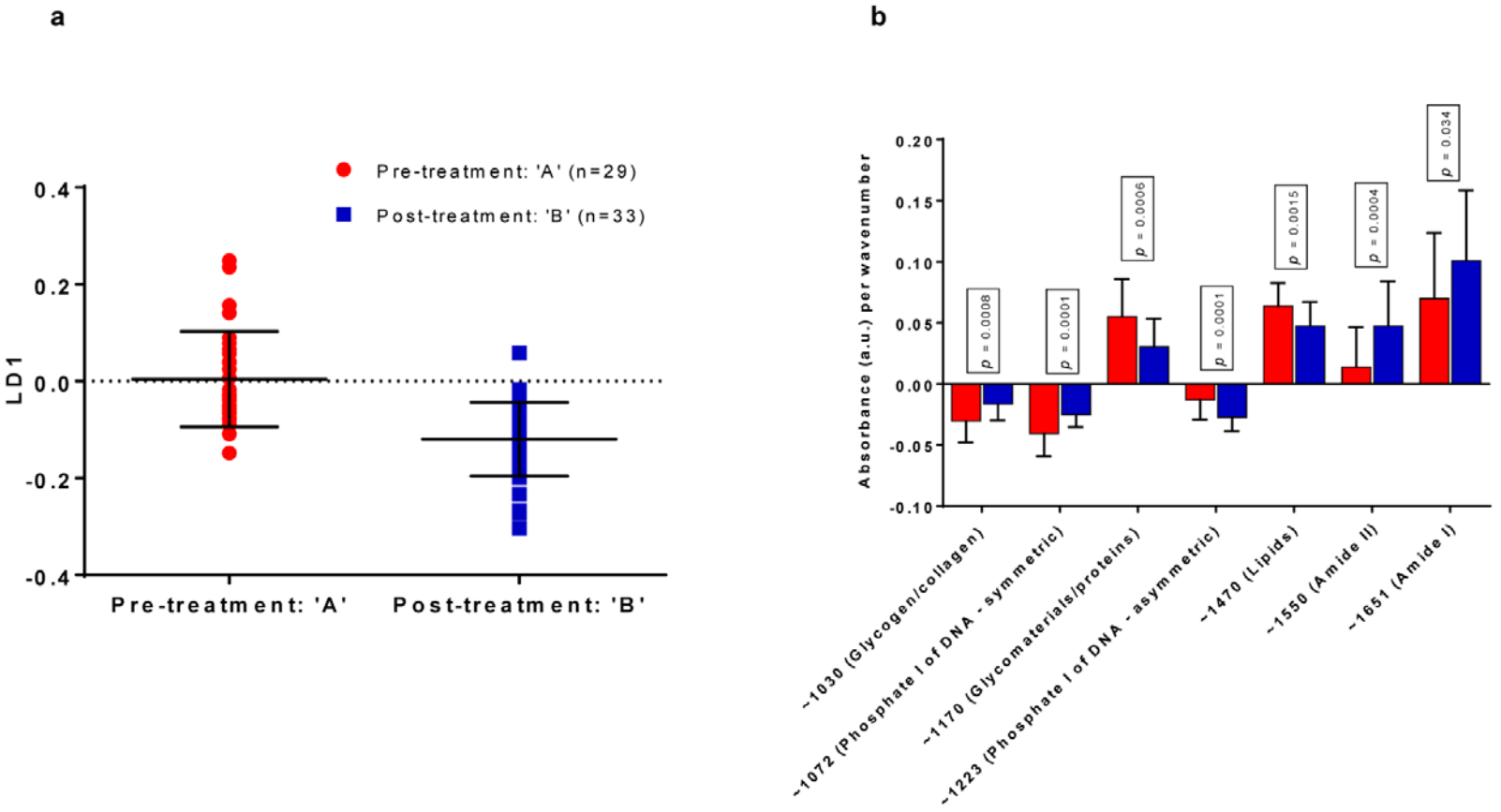

a
Pre-treatment: 'A' (n=29)
Post-treatment: 'B' (n=33)
LD1
0.4
0.2
0.0
-0.2
-0.4
Pre-treatment: 'A'
Post-treatment: 'B'
b
Absorbance (a.u.) per wavenumber
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
p = 0.0008
p = 0.0001
p = 0.0006
p = 0.0001
p = 0.0015
p = 0.0004
p = 0.034
~1030 (Glycogen/collagen)
~1072 (Phosphate I of DNA - symmetric)
~1170 (Glycomaterials/proteins)
~1223 (Phosphate I of DNA - asymmetric)
~1470 (Lipids)
~1550 (Amide II)
~1651 (Amide I)

Figure 3

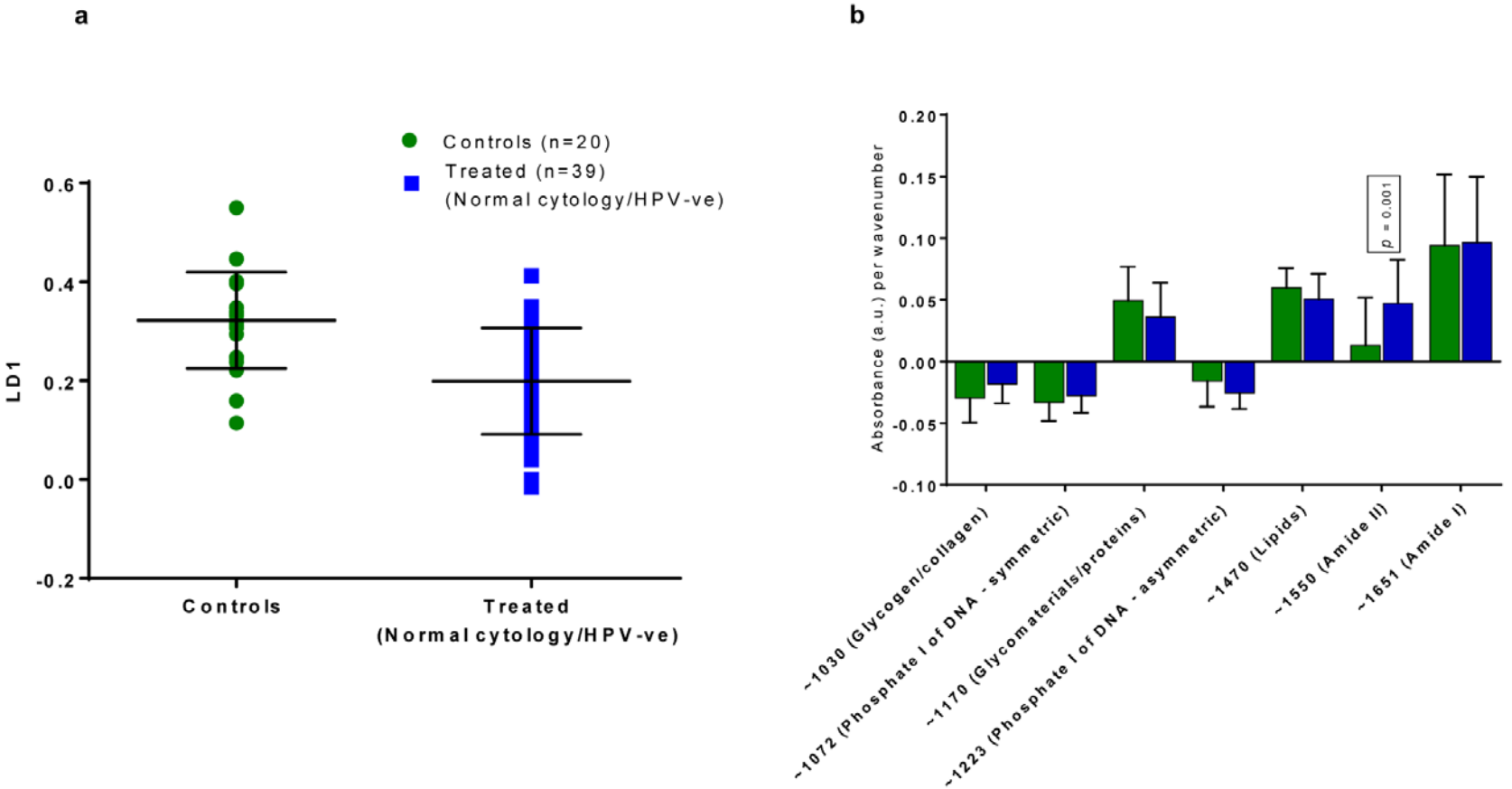

a
Controls (n=20)
Treated (n=39)
(Normal cytology/HPV-ve)
LD1
0.6
0.4
0.2
0.0
-0.2
Controls
Treated
(Normal cytology/HPV-ve)
b
Absorbance (a.u.) per wavenumber
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
p = 0.001
~1030 (Glycogen/collagen)
~1072 (Phosphate I of DNA - symmetric)
~1170 (Glycomaterials/proteins)
~1223 (Phosphate I of DNA - asymmetric)
~1470 (Lipids)
~1550 (Amide II)
~1651 (Amide I)

Figure 4

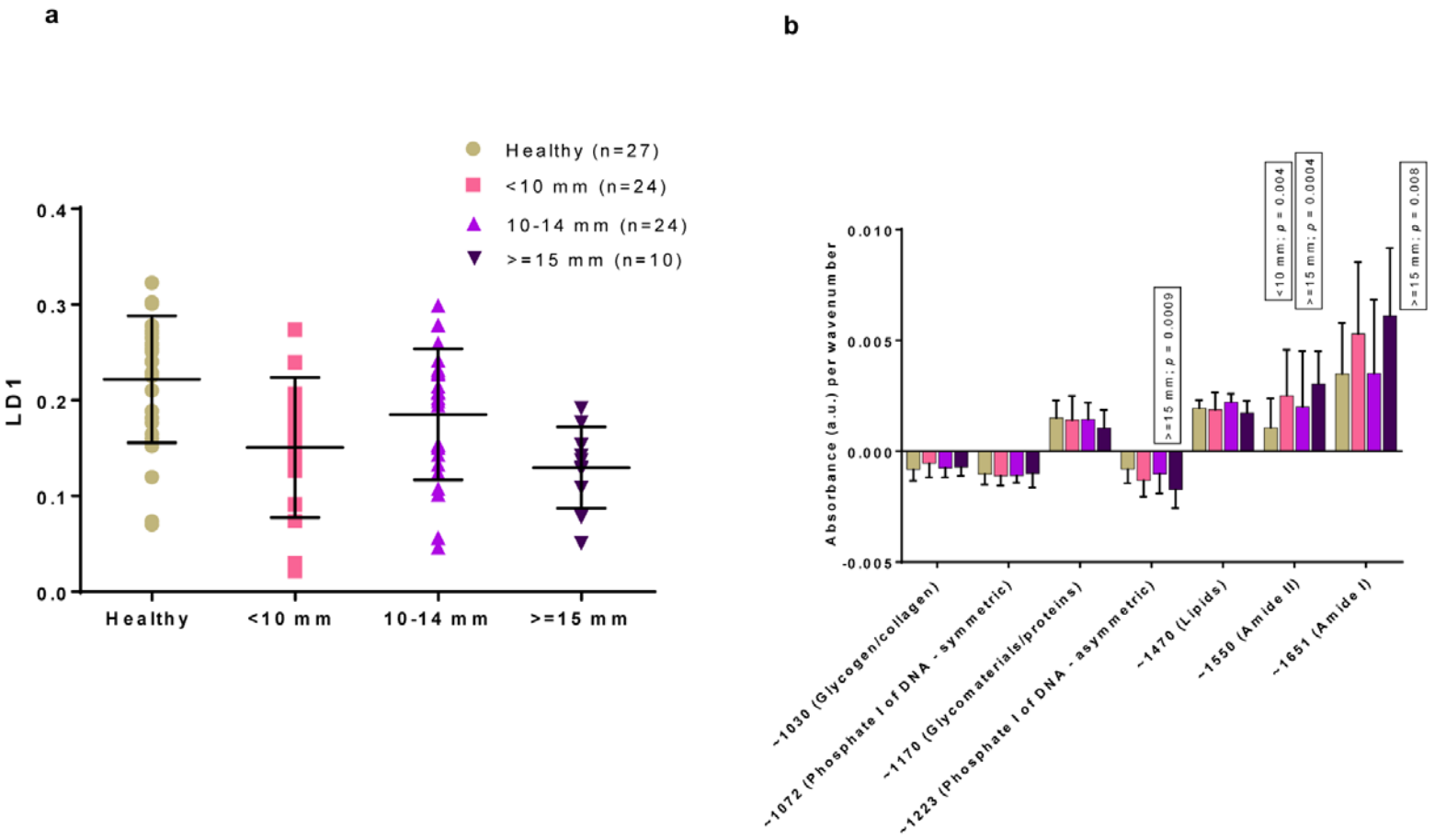
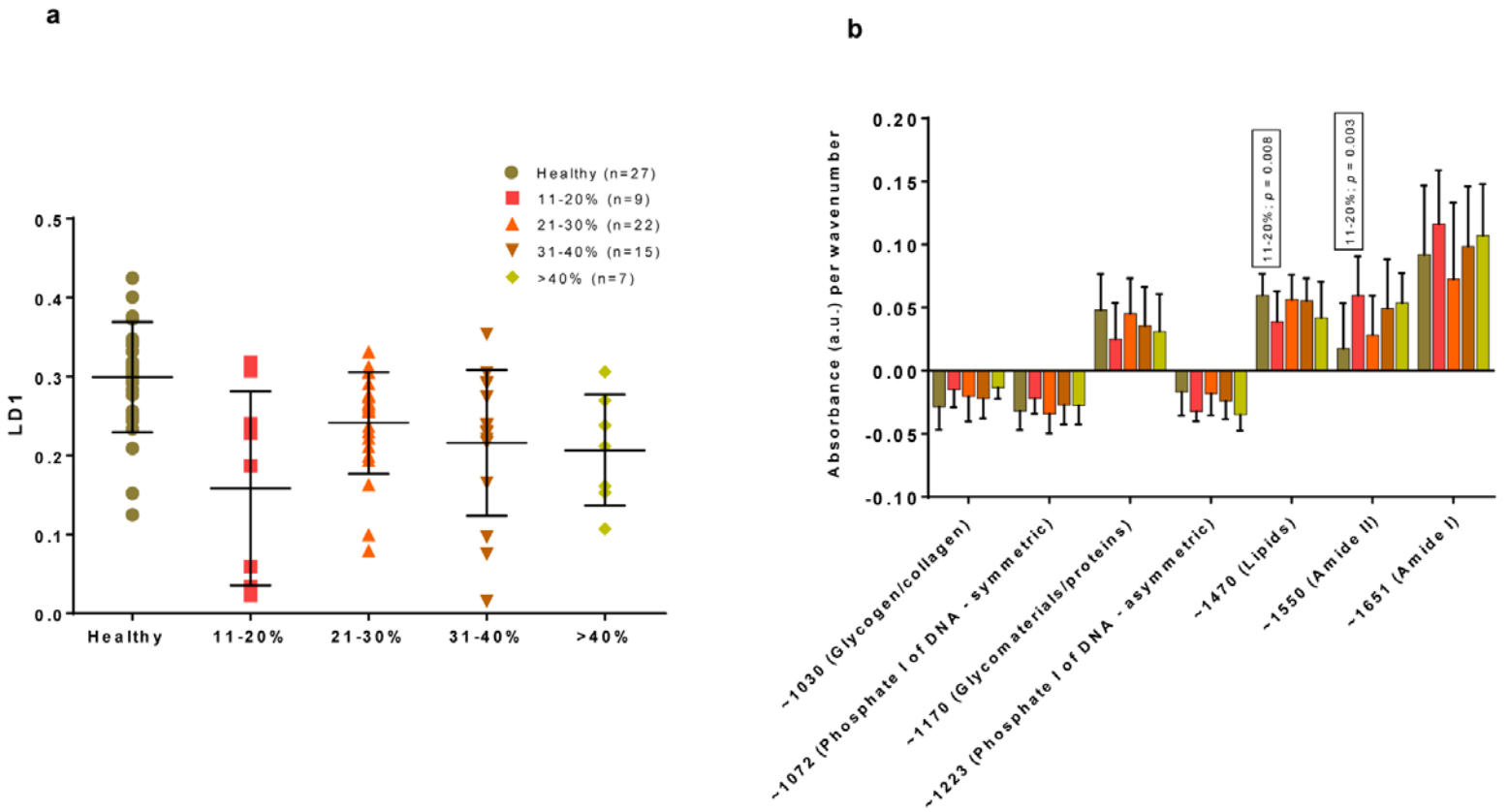

a
Healthy (n=27)
<10 mm (n=24)
10-14 mm (n=24)
>=15 mm (n=10)
LD1
0.4
0.3
0.2
0.1
0.0
Healthy
<10 mm
10-14 mm
>=15 mm
b
Absorbance (a.u.) per wavenumber
0.010
0.005
0.000
-0.005
>=15 mm; p = 0.0009
<10 mm; p = 0.004
>=15 mm; p = 0.0004
>=15 mm; p = 0.008
~1030 (Glycogen/collagen)
~1072 (Phosphate I of DNA - symmetric)
~1170 (Glycomaterials/proteins)
~1223 (Phosphate I of DNA - asymmetric)
~1470 (Lipids)
~1550 (Amide II)
~1651 (Amide I)

Figure 5

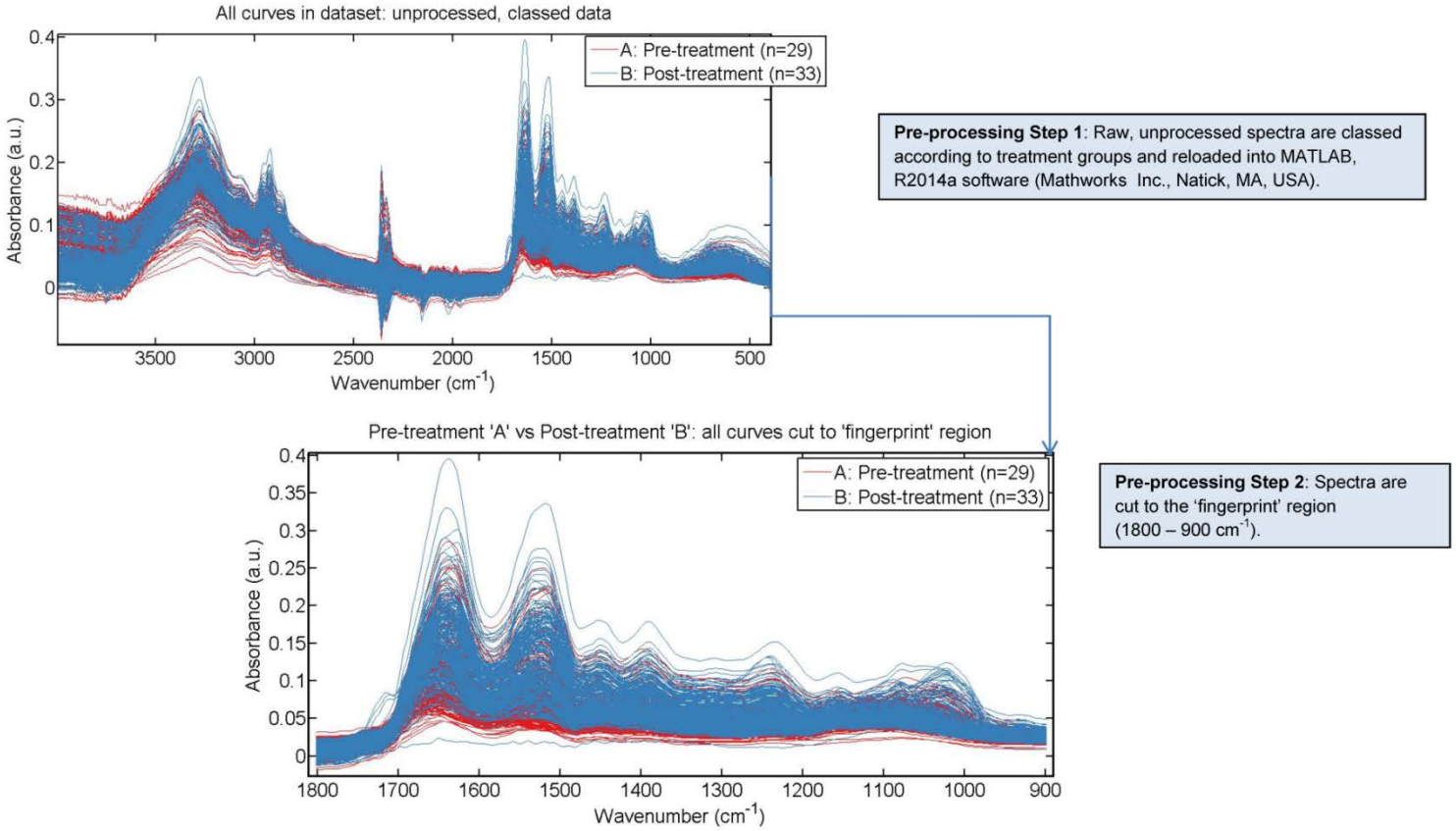

a
Healthy (n=27)
11-20% (n=9)
21-30% (n=22)
31-40% (n=15)
>40% (n=7)
LD1
0.5
0.4
0.3
0.2
0.1
0.0
Healthy
11-20%
21-30%
31-40%
>40%
b
Absorbance (a.u.) per wavenumber
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
11-20%; p = 0.008
11-20%; p = 0.003
~1030 (Glycogen/collagen)
~1072 (Phosphate I of DNA - symmetric)
~1170 (Glycomaterials/proteins)
~1223 (Phosphate I of DNA - asymmetric)
~1470 (Lipids)
~1550 (Amide II)
~1651 (Amide I)

Nature Communications

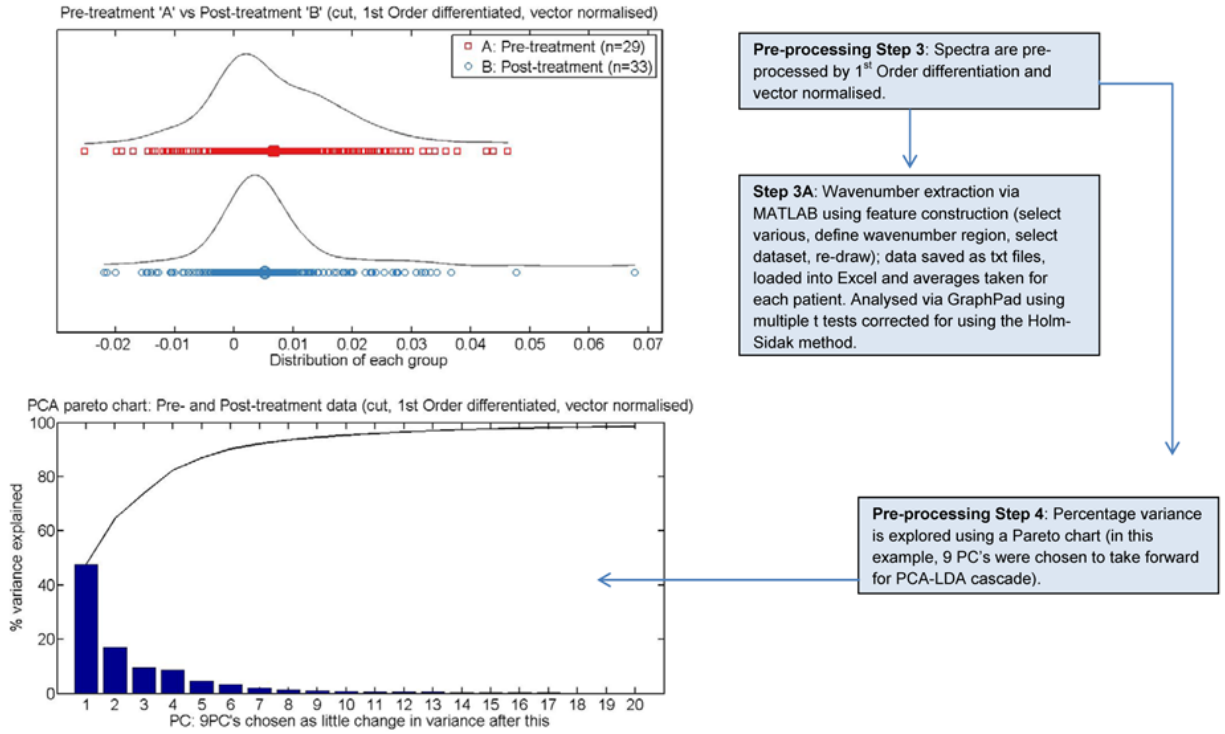
Electronic Supplementary Information

Tracking the Impact of Excisional Cervical Treatment on the Cervix using Biospectroscopy: Supplementary information

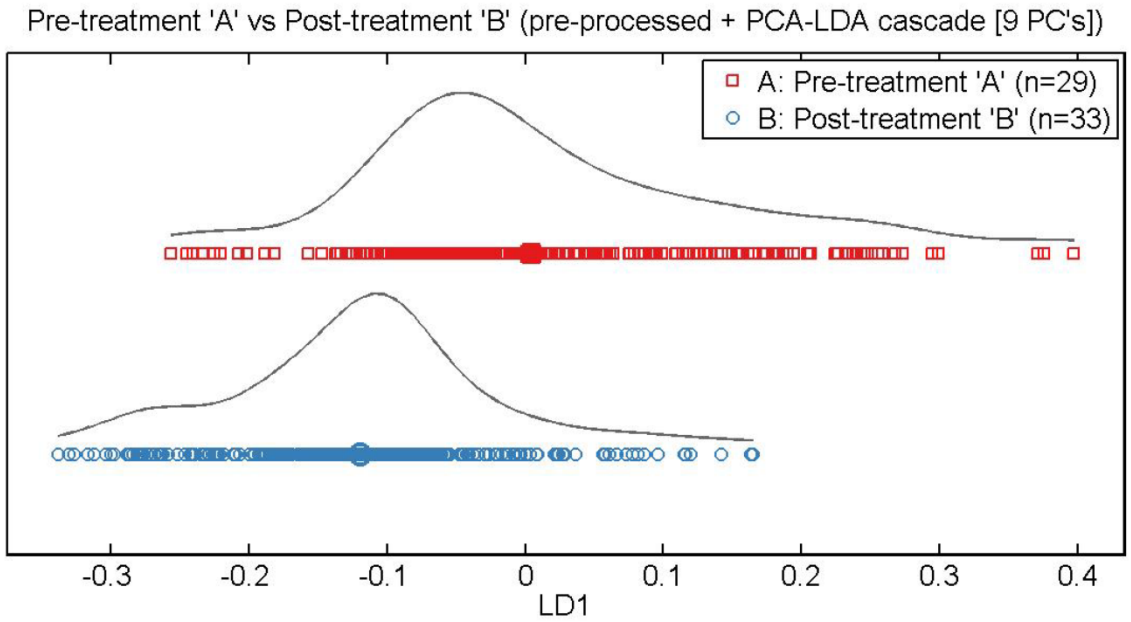
Diane E. Halliwell,[a] Maria Kyrgiou,[b,c] Anita Mitra,[b,c] Ilkka Kalliala,[b,c] Evangelos Paraskevaidis,[d] Georgios Theophilou,[e] Pierre L. Martin-Hirsch,[f] Francis L. Martin[a]



Supplementary Figure 1. Processing steps for spectra obtained using ATR-FTIR spectroscopy. Step 1: Classifying the data by treatment group; **Step 2:** Cutting the spectra to the fingerprint region (1800-900 cm^{-1}). ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared.



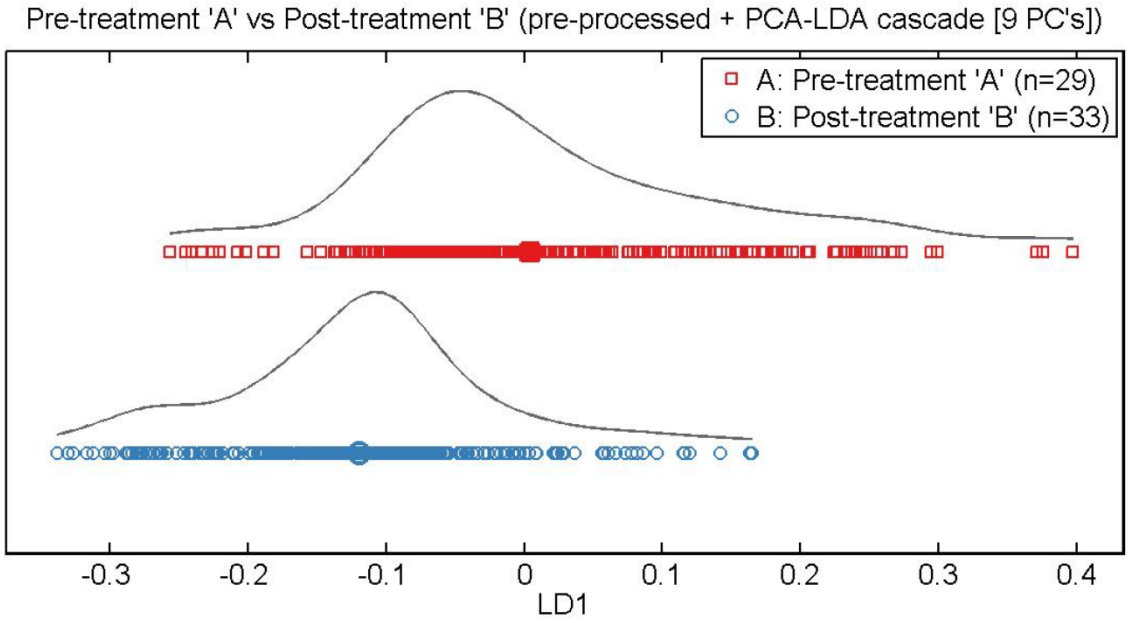
Supplementary Figure 2. Processing steps for spectra obtained using ATR-FTIR spectroscopy. Step 3: 1st order differentiated, followed by vector normalisation; **Step 3A:** Wavenumber extraction using previous data; **Step 4:** Calculation of percentage variance (i.e., number of Principal Components [PCs] to take forward for PCA-LDA cascade) using the Pareto chart. ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared. PCA-LDA: Principal Components Analysis coupled to Linear Discriminant Analysis.



Step 5: PCA-LDA cascade applied using the number of PC's determined by Pareto.

Step 6: PCA-LDA output saved as txt file, loaded into Excel; averages taken for each patient and analysed via GraphPad Prism 6 (GraphPad Software Inc., La Jolla, CA, USA).

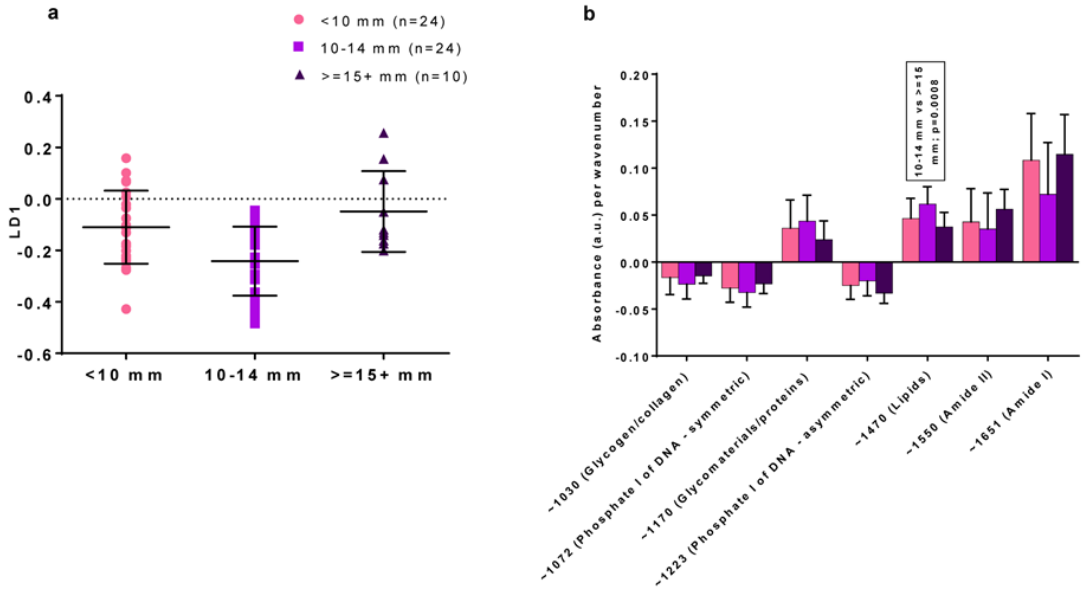
Supplementary Figure 3. Processing steps for spectra obtained using ATR-FTIR spectroscopy. Step 5: Applying PCA-LDA-Cascade using the previously defined number of principal components identified from Pareto plotting**.** ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared. PCA-LDA: Principal Components Analysis coupled to Linear Discriminant Analysis; LD: Linear Discriminant.



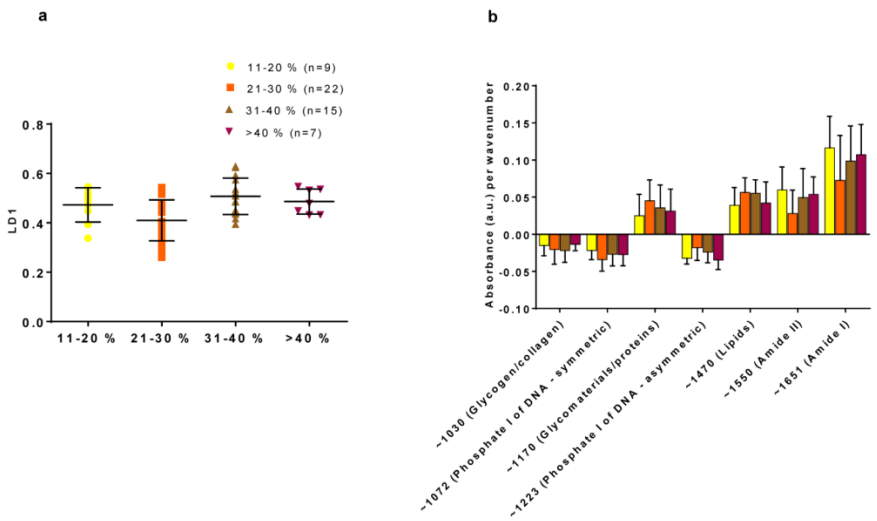
Step 5: PCA-LDA cascade applied using the number of PC's determined by Pareto.

Step 6: PCA-LDA output saved as txt file, loaded into Excel; averages taken for each patient and analysed via GraphPad Prism 6 (GraphPad Software Inc., La Jolla, CA, USA).

Supplementary Figure 3. Processing steps for spectra obtained using ATR-FTIR spectroscopy. Step 5: Applying PCA-LDA-Cascade using the previously defined number of principal components identified from Pareto plotting**.** ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared. PCA-LDA: Principal Components Analysis coupled to Linear Discriminant Analysis; LD: Linear Discriminant.



Supplementary Figure 4. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1: Cone depth (a) together with absorbance per wavenumber (b). Comparison of <10 mm with >=15 mm was not significantly different along LD1 (Mean/SD (**a**): -0.11/0.14 for <10 mm; -0.05/0.16 for >=15 mm; p=0.489, 95% CI = -0.19 to 0.07). Comparison of <10 mm with 10-14 mm was significantly different along LD1 (Mean/SD: -0.11/0.14 for <10 mm; -0.24/0.13 for 10-14 mm; p=0.006). Comparison of 10-14 mm with >=15 mm was significantly different along LD1 (Mean/SD: -0.24/0.13 for 10-14 mm; -0.05/0.16 for >=15 mm; p=0.002). Absorbance associated lipids were shown to have a significant positive rate of change for the 10-14 mm group compared with the >=15 mm group, evidencing their higher bioavailability. No significant differences were detected for wavenumbers associated with other biomarkers between all 3 groups **(b)**. ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared; CI: Confidence interval; SD: Standard deviation.



Supplementary Figure 5. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1: Percentage Excision (a) together with absorbance per wavenumber (b).

Comparison of 11-20% with 21-30% was not significantly different along LD1 (Mean/SD (**a**): 0.47/0.07, 0.41/0.08 respectively; p=0.16; 95 CI -0.02 to -0.14). Comparison of 11-20% with 31-40% was not significantly different along LD1 (Mean/SD: 0.47/0.07, 0.51/0.07 respectively; p=0.69; -0.12 to 0.05). Comparison of 11-20% with >40% was not significantly different along LD1 (Mean/SD: 0.47/0.07, 0.49/0.05 respectively; p=0.98; 95 CI -0.11 to 0.09). Comparison of 21-30% with >40% was not significantly different along LD1 (Mean/SD: 0.41/0.08; 0.49/0.05 respectively, p=0.10; 95 CI -0.16 to 0.01). Comparison of 31-40% with >40% was not significantly different along LD1 (Mean/SD: 0.51/0.07, 0.49/0.05 respectively; p=0.92; CI -0.07 to 0.11). A significant difference was detected along LD1 for 21-30% vs 31-40% (Mean/SD: 0.41/0.08, 0.51/0.07, respectively; p=0.0016; 95 CI -0.16 to -0.03). No significant differences were detected for wavenumbers associated with the 7 biomarkers between all 4 groups (**b**). ATR-FTIR: Attenuated total reflectance, Fourier-transform Infrared; CI: Confidence interval; SD: Standard deviation.

Supplementary Table 1. Important biomarkers

Tentative assignment of Biomarkers[a]	**Wavenumber (cm^{-1})**
Amide I (of proteins predominantly in α helix conformation)	~1651
Amide II (of proteins predominantly in β sheet conformation)	~1550
Methylene chains in lipids	~1470
Phosphate I - asymmetric (DNA)	~1223
C-O bands from glycomaterials and proteins	~1170
Phosphate I - symmetric (DNA)	~1072
Glycogen & collagen	~1030

[a] Movasaghi *et al*, 2008[36]; N.B.: The signal at a particular wavenumber could have contributions from more than one biomarker. Amides I and II are linked to the secondary structure of proteins and are indicative of their bioavailability.

Supplementary Table 2: Patient characteristics for the different cone depth and cervical proportion excised clinical groups (Comparison 3)

Characteristics	**Normal controls (n=27)**	**Cone Depth[a] (n=58)**			**Percentage Excision[b] (n=53)**			
		Category 1: <10 mm (n=24)	**Category 2: 10-14 mm (n=24)**	**Category 3: >= 15 mm (n=10)**	**Category 2: 11-20% (n=9)**	**Category 3: 21-30% (n=22)**	**Category 4: 31-40% (n=15)**	**Category 5: >40% (n=7)**
Age, years								
Mean (SD, range)	30.0 (4.4, 22-37)	30.6 (5.2, 25-42)	29.8 (3.1, 25-36)	33.5 (4.5, 25-43)	31.1 (4.0, 26-38)	30.3 (4.4, 25-42)	30.6 (5.1, 25-43)	33.0 (3.6, 25-38)
Ethnicity, n/N (%)								
Caucasian	18/27 (67)	19/24 (79)	20/24 (83)	7/10 (70)	5/9 (56)	19/22 (86)	14/14 (93)	6/7 (86)
Asian	4/27 (15)	4/24 (17)	3/24 (13)	2/10 (20)	4/9 (44)	2/22 (9)	0/15 (0)	0/7 (0)
Black	5/27 (18)	1/24 (4)	1/24 (4)	1/10 (10)	0/9 (0)	1/22 (5)	1/14 (7)	1/7 (14)
Smoking status, n/N (%)								
Non-smoker	23/27 (85)	19/24 (79)	15/24 (63)	7/10 (70)	7/9 (78)	17/22 (77)	8/15 (53)	5/7 (71)
Current smoker	4/27 (15)	5/24 (21)	9/24 (37)	3/10 (30)	2/9 (22)	5/22 (23)	7/15 (47)	2/7 (29)
Contraception, n/N (%)								
Nil	16/27 (59)	6/24 (25)	8/24 (33.5)	2/10 (20)	3/9 (33)	7/22 (32)	4/15 (26)	0/7 (0)
Condoms	4/27 (15)	4/24 (17)	2/24 (8.5)	2/10 (20)	1/9 (11)	2/22 (9)	3/15 (20)	2/7 (29)
COCP	4/27 (15)	12/24 (50)	12/24 (50)	4/10 (40)	5/9 (56)	12/22 (54)	6/15 (40)	4/7 (57)
POP	1/27 (4)	1/24 (4)	0/24 (0)	1/10 (10)	0/9 (0)	0/22 (0)	1/15 (7)	0/7 (0)
Implant	1/27 (4)	0/24 (0)	0/24 (0)	0/10 (0)	0/9 (0)	0/22 (0)	0/15 (0)	0/7 (0)
Mirena IUS	1/27 (4)	1/24 (4)	1/24 (4)	0/10 (0)	0/9 (0)	1/22 (5)	0/15 (0)	0/7 (0)
Copper IUD	0/27 (0)	0/24 (0)	0/24 (0)	1/10 (10)	0/9 (0)	0/22 (0)	0/15 (0)	1/7 (14)
Vaginal ring	0/27 (0)	0/24 (0)	1/24 (4)	0/10 (0)	0/9 (0)	0/22 (0)	1/15 (7)	0/7 (0)
Parity, n/N (%)								
Nulliparous	22/27 (81)	19/24 (79)	20/24 (83)	6/10 (60)	8/9 (89)	18/22 (82)	12/15 (80)	5/7 (71)
Parous	5/27 (19)	5/24 (21)	4/24 (17)	4/10 (40)	1/9 (11)	4/22 (18)	3/15 (20)	2/7 (29)
Time since last intercourse, n/N (%)								
>48 hours	22/27 (81)	19/24 (79)	24/24 (100)	9/10 (90)	6/9 (67)	21/22 (95)	15/15 (100)	7/7 (100)
<48 hours	5/27 (19)	5/24 (21)	0/24 (0)	1/10 (10)	3/9 (33)	1/22 (5)	0/15 (0)	0/7 (0)

Characteristics	Normal controls (n=27)	Cone Depth[a] (n=58)			Percentage Excision[b] (n=53)			
Phase of menstrual cycle, n/N (%)								
Luteal	7/27 (26)	14/24 (58)	9/24 (38)	6/10 (60)	6/9 (67)	12/22 (55)	6/15 (40)	3/7 (43)
Follicular	14/27 (52)	8/24 (34)	12/12 (50)	3/10 (30)	3/9 (33)	10/22 (45)	6/15 (40)	3/7 (43)
Unknown	6/27 (22)	2/24 (8)	3/24 (12)	1/10 (10)	0/9 (0)	0/22 (0)	3/15 (20)	1/7 (14)
Vaginal pH								
<4.5	15/27 (56)	12/24 (50)	9/24 (38)	3/10 (30)	4/9 (44)	9/22 (41)	8/15 (53)	2/7 (29)
>=4.5	10/27 (37)	11/24 (46)	14/24 (58)	6/10 (60)	5/9 (56)	12/22 (55)	6/15 (40)	4/7 (57)
Unknown/NA	2/27 (7)	1/24 (4)	1/24 (4)	1/10 (10)	0/9 (0)	1/22 (4)	1/15 (7)	1/7 (14)
Bacterial vaginosis, n/N (%)								
No	21/27 (78)	22/24 (92)	22/24 (92)	8/10 (80)	8/9 (89)	22/22 (100)	14/15 (94)	5/7 (71)
Yes	3/27 (11)	2/24 (8)	1/24 (4)	2/10 (20)	1/9 (11)	0/22 (0)	0/15 (0)	2/7 (29)
Unknown	3/27 (11)	0/24 (0)	1/24 (4)	0/10 (0)	0/9 (0)	0/22 (0)	1/15 (6)	0/7 (0)
Follow up cytology/HPV, n/N (%)								
Normal & HPV -ve	NA	14/24 (58)	19/24 (79)	6/10 (60)	8/9 (89)	10/22 (45)	14/15 (93)	4/7 (58)
Normal & HPV +ve	NA	4/24 (17)	4/24 (17)	1/10 (10)	1/9 (11)	7/22 (32)	0/15 (0)	1/7 (14)
LSIL & HPV –ve	NA	3/24 (13)	1/24 (4)	2/10 (20)	0/9 (0)	3/22 (13)	1/15 (7)	1/7 (14)
LSIL & HPV +ve	NA	1/24 (4)	0/24 (0)	1/10 (10)	0/9 (0)	1/22 (5)	0/15 (0)	1/7 (14)
HSIL & HPV -ve	NA	1/24 (4)	0/24 (0)	0/10 (0)	0/9 (0)	1/22 (5)	0/15 (0)	0/7 (0)
HSIL & HPV +ve	NA	1/24 (4)	0/24 (0)	0/10 (0)	0/9 (0)	0/22 (0)	0/15 (0)	0/7 (0)
HPV DNA test, n/N (%)								
Negative	20/27 (74)	18/24 (75)	20/24 (83)	8/10 (80)	8/9 (89)	14/22 (64)	15/15 (100)	5/7 (72)
Positive	7/27 (26)	6/24 (25)	4/24 (17)	2/10 (20)	1/9 (11)	8/22 (36)	0/15 (0)	2/7 (28)

[a] Patients were included in the dimensional analyses if they had had treatment and had follow up data at 6 months after treatment. [b] There was insufficient patients in Category 1 to include in the analysis. No segregation was made based on histology, cytology or HPV testing. COCP: Combined oral contraceptive pill: IUD: Intrauterine device; IUS: Intrauterine system; NA: Not applicable; POP: Progesterone-only pill; SD: Standard deviation.

3 Project Two

This paper has been accepted for publication by *Scientific Reports* as:

"Imaging cervical cytology with scanning near-field microscopy (SNOM) coupled with an IR-FEL"

Diane E. Halliwell, Camilo L. M. Morais, Kássio M. G. Lima, Julio Trevisan, Michele R. F. Siggel-King, Tim Craig, James Ingham, David S. Martin, Kelly Heys, Maria Kyrgiou, Anita Mitra, Evangelos Paraskevaidis, Georgios Theophilou, Pierre L. Martin-Hirsch, Antonio Cricenti, Marco Luce, Peter Weightman, and Francis L. Martin.

Contribution:

I collaborated with the co-authors on this project by preparing the samples for imaging; acting as biological consultant during experimental shifts; acting as Second Commissioner during experimental shifts at the ALICE facility at Daresbury during 2015; and helping to prepare the manuscript for publication. This work was performed under the supervision of Professor Francis L. Martin.

…… ……

Diane E. Halliwell Professor Francis L. Martin

Scientific Reports

Dear Prof Martin,

We are delighted to accept your manuscript entitled "Imaging cervical cytology with scanning near-field optical microscopy (SNOM) coupled with an IR-FEL" for publication in Scientific Reports. Thank you for choosing to publish your work with us.

You should have just received another email from scientificreports@nature.com with instructions for the next step, which is to complete your publication agreements. To continue with your publication agreements you will need to create a new account on this new system. Please complete these as soon as possible so we can start preparing your manuscript for publication. The agreements include the licence, which defines the terms of publication, and billing information for your Open Access article.

After we've prepared your paper for publication, you will receive a PDF proof for checking. At that point, please check the author list and affiliations to ensure that they are correct. For the main text, only errors that have been introduced during the production process or those that directly compromise the scientific integrity of the paper may be corrected at this stage. Please ensure that only one author communicates with us and that only one set of corrections is returned. The corresponding (or nominated) author is responsible on behalf of all co-authors for the accuracy of all content, including spelling of names and current affiliations.

To ensure prompt publication, your proofs should be returned within two working days; please let us know immediately if there is any period within the next two weeks in which you (or the nominated author) won't be available.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (seehttp://www.nature.com/srep/policies/index.html). In particular, your manuscript must not be published elsewhere and there must be no announcement of this work to any media outlet until the publication date is confirmed. We will inform you by email as soon as your manuscript is scheduled for publication, which will be after we have received and approved your proof corrections. Advice about media relations is available from NPG's press office at press@nature.com.

Your article will be open for online commenting on the Scientific Reports website. You may use the report facility if you see any comments which you consider inappropriate, and of course, you can contribute to discussions yourself. If you wish to track comments on your article, please register for this service by visiting the 'Comments' section in the full text (HTML) version of your paper.

A form to order reprints of your article is available at http://www.nature.com/reprints/author-reprints.html. To obtain the special author reprint rate, orders must be made within a month of the publication date. After that, reprints are charged at the normal (commercial) rate.

We look forward to publishing your article.

Best regards,

Christian Eggeling
Editorial Board Member
Scientific Reports

Scientific Reports

Imaging cervical cytology with scanning near-field optical microscopy (SNOM) coupled with an IR-FEL

Diane E. Halliwell,[a] Camilo L. M. Morais,[b] Kássio M. G. Lima,[b] Julio Trevisan,[c] Michele R. F. Siggel-King,[de] Tim Craig,[d] James Ingham,[d] David S. Martin,[d] Kelly Heys,[a] Maria Kyrgiou,[f,g] Anita Mitra,[f,g] Evangelos Paraskevaidis,[h] Georgios Theophilou,[i] Pierre L. Martin-Hirsch,[aj] Antonio Cricenti,[k] Marco Luce,[k] Peter Weightman,[d] and *Francis L. Martin[a]

[a] Centre for Biophotonics, LEC, Lancaster University, Lancaster, UK; [b] Biological Chemistry and Chemometrics, Institute of Chemistry, Federal University of Rio Grande do Norte, Natal 59072-970, RN, Brazil; [c] Institute of Astronomy, Geophysics and Atmospheric Sciences University of São Paulo, Brazil; [d] Department of Physics, University of Liverpool, Oliver Lodge Building, Liverpool, UK; [e] Accelerator Science and Technology Centre (ASTEC), STFC Daresbury Laboratory, UK; [f] Institute of Reproductive and Developmental Biology, Department of Surgery & Cancer, Faculty of Medicine, Imperial College, London, UK; [g] West London Gynaecological Cancer Centre, Imperial College NHS Healthcare, London, UK; [h] Department of Obstetrics and Gynaecology, University of Ioannina, Ioannina, Greece; [i] St James Hospital, Leeds, West Yorkshire, UK; [j] Department of Obstetrics and Gynaecology, Lancashire Teaching Hospitals NHS Trust Foundation, Preston, UK. [k] Istituto di Struttura della Materia, CNR, via del Fosso del Cavaliere 100, Rome, Italy.

*Correspondence and request for materials should be addressed to Professor Martin, Centre for Biophotonics, LEC, Lancaster University, Lancaster LA1 4YQ, UK; Email: f.martin@lancaster.ac.uk; Tel: +44(0)1524 51020.

Abstract

Cervical cancer remains a major cause of morbidity and mortality among women, especially in the developing world. Increased synthesis of proteins, lipids and nucleic acids is a condition for the rapid proliferation of cancer cells. We show that scanning near-field optical microscopy in combination with an infrared free electron laser (SNOM-IR-FEL), is able to distinguish between normal and squamous low-grade and high-grade dyskaryosis, and between normal and mixed squamous/glandular pre-invasive and adenocarcinoma cervical lesions, at designated wavelengths associated with DNA, amides I and II and lipids. These findings evidence the promise of the SNOM-IR-FEL technique in obtaining chemical information relevant to the detection of cervical cell abnormalities and cancer diagnosis at spatial resolutions below the diffraction limit (≥0.2 μm). We compare these results with analyses following attenuated total reflectance Fourier-transform infrared (ATR-FTIR) spectroscopy; although this latter approach has been demonstrated to detect underlying cervical atypia missed by conventional cytology, it is limited by a spatial resolution of ~3 μm to 30 μm.

Keywords: Biomarkers, Biospectroscopy, Cervical intra-epithelial neoplasia, IR-FEL, SNOM

Cervical cancer is associated with the persistent infection of high-risk types of Human papillomavirus (HPV), together with other socioeconomic co-factors[1]. Screening involves cytological and histological classification of cervical cells. In the UK, cytological examination of cervical squamous cells is classified as normal, borderline or mild dyskaryosis, moderate or severe dyskaryosis and invasive cervical cancer (ICC). Histology is defined as normal, cervical intra-epithelial neoplasia (CIN): CIN1, CIN2, CIN3, or invasive cervical cancer. For atypical cells found in the glandular cells of the cervix, the pre-invasive lesion of adenocarcinoma is defined by changes termed cervical glandular intraepithelial neoplasia (CGIN), and sub-classified as low-grade cervical intra-epithelial glandular neoplasia (LGCGIN) and high-grade cervical intra-epithelial glandular neoplasia (HGCGIN). Squamous and glandular lesions may co-exist together and are defined by the level of CIN together with either LGCGIN or HGCGIN. Conventional screening is flawed as it is dependent on the subjective visual inspection of cytology; this often results in mis-diagnoses when grading samples[2].

Attenuated total reflectance, Fourier-transform infrared (ATR-FTIR) spectroscopy has shown potential over conventional screening methods, demonstrating it can segregate grades of cervical cytology more accurately than conventional cytological screening[3,4], classify cervical cytology based on HPV infection with low- or high-risk types[5], and can diagnose underlying disease more accurately that conventional cytology screening[2]. However, FTIR spectroscopy is limited in spatial resolution by the effect of diffraction, defined as the interference of waves when they hit an obstacle or slit. This effect restricts the spatial resolution of FTIR to about half the wavelength of light or ~3 μm to 30 μm[6], with the resolution being a measure of how closely the lines of an image can be resolved (i.e., the number of independent pixels per value per unit length).

Scanning near field optical microscopy (SNOM) belongs to a family of nanoscopic techniques that have shown potential in providing detailed information on cell topography and cytoplasmic structures.

SNOM has a clear advantage over conventional infrared IR microscopy in terms of spatial resolution because it is able to overcome the diffraction limit; this is achieved by the use of an apertured fibre optic scanning tip. The SNOM technique requires relatively high photon intensities such as those provided by an IR free electron laser (IR-FEL). The SNOM-IR-FEL enables the simultaneous collection of topography and optical features at scales not normally achieved with conventional IR techniques to produce high quality, chemically-rich images at designated wavelengths with a spatial resolution of ≥0.2 μm[7,8].

Increased synthesis of proteins, lipids and nucleic acids is a condition for the rapid proliferation of cancer cells[9], and changes in the bioavailability of these biomarkers can reveal important patterns of intracellular change. The IR-FEL on the ALICE accelerator at Daresbury Laboratory (Warrington, UK) is tuneable over the range of 5.5 μm to 8.8 μm (~1818 cm^{-1} to ~1136 cm^{-1}), which includes a number of biologically important biomarkers[10] at designated wavenumbers or wavelengths (Table 1). These biomarkers have previously been used to separate normal, low- and high-grade dyskaryosis and cancer cells from each other[2].

SNOM has been used to investigate the localisation of molecules within cell membranes of prostate cancer cells[11]. Further research demonstrated SNOM can accurately define both the cell surface and internal structures in both healthy and anomalous sperm, including the acrosome, nucleus and the organisation of mitochondria[12], and has demonstrated potential for single molecule imaging[13]. The application of SNOM to oesophageal cancer tissue studies provided evidence of its potentiality for cancer diagnosis[8]. The increased spatial resolution of SNOM has the potential to reveal and quantify highly localised cancer-related changes in cervical cells at the sub-cellular level (1-0.1 μm), and more accurately and precisely than

conventional IR techniques. The above described IR-SNOM studies were all carried out in reflection mode. To be best of our knowledge this is the first publication reporting data obtained using IR-SNOM in transmission mode, and using IR-SNOM to image whole cervical cells.

The aim of this pilot study was to assess the potential of SNOM in combination with an IR-FEL in the detection of the biophysical properties of cervical cell abnormalities. Spectra were also collected using traditional ATR-FTIR biospectroscopy to investigate the differences between techniques.

Results

We recruited 5 patients into this pilot study; the youngest aged 25 years (squamous & glandular pre-invasive lesions [CIN2, HGCGIN]) and the oldest 42 years (squamous lesion; low-grade dyskaryosis). Table 2 shows the characteristics for each patient; limited demography was available for the patient diagnosed with high-grade dyskaryosis.

Two out of the 5 patients were current smokers (high-grade dyskaryosis and CIN2 HGCGIN), and 1 patient was taking antibiotics (CIN2, HGCGIN). Four patients tested positive for HPV; none for HPV 18 and 2 for HPV 16. Both normal and high-grade dyskaryosis tested positive for HPV 'other' type (i.e., not high-risk HPV types 16 or 18); only normal had an abnormal high vaginal swab. All patients were tested for bacterial vaginosis and were normal.

A total of 34 cells were included in the SNOM images. The number of cells was evenly distributed for low-grade dyskaryosis, CIN2, HGCGIN and adenocarcinoma stage 1B1 (6, 5, and 5 cells each, respectively). Sixteen cells were imaged for normal and 2 for high-grade dyskaryosis, which was limited by the number of acceptable cells on the slide. Each SNOM scan comprised topographic, raw transmission (SNOM light) and IR (light) intensity reference images all collected simultaneously at a fixed wavelength. Example SNOM-IR-FEL

topography and associated transmission images for the pre-invasive lesion (CIN2, HGCGIN) is presented in Figure 6 (Methods; computational analysis). The topography and associated transmission images for the other 4 cells types is presented in Supplementary Figs: 1-4.

Changes in biomarkers at different grades of cervical dyskaryosis detected by SNOM-IR-FEL. When applied to spectra, Principal Component Analysis (PCA) identifies common sources of variance across spectra and collates them into a small number of dimensions (i.e., Principal Components [PCs]). Similar behaviour between samples 'nests' them closer together. PC score plots (Figure. 1) showed adequate separation for amide I between normal, high-grade dyskaryosis and CIN2, HGCGIN, with PC1 representing 84.19% and PC2 8.22% of the data variance (Figure. 1 [a]). The overlap observed for low-grade dyskaryosis and adenocarcinoma Stage 1B1 in the amide I band indicates there was insufficient information to achieve complete class segregation. Clean separation was observed for amide II between all samples, with the exception of normal and high-grade dyskaryosis, with PC1 and PC2 signifying 74.99% and 17.41%, respectively of the data variation (Figure. 1 [b]). There was considerable overlap for lipids for low- and high-grade dyskaryosis and adenocarcinoma Stage 1B1, with good separation observed between normal and CIN2, HGCGIN; PC1 was 76.39%, and PC2 was 16.75% (Figure. 1 [c]). All five cell types were distinguishable for DNA, although low- and high-grade dyskaryosis were very close to each other, with PC1 representing 84.28% and PC2 7.81% of the data variance (Figure. 1 [d]). Hotelling T^2 versus Q Residuals were plotted to assess how well the model described the samples with the optimal score for Hotelling T^2 being 100%, and the optimum score for Q Residuals being 0%. These plots show that all the samples fell with the 95% confidence limits and that no outliers were detected (Supplementary Fig.5). Validation of the PCA model was performed using Q Residuals to measure variation outside the PCA model for each sample according each biomarker response (Supplementary Fig. 6); and Hotelling T^2 was used to

measure variation within the PCA model for each sample according each biomarker response (Supplementary Fig. 7). All values fell within the 95% confidence limit and show there were no outliers. The T^2 and Q residuals plots show the data fits the model well.

The area of absorbance for each biomarker for each cell type is shown in Figure 2, with the percentage of area variation compared with normal shown in Table 3. For low-grade dyskaryosis, amide I and lipids were lower than normal cells (-73% and -31%, respectively) with large increases detected for amide II and DNA (143% and 111%, respectively). This pattern of decreased amide I and lipids was also observed for high-grade dyskaryosis (amide I: -94%, lipids: -78%), and similarly, amide II and DNA were higher (40% and 132%, respectively). All four biomarkers were higher for the pre-invasive squamous/glandular cells (CIN2, HGCGIN) than normal cells, with dramatic increases observed for amide II (509%) and DNA (1272%). Amide I was 38% higher and lipids were 93% higher. Conversely, the profile for adenocarcinoma Stage 1B1 was similar to low-grade dyskaryosis for amide I (-66%) and lipids (-47%). Adenocarcinoma Stage 1B1 was the only cell type in which a decrease in amide II was detected (-46%), and DNA availability was approximately half that detected for CIN2, HGCGIN (585%).

Spectra collected by ATR-FTIR spectroscopy for each cell type were very similar to each other, where the signal's difference appeared to be close to the instrument noise (Supplementary Fig. 8). However, the application of PCA was able to discriminate each cell type by class (Supplementary Fig. 9). PCA alone is often not enough to segregate out data classes or clusters sufficiently. By applying a supervised technique such as Linear Discriminant Analysis (LDA) to the PCA output as above (PCA-LDA), it promotes inter-class variation to be identified whilst preventing over-fitting of the data. PCA-LDA revealed good separation of classes, although 2 spectra from the normal set appeared within the CIN2, HGCGIN class (Figure 3).

Successive Projections Algorithm (SPA) is a variable selection technique that can produce models with good prediction ability. Used together with LDA (SPA-LDA), this technique produced similar results to PCA-LDA, although the clustering of cell types was more acute (Figure. 4), with discriminant wavenumbers being 1022, 1157, 1184, 1234, 1331, 1512, 1566, 1662, 2345 and 2939 cm^{-1}, and 8 of these occupying the fingerprint region (1800-900 cm^{-1}); (Figure 5). The tentative assignment of these wavenumbers in the fingerprint region to associated biomarkers is given in Table 4.

Discussion

Although the number of cells per patient per cell type is small, the results evidence the promise of the SNOM-IR-FEL technique in obtaining chemical information relevant to the detection of cervical cell abnormalities and cancer diagnosis at high spatial resolution. Clear trends in increased bioavailability of DNA were seen across all four disease cell types comparative to normal, with dramatic increases observed for the pre-invasive lesion of squamous/glandular neoplasia and for adenocarcinoma Stage 1B1. Both of these patients were infected with HPV 16, which is known to integrate into the host's DNA and produce a range of proteins that accelerate biochemical cascades that result in an overexpression of proto-oncogenes, stimulate rapid cell growth and increase the expression of proteins necessary for DNA replication[14,15]. It is possible that the increased DNA expressed in these cell types may be a combination of increased human and viral DNA. The dramatic increase in all biomarkers for the pre-invasive squamous/glandular lesion suggests a 'commitment' to carcinogenesis, whilst the mechanism behind the downregulation in proteins and lipids in adenocarcinoma is unclear, and may reflect the tumour achieving some form of steady-state, or that energy supply is exhausted. It should also be noted that 'normal' cells were infected with HPV 'other' type (i.e., not high-risk HPV type 16 or 18), and this patient had an abnormal high vaginal swab which may have influenced the profile observed here.

For the squamous lesions (low- and high-grade dyskaryosis), the results suggests the production of lipids and specific proteins featuring amide I bonding has been down-regulated. An amide bond is formed when two amino acids are joined at the C=O and N-H junction. Amide I is associated with the stretching of C=O bonds in polypeptides, whilst amide II is associated with the bending of the N-H bond. Previous vibrational work has established correlations between the frequencies of amides I/II to the secondary structure of polypeptides, which include α helices, β sheets, turns and undefined structure[16]. The increases in amide II for low-grade and high-grade dyskaryosis and the pre-invasive squamous/glandular lesion may be due to the increased production of specific polypeptides that feature amide II bonding in their secondary structures, and may represent a singular biomarker of importance. Previous work has shown the presence of proteins in the β sheet conformation has been linked in formation of the protein aggregates and fibrils observed in many human diseases, notably the amyloids seen in Alzheimer's disease[17].

Whilst lipids are associated with the proliferation of cancer cells[9], we detected increases only for the mixed pre-invasive squamous/glandular lesion. However, it should be noted that the wavelength for lipids at 5.71 μm lies near the limit of the IR-FEL where the beam intensity was low and shot-to-shot stability less good than for other wavelengths, which resulted in SNOM transmission images that had a lower signal-to-noise ratio. Overall, although the SNOM-IR-FEL technique was able to distinguish different cell types according to biomarkers, the delineation was less distinctive between low- and high-grade dyskaryosis than for normal, pre-invasive squamous/glandular lesion and adenocarcinoma Stage 1B1, which may be influenced by the small number of cells imaged. Given that the underlying normal cellular activity of squamous cells is different to mucus-producing glandular cells, it is not unreasonable to assume that the biological processes involved in the development of squamous cervical lesions may be different to glandular lesions.

As expected, spectra collected the traditional way using ATR-FTIR coupled with multivariate analysis, was able to segregate the cell types clearly into classes. Spectra for the normal sample were collected the day before those collected for the pre-invasive squamous/glandular sample; thus the appearance of two spectra within the pre-invasive class is likely due to natural variation of the normal sample rather than cross-contamination. The range of biomarkers available for investigation with ATR-FTIR was broader than that available for the SNOM-IR-FEL measurements in this experiment and encompasses the whole of the fingerprint region. Nonetheless, a large portion of this region was still available for investigation with SNOM-IR-FEL, where the wavelength range is dictated by the ALICE accelerator beam energy and FEL undulator gap settings.

Although SNOM has been shown to reveal cytoplasmic structures in previous studies[12], the SNOM-IR-FEL images obtained in this study were difficult to interpret in terms of structures. This was due to the use of a cleaved fibre for the SNOM imaging, which added significant tip artefacts to the topographical image and often off-set the topographic from the transmission images. Additionally, the method used to prepare the slides (cytospinning) resulted in many cells rupturing upon impact with the BaF_2 slide and left very few whole cells that were free of debris. However, care was taken to avoid any debris within the field of images as much as possible, and to ensure that all images contained the whole cells.

In terms of a diagnostic tool for use in routine patient screening, the technique of ATR-FTIR has low running costs and inexpensive consumables, together with a turnaround of 15-20 minutes per sample. The SNOM-IR-FEL technique has higher running (estimated at £250/hour) and consumables costs, and requires specialised fibre optics that need to be replaced regularly. In this experiment, the time required to collect each SNOM scan, at each wavelength, was approximately 80-100 minutes, depending upon the number of pixels obtained per image. Although building small accelerators on hospital sites is achievable, their

installation would require specialised personnel and a safe location away from the main hospital thoroughfare. Therefore, between the two techniques studied here, ATR-FTIR spectroscopy is the technique of choice for routine screening applications.

The results presented here demonstrate that the SNOM-IR-FEL technique is able to correctly identify cervical cell abnormalities using whole cells. In this study, the SNOM technique was used in 'low resolution' mode, which enabled direct comparison with the ATR-FTIR method. SNOM has a clear advantage over ATR-FTIR in terms of being a 'precision tool' that can be used to identify the location of biomarkers within the cell, leading to further understanding of how cancer develops and in identifying targets of therapeutic potential. It can be tuned to specific wavenumbers/wavelengths, which may help to exclude the 'noise' of other biomarkers that lie within close proximity, and makes the extraction of specific biomarkers more accessible.

Methods

Study population

Ethical approval was obtained from the National Research Ethics Service Committee London – Fulham (Approval number 13/LO/0126). This study was conducted according to the principles of the Declaration of Helsinki and all other applicable national or local laws and regulations. All patients gave written informed consent before any protocol-specific procedure was performed.

Patients were selected from a larger cohort of patients taking part in a larger study and were chosen based on their cytology and histology typing (worse grade) to match a diagnosis of 'normal', squamous lesions (low-grade dyskaryosis and high-grade dyskaryosis), pre-invasive mixed lesions involving both squamous and glandular cells (CIN2, HGCGIN), or developed glandular lesions (adenocarcinoma). We selected pre-menopausal, non-pregnant women of reproductive age (18-45 years of age) who were scheduled, if necessary, to undergo

local cervical treatment at Imperial College NHS Healthcare Trust. All samples were collected prior to treatment. The recruitment commenced in May 2013 and was completed in May 2015.

Patients were anonymised and assigned a unique identifier. We collected patient characteristics that included ethnicity, parity, smoking habits, antibiotic use within the last 2 weeks, phase in their cycle and use of contraception. The type of contraception and the time of their cycle (follicular or luteal) were documented. Medical and gynaecological history was collected for each patient including time since last sexual intercourse. For each patient, we collected data on the cytology, HPV DNA test and typing and histology, if available. Ethnicity was self-reported as Caucasian, Asian or Black.

Women who were HIV or hepatitis B/C positive, women with autoimmune disorders, and women that received pessaries within 14 days of sampling were excluded. Women with a previous history of cervical treatment were also excluded.

Sample collection

A sterile, disposable speculum was inserted, without lubricant, and a cervical sample of ThinPrep, liquid-based cytology (LBC) was taken from the cervix (ThinPrep, HOLOGIC Inc., Bedford, USA). This was analysed for cytological diagnosis and HPV DNA test and typing. HPV DNA test and 16/18 genotyping was carried out according to manufacturer's guidelines using the Abbott RealTime High Risk (HR) HPV assay on Abbott M2000 platform; a clinically validated in vitro polymerase chain reaction (PCR) assay with identification of HPV-16, -18 and 12 other HR HPV subtypes (31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68)[18]. From the remaining methanol-based fluid, 1 ml was stored at 4° centigrade at the Centre for Biophotonics, Lancaster University, England, until preparation for SNOM-IR-FEL and ATR-FTIR analysis.

Slide preparation

Each sample was agitated to disperse the cell pellet, and then a 500 µl aliquot was collected from the 1 ml sample into a clean micro tube. The 500 µl aliquots were centrifuged at 2000 rpm for 5 minutes and the ThinPrep supernatant was aspirated from above the pellet to remove its spectral signature (i.e., the methanol fixative). Each sample was re-suspended in 500 µl of distilled H_2O, agitated and centrifuged again. The supernatant was removed again and the wash step was repeated once more. For ATR-FTIR analysis, the final pellet was immersed in 100 µl of distilled H_2O, agitated and dispensed onto IR-reflective glass slides (Low-E; Kevley Technologies Inc., Chesterland, OH, USA) in a uniform spread of whole cells and allowed to bench dry for a minimum of 24 hours. Samples were then stored in a desiccator for a minimum of 48 hours to remove any residual water before spectral analysis.

For SNOM-IR-FEL analysis, the remaining 500 µl aliquot was washed as described above. If the final pellet was small, it was suspended in 500 µl of distilled H_2O, and larger pellets in 1000 µl of distilled H_2O. Each suspension was then agitated to disperse the pellet, and 5-6 drops added to a cytofunnel held in a cytoclip that had been pre-loaded with a barium fluoride (BaF_2) slide; (Crystan Ltd, Dorset, UK). Samples were spun at 3000 rpm for 5 minutes in a Cytospin™ 4 Cytocentrifuge (Thermo Fisher Scientific Inc., MA, USA) to disperse the cells in a single layer onto the slide. Slides were then housed in slide cartridges and kept in a desiccator until required.

SNOM and IR-FEL experimental set-up

The experiments were performed on the IR-FEL beamline at the ALICE energy recovery linear accelerator at Daresbury Laboratory[19, 20]. The wavelength of light from the FEL was selected by changing the undulator gap and, at the present accelerator settings, could be varied continuously from about 5.5 µm to 8.8 µm (~1818 cm^{-1} to ~1136 cm^{-1}), a range which covers a number of biologically important absorption bands. The IR-FEL operates at a macro-pulse repetition rate of 10 Hz, which limits, and determines, the rate of data collection. The IR light

from the FEL was transported to the experimental area via an evacuated beamline and exited the beamline through a KBr window. The intensity of the FEL radiation was attenuated using a set of polarisers and focussed onto the sample. A CaF_2 beam-splitter enables the FEL radiation to be split so that approximately 80% went to the SNOM and 20% was used as a reference signal. The reference signal was monitored with a single-element pyro detector.

The general principal of operation for the SNOM used in these experiments has been previously described[7]. In brief, the scanning tip is a specially prepared infrared-transmitting Chalcogenide glass fibre, where one end is etched to a sharp tip. Gold is then evaporated onto the tip so that it covers all but the very end, forming an aperture of 0.1-1 μm in diameter through which the light is collected. The fibre tip is then rastered over the surface of the sample, keeping the tip-to-sample distance constant with shear-force feedback. A single IR-FEL macro-pulse is used for each pixel of the images. The standard mode of operation for IR-SNOM is reflection, where the light approaches the sample at a grazing incidence angle of approximately 15° and the reflected light is collected by the fibre, transmitted through the fibre and detected using a liquid nitrogen cooled mercury-cadmium-telluride (MCT) detector.

Here we report the first measurements made in transmission mode, where the sample was illuminated through the slide; the light that was transmitted through the sample was collected by the fibre. For the measurements reported here, the fibre was cleaved and the entire 6-mm diameter fibre core was used to collect the infrared light signal so that a direct comparison could be made with standard IR techniques such as ATR-FTIR. The SNOM was incorporated into an inverted optical microscope, which was used to locate specific cells of interest on the sample and to position them within the SNOM scan area.

A BaF_2 slide containing the cells was mounted onto the SNOM and scans acquired at fixed wavelengths of 5.71 μm/~1750 cm-^{1} (lipids), 6.06 μm/~1650 cm^{-1} (amide I), 6.46 μm/~1550 cm^{-1} (amide II) and 8.16 μm/~1225 cm^{-1} (DNA-asymmetric phosphate) for each set of cells.

Topographic, raw transmission and intensity reference data were collected simultaneously, at a fixed wavelength, for each SNOM image scan.

Atomic force microscopy (AFM) imaging of cells

To further evidence that whole cervical cells had been used to collect the SNOM-IR-FEL data, atomic force microscopy (AFM) was performed on the adenocarcinoma stage 1B1 sample, using a Bruker Innova AFM in contact mode using silicon nitride probes of nominal spring constant 0.07 N/m (Supplementary Fig. 10). Topography and deflection (error signal) channels were recorded simultaneously. The contact force of the AFM tip on the cells was minimised to optimise image quality.

ATR-FTIR spectroscopy

Spectra were acquired using a Tensor 27 FTIR spectrometer with a Helios ATR attachment (Bruker Optik GmbH). Each spectrum comprised 32 scans at 8 cm-1 wavenumber spacing with 2 x interferogram zero-filling. Before the spectra were taken, the crystal was cleaned with distilled H_2O and inspected by video camera to be free of any contaminants. A background spectrum was acquired before the sample slide was mounted and the stage moved to bring the cervical cells in contact with the diamond. Spectra were collected from ten random sites on the slide. Spectra were converted to absorbance by Bruker OPUS software (Bruker Inc., Billerica, MA, USA).

Pre-processing of SNOM-IR-FEL images

The raw forward and backward SNOM transmission images were loaded into the freely available software Gwyddion 2.40, available at http://gwyddion.net/, and converted into text files ready for importing into MATLAB. A second set of raw data files and topographical images were converted into jpgs for image enhancement. No other pre-processing was performed other than file conversion.

SNOM-IR-FEL image enhancement

The images presented in Figure 6 and Supplementary Figures 1-4 were processed for presentation using Gwyddion 2.40 and a median height line correction in the horizontal (fast scan) axis, followed by the removal of high frequency noise using a two-dimensional Fourier transform. The computational analysis was performed using raw data.

Computational analysis:

SNOM-IR-FEL transmission images

The SNOM-IR-FEL transmission images were processed using MATLAB software 2014a and PLS Toolbox version 7.9.3 (Eigenvector Research, Inc., WA, USA). Each SNOM-IR-FEL data set (transmission images) comprised four matrixes with size of 150 x 150 corresponding to each biomarker response (Figure. 6). To obtain a spectrum-like signal profile from the biomarker response, the biomarker data matrix was converted into a vector by the mean calculation of the matrix in the column-mode direction (Equation. 1), where, s_j is an element of the row-vector s {1 x 150}, corresponding to the spectrum-like signal; *m* is the size of the image on column-mode direction; and x_{ij} is an element of the biomarker matrix X.

$$s_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}$$

Equation. 1

Thereafter, the spectrum-like signal was normalized by mean-centring and absolute value. The bar charts were made with the area of the spectrum-like signal integrated into an interval of spatial distribution according to the cell content position (Figure. 6).

PCA was performed with the whole spectrum-like signal using only mean-centring and absolute value as pre-processing. A summary of the computational steps in processing the data is given in Supplementary Fig. 11. PCA is an unsupervised technique commonly used as the first step in analysing large, multivariate data sets. Unsupervised techniques require no

information from the user but rely instead on an internal criterion to guide learning. In unsupervised learning, the system forms clusters (groupings, regions of data space). In general terms, PCA reduces the dimensionality of large data sets and using mathematical projection, the original data set which may have involved many variables, can often be interpreted in just a few variables (the principal components; PCs). This reduced dimensional data set will allow the user to spot trends, patterns and outliers in the data, far more easily than would have been possible without performing the PCA. When applied to spectra, PCA identifies common sources of variance across spectra and collates them into a small number of dimensions. PCA is often not enough to segregate out data classes or clusters sufficiently. By applying a supervised technique such as LDA to the PCA output, it promotes inter-class variation to be identified whilst preventing over-fitting of the data.

PCA was executed using the average signal of each biomarker (triplicate) for five samples, one for each type of cell morphology: normal, low-grade dyskaryosis, high-grade dyskaryosis, CIN2, HGCGIN and adenocarcinoma Stage 1B1. Additionally, the area for each biomarker for each cell type was determined, as was the percentage area variation from 'normal' for each biomarker for each cell type.

ATR-FTIR spectra

The ATR-FTIR data were analysed using multivariate techniques of PCA for preliminary data reduction, and the output was processed using LDA and a variable selection technique employing Successive Projections Algorithm (SPA)[21], in conjunction with LDA for selecting an appropriate subset of wavenumbers for classification purposes. SPA is a variable selection technique specifically designed to improve the conditioning of multiple linear regression by minimizing collinearity effects in the calibration data set and can result in models with good prediction ability[22].

The classic Kennard–Stone (KS) uniform sampling algorithm[23] was adopted to divide the available samples into training (70%), validation (15%) and prediction sets (15%) for construction and validation of the PCA-LDA and SPA-LDA models. The training set was used to obtain model parameters (including variable selection for LDA), and the validation set was employed to choose the best number of the PCs for PCA model and to guide the variable selection. The optimum number of variables for SPA–LDA was used to select variables employing the G function as cost function[23].

References

37. Castellsagué, X., Bosch, F. X., & Muñoz, N. Environmental co-factors in HPV carcinogenesis. *Virus Res.* **89,** 191-199 (2002).

38. Gajjar, K. *et al*. 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,** e82416: doi: 10.1371/journal.pone.0082416 (2014).

39. Purandare, N. C. *et al*. Infrared spectroscopy with multivariate analysis segregates low-grade cervical cytology based on the likelihood to regress, remain static or progress. *Anal. Methods.* **6,** 4576-4584 (2014).

40. Purandare, N. C. *et al*. Biospectroscopy insights into the multi-stage process of cervical cancer development: probing for spectral biomarkers in cytology to distinguish grades. *Analyst.* **138,** 3909-3916 (2013).

41. Lima, K. M. G. *et al*. Classification of cervical cytology for human papilloma virus (HPV) infection using biospectroscopy and variable selection techniques. *Anal. Methods.* **6,** 9643-9652 (2014).

6. Harrison, A. J., Bilgili, E. A., Beaudoin, S. P., & Taylor, L. S. Atomic force microscope infrared spectroscopy of griseofulvin nanocrystals. *Anal. Chem.* **85,** 11449-11455 (2013).

7. Cricenti, A. *et al.* Very high resolution near-field chemical imaging using an infrared free electron laser. *Phys. Chem. Chem. Chem.* **4,** 2738–2741 (2002).

8. Smith, A. D. *et al*. Near-field optical microscopy with an infra-red free electron laser applied to cancer diagnosis. *Appl. Phys. Lett.* **102**, 053701 (2013); doi: 10.1063/1.4790436.

9. Baenke, F., Peck, B., Miess, H., & Schulze, A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Dis. Model Mech.* **6,** 1353-1363 (2013).
10. Movasaghi, Z., Rehman, S., & ur Rehman, I. Fourier transform infrared (FTIR) spectroscopy of biological materials. *Appl. Spectosc. Rev.* **43,** 134-179 (2008).
11. Walker, K-A. D., Morgan, C., Doak, S. H., & Dunstan, P. R. Quantum dots for multiplexed detection and characterisation of prostate cancer cells using a scanning near-field optical microscope. *PLoS ONE*. **7,** e31592. doi:10.1371/journal.pone.0031592 (2012).
12. Andolfi, L. *et al*. The application of scanning near field optical imaging to the study of human sperm morphology. *J Nanobiotechnology*.**13,** doi: 10.1186/s12951-014-0061-5 (2015).
13. Zhong, L., Wentao, L., Xiaopin, W., & Cai, J. Detection the specific marker of CD3 molecules of human peripheral T lymphocytes using SNOM and quantum dots. *Colloids and Surfaces A: Physiochem. Eng. Aspects.* **313-314,** 642-646 (2008).
14. Tsai, T-C, & Chen, S-L. The biochemical and biological functions of human papillomavirus type 16 E5 protein. *Arch Virol.* **148,** 1445-1453 (2003).
15. Munger, K. *et al*. Mechanisms of Human papillomavirus-induced oncogenesis. *Virol.* **78,** 11451-11460 (2004).
16. Krimm, S., & Jr. Reisdorf, W.C. Understanding normal modes of proteins. *Faraday Discuss.* **99,** 181-197 (1994).
17. Nelson, R. *et al*. Structure of the cross-beta spine of amyloid-like fibrils. *Nature*. **435**, 773–778 (2005).
18. Coutlée, F., Rouleau, D., Ferenczy, A., & Franco, E. The laboratory diagnosis of genital human papillomavirus infections. *Can J Infect Dis Med Microbiol*. **16,** 83-91 (2005).

19. Thompson, N.R. *et al*. First lasing of the ALICE infra-red free-electron laser. *Nucl. Instrum. Methods Phys. Res.* **A 680,** 117-123 (2012).
20. Thompson, N.R. *et al*. Status of the ALICE IR-FEL: from ERL demonstrator to user facility. *International Free Electron Laser Conference - FEL 2015*, Daejeon, Korea, 2014: TUP015.
21. Theophilou, G., Lima, K.M.G., Martin-Hirsch, P.L., Stringfellow, H.F., & Martin, F.L. ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer. *Analyst*. **141**, 585-594 (2016).
22. Galvão, R.K.H. *et al*. Cross-validation for the selection of spectral variables using the successive projections algorithm. *J. Braz. Chem. Soc*. **18,** 1580-1584 (2007).
23. Kennard, R.W., & Stone, L.A. Computer aided design of experiments. *Technometrics*, **11**, 137-148 (1969).

Acknowledgements

We thank all participants in this study. Our work was supported by the British Society of Colposcopy Cervical Pathology (Jordan & Singer Award), the Imperial College Healthcare Charity, Genesis Research Trust and the Imperial Healthcare NHS Trust NIHR Biomedical Research Centre. This work was supported by The Engineering and Physical Sciences Research Council (EPSRC; Grant No: EP/K023349/1). We would like to thank colleagues from the Cockcroft Institute, STFC's Daresbury Laboratory and the A.L.I.C.E operations team. The authors wish to express their thanks to the staff at the Pathology Department, Royal Preston Hospital, Lancashire, for their kind support in the preparation of the slides. C.L.M. Morais would like to acknowledge the financial support from CAPES-Brazil. K.M.G. Lima acknowledges the CNPq (Grant 305962//2014–4) for financial support.

Author Contributions

D.E.H and G.T. prepared the samples and slides for the SNOM-IR-FEL experiment. C.L.M.M. and K.M.G.L. developed the mathematical and computational analyses for both the SNOM-IR-FEL and ATR-FTIR data. J.T. reviewed the original raw data and provided early guidance on potential approaches for analysis. M.R.F.S-K., T.C., J.I., D.S.M., and P.W. performed the SNOM-IR-FEL experiment and were also second commissioners along with D.E.H and K.H. at the A.L.I.C.E facility. D.S.M performed the AFM measurements. A.C. and M.L. originally developed the SNOM-IR-FEL technique and continue to develop its capability. D.E.H prepared the samples for ATR-FTIR spectral acquisition and collected the patient spectra. M.K and A.M. recruited patients and collected clinical information for the study. M.K., E.P. and P.L.M-H contributed the clinical evaluation of the results. All authors critically reviewed the paper. F.L.M conceived and directed the project and wrote the manuscript with the help of D.E.H.

Additional information

Supplementary Information accompanies this paper.

Competing financial interests: The authors declare no competing financial interests.

Figure legends

Figure 1. Transmission SNOM-IR-FEL: PC score plots. Scores for 1st and 2nd PCs for the type of cells according to each biomarker response: (**a**) Amide I; (**b**) Amide II; (**c**) Lipids; (**d**) DNA. Dotted line indicates 95% confidence limits and shows there were no outliers. CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; PC: Principal components; SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.

Figure 2. **Transmission SNOM-IR-FEL: bar chart for each type of cell according to the biomarker responses.** CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.

Figure 3. ATR-FTIR Spectroscopy: Discriminant Function (DF) plot for PCA-LDA (6 PC's). This technique promoted a better clustering of the cell types than using PCA alone (Supplementary Fig. 9). Dotted regions indicate each class. CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; PCA-LDA: Principal component analysis coupled to linear discriminant analysis.

Figure 4. ATR-FTIR spectroscopy: Discriminant Function (DF) plot for SPA-LDA. Using this technique, cells types were clustered more acutely than that observed for PCA-LDA. Dotted regions indicate each class. CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; SPA-LDA: Successive projections algorithm in conjunction with linear discriminant analysis.

Figure 5. ATR-FTIR spectroscopy: average raw spectrum and its selected variables (circled) by SPA-LDA. SPA-LDA: Successive projections algorithm in conjunction with linear discriminant analysis.

Figure 6. Transmission IR-SNOM images (400 µm x 400 µm) of the same cell sampled from pre-invasive lesion (CIN2, HGCGIN) at the different biomarker wavelengths: (**a**) Amide I - 6.06 µm (~1650 cm^{-1}), the horizontal line in (**a**) is shown in cross section in (**b**), (**c**) Amide II - 6.46 µm (~1550 cm^{-1}), (**d**) DNA - 8.16 µm (~1225 cm^{-1}) and (E) Lipids - 5.71 µm (1750 cm^{-1}). The colour scale bar arrow indicates increasing biomarker absorption. The shaded region in (**b**) corresponds to the interval selected for area calculation according to the cell content. CIN2, HGCGIN: Cervical intra-epithelial neoplasia, high-grade cervical glandular intraepithelial neoplasia.

Tables

Wavenumber (cm^{-1}) and associated biomarker	Wavelength (µm)
~1225 (DNA – asymmetric phosphate)[a]	8.16
~1650 (Amide I of proteins predominantly in α helix conformation)[a]	6.06
~1550 (Amide II of proteins predominantly in β sheet conformation)[a]	6.46
~1750 (Lipids)[a]	5.71

Table 1. Important biomarkers (wavenumbers and associated wavelength). [a] Movasaghi *et al,* 2008[10];

N.B.: The signal at a particular wavenumber could have contributions from more than one biomarker. Amides I and II are linked to the secondary structure of proteins and are indicative of their bioavailability.

Characteristics	**Normal** (n=1)	**Low-grade dyskaryosis** (Squamous) (n=1)	**High-grade dyskaryosis[a]** (Squamous) (n=1)	**CIN2, HGCGIN Pre-invasive (Squamous & glandular)** (n=1)	**Adenocarcinoma Stage 1B1** (Glandular) (n=1)
No of cells imaged	**16**	**6**	**2**	**5**	**5**
Age	31	42	Unknown	25	36
Ethnicity	Caucasian	Caucasian	Unknown	Caucasian	Caucasian
Smoker	No	No	Yes	Yes	No
Co-morbidities	Reflux	No	No	Acne	No
Current medications	Omeprazole	Nil	Unknown	Erythromycin	Nil
Recent antibiotics (within last 2 weeks)	No	No	No	Yes	No
Recent pessaries	No	No	No	No	No
Contraception	Nil	Copper IUD	COCP	COCP	Condoms
48 hours since last intercourse	Yes	Yes	Unknown	Yes	Yes
Phase of menstrual cycle	Follicular	Follicular	Luteal	Luteal	Unknown
Vaginal pH	4.4	5	Unknown	4.4	4.4
Mid-stream specimen of urine	Mixed growth	Mixed growth	No growth	No growth	No growth
High vaginal swab	Abnormal	Normal	Normal	Normal	Normal
Cytology, histology, HPV					
Referral smear	Negative	Moderate	Unknown	Severe	Severe
Biopsy	NA	CIN1	CIN2	Micro-invasive SMILE	HGCGIN
Cone	NA	HPV	CIN3	CIN2, HGCGIN	Adenocarcinoma Stage 1B1
HPV positive test	Positive	Unknown	Positive	Positive	Positive
HPV 18	No	Unknown	No	No	No
HPV 16	No	Unknown	Unknown	Yes	Yes
HPV other type	Yes	Unknown	Yes	No	No

Table 2. Patient characteristics. [a] There was limited data available for the patient diagnosed with High-grade dyskaryosis. COCP: Combined oral contraceptive pill: HPV: Human Papillomavirus; CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; IUD: Intrauterine device; POP: Progesterone-only pill: NA: Not applicable; SMILE: Stratified Mucinous Intraepithelial Lesion.

Cell type	ΔA (%)			
	Amide I	Amide II	Lipids	DNA
Normal	--	--	--	--
Low-grade dyskaryosis	-73	143	-31	111
High-grade dyskaryosis	-94	40	-78	132
CIN2, HGCGIN	38	509	93	1272
Adenocarcinoma Stage 1B1	-66	-46	-47	585

Table 3. Percentage of area variation (ΔA (%)) from the 'normal' cell morphology for each biomarker for the type of cell. CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia.

Wavenumber (cm^{-1})	Corresponding wavelength (μm)	Tentative assignment of biomarkers[a]
~1022	9.8	Glycogen
~1157	8.6	C-O Proteins and carbohydrates
~1184	8.4	Amide III; deoxyribose
~1234	8.1	Amide III as well as phosphate
~1331	7.5	vibration of nucleic acids
~1512	6.6	Polysaccharides; collagen
~1566	6.4	Amide II
~1662	6.02	Amide I; ring base

Table 4. Tentative assignment of biomarkers to wavenumbers. [a] Movasaghi *et al*, 2008[9]; N.B.: The signal at a particular wavenumber could have contributions from more than one biomarker.

Figure 1

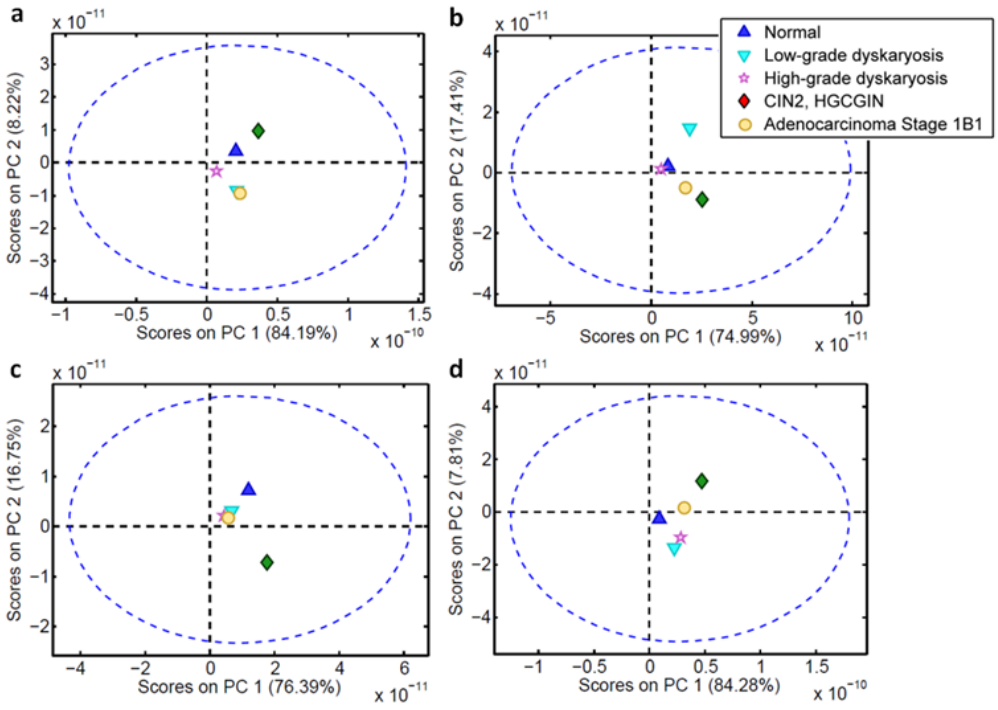


Figure 2

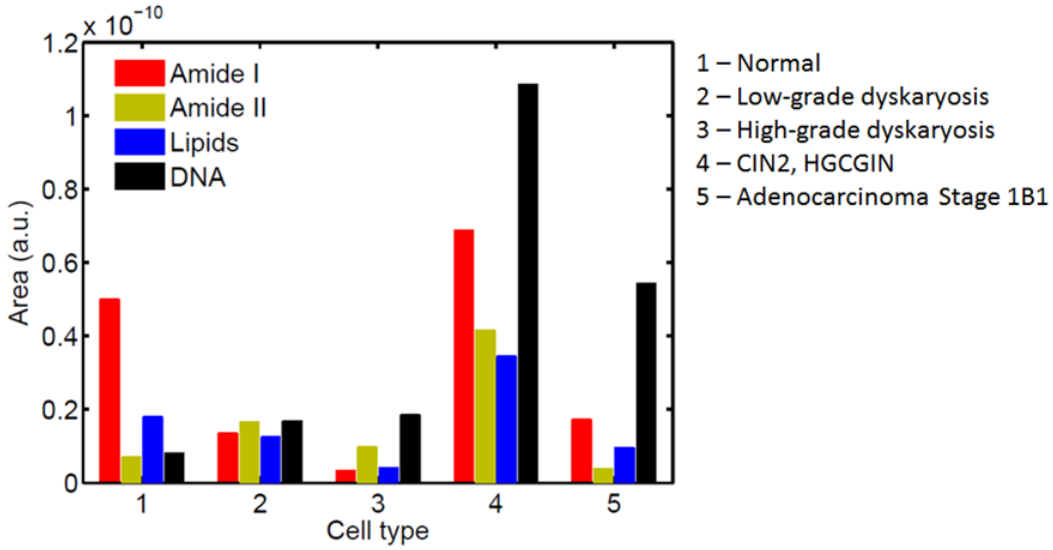


Figure 3

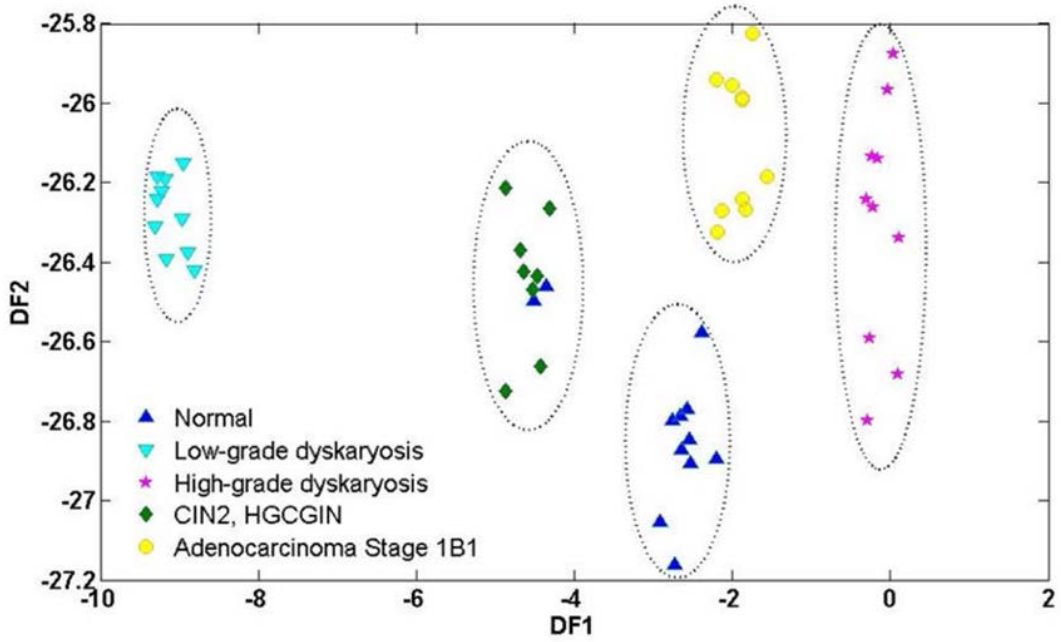


Figure 4

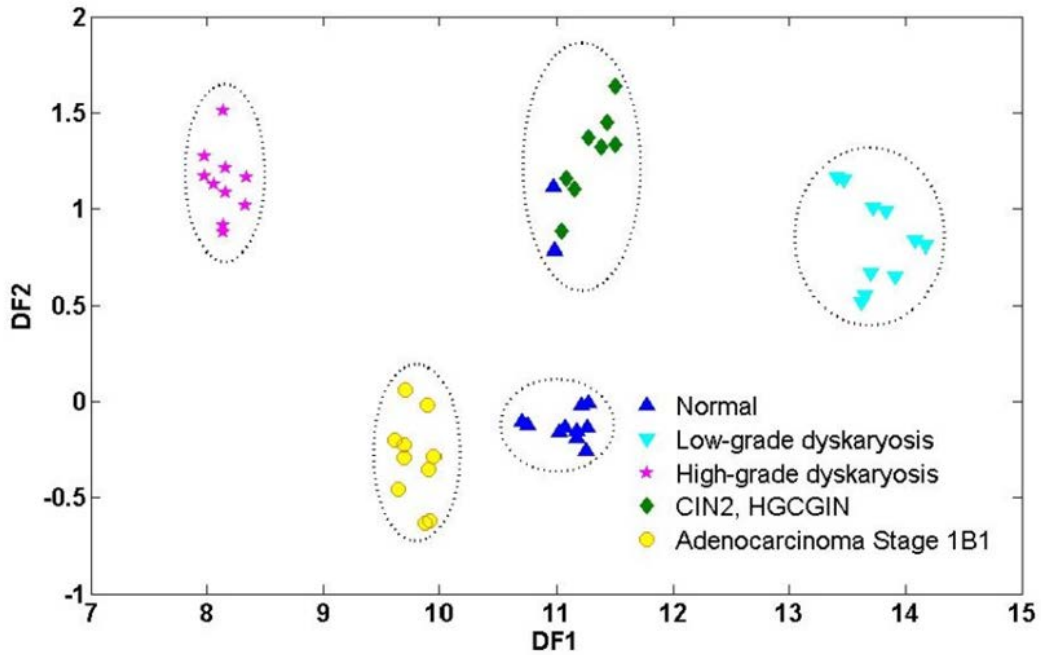


Figure 5

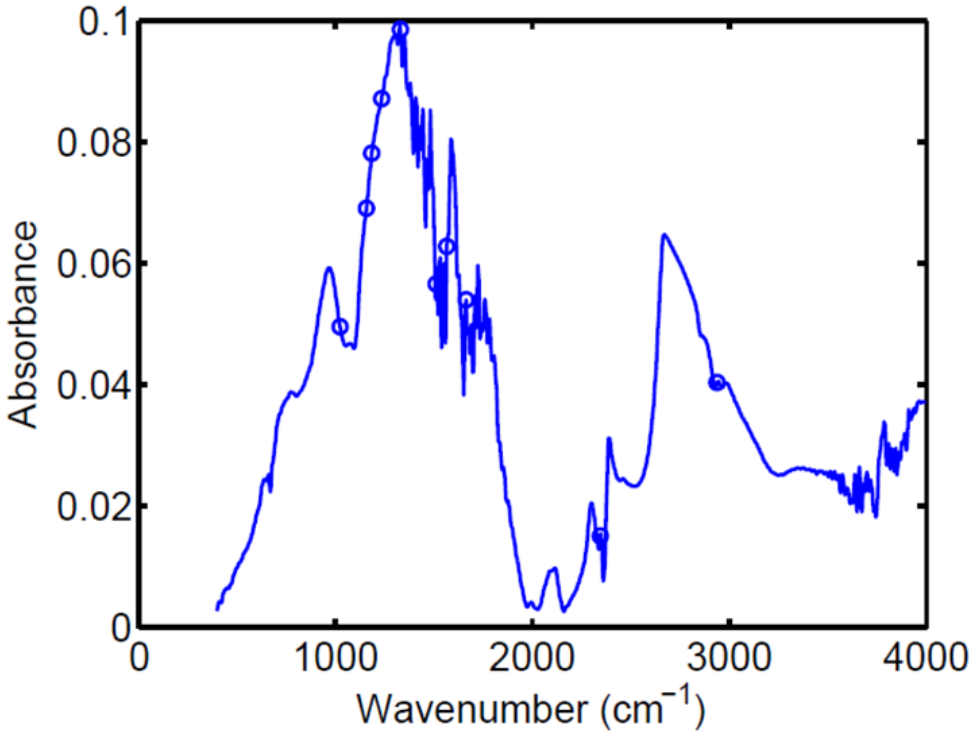
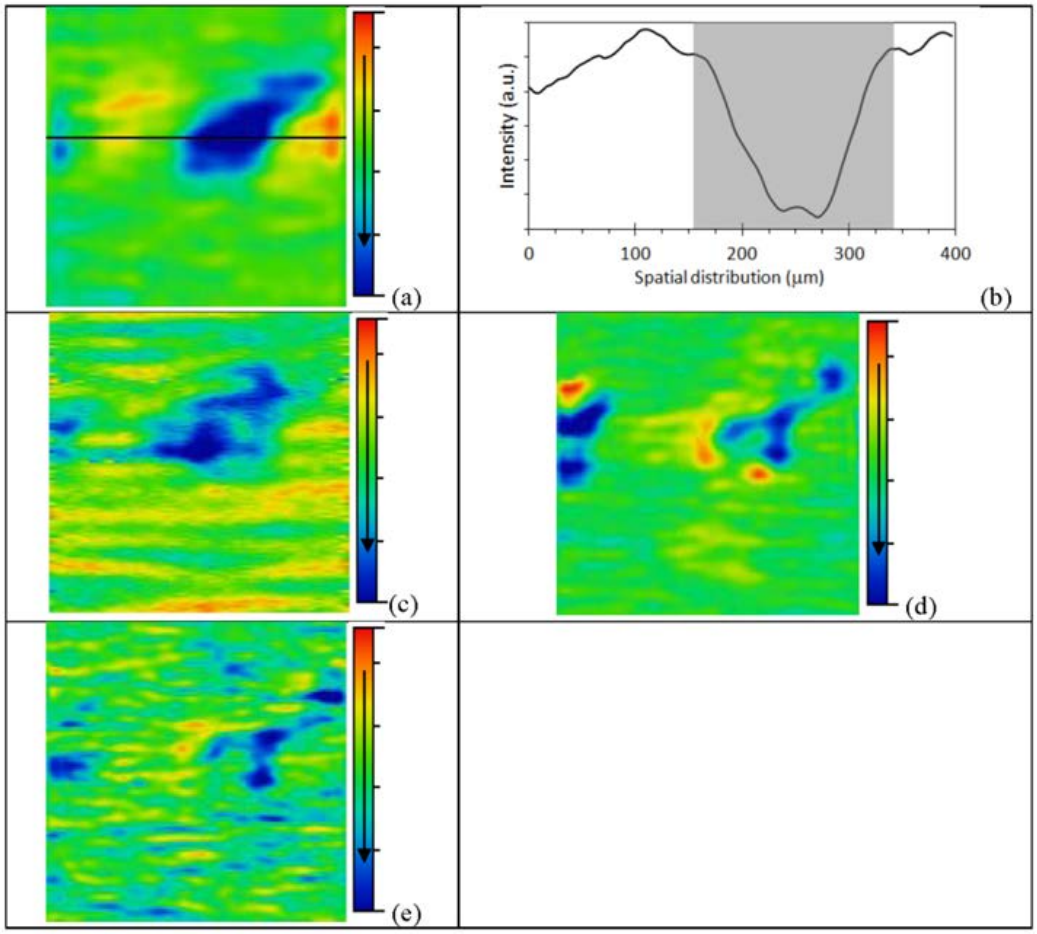

0.1
0.08
0.06
0.04
0.02
0
Absorbance
0
1000
2000
3000
4000
Wavenumber (cm^{-1})

Figure 6

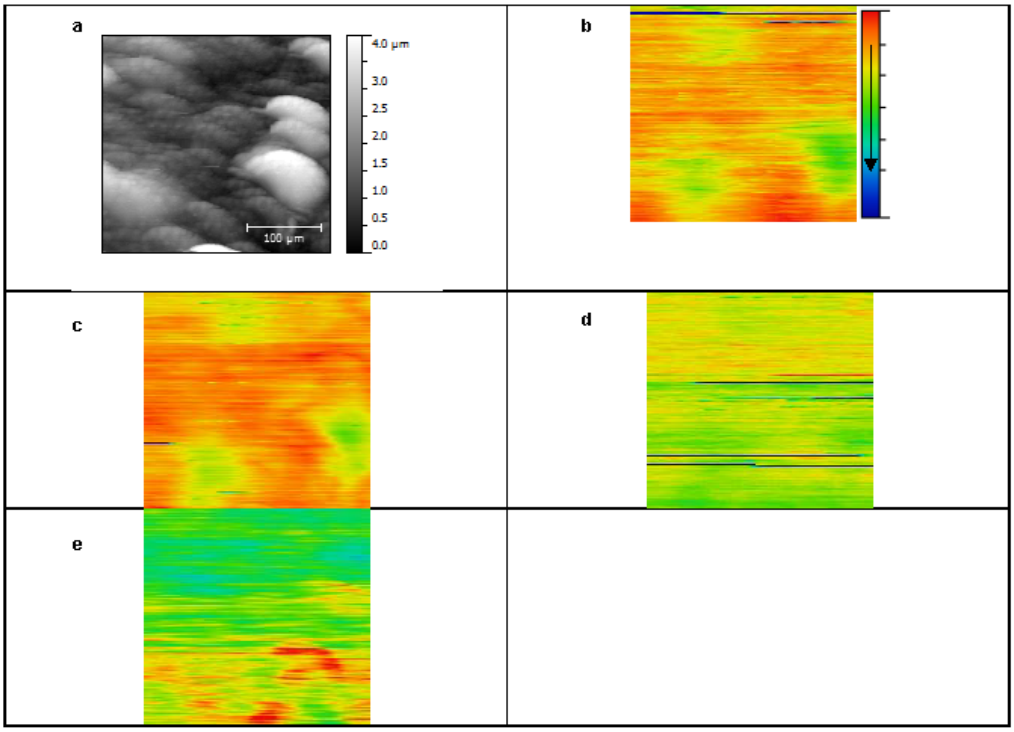

Intensity (a.u.)
0
100
200
300
400
Spatial distribution (μm)
(a)
(b)
(c)
(d)
(e)

Scientific Reports

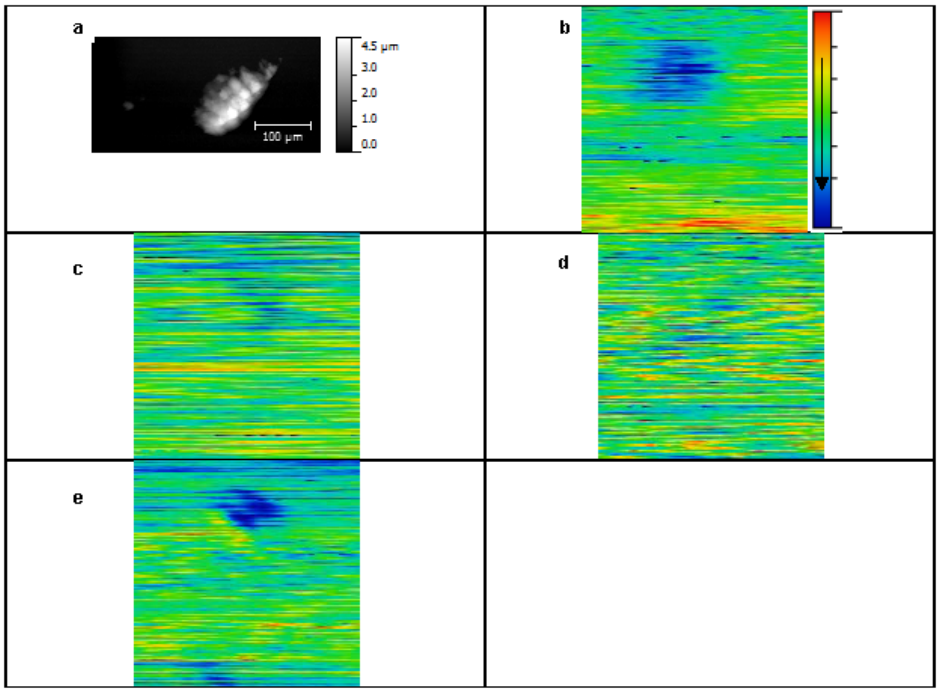
Electronic Supplementary Information

Imaging cervical cytology with scanning near-field optical microscopy (SNOM) coupled with an IR-FEL

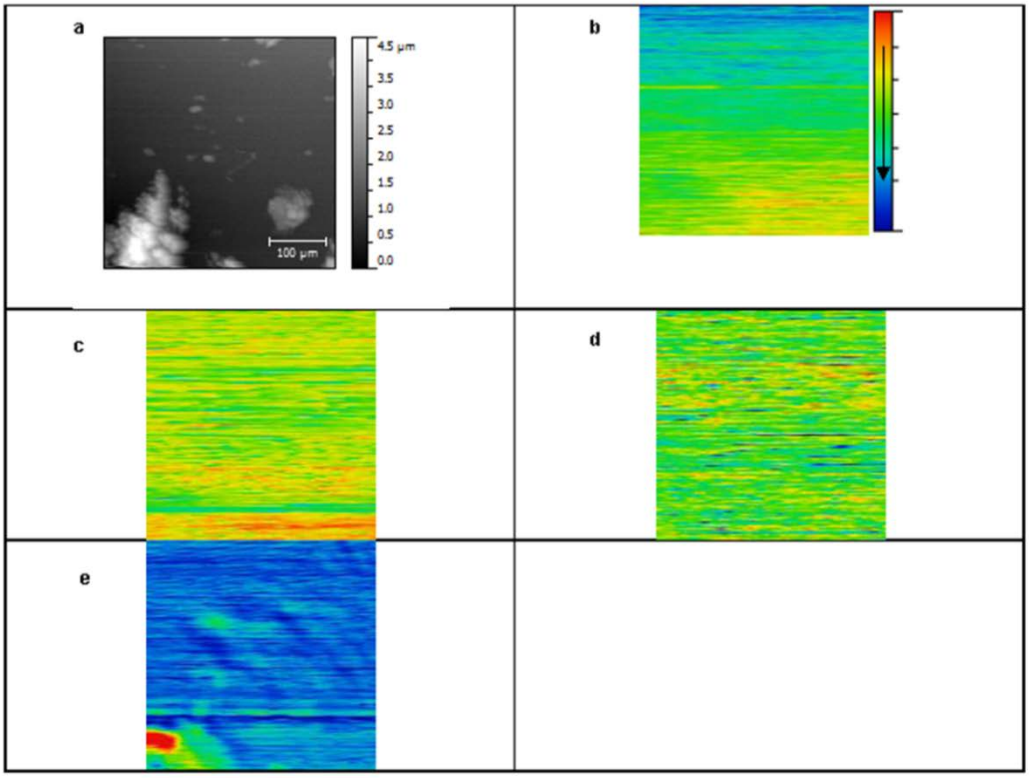
Diane E. Halliwell,[a] Camilo L. M. Morais,[b] Kássio M. G. Lima,[b] Julio Trevisan,[c] Michele R. F. Siggel-King,[de] Tim Craig,[d] James Ingham,[d] David S. Martin,[d] Kelly Heys,[a] Maria Kyrgiou,[f,g] Anita Mitra,[f,g] Evangelos Paraskevaidis,[h] Georgios Theophilou,[i] Pierre L. Martin-Hirsch,[a,j] Antonio Cricenti,[k] Marco Luce,[k] Peter Weightman,[d] and *Francis L. Martin[a]



Supplementary Figure 1. **SNOM-IR-FEL images of normal cells:** (**a**) topography; transmission images: (**b**) Amide I; (**c**) Amide II; (**d**) Lipids; (**e**) DNA. The colour scale bar arrow in (**b**) applies to (**b-e**) and indicates increasing biomarker absorption. SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.

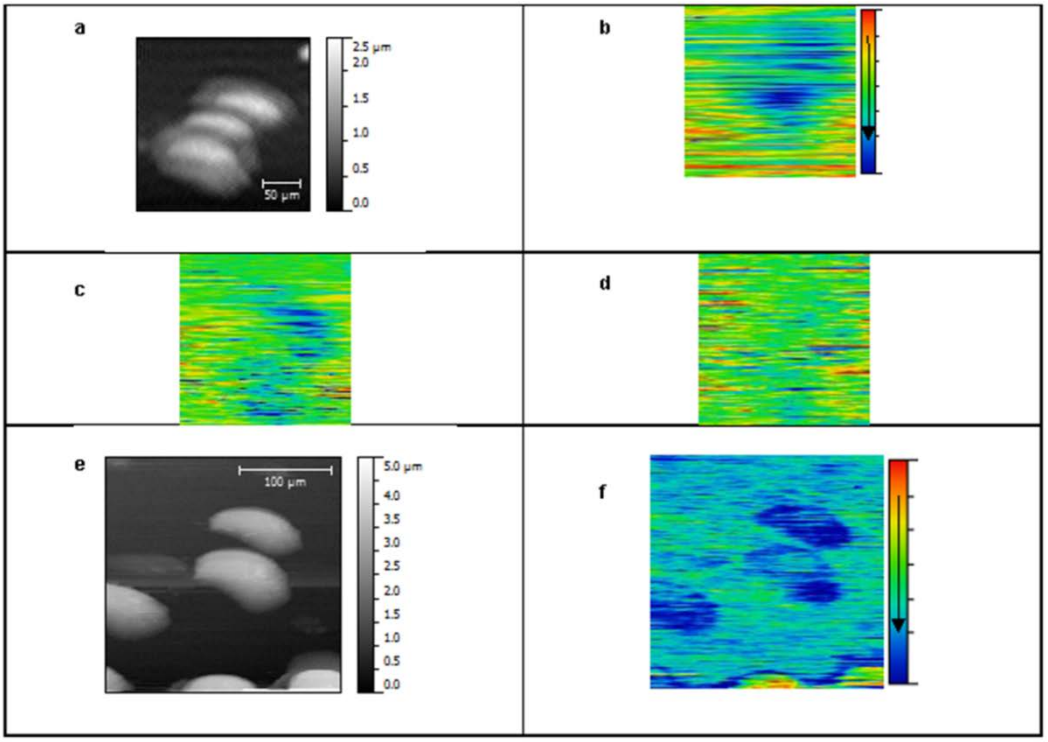


Supplementary Figure 2. **SNOM-IR-FEL images of low-grade dyskaryosis:** (**a**) topography; transmission images: (**b**) Amide I; (**c**) Amide II; (**d**) Lipids; (**e**) DNA. The colour scale bar arrow in (**b**) applies to (**b-e**) and indicates increasing biomarker absorption. SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.

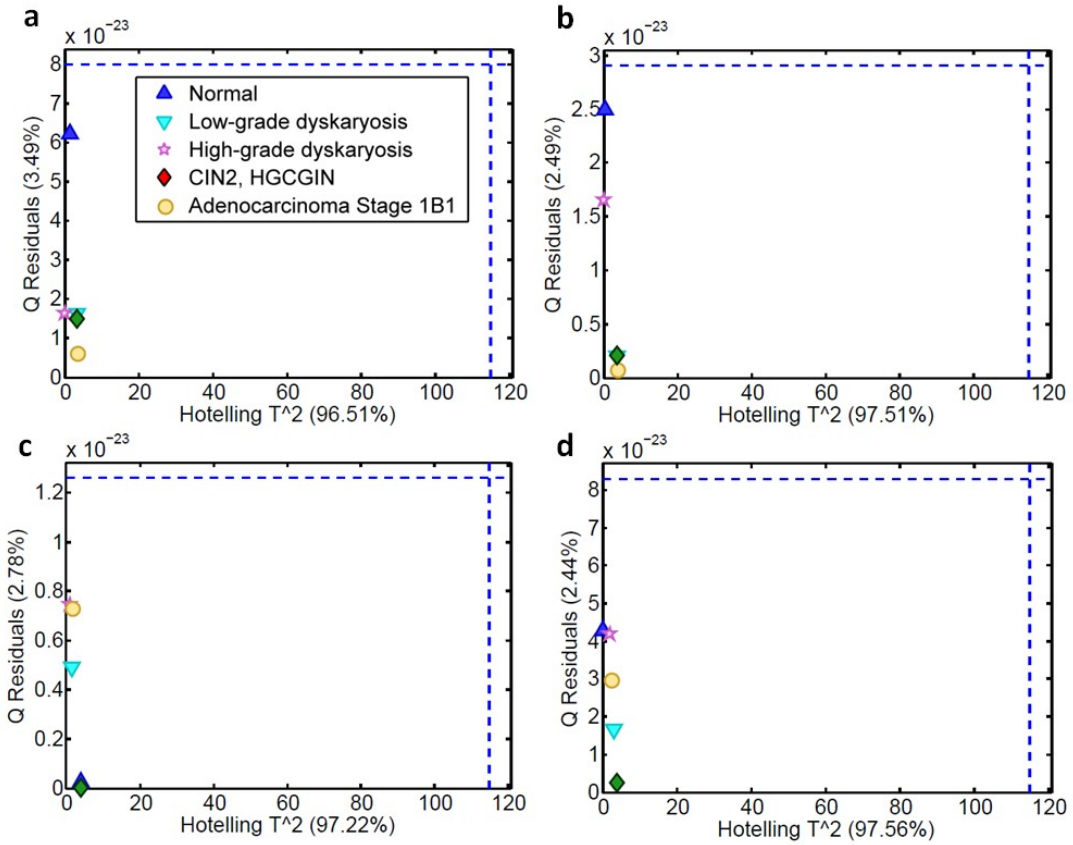


Supplementary Figure 3. **SNOM-IR-FEL images of high-grade dyskaryosis**: (**a**) topography; transmission images: (**b**) Amide I; (**c**) Amide II; (**d**) Lipids; (**e**) DNA. The colour scale bar arrow in (**b**) applies to (**b-e**) and indicates increasing biomarker absorption. SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.

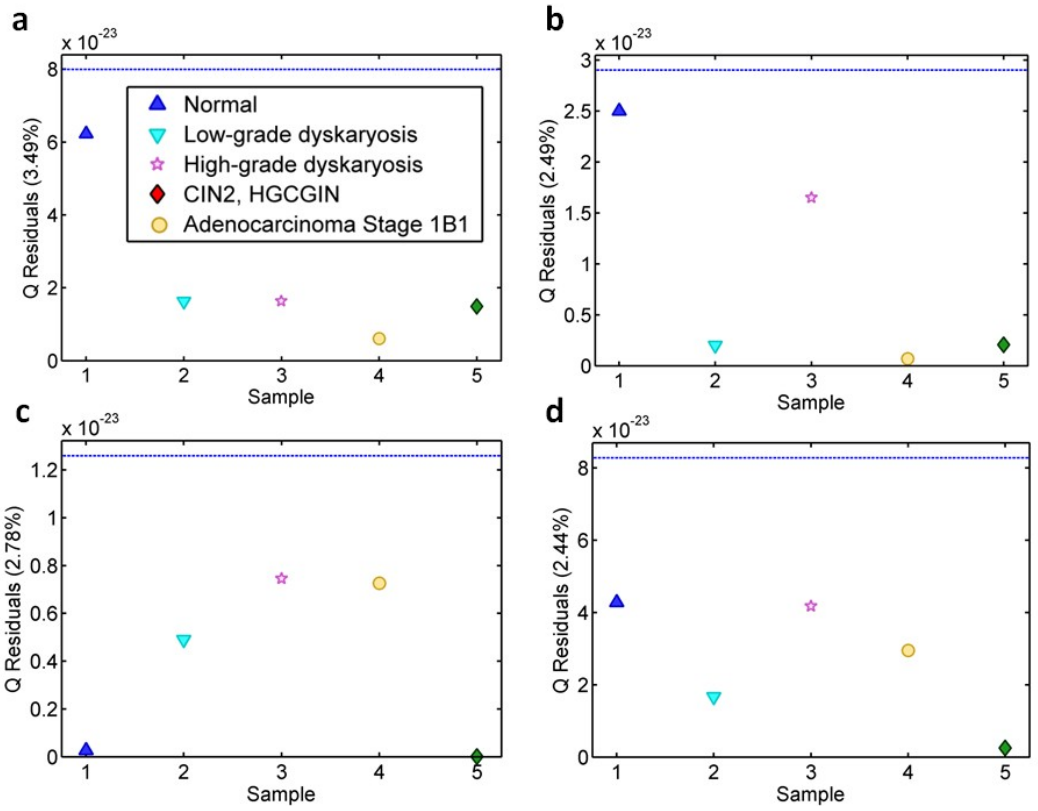
The SNOM-IR-FEL images and associated topography of the pre-invasive lesion (CIN2, HGCGIN) are presented in the main body of the text (see Figure 6).



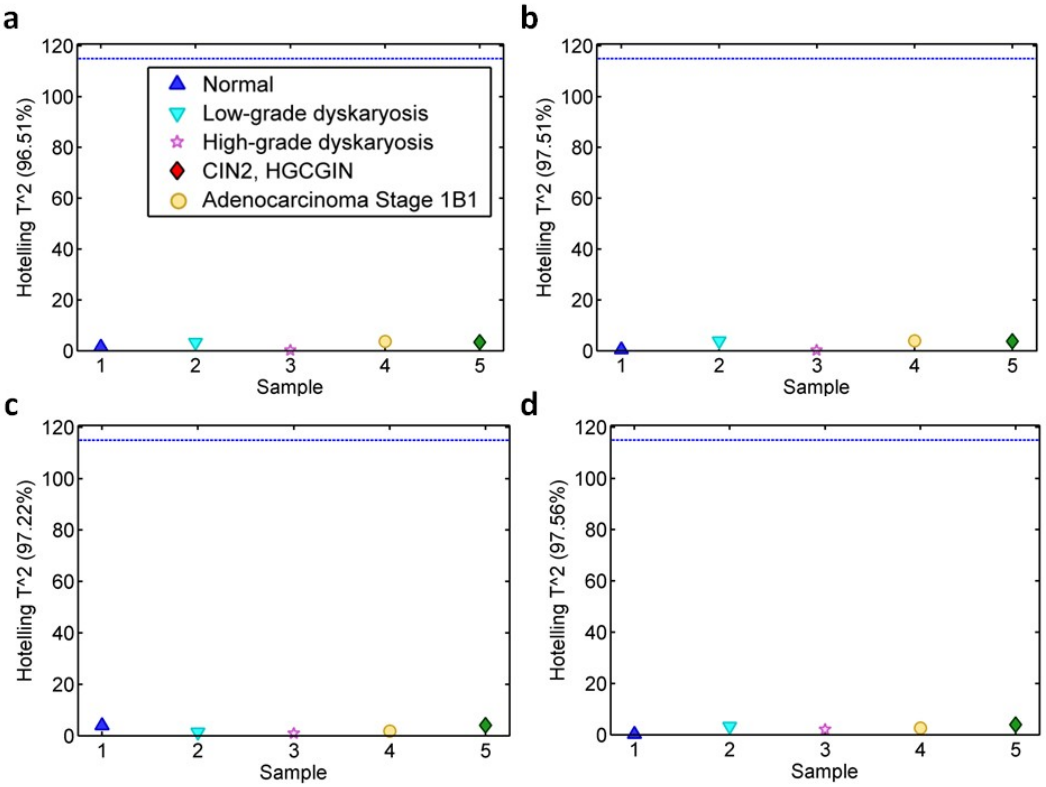
Supplementary Figure 4. **SNOM-IR-FEL images of adenocarcinoma Stage 1B1:** (**a**) topography; transmission images: (**b**) Amide I (imaged from different site to topography shown here); (**c**) Amide II; (**d**) Lipids. (**e**) Topography of cells from a second area and (**f**) the corresponding SNOM transmission image for the DNA biomarker. The colour scale bar arrow in (**b**) applies to (**b-d, f**) and indicates increasing biomarker absorption. SNOM-IR-FEL: Scanning near-field optical microscopy coupled to an infrared-free electron laser.



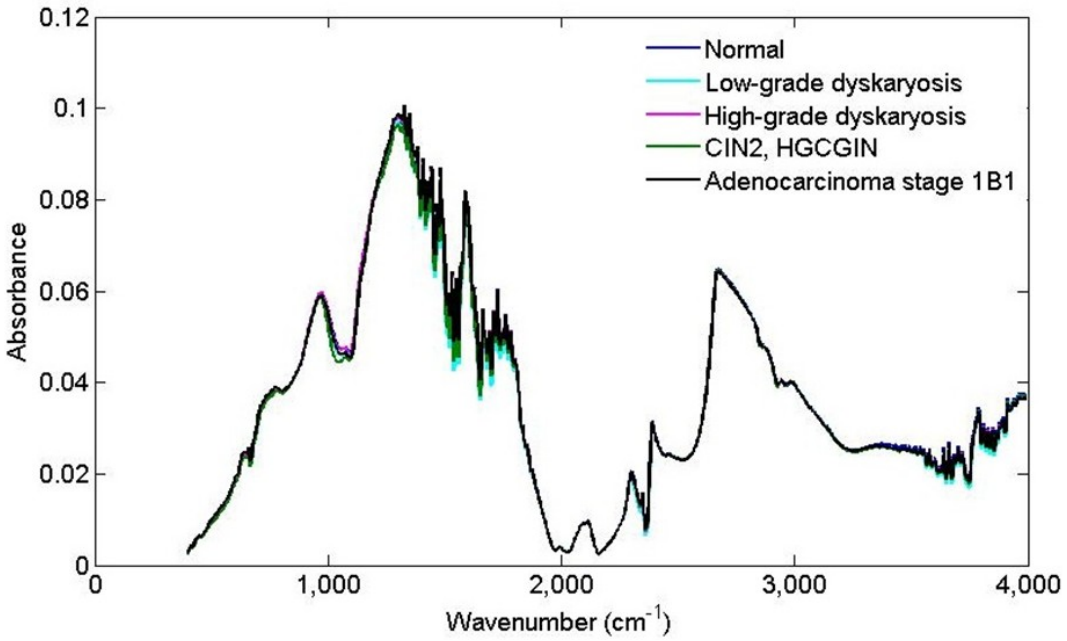
Supplementary Figure 5. Transmission SNOM-IR-FEL: Hotelling T^2 versus Q Residuals graphs for the type of cells according to each biomarker response: (**a**) Amide I; (**b**) Amide II; (**c**) Lipids; (**d**) DNA. All 5 samples fell within the 95% confidence limits (blue dotted line), and shows there were no outliers. The score for Hotelling T^2 ranged from 96.51% to 97.56%; whilst the score for Q residuals ranged from 2.44% and 3.49%. CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.



Supplementary Figure 6. Transmission SNOM-IR-FEL: Validation of the PCA model using Q Residuals to measure variation outside the PCA model for each sample according each biomarker response: (**a**) Amide I; (**b**) Amide II; (**c**) Lipids; (**d**) DNA. The optimal score for Q Residuals is 0% and here ranged from 2.44% to 3.49%. All 5 samples fell within the 95% confidence limits (blue dotted line), shows there were no outliers and that the data fits the model well. CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.

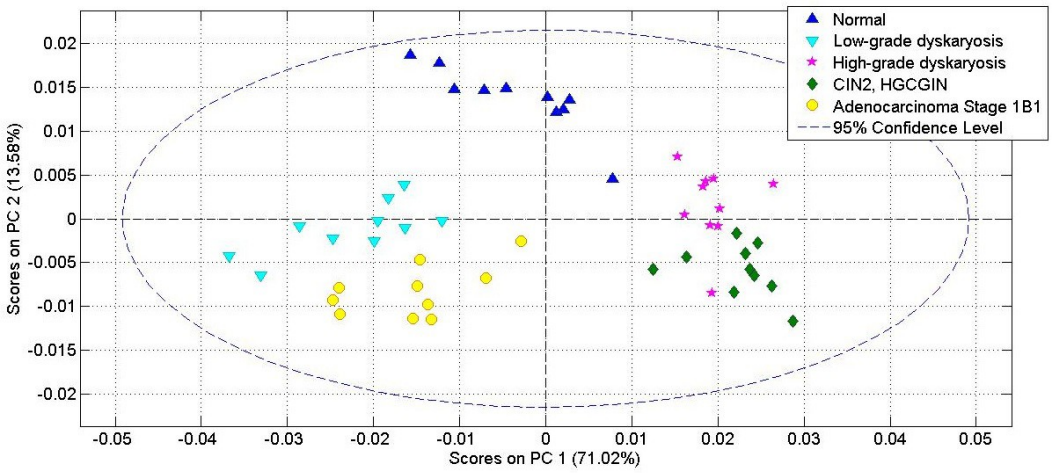


Supplementary Figure 7. **Transmission SNOM-IR-FEL:** Validation of the PCA model using Hotelling T^2 to measure variation within the PCA model for each sample according each biomarker response: (**a**) Amide I; (**b**) Amide II; (**c**) Lipids; (**d**) DNA. The optimal score for Hotelling T^2 is 100% and here ranged from 96.51% to 97.56%. All 5 samples fell within the 95% confidence limits (blue dotted line), shows there were no outliers and that the data fits the model well. CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; SNOM-IR-FEL: Scanning near-field optical microscopy coupled with an infrared-free electron laser.

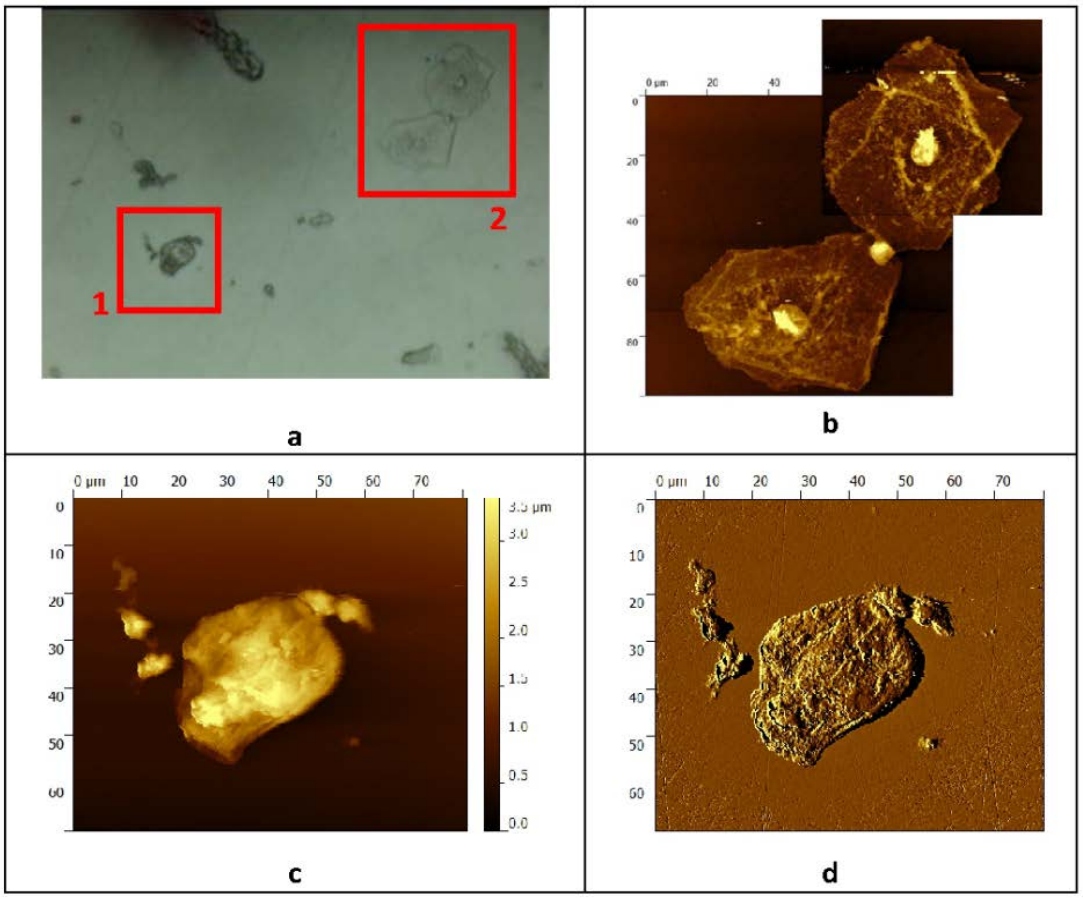


Supplementary Figure 8: ATR-FTIR spectroscopy: Average infrared spectra of cell types.

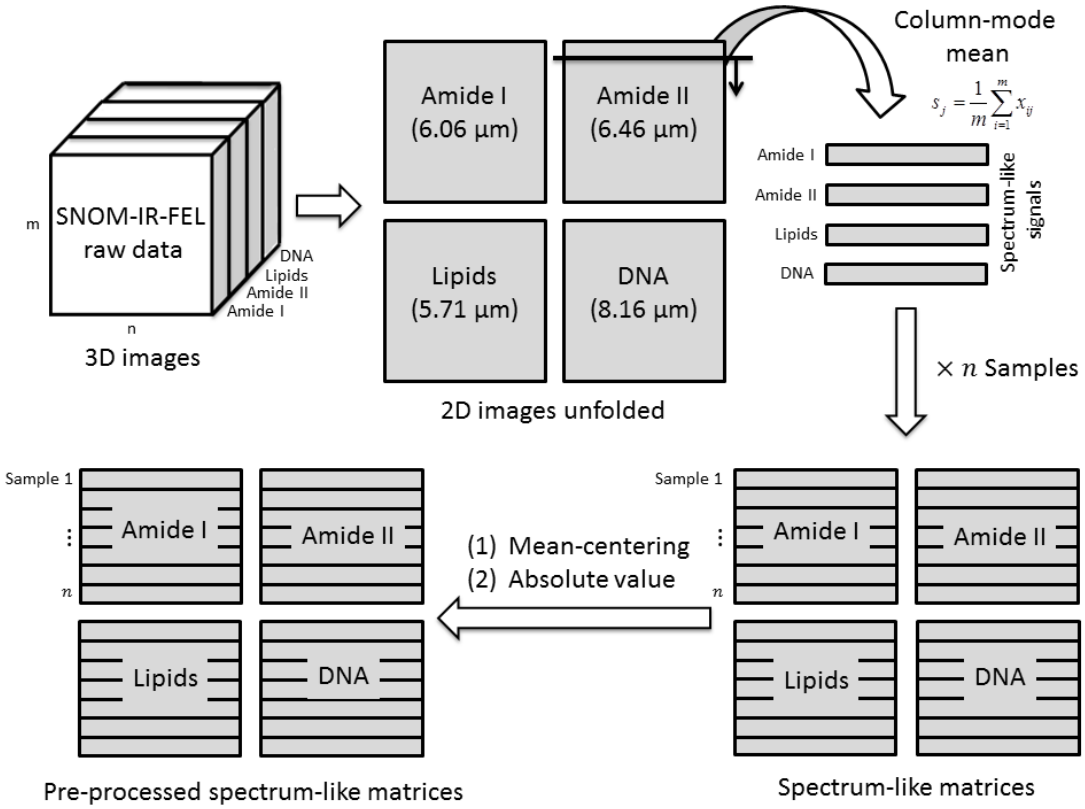
ATR-FTIR spectroscopy: Attenuated total reflectance, Fourier-transform infrared spectroscopy; CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia.



Supplementary Figure 9. **ATR-FTIR spectroscopy:** Scores plot of 1st and 2nd principal components at a 95% confidence level. ATR-FTIR spectroscopy: Attenuated total reflectance, Fourier-transform infrared spectroscopy; CIN2, HGCGIN: Cervical intraepithelial neoplasia 2, high-grade cervical glandular intraepithelial neoplasia; Principal components.



Supplementary Figure 10. AFM imaging of adenocarcinoma Stage 1B1: (**a**) Optical image (x10 magnification) identifying cells for investigation by AFM. (**b**) AFM topography image of two intermediate glandular cells [area 2 in (**a**)], the lower cell has two nuclei. The cells exhibit a long axis of ~75 microns. The cell thickness was measured at ~200 nm, whereas the nuclei protruded ~1 micron in height from thesubstrate. (**c**) AFM topography and (**d**) deflection image of a cell identified [area 1 in (**a**)] as having a single enlarged nucleus separated from the rest of the cell by a halo. AFM: Atomic force microscopy.



Supplemental Figure 11. The computational steps taken in processing the data.

4 General Discussion

4.1 Project One: Tracking the impact of excisional treatment

Many women undergoing excisional treatments for pre-invasive CIN are of child-bearing age (Jakobsson and Bruinsma, 2008), yet considerable evidence suggests that these types of treatment intervention have an impact on future pregnancy outcomes. Cold knife conisation has been shown to be significantly associated with perinatal mortality, low birth weight and severe preterm delivery, including Caesarean Section (Arbyn *et al*., 2008; Kyrgiou *et al*., 2006). LLETZ is associated with a significant increased risk of premature rupture of the membranes (Kyrgiou *et al*., 2006; Sadler *et al*., 2004), and miscarriage in the second trimester (Kyrgiou *et al*., 2015b), although the latter meta-analysis was compromised by the inclusion of non-randomised trials, some of which were deemed to be of low quality. Ablative treatments have been shown to be associated with fewer adverse sequelae than excisional treatments (Arbyn *et al*, 2008; Bruinsma *et al*., 2007; Kyrgiou *et al*., 2006). Therefore, the amount of tissue excised appears to have a critical impact on the structure and function of the cervix during future pregnancies.

The findings of our study using biospectroscopy to track the impact of excisional treatment for pre-invasive CIN (Project One), showed that treatment significantly alters the biochemistry of the cervix, compared with women who

have not had local treatment of the cervix. These changes were directly due to excision of cervical tissue, rather than the removal of disease.

It was hoped that different cone depth and/or percentage excisions would be reflected in a matched pattern of change in the spectral absorbance and associated biomarkers. This pattern of change could then be used to identify women at high risk of adverse pregnancy outcomes, who would then benefit from intensive antenatal surveillance when pregnant. The original premise of this study was to attempt to identify a clinical cut off for excision by comparing one cone depth variable with another, and likewise for percentage excision. Where a comparison was found not to be different between a higher dimension compared with a lower dimension, the lower dimension was proposed to be the optimal clinical cut off (i.e., where excising more tissue conferred no additional clinical benefit). Later discussions led to comparisons of absorbance for each dimension against that obtained for a healthy cervix population, in addition to the multi-comparisons described above. Although the authors concluded that absorbance did not seem to correlate to the cone depth or proportion of cervical length excised, there are several factors that are likely to have influenced the accuracy of the data. Some of these are obvious, such as small sample size; others are more confounding and relate to the mechanisms behind underlying patient characteristics.

4.1.1 Sample size and variable selection

The selection of the amount of cone depth removed as a variable for comparison was based upon previously published data that evidenced this dimension correlated to the frequency and severity of the adverse events (Arbyn *et al*., 2008). The selection of

percentage of tissue excised as a second variable was based upon the fact that pre-treatment cervical dimensions vary amongst women (Kyrgiou *et al*., 2015b), and the belief that percentage excision may provide a more accurate cut-off for excision than absolute dimensions (Founta *et al*., 2010). However, our study was not sufficiently powered to enable comparisons across all the percentage excision groups, leading to the 0-20% group being excluded. Secondly, the number of patients in the remaining four groups ranged from 7 to 22 patients. Although the number of patients in the groups used in the comparison by cone depth removed was marginally better, with 24, 24 and 10 patients in the <10 mm, 10-14 mm, and ≥ 15 mm groups, respectively, it was acknowledged that the sample sizes were small.

Scott (2012) stated that 'even the most rigorously executed study may fail to answer its research question if the sample size is too small.' The overall goal of calculating sample size is to estimate a suitable number of subjects that will determine the answer of the clinical question(s) of a study. The author of this dissertation was not involved in the development of the clinical protocol or the determination of sample size, and it is clear that the study was underpowered and unable to address some of the clinical questions.

Furthermore, this study collected additional dimensional data obtained by ultrasound, including cone volume removed (cm^3), 2D and 3D volume of cervix, and 2D and 3D percentage excision. It is feasible that additional analyses using these variables may have contributed to defining a more explicit picture with regards to optimal excision and off-set some of the limitations of sample size. Potential complementary analyses to support the current ones reported here are suggested in Table 8.

The limitations of study size should be considered alongside the findings of the post-hoc analyses of selected underlying patient characteristics (see Section 4.1.2).

Current analyses	Suggested complementary analyses
Healthy Cervix versus Treated by Cone Depth Removed (mm) (Individual dimensional groups by category)	Healthy Cervix versus Treated by Cone Volume Removed (cm^3) (Individual dimensional groups by category)
Not performed	Healthy Cervix 2D Volume versus Treated by 2D Volume (Individual dimensional groups by category)
Not performed	Healthy Cervix 3D Volume versus Treated by 3D Volume (Individual dimensional groups by category)
Healthy Cervix versus Treated by Percentage Excision (Individual dimensional groups by category)	Healthy Cervix versus Treated by 2D Percentage Excision (Individual dimensional groups by category)
	Healthy Cervix versus Treated by 3D Percentage Excision (Individual dimensional groups by category)

Table 8. Additional analyses considered complementary to the current dimensional analyses.

4.1.2 The mechanisms of underlying patient characteristics

Several groups have found evidence that an increased risk of pre-term delivery is associated with a diagnosis of precancerous changes together with a positive HPV status, even in women with an untreated cervix (Huang *et al.*, 2014; Bruninsma *et al.*, 2007), suggesting that other factors, including patient-specific characteristics and/or the HPV infection itself, may contribute to adverse pregnancy outcomes and that excisional treatments may exacerbate them.

Known risk factors for pre-term birth in women without pre-malignant CIN, include a history of induced abortion or miscarriage, previous pre-term birth, infertility treatment, current or previous infection with a sexually transmitted disease, being a single mother, intimate partner violence, physical workload and those with jobs that involves standing for greater than 6 hours a day, alcohol, drug use and cigarette smoking (Vicedo-Cabrera *et al*., 20016; Lumley, 2005; Lumley, 1993). A number of these factors, including a current smoking habit, are also associated with an increased risk for the development of pre-malignant CIN if present with persistent infection of high-risk HPV types (Castellsagué, *et al*., 2002). It has been suggested that any study that does adequately account for these factors in the study design, then any detected association with treatment may be due to these other factors rather than the treatment itself (Bruinsma *et al*., 2007).

Considering the findings of Project One, it was reasoned that a current smoking habit may alter the absorbance profile that is distinguishable from non-smokers, and thus may have the potential to affect the excisional outcomes. Other patient characteristics are known to directly affect the size of the cervix and include age, menstrual phase, hormonal status, and parity (IRAC, 2016a).

Although previous research has shown that Raman spectroscopy can improve the classification of LSIL stratified by menopausal and hormonal status from 74% to 97% (Kanter *et al*., 2009), and that ATR-FTIR spectroscopy can discern women infected with HPV high-risk types 16 or 18 based upon their age (Kelly *et al*., 2010), no studies have been conducted using ATR-FTIR spectroscopy to determine if the absorbance

profile within the ‘fingerprint region’ is different according to a patient’s smoking status, menstrual phase, combined oral contraceptive pill (COCP) use, or parity.

Four post-hoc analyses were performed using the treated group of women used in the analysis of cone depth. Patients were classed according to each of the four patient characteristics (smoking habit, parity, menstrual phase and COCP use). The analyses also included the extraction and comparison of the seven biomarkers used in the main analyses. The statistical analyses used have been previously described (see Project One); the results are presented in Appendix A.

4.1.2.1 Current smoking habit

A post-hoc analysis of treated women detected a significant difference between LD1 for smokers, compared with non-smokers (Supplementary Figure 1 [A]). No significant differences were detected between smokers and non-smokers for absorbance associated with the seven biomarkers selected (Supplementary Figure 1 [B]).

Fisher’s exact test had previously detected that the number of women in each group who were either non-smokers or current smokers was statistically different only for Comparison 2 (i.e., treated women with negative cytology and negative HPV versus untreated controls with negative cytology and negative HPV); (Supplementary Table 1). This comparison sought to prove that the difference observed between pre- and post-treated patients was due to excisional treatment. The groups were matched by cytology and HPV status, but not smoking status. The number of smokers in the treated group (segregated by dimensional groups; Supplementary Table 2), ranged from 21% to 47%, and all were higher than normal controls (15%).

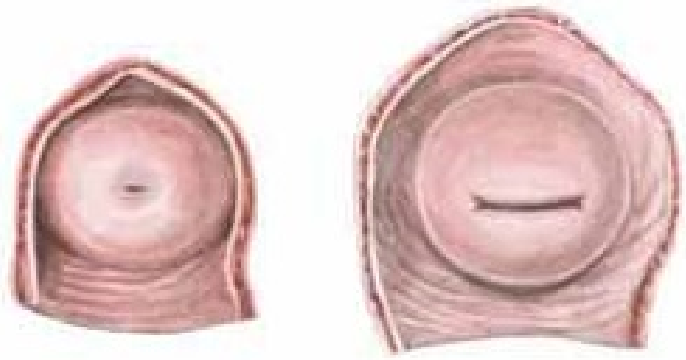
Whilst smoking status is unlikely to affect the physical dimensions of a treated cervix, the data suggests that a current smoking habit, although not detected in the seven biomarkers selected, alters cervical cell biochemistry. In healthy female smokers, nicotine levels are increased by up to 40 times in cervical mucus, with cotinine, nicotine's major metabolite, increased by 4 times, compared with serum levels (Sasson *et al*., 1985). Benzo-alpha-pyrene, an aromatic hydrocarbon found in cigarette smoke known to affect DNA repair mechanisms, is listed as a 'Group 1 carcinogen' by IRAC (2015), and has also been detected in cervical tissue (Melikian *et al*., 1999). Spectra obtained from samples collected from treated women who are current smokers are likely to include spectral signatures associated with these compounds. Therefore, two confounders are at work here with the potential to affect Comparison 2, and well as the dimensional comparisons to normal controls and the individual dimensional comparisons to each other: the presence of tobacco-related carcinogens that are likely to have an absorbance profile peculiar to them; and the associated effects of these carcinogens on cervical cell functionality.

Analyses to compare smokers and non-smokers within normal controls and in the Pre-treatment 'A' group, would have been preferable to evidence any difference in the absence of treatment or pre-invasive CIN. However, there were insufficient numbers of patients with a current smoking habit in either group to power the comparisons.

4.1.2.2 Parity

Parity, defined as the number of births with >20 weeks gestation (viable and non-viable [i.e., still born]), has a direct impact on the physical properties of the cervix

(IARC, 2016a), changing the physical appearance that can be visualised directly (Figure 21).



A B

Atlas Human Anatomy (2016).

Figure 21. The appearance of the external ostium in A: nulliparous women; B: parous women.

In nulliparous women, the external ostium of the cervix appears as a small external opening. In parous women, the cervix is more bulky and the external ostium is transformed into a transverse gaping slit.

A statically significant difference was detected in LD1 between treated nulliparous and treated parous women; the result was more significant than that detected for smoking status within the same group (Supplementary Figure 2 [A]). However, the number of women who were nulliparous in this comparison was >3 times the number of women in the parous group; thus significance it likely to be influenced by the unequal populations compared. Whilst there were no significant differences detected between treated nulliparous and treated parous women for absorbance associated with the seven biomarkers (Supplementary Figure 2 [B]), the data suggests parity may

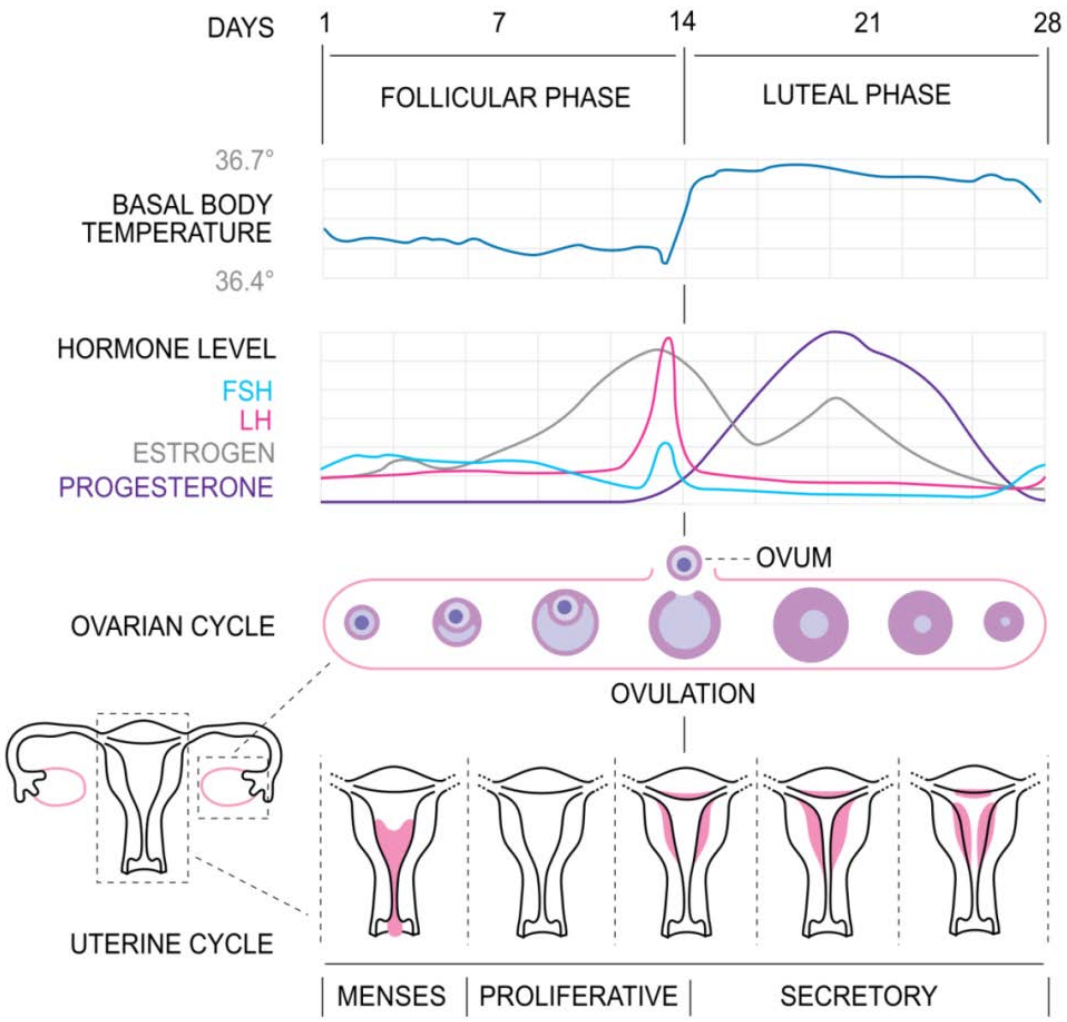
permanently alter the function of cervical cells, if not the entire uterus, that is still evident even after treatment for pre-invasive CIN.

Fisher's exact test found no differences across all three comparisons when comparing the number of women in each group who were for parous/nulliparous (Supplementary Table 1). However, the number of parous women in the treated group (segregated by dimensional groups; Supplementary Table 2), ranged from 17% to 40%, and was 19% for normal controls. Therefore, although the three comparisons were well balanced in terms of parous/nulliparous women, the comparison of individual dimensions to each other and to normal controls is likely to be affected by the uneven numbers across the groups.

No comparison between nulliparous/parous women within normal controls or the Pre-treatment 'A' group was possible due to an insufficient number of parous women in both groups. Therefore, the impact of parity in the absence of pre-invasive CIN or treatment, could not be determined.

4.1.2.3 Menstrual phase

Menstruation signals the beginning of the follicular phase which culminates in ovulation. During this phase, the pituitary gland secretes Follicle Stimulating Hormone (FSH) and Luteinizing Hormone (LH), which promotes the maturation of around 10-20 follicles in the ovaries. In turn, the maturing follicles secrete oestrogen (estradiol) and, as the levels of oestrogen rise over the next 7 days, this causes the lining of the uterus to thicken. One follicle becomes dominant and as oestrogen levels reach a threshold, the pituitary gland releases a surge of LH, causing the dominant follicle to erupt, releasing a mature ovum (ovulation); (Figure 22).



FSH: Follicle Stimulating Hormone; LH Luteinizing Hormone.

Isometrik, 2009

Figure 22. Phases of the menstrual cycle.

During the lead up to ovulation, oestrogen causes the ligaments of the uterus to tighten which draws the cervix deeper into the vagina. The cervix may appear more centrally aligned, feel softer with the external ostium slightly parted. The erupted follicle (corpus luteum) begins to produces significant amounts of progesterone, signalling the beginning of the luteal phase which lasts for about 14 days. The released ovum is

swept into the oviduct, where it will survive for approximately 24 hours. If the ovum is not fertilised, the corpus luteum shrinks after 14 days, triggering a sharp decline in both oestrogen and progesterone. The cervix returns to its original position, feels firmer to the touch with the external ostium closed until the onset of menses. Thus, the menstrual phase influences the physical size and location of the cervix. It is unknown whether menstrual phase may have affected the determination of cervical dimensions by ultrasound as collected in our study.

Additionally, a recently published study showed that the differential expression of 110 genes located in endocervix (glandular) tissues is dependent upon the phase of menstrual cycle (Yildiz-Arlan *et al*., 2016); (Table 9).

Associated functions of genes expressed in endocervical tissues during the follicular phase[a] (Estrogen/pre-ovulation)	**Associated functions of genes expressed in endocervical tissues during the luteal phase**[a] (Progesterone/post-ovulation)
Extracellular matrix remodelling	Chromatin re-modelling
Cell-matrix interactions	Inflammation
Amino acid and lipid metabolism	Angiogenesis
Immune regulation	Oxidative stress
	Immune cell regulation

[a] Yildiz-Arlan *et al*., 2016.

Table 9. Associated function of predominant genes expressed during the menstrual cycle.

The extracellular matrix of an organ represents the dynamical system that cooperates with cells to regulate a range of diverse functions, including proliferation, migration and differentiation (Bonnans *et al*., 2014). Remodelling of the extracellular matrix is crucial in many tissues. Dysregulation of the extracellular matrix, including

composition, stiffness and structure, are important in several diseases, including invasive cancer.

The follicular phase is predominantly controlled by oestrogen under the influence of FSH and LH. Genes expressed by glandular cells during this phase will control re-modelling of matrix, cell-matrix interactions, immune function and lipid and amino acid metabolism. By contrast, the luteal phase is predominantly controlled by progesterone. Genes preferentially expressed during this phase control the remodelling of tightly bound DNA (chromatin) to enable gene expression, the development of new blood vessels (angiogenesis), together with the management of oxidative stress, inflammation and immune cell regulation.

A statistically significant difference was detected for LD1 between treated patients in the luteal phase, compared with treated patients in the follicular phase; this difference was the most significant of all four patient characteristics analysed (Supplementary Figure 3 [A]). No significant differences were detected between the groups for absorbance associated with the seven biomarkers (Supplementary Figure 3 [B]).

Fisher's exact test found no significant differences when the comparison groups were compared by the number of women in either the luteal or follicular phase (Supplementary Table 1). However, the number of women in either the luteal/follicular phase in the treated group (segregated by dimensional groups; Supplementary Table 2), ranged from 38% to 67%, and was 26% for normal controls.

Since the number of glandular cells has been found to be significantly lower following treatment with LLETZ (Maguire *et al.*, 2008), it is reasonable to assume that the number of glandular cells collected during our study at 6 months follow-up after

excisional treatment, are likely to contain fewer glandular cells. Although a literature search did not find any publication detailing differential gene expression during the menstrual phase for squamous cells, it is plausible that cellular function of these cells also differs according to each phase.

Whilst the groups of women in the follicular/luteal phases were reasonably well balanced across the three comparisons (Supplementary Table 1); the number of women in each phase fluctuated greatly across the treated group (segregated by dimensions, Supplementary Table 2), compared to normal controls. Therefore, the comparison of dimensional groups to normal controls and to each other, are likely to be affected several factors: the effect of circulating hormones on cell activity as determined by menstrual phase; the differential gene expression of different cell types within each phase; the biochemical behaviour of the tissue and cells as a result of different gene expression; and the reduced number of glandular cells collected post treatment.

It was not possible to compare the groups within the normal or Pre-treatment ‘A’ group due to insufficient numbers of patients in groups segregated by menstrual phase.

4.1.2.4 Combined oral contraceptive use

During the normal menstrual cycle, oestrogen dominates the follicular phase and progesterone dominates the luteal phase (Figure 22). The COCP provides an exogenous supply of both hormones that together suppress gonadotrophins and inhibit the release of FSH, thereby preventing the maturation of follicles in the ovaries and ovulation. In doing so, the major effect of progesterone keeps the cycle in an artificial ‘luteal-like’ phase. It is plausible that genes controlled by progesterone during this

phase in glandular cells and potentially squamous cells, would be expressed for longer. However, the effect of both hormones combined upon gene expression in either cell is currently unknown.

A statistically significant difference was detected for LD1 between treated women who were not taking the COCP, compared with women who used this form of contraception (Supplementary Figure 4 [A]). No significant differences were detected between the two groups for absorbance associated with the seven biomarkers selected (Supplementary Figure 4 [B]).

The number of women in each group who were non-COCP users or current COCP users was found to be statistically different for Comparison 2 (i.e., treated women with negative cytology and negative HPV versus untreated controls with negative cytology and negative HPV); (Supplementary Table 1); and for Comparison 3. The number of current COCP users in the treated group (segregated by dimensional groups; Supplementary Table 2), ranged from 40% to 57%, and all were higher than normal controls (15%). These findings suggest that COCP use possibly alters differential gene expression and related cellular and extra-cellular functions; thus impacting the excisional outcomes of our study.

It was not possible to compare the groups within the normal or Pre-treatment 'A' group due to insufficient numbers of patients in groups segregated by COCP use.

4.1.2.5 Other factors: genetics of exfoliated cells and cervical tissue

The number of genes of exfoliated squamous cervical cells collected from the terminally differentiated superficial layers of the cervix has been found to be lower than in underlying tissue (Steinau *et al*., 2005). Of the 25,353 genes of the transcribed

human genome, 57% of these genes were found in cervical tissue and 40% were found in exfoliated cells, with 7320 genes found in both tissue and exfoliated cells, suggesting that exfoliated squamous cells have a more conservative genetic makeup than associated tissue. Genes found only in tissue were associated with a range of molecular functions, including those of the extracellular matrix. Genes found only in exfoliated cells were also associated with a range of functions, including protein phosphatase activity.

Although the healing process following excisional treatments like LLETZ is considered to be almost complete around 6 months after treatment (Paraskevaidis *et al*., 2002), cervical regeneration is dependent upon both the percentage of tissue remaining immediately after treatment, and the percentage of cone volume removed (Papoutsis *et al*., 2012); less remaining tissue and larger cone removal are associated with less regeneration. The effect of treatment on the genetic makeup of cells at the superficial layers of the remaining cervix is unknown, as is the effect of treatment upon the genetic makeup of deeper tissue and may directly impact on the bioavailability of biomarkers. Therefore, the results of our study are complicated further by unknown genetic traits of treated tissue.

4.1.3 Summary

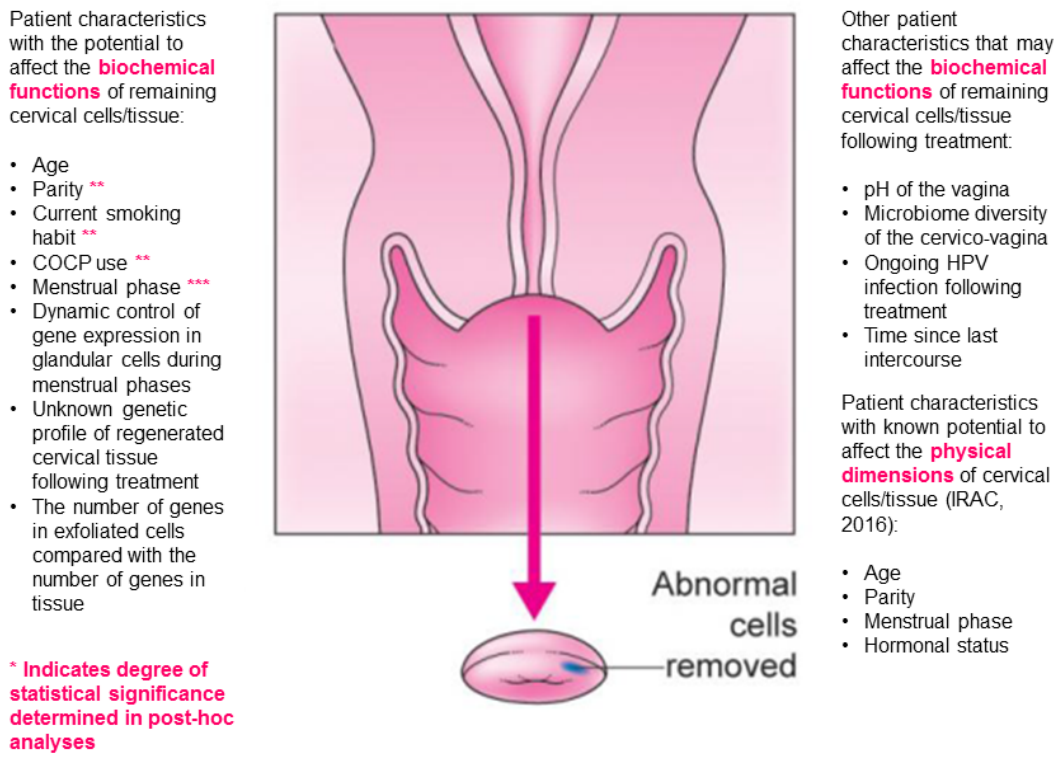
Although the groups were reasonably well matched in Comparison 1 (same patients pre- and post-treatment) for smoking status, parity, menstrual phase and COCP use, the groups were not well matched for smoking status and COCP use in Comparison 2. This comparison sought to evidence that the difference in absorbance (LD1) was due to the excisional treatment. Comparison 3, which compared dimensions against

normal controls and with each another, was not well matched for COCP use (Supplementary Table 1).

The findings of the post-hoc analyses suggest there is a complex interplay of underlying patient characteristics involving genetics, carcinogens, hormones, and the features of regenerated cervical tissue, in addition to other factors that can affect the physical aspects of the cervix such as parity. Of note, was that smoking, previous parity and COCP use all *depressed* LD1; the reasons are unknown. In Comparison 2, the LD1 of treated patients was lower than normal controls. It is likely that the effects of smoking and COCP use in this group contributed to the result.

Future studies should aim to investigate the full range of patient characteristics that can influence cell biochemistry and function, and contribute to spectral absorbance. Comparison groups should be well matched by characteristics.

A summary of potential factors that may influence the biochemical functions of remaining cervical tissue following treatment is summarised in Figure 23.



COCP: Combined oral contraceptive pill; IRAC: International Agency for Reach on Cancer.

Jo's Cervical Trust, 2016 (modified by D. E. Halliwell)

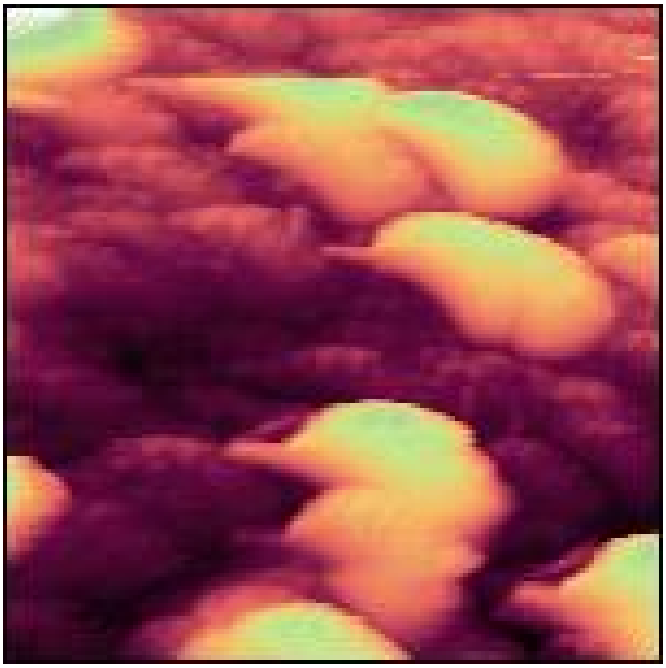
Figure 23. Patient characteristics that may affect excisional outcomes determined in Project One.

4.2 Project Two: The future role of SNOM-IR-FEL imaging in cancer studies

The results of the SNOM-IR-FEL pilot study were promising in terms of tracking the changes associated with various degrees of cervical dyskaryosis, although limited by the number of patients and individual cells imaged. However, at the time of the experimental period (February to May, 2015), the effect of the underlying patient characteristics had not been determined.

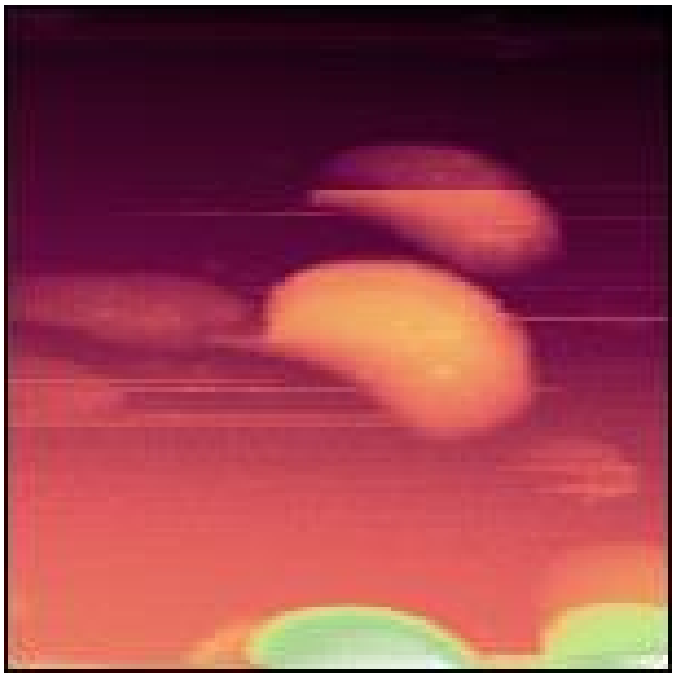
The samples had been collected from one normal patient and four patients with increasing grades of dyskaryosis. Although the last four samples were collected prior to treatment, it is reasonable to assume that the same patient characteristics investigated post-hoc in treated patients, also affect the spectra of untreated patients and may, in the presence of HPV infection, be more exaggerated. Two of the five patients were current smokers (high-grade dyskaryosis and the mixed pre-invasive lesion). Only one patient was parous (low-grade dyskaryosis); two patients were sampled in the luteal phase of their cycle (high-grade dyskaryosis and the mixed pre-invasive lesion), and two patients had been sampled in the follicular phase of their cycle; the other was unknown. Finally, two patients were using the COCP (high-grade dyskaryosis and the mixed pre-invasive lesion). Although, the small number of patients in this study was acceptable for a pilot study, it is clear that the chemical images collected are likely to be affected by differences in smoking habit, parity history, menstrual phase and COCP use. Other, currently undetermined patient characteristics such as vaginal pH and the microbiome of the cervicovagina, are also likely to influence the chemistry profiles of different lesions.

Previous work using SNOM-IR-FEL used in reflection mode, revealed the organisation of mitochondria within the mid-piece region of the tail of human sperm as well as acrosome and nucleus (Andolfi *et al*., 2015). However, to an untrained eye these individual structures are not obvious. Indeed, the SNOM topography obtained during Project Two revealed little in terms of cellular structures and was affected by significant tip artefacts (Figure 24 and Figure 25).



Project Two

Figure 24. Topography of normal cells: ‘dolphin-nosed’ cells considered to be tip artefacts



Project Two

Figure 25. Topography of adenocarcinoma cells: ‘dolphin-nosed’ cells considered to be tip artefacts

Despite the less than ideal topography, the potential of SNOM-IR-FEL is that the IR-FEL on the ALICE accelerator at Daresbury Laboratory (Warrington, UK) is tuneable over the range of 5.5 μm to 8.8 μm (~1818 cm^{-1} to ~1136 cm^{-1}), which includes a number of biologically important biomarkers (Movasaghi *et al*., 2008). Although no formal statistical tests were performed on the data due to the small data set, clear differences in each biomarker was evident and appeared dependent on the biochemistry of each cell type. However, any future studies using this technology should aim to match samples at least by smoking status, parity, menstrual phase and COCP use.

Further studies using the Daresbury facility should aim to address several issues, including achieving a higher resolution than that achieved in this pilot study (currently ~6 μm); the refinement of tips to minimise tip artefacts; and preparation of samples to minimise the loss of available cells to image. The latter may be achieved by experimenting with manual mounting onto slides rather than using the cytospinner. Researchers should aim to image a consistent number of cells per patient and if time allows, a higher number of cells per patient, and at the same wavenumbers and associated wavelengths documented here to build up a more explicit picture of chemical changes within various grades of cervical dyskaryosis.

All technologies are entitled to their birth right, and SNOM-IR-FEL is no exception. The technique, although fraught with early development limitations, is tuneable to specific wavelengths, which offers researchers a unique opportunity to investigate chemical changes within cells with precision. Whilst only four biomarkers were investigated during the pilot study, the range available with the IR-FEL on the ALICE

accelerator offers the opportunity to investigate other biomarkers of significance, including the N-H bond of thymine (~1272 cm^{-1}), the C=N bond of guanine (~1528 cm^{-1}), the C=N bond of adenine (~1571 cm^{-1}), and the C=N bond of cytosine (~1602 cm^{-1}). Investigation of these biomarkers may reveal patterns of change within the structure of DNA. The value of Project Two is strategic in that the published paper from the 2015 beam time will be used to support a new grant application to fund further studies.

4.3 Future work: meta-analysis

Despite the limitations of the two projects, the value of Project One lies in its potential to contribute to a future meta-analysis. Previous research conducted by Martin's Biophotonics Group has amalgamated spectra obtained from cervical cells collected from over 1300 patients (Table 10).

Number of patients (spectra[a])	Associated publication
220 (220)	Present study
350 (322)	Gajjar *et al*., 2014
67 (67)	Purandare *et al*., 2014
529 (529)	Purandare *et al*., 2013
147 (147)	Kelly *et al*., 2010a
30 (30)	Kelly *et al*., 2010b
20 (20)	Walsh *et al*., 2007
Total: 1363 (1268)	

[a] Ten spectra are typically collected per patient and then an average of these taken.

Table 10. Number of patients (spectra) collected and stored on the Biophotonics Server.

These data provide a rich opportunity to overcome the limitations of sample size, investigate a range of patient characteristics, and provide a substantial study population where patterns of cellular change can be more accurately matched to excisional data. However, the demographic data collected for Project One was considerable (63 variables). The other studies listed in Table 10 are thought to have more limiting demographic data associated with them, but if the data can be collected retrospectively, it would provide a substantial dataset to support addressing key clinical questions.

Suggested approaches for the meta-analysis include:

- Initially, define patient characteristics within a normal population, including the 'normal' microbiome diversity of the cervicovagina

Then determine the changes within the following populations:

- Women with HPV infection but without any obvious cytological change
- Women with low-grade dyskaryosis
- Women with high-grade dyskaryosis
- Women before and after excisional treatment for pre-invasive CIN
- Women with invasive cervical carcinomas and adenocarcinomas.

4.4 Conclusions

ATR-FTIR biospectroscopy detected changes in cervical cells pre- and post-treatment following excisional intervention for pre-invasive CIN. Although it was concluded that these changes were directly due to excisional treatment, and the spectra did

not seem to correlate to the cone depth or proportion (percentage) of cervical length excised, the study findings are hampered by the limitation of sample size and the previously undetermined effects of underlying patient characteristics. However, the extensive database of cervical cell spectra could support a meta-analysis involving over 1300 patients that would enable many of these problems to be overcome.

The future of SNOM-IR-FEL lies in the precise investigation of chemical changes in cervical cells in different stages of dyskaryosis, with an emphasis on changes within the four previously selected biomarkers (DNA, lipids and amide I and amide II bands), with potential explorations investigating changes in the four nucleobases of DNA.

References (paper, web pages and images)

Abdul-Karim, A., Fu, F.W., Reagan, J.W. and Wentz, B.W. (1982). Morphometric study of intraepithelial neoplasia of the uterine cervix. *Obstetrics and Gynecology.* **60**(2), pp.210-214.

Adobe Systems Incorporated. (2015). *Human papillomavirus*. [Image]. Purchased under licence by D. E. Halliwell. [No: AD003315597UK; 15 November, 2015].

American Cancer Society. (2013). *Working to Reduce the Burden of Cervical Cancer Globally.* [Online]. [Accessed 19 November 2015]. Available from: http://www.cancer.org/acs/groups/content/@internationalaffairs/documents/document/acspc-037351.pdf

American Cancer Society and LIVESTRONG. (2010). *The Global Economic Cost of Cancer (no. 005444).* [Online]. [Accessed 5 October 2015]. Available from: http://www.cancer.org/acs/groups/content/@internationalaffairs/documents/document/acspc-026203.pdf

Andolfi, L., Trevisan, E., Troian, B., Prato, S., Boscolo, R., Giolo, E., Luppi, S., Martinelli, M., Ricci, G. and Zweyer, M. (2015). The application of scanning near field optical imaging to the study of human sperm morphology. *Journal of Nanobiotechnology*. **13**(2). doi: 10.1186/s12951-014-0061-5.

Anolue, F.C., Ojiyi, E.C., Dike, E.I., Okeudo, C. and Ejikeme, C.E. (2014). Carcinoma of the cervix at a university teaching hospital in Eastern Nigeria. *Nigerian Journal of Surgical Sciences.* **24**(2), pp.49-52. doi: 10.4103/1116-5898.149603.

Arbyn, M., Kyrgiou, M., Gondry, J., Petry, K.U. and Paraskevaidis, E. (2014). Long term outcomes for women treated for cervical precancer. *British Medical Journal*. **348**(f7700). doi:10.1136/bmj.f7700.

Arbyn, M., Kyrgiou, M., Simoens, C., Raifu, A.O., Koliopoulos, G., Martin-Hirsch, P., Prendiville, W. and Paraskevaidis, E. (2008). Perinatal mortality and other severe adverse outcomes associated with treatment of cervical intraepithelial neoplasia: meta-analysis. *British Medical Journal*. **337**(a1284). doi:10.1136/bmj.a1284.

Atlas Human Anatomy. (no date). *In Fig. 189. Vaginal portion of the uterus.* [Online]. [Accessed 27 May 2016]. Available from: http://anthropotomy.com/the-internal-organs-and-systems/genitals/female-genital-organs

Baker, M.J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H.J., Dorling, K.M., Fielden, P.R., Fogarty, SW., Fullwood, N.J., Heys, K.A., Hughes, C., Lasch, P., Martin-Hirsch, B., Obinaju, B., Sockalingum, G.D., Sulé-Suso, J., Strong, R.J., Walsh, M.J., Wood, B.R., Gardener, P. and Martin, F.L. (2014). Using Fourier

transform IR spectroscopy to analyse biological materials. *Nature Protocols.* **9**(8), pp.1771-1791. doi:10.1038/nprot.2014.110.

Bate, A. and Baker, C. (2015). *House of Commons. Briefing Paper. Cancer Statistics: In Detail. Number SN02677.* [Online]. [Accessed 12 February 2016]. Available from: http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN02677

Bellisola, G. and Sorio, C. (2012). Infrared spectroscopy and microscopy in cancer research and diagnosis. *American Journal of Cancer Research*. **2**(1), pp.1-21.

Berrington de González, A., Sweetland, S. and Green, J. (2004). Comparison of risk factors for squamous cell and adenocarcinomas of the cervix: a meta-analysis. *British Journal of Cancer*. **90**(9), pp.1787-1791.

Bliss, D. (2004). *Cells of the cervix.* [Online]. [Accessed 14 November 2015]. Available from: https://commons.wikimedia.org/wiki/File:Cells_of_the_cervix.jpg?uselang=en-gb

Bonnans, C., Chou, J. and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. *Natures Reviews Molecular Cell Biology.* **15**(12), pp.786-801. doi:10.1038/nrm3904.

Boone, J.D., Erickson B.K., and Huh, W.K. (2012). New insights into cervical screening. *Journal of Gynecologic Oncology*. **23**(4), pp.282-287. doi.org/10.3802/jgo.2012.234.282.

Bosch, F.X., Manos, M.M., Muñoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R. and Shah, K.V. (1995). Prevalence of Human papillomavirus in cervical cancer: a worldwide perspective. *Journal of the National Cancer Institute.* **87**(11), pp.796–802.

Boskey, E.R., Telsch, K.M., Whaley, K.J., Moench, T.R. and Cone, R.A. (1999). Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. *Infection and Immunity.* **67**(10)**,** pp.5170–5175.

Brown, L. (no date). *The Diagnosis of Cervical Glandular Intraepithelial Neoplasia (CGIN); Bristol Symposium (handout). British Division of the International Academy of Pathology*. [Online]. [Accessed 3 December 2015]. Available from: http://www.bdiap.org/Bristol03/Bristol%20Symposia%20Handouts/3.7.03%20Gynae%20Symp%20-%2010.00%20hrs%20Dr%20L%20Brown.htm

Bruinsma, F., Lumley, J., Tan, J. and Quinn, M. (2007). Precancerous changes in the cervix and the risk of subsequent preterm birth. *British Journal of Obstetrics and Gynaecology.* **114**(1), pp.70-80. doi:10.1111/j.1471-0528.2006.01107.x.

Buckley, C.H., Butler, E.B. and Fox, H. (1982). Cervical intraepithelial neoplasia. *Journal of Clinical Pathology*. **35**(1), pp.1-13.

Burghardt, E. (1986). Natural history of cervical lesions. In: Peto, R. and Zur Hausenm H. eds. *Viral Etiology of Cervical Cancer*. New York: Cold Spring Harbour Laboratory Press, pp.81-82.

Cancer Research UK. (2015). *Cervical cancer incidence statistics.* [Online]. [Accessed 12 January 2016]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/incidence#heading-Zero

Cancer Research UK. (2014a). *Cervical cancer mortality statistics.* [Online]. [Accessed 12 January 2015]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/mortality#heading-Zero

Cancer Research UK. (2014b). *Diagram showing the transformation zone on the cervix CRUK 375.svg.* [Online]. [Accessed 14 November 2015]. Available from: https://commons.wikimedia.org/wiki/File:Diagram_showing_the_transformation_zone_on_the_cervix_CRUK_375.svg?uselang=en-gb

Cancer Research UK. (2014c). *Treatment if you have abnormal cancer cells.* [Online]. [Accessed 10 December 2015]. Available from: http://www.cancerresearchuk.org/about-cancer/type/cervical-cancer/smears/treatment-if-you-have-abnormal-cervical-cells

Carcopino, X., Mancini, J., Charpin, C., Grisot, C., Maycock, J.A., Houvenaeghel, G, Agostini, A., Boubli, L. and Prendiville, W. (2013). Direct colposcopic vision used with the LLETZ procedure for optimal treatment of CIN: results of joint cohort studies. *Archives of Gynecology and Obstetrics.* **288**(5), pp.1087-1094. doi:10.1007/s0404-013-2882-0.

Castellsagué, X., Bosch, F.X. and Muñoz, N. (2002). Environmental co-factors in HPV carcinogenesis. *Virus Research*. **89**(2), 191-199.

Castle, P.E. and Giuliano, A.R. (2003). Chapter 4: Genital tract infections, cervical inflammation, and antioxidant nutrients—assessing their roles as Human papillomavirus cofactors. *Journal of the National Cancer Institute. Monographs.* **31**, pp.29–34.

Clifford, G.M., Smith, J.S., Plummer, M., Muñoz, N. and Franceschi, S. (2003). Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *British Journal of Cancer*. **88**(1), pp.63-73.

Continuing Medical Education. (2014). *Fig. 2. Examples of LBC (left) and conventional (right) cytology slides after staining and mounting.* [Online]. [Accessed 12 December 2015]. Available from: http://www.cmej.org.za/index.php/cmej/article/view/2301/2186

Cooley, J.W. and Tukey, J.W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*. **19**(90), pp.297–301.

Cooper, G.M. (2000). *The Cell: a molecular approach*. 2nd ed. Washington (DC): ASM Press.

Cricenti, A., Generois, R., Luce, M., Perfetti, P., Margaritondo, G., Talley, D., Sanghera, J.S., Aggarwal, I.D. and Tolk, N.H. (2002). Very high resolution near-field chemical imaging using an infrared free electron laser. *Physical Chemistry Chemical Physics*. **4**(2), 2738–2741. doi:10.1039/b109279k.

Cuzick, J., Clavel, C., Petry K-U., Meijer C.J.L.M., Hoyer, H., Ratnam, S., Szarewski, A., Birembaut, P., Kulasingham, S., Sasieni, P. and Iftner, T. (2006). Overview of the European and North American studies on HPV testing in primary cervical cancer screening. *International Journal of Cancer*. **119**(5), pp.1095-1101.

Diem, M., Romeo, M., Boydston-White, S., Miljković, M. and Matthäus C. (2004). A decade of vibrational micro-spectroscopy of human cells and tissue. *The Analyst*. **129**(10), pp.880-885.

Doorbar, J. (2005). The papillomavirus life cycle. *Journal of Clinical Virology*. **32**(Suppl 1), pp.S7-S15.

Edwards, R.P., Taylor, S.E., Karjane, N.W. and Rivlin, M.E. (2015). *Cervical Screening.* [Online]. [Accessed 6 December 2015]. Available from: http://emedicine.medscape.com/article/1618870-overview#a1

Eurocytology. (2015). *The squamocolumnar junction, metaplastic change and the transformation zone (/en/course/465).* [Online]. [Accessed 6 January 2016]. Available from: http://lnx.eurocytology.eu/en/course/465

Faridi, R., Zahra, A., Khan, K. and Idress, M. (2011). Oncogenic potential of Human Papillomavirus (HPV) and its relation with cervical cancer. *Virology Journal*. **8**(1), 269. doi: 10.1186/1743-422X-8-269.

Fellgett, P. (1949). *Theory of Infra-Red Sensitivities and Its Application to Investigations of Stellar Radiation in the Near Infra-Red by infrared spectroscopy.* Masters dissertation. University of Cambridge.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D. and Bray, F. (2013a). *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013.* [Online]. [Accessed 5 October 2015]. Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.asp

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W., Comber, H., Forman, D. and Bray, F. (2013b). Cancer incidence and mortality

patterns in Europe: Estimates for 40 countries in 2012. *European Journal of Cancer*. **49**(6), pp.1374-1403. doi: 10.1016/j.ejca.2012.12.027.

Ferlay, J., Shin, H-R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*. **127**(12), pp. 2893–2917. doi: 10.1002/ijc.25516.

Foley, G., Alston, R., Geraci, M., Brabin, L., Kitchener, H. and Birch, J. (2011). Increasing rates of cervical cancer in young women in England: an analysis of national data 1982-2006. *British Journal of Cancer*. **105**(1), pp.177-184. doi: 10.1038/bjc.2011.196.

Founta, C., Arbyn, M., Valasoulis, M., Kyrgiou, M., Tsili, A., Martin-Hirsch, P., Dalkalitsis, N., Karakitsos, P., Kassanos, D., Prendiville, W., Loufopoulos, A. and Paraskevaidis E. (2010). Proportion of excision and cervical healing after large loop excision of the transformation zone for cervical intraepithelial neoplasia. *British Journal of Obstetrics & Gynecology*. **117**(12), pp.1468-1474. doi: 10.1111/j.1471-0528.2010.02709.x.

Gajjar, K., Ahmadzai, A.A., Valasoulis, G., Trevisan, J., Founta, C., Nasioutziki, M., Loufopoulos, A., Kyrgiou, M., Stasinou, S.M., Karakitisos, P, Paraskevaidis, E., D Gama-Rose, B., Martin-Hirsch, P.L. and 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), e82416:1-1. doi:10.1371/journal.pone.0082416.

García-Closas, R., Castellsagué, X., Bosch, X. and González, C.A. (2005). The role of diet and nutrition in cervical carcinogenesis: a review of recent evidence. *International Journal of Cancer*. **117**(4), pp.629–937.

Gok, S., Aydin O.Z., Sural, Y.S., Zorlu, F., Bayol, U. and Severcan, F. (2016). [Forthcoming]. Bladder cancer diagnosis from bladder wash by Fourier transform infrared spectroscopy as a novel test for tumor recurrence. *Journal of Biophotonics.* [Online]. [Accessed 8 June 2016] Available from: http://onlinelibrary.wiley.com/doi/10.1002/jbio.201500322/abstract;jsessionid=1D138643BE9069A621255E9118C549A6.f04t02

Griffiths, P. and De Haseth J.A. (2007). *Fourier Transform Infrared Spectrometry.* 2nd edn. Chichester (West Sussex): John Wiley & Sons Ltd.

Halliwell, D.E. (2016a). *Unstained normal squamous epithelial cells of the superficial layer of the ecotcervix.* [Online]. [Accessed 4 June 2016]. Available from: https://commons.wikimedia.org/wiki/File:Unstained_normal_squamous_epithelial_cells_of_the_superficial_layer_of_the_ectocervix.tif

Halliwell, D.E. (2016b). *Unstained cervical cells sampled from a patient with cervical adenocarcinoma Stage 1B1*. [Online]. [Accessed 12 January 2016]. Available from:

https://commons.wikimedia.org/wiki/File:Cervical_Adenocarcinoma_Stage_1B1.tif#.7B.7Bint:filedesc.7D.7D

Halliwell, D.E. (2016c). *Changes in dipole moments after being irradiated with infrared light.* [Online]. [Accessed 10 April 2016]. Available from: https://commons.wikimedia.org/wiki/File:Potential_changes_in_the_dipole_moments_of_a_heteronuclear_molecule.tif

Halliwell, D.E. (2016d). *An example of an IR-inactive molecule.* [Online]. [Accessed 10 April 2016]. Available from: https://commons.wikimedia.org/wiki/File:An_example_of_an_IR-inactive_molecule.tif

Hands, J.R., Clemens, G., Stables, R., Ashton, K., Brodbelt, A., Davies, C., Dawson, T.P., Jenkinson, M.D., Lea, R.W., Walker, C., Baker, M.J. (2016). Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy. *Journal of Neuro-oncology.* **127**(3):463-72. doi: 10.1007/s11060-016-2060-x.

Harrison, A.J., Bilgili, E.A., Beaudoin, S.P. and Taylor, L.S. (2013). Atomic force microscope infrared spectroscopy of griseofulvin nanocrystals. *Analytical Chemistry*. **85(**23)**,** pp. 11449-11455. doi:10.1021/ac4025889.

Hein, M., Helmig, R.B., Schønheyder, H.C., Ganz, T. and Uldbjerg, N. (2001). An in vitro study of antimicrobial properties of the cervical mucus plug in pregnancy. *American Journal of Obstetrics and Gynaecology*. **185**(3), pp.586-592. doi:10.1067/mob.2001.116685.

Honduras Health. [no date]. *Different stages of cervical cancer.* [Online]. [Accessed 12 September 2015]. Available from: http://hondurashealth.wikidot.com/cervical-cancer

Huang, Q.T., Zhong, M., Gao, Y.F., Huang, Q., Wang, W., Wang, Z.J. and Yu, Y.H. (2014). Can HPV vaccine have other health benefits more than cancer prevention? A systematic review of association between cervical HPV infection and preterm birth. *Journal of Clinical Virology.* **61**(3), pp. 321-328. doi:10.1016/j.jcv.2014.09.002.

International Agency for Research on Cancer (IARC). (2016a). *Chapter 1: An introduction to the anatomy of the uterine cervix.* [Online]. [Accessed 25 may 2016]. Available from: http://screening.iarc.fr/colpochap.php?chap=1

International Agency for Research on Cancer (IARC). (2016b). *Figure 1.2: Stratified squamous epithelium (x 20).* [Online]. [Accessed 1 June 2016]. Available from: http://screening.iarc.fr/colpochap.php?chap=1

International Agency for Research on Cancer (IARC). (2016c). *FIGURE 1.3: Columnar epithelium (x 40).* [Online]. [Accessed 5 June 2016]. Available from: http://screening.iarc.fr/colpochap.php?lang=1&chap=2

International Agency for Research on Cancer (IARC). (2016d). *Chapter 2: An introduction to cervical intraepithelial neoplasia* (CIN). [Online]. [Accessed 5 June 2016]. Available from: http://screening.iarc.fr/colpochap.php?chap=1

International Agency for Research on Cancer (IARC). (2015). Agents Classified by the IARC Monographs, Volumes 1–112. [Online]. [Accessed 15 June 2016]. Available from: http://monographs.iarc.fr/ENG/Classification/ClassificationsAlphaOrder.pdf

Information Services Division, Scotland. (2015). *Cervical Cancer Screening Programme Statistics 2014-15*. [Online]. [Accessed 12 December 2015]. Available from: http://www.isdscotland.org/Health-Topics/Cancer/Cervical-Screening/

Isometrik. (2009). *Diagram of the menstrual cycle (based on several different sources)*. [Online]. [Accessed 11 June 2016]. Available from: https://commons.wikimedia.org/wiki/File:MenstrualCycle2_en.svg

Jakobsson, M. and Bruinsma, F. (2008). Adverse pregnancy outcomes after treatment for cervical intraepithelial neoplasia. *British Medical Journal.* **337**(a1350). doi:10.1136/bmj.a1350.

Jo's Cervical Cancer Trust. (2016). *Large loop excision of the transformation zone (LLETZ).* [Online]. [Accessed 11 May 2016]. Available from: http://www.jostrust.org.uk/about-cervical-cancer/cervical-cancer/treatments/surgery/lletz
This image was re-used with kind per permission from Sanger, K (Communications Manager, Jo Cervical Cancer Trust). (2016). Email to Diane E Halliwell, 8 May.

Kanter, E.M., Majumder, S., Kanter, G.J., Woeste, E.M. and Mahadevan-Jansen., A. (2009). Effect of hormonal variation on Raman spectra for cervical disease detection. *American Journal of Obstetrics and Gynecology.* **200**(5), pp. 512.e1–512.e5. doi:10.1016/j.ajog.2008.11.024.

Kelly, J.G., Cheung, K.T., Martin, C., O'Leary, J.J., Prendiville, W., Martin-Hirsch, P.L. and Martin, F.L. (2010a). A spectral phenotype of oncogenic human papillomavirus-infected exfoliative cervical cytology distinguishes women based on age. *Clinica Chimica Acta.* **411**(15-16), pp.1027-1033. doi: 10.1016/j.cca.2010.03.029.

Kelly, J.G., Angelov, P.P., Trevisan, J., Vlachopoulou, A., Paraskevaidis, E., Martin-Hirsch, P.L. and Martin, F.L. (2010b). Robust classification of low-grade cervical cytology following analysis with ATR-FTIR spectroscopy and subsequent application of self-learning classfier eClass. *Analytical and Bioanalytical Chemistry.* **398**(5), pp. 2191-2201. doi: 10.1007/s00216-010-4179-5.

Khalid, S., Dimitriou, E., Conroy, R., Paraskevaidis, E., Kyrgiou, M., Harrity, C., Arbyn, M. and Prendiville, W. (2012). The thickness and

volume of LLETZ specimens can predict the relative risk of pregnancy-related morbidity. *British Journal of Obstetrics and Gynaecology*. **119**(6), pp.685-681. doi:10.1111/j.14710528.2011.03252.x.

Kitchener, H.C., Almonte, M., Gilham, C., Dowie, R., Stoykova, B., Sargent, A., Roberts, C., Desai, M. and Peto, J. (2009). ARTISTIC: a randomised trial of human papillomavirus (HPV) testing in primary cervical screening. *Health Technology Assessment.* **13**(51), pp.1-150, iii-iv. doi:10.3310/hta13510.

Kyrgiou, M., Valasoulis, G., Stasinou, S-M., Founta, C., Athanasiou, A., Bennett, P. and Paraskevaidis, E. (2015a). Proportion of cervical excision for cervical intraepithelial neoplasia as a predictor of pregnancy outcomes. *International Journal of Gynecology and Obstetrics*. **128**(2), pp.141-147. doi: 10.1016/j.ijgo.2014.07.038.

Kyrgiou, M., Mitra, A., Arbyn, M., Paraskevaidis, M., Athanasiou, A., Martin-Hirsch, P.P., Bennett, P. and Paraskevaidis, E. (2015b). Fertility and early pregnancy outcomes after treatment for cervical intraepithelial neoplasia. *The Cochrane Database Systematic Reveiws.* **9**(CD008478). doi:10.1002/14651858.CD008478.pub2.

Kyrgiou, M., Mitra, A., Arbyn, M., Stasinou, S-M., Martin-Hirsch, P., Bennett, P. and Paraskevaidis, E. (2014). Fertility and early pregnancy outcomes after treatment for cervical intraepithelial neoplasia: systemic review and meta-analysis. *British Medical Journal.* **349**(g6192), pp.1-17. doi:10.1136/bmj.g6192.

Kyrgiou, M. and Shafi, M.I. (2014). Colposcopy and cervical intraepithelial neoplasia. *Obstetrics, Gynaecology and Reproductive Medicine.* **24**(7), pp.204-214. doi:10.1016/j.ogrm.2014.05.002.

Kyrgiou, M. and Shafi, M.I. (2013). Invasive cancer of the cervix. *Obstetrics, Gynaecology and Reproductive Medicine.* **23**(11), pp. 343-351. doi:10.1016/j.ogrm.2013.08.005.

Kyrgiou, M., Arbyn, M., Martin-Hirsch, P. and Paraskevaidis, E. (2012). Increased risk of preterm birth after treatment for CIN. *British Medical Journal.* **345(**e5847). doi:10.1136/bmj.e5847.

Kyrgiou, M., Koliopoulos, G., Martin-Hirsch, P., Arbyn, M., Prendiville, W. and Paraskevaidis, E. (2006). Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. *Lancet.* **367**(9509), pp.489-498.

Lancucki, L., Sasieni, P., Patnick, J., Day, T.J. and Vessey, M.P. (2012). The impact of Jade Goody's diagnosis and death on the NHS Cervical Screening Programme. *Journal of Medical Screening*. **19**(2), pp.89-93. doi:10.1258/jms.2012.012028.

Lasch, P. (2012). Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems.* **117**(8), pp.100–114. doi:10.1016/j.chemolab.2012.03.011.

Lima, C.A., Goulart, V.P., Côrrea, L., Pereira, T.M. and Zezell, D.M. (2015). ATR-FTIR spectroscopy for the assessment of biochemical changes in skin due to cutaneous squamous cell carcinoma. *International Journal of Molecular Science.* **16**(4):6621-6630. doi: 10.3390/ijms16046621.

Lima, K.M.G., Gajjar, K., Vasasoulis, G., Nasioutziki, M., Kyrgiou, M., Karakitsos, P., Paraskevaidis, E, Martin-Hirsch, P.L. and Martin, F.L. (2014). Classification of cervical cytology for human papilloma virus (HPV) infection using biospectroscopy and variable selection techniques. *Analytical Methods*. **6**(24), pp.9643-9652. doi:10.1039/C4AY01736F.

Lumley, J. (2005). Recent work on the epidemiology of preterm birth. *Acta Obstetricia et Gynecologica Scandinavica.* **84**(6), pp. 541-542.

Lumley, J. (1993). The epidemiology of preterm birth. *Baillieres Clinical Obstetrics and Gynaecology.* **7**(3), pp. 477-498.

Maguire, A., Turner, L., Magee, D. and Gibbons, D. (2008). Decrease in numbers of glandular cells groups in post-LLETZ liquid-based cytology preparations. *Cytopathology.* **19**(1), pp.44-47. doi:10.1111/j.1365-2303.2007.00519.x.

Martin, F.L., Kelly, J.G., Llabjani, V., Martin-Hirsch, P.L., Patel, I.I., Trevisan, J., Fullwood, N.J. And Walsh, M.J. (2010). Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nature Protocols.* **5**(11), pp.1748-1760. doi:10.1038/nprot.2010.133.

Martyn, F., McAuliffe, F.M. and Wingfield, M. **(2014).** The role of the cervix in fertility: is it time for a reappraisal? *Human Reproduction.* **29**(10), pp.2092–2098. Doi:10.1093/humrep/deu195.

Mazouni, C., Bretelle, F., Blanc, K., Heckenroth, H., Haddad, O., Agostini, A., Cravello, L., Blanc, B. and Gamerre, M. **(2005).** Transvaginal sonographic evaluation of cervix length after cervical conization. *Journal of Ultrasound Medicine*. **24**(11), pp.1483-1486.

Mehta, V., Vasanth, V. and Balachandran, C. (2009). Pap smear. *Indian Journal Dermatology Venereology Leprology.* **75**(2), pp.214-216.

Melikian, A.A., Sun, P., Prokopczyk, B., El-Bayoumy, K., Hoffman, D., Wang, X. and Waggoner, S. (1999). Identification of benzo(a)pyrene metabolites in cervical mucus and DNA adducts in cervical tissues in humans by gas chromatography-mass spectrometry. *Cancer Letters.* **146**(2), pp. 127-134.

Movasaghi, Z., Rehman, S. and ur Rehamn, I. (2008). Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, **43**(2), pp.134-179. doi:10.1080/05704920701829043.

Muka, T., Imo, D., Jaspers, L., Colpani, V., Chaker, L., van der Lee, S.J., Mendis, S., Chowdhury, R., Bramer, W.M., Falla, A., Pazoki, R. and Franco, O.H. (2015). The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. *European Journal of Epidemiology*. **30**(4), pp.251-277. doi:10.1007/s10654-014-9984-2.

Münger, K., Baldwin, A., Edwards, K.M., Hayakawa, H., Nguyen, C.L., Owens, M., Grace, M. and Huh, K.W. (2004). Mechanisms of human papillomavirus-induced oncogenesis. *Journal of Virology*. **78**(21), pp.11451-11460. doi:10.1128/JVI.78.21.11451-11460.2004.

Muñoz, N., Castellsagué, X., Berrington de González, A. and Gissmann, **L. (2006).** Chapter 1: HPV in the etiology of human cancer. *Vaccine*. **24**(Suppl 3), pp.S3/1–S3/10. doi:10.1016/j.vaccine.2006.05.115.

Nanda, K., McCrory, D.C., Myers, E.R., Bastian, L.A., Hasselblad, V., Hickey, J.D. and Matchar, D.B. (2000). Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Annals of Internal Medicine.* **132**(10), pp.810-819.

NASA. (2007). *The Electromagnetic spectrum.* [Online]. [Accessed 12 December 2015]. Available from: http://mynasadata.larc.nasa.gov/images/EM_Spectrum3-new.jpg

National Cancer Institute. (2001). *Besthesda System 2001***.** [Online]. [Accessed 10 December 2015]. Available from: http://nih.techriver.net/

National Cancer Intelligence Network (NCIN) and Cancer Research UK. (2009). *Cancer Incidence and Survival by Major Ethnic Group, England, 2002-2006.* [Online]. [Accessed 12 October 2015]. Available from: http://publications.cancerresearchuk.org/downloads/Product/CS_REPORT_INCSURV_ETHNIC.pdf

National Cancer Intelligence Network (NCIN). (2005). *Cancer Incidence and Mortality by Cancer Network*, UK, 2005. London: NCIN.

NHS Cervical Screening Programme. (2013). *Achievable standards, Benchmarks for reporting, and Criteria for evaluating cervical cytopathology*. 3rd ed. Sheffield (Yorkshire): NHS Cancer Screening Programme.

Office for National Statistics. (2015). *Ageing Interactives: Ageing in the UK, 1992 to 2037.* [Online]. [Accessed 10 December 2015]. Available from: http://www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc248/index.html

OpenStax College. (2013). *In most cases, cells infected with the HPV virus heal on their own. In some cases, however, the virus continues to spread and becomes an invasive cancer.* Available from: (https://commons.wikimedia.org/wiki/File:Figure_28_02_08.JPG?uselang=en-gb)

Papanicolaou, G.N. and Traut, H.F. **(1997).** The diagnostic value of vaginal smears in carcinoma of the uterus. 1941. *Archives of Pathology and Laboratory Medicine*. **121**(3), pp.211-224.

Papoutsis, D., Rodolakis, A., Mesogitis, S., Sotiropoulou, M. and Antsaklis, A. (2012). Regeneration of uterine cervix at 6 months after large loop excision of the transformation zone for cervical intraepithelial neoplasia. *British Journal of Obstetrics and Gynaecology.* **119**(6), pp. 678-684. doi:10.1111/j.1471-0528.2012.03275.x.

Paraskevaidis, E., Bilirakis, H., Koliopoulos, G., Kalantaridou, S., Paschopulos, M., Plachouras, N., Malamou-Mitsi, V. and Kitchener, H.C. (2002). Cervical regeneration after diathermy excision of cervical intraepithelial neoplasia as assessed by transvaginal sonography. *European Journal of Obstetrics, Gynecology and Reproductive Biology.* **102**(10), pp.88-91.

Pardo, J., Yogev, Y., Ben-Haroush, A., Peled, Y., Kaplan, B. and Hod, M. (2003). Cervical length evaluation by transvaginal sonography in nongravid women with a history of preterm delivery. *Ultrasound Obstetrics & Gynecology.* **21**(5), pp.464-466. doi:10.1002/uog/116.

Peto, J., Gilham, C., Fletcher, O. and Matthews, F.E. (2004). The cervical cancer epidemic that screening has prevented in the UK. *Lancet*. **364**(9430), pp.249-256.

Phadnis, S.V., Atilade, A., Bowring, M., Kyrgiou, M., Young, M.P., Evans, H., Paraskevaidis, E. and Walker, P. (2011). Regeneration of cervix after excisional treatment for cervical intraepithelial neoplasia: a study of collagen distribution. *British Journal of Obstetrics and Gynaecology.* **118**(13), pp.1585-1591. doi:10.1111/j.1471-0528.2011.03085.x

PowerPoint. (Office 2013). *Papillomavirus (ClipArt).* This image is licenced for re-use by Microsoft: Richmond, WA: Microsoft.

Prendiville, W., Cullimore, J. and Norman, S. (1989). Large loop excision of the transformation zone (LLETZ). A new method of management for women with cervical intraepithelial neoplasia. *British Journal of Obstetrics and Gynaecology.* 96(9), pp.1054-1060.

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. and 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), pp.4576-4584. doi:10.1039/c3ay42224k.

Purandare, N.C, Patel, I.I., Trevisan, J., Bolger, N., Kelehan, R., von Bünau, G., Martin-Hirsch, P.L., Prendiville, W.J. and Martin, F.L. (2013). Biospectroscopy insights into the multi-stage process of cervical cancer development: probing for spectral biomarkers in cytology to distinguish grades. *Analyst*. **138**(14), pp.3909-3916. doi:10.1039/c3an36527a.

Ramakrishnan, S., Partricia, S. and Mathan, G. (2015). Overview of high-risk HPV's 16 and 18 infected cervical cancer: Pathogenesis to prevention. *Biomedicine & Pharmacotherapy*. **70**, pp.103–110. doi:10.1016/j.biopha.2014.12.041.

Reproductive Health Technologies Project. (2000). *The role of the cervix in reproductive health.* [Online]. [Accessed 12 November 215]. Available from: http://www.rhtp.org/fertility/cervix/RoleofCervix.asp

Richart RM. (1967). Natural history of cervical intraepithelial neoplasia. *Clinical Obstetrics and Gynecology*. **10**, pp.748-784.

Ronco, G., Cuzick, J., Pierotti, P., Cariaggi, M.P., Palma, P.D., Naldoni, C., Ghiringhello, B., Giorgi-Rossi, P., Minucci, D., Parisio, F., Pojer, A., Schiboni, M.L., Sintoni, C., Zorzi, M., Segnan, N. and Confortini, M. (2007). Accuracy of liquid based versus conventional cytology: overall results of new technologies for cervical cancer screening: randomised controlled trial. *British Medical Journal*. **335**(7609), pp.28-31.

Sadler, L., Saftlas, A., Wang, W., Exeter, M., Whittaker, J. and McCowan, L. (2004). Treatment for cervical intraepithelial neoplasia and risk of preterm delivery. *Journal of the American Medical Association.* **291**(17), pp.2100-2106.

Safaeian, M., Solomon, D. and Castle, P.E. (2007). Cervical cancer prevention-cervical screening: science in evolution. *Obstetrics and Gynecology Clinics of North America.* **34**(4), pp.739-760. doi:10.1016/j.ogc.2007.09.004.

Salter J. **(2014).** *Revealing the true costs of cervical cancer*. London: Demos.

Sasieni, P. and Adams, J. (2001). Changing rates of adenocarcinoma and adenosquamous carcinoma of the cervix in England. *Lancet*. **357**(9267):1490-1493.

Sasson, I.M., Haley, N.J., Hoffman, D., Wynder, E.L., Hellberg, D. and Nilsson, S. (1985). Cigarette smoking and neoplasia of the uterine cervix: smoke constituents in cervical mucus. *New England Journal of Medicine.* **312**(5), pp. 315-516.

Scheiden, R., Wagener, C., Knolle, U., Dippel, W. and Capesius, C. **(2004).** Atypical glandular cells in conventional cervical smears: incidence and follow-up. *BioMed Central Cancer*. **4**(37). doi:10.1186/1471-2407-4-37.

Scott, T.A. (2012). *Sample Size Planning, Calculation, and Justification.* [Online]. [Accessed 10 June 2016]. Available from:

http://biostat.mc.vanderbilt.edu/wiki/pub/Main/TheresaScott/SampleSize.TAScott.handout.pdf

Screening and Immunisations team. **(2014).** *Cervical Screening Programme. England: Statistics for 2013-14, V1.0.* London: Health & Social Care Information Centre.

Sherman, S.M., Castanon, A., Moss, E. and Redman, C.W.E. (2015). Cervical cancer is not just a young women's disease. *British Medical Journal*. **350**(h2729). doi:10.1136/bmj.h2729.

Smith, A.D., Siggel-King, M.R.F., Holder, G.M., Cricenti, A., Luce, M., Harrison, P., Martin, D.S., Surman, M., Craig, T., Barrett, S.D., Wolski, A., Dunning, D.J., Thompson, N.R., Saveliev, Y., Pritchard, D.M., Varro, A., Chattopadhyay, S. and Weightman, P. (2013). Near-field optical microscopy with an infra-red free electron laser applied to cancer diagnosis. *Applied Physics Letters.* **102**(053701). doi:10.1063/1.4790436.

Smith, J.S., Melendy, A., Rana, R.K. and Pimenta, J.M. (2008). Age-specific prevalence of infection with human papillomavirus in females: A global review. *Journal of Adolescent Health.* **43**(4 Suppl), pp.S5-25.doi:10.1016/j.jadohealth.2008.07.009.

Smith, H.O., Tiffany, M.F., Qualls, C.R. and Key, C.R. (2000). The rising incidence of adenocarcinoma relative to squamous cell carcinoma of the uterine cervix in the United States—A 24-year population-based study. *Gynecologic Oncology.* **78**(2), pp.97–105. doi:10.1006/gyno.2000.5826.

Sobota, R.S., Ramogola-Masire, D., Williams, S.M. and Zetola, N.M. (2014). Co-infection with HPV types from the same species provides natural cross protection from progression to cervical cancer. *Infectious Agents and Cancer*. **9**(26). doi:10.1186/1750-9378-9-26.

Spence, A.R., Goggin, P. and Franco, E.L. (2007). Process of care failures in invasive cervical cancer: systematic review and meta-analysis. *Preventative Medicine*. **45**(2-3), pp.93-106.

Steinau, M., Lee, D.R., Rajeevan, M.S., Vernon, S.D., Ruffin, M.T. and Unger, E.R. (2005). Gene expression profile of cervical tissue compared to exfoliated cells: impact on biomarker discovery. *BioMed Central Genomics.* **6**(64), pp.1-14. doi:10.1186/1471-2164-6-64.

Stout, M.J., Frey, H.A., Tuuli, M.G., Cahill, A.G., Odibo, A.O., Roehl, K.A. and Macones, G.A. (2014). Loop electrosurgical excision procedure and risk of vaginal infections during pregnancy: an observational study. *British Journal of Obstetrics and Gynaecology.* **122**(4), pp.545-551. doi:10.1111/1471-0528.13252.

Stuart B. **(2004).** *Infrared Spectroscopy: Fundamentals and Applications.* Chichester: John Wiley & Sons Ltd.

Sun, Q., Tsutsumi, K., Kelleher, B., Pater, A. and Pater, M.M. (1992). Squamous metaplasia of normal and carcinoma in situ of HPV 16-immoralized Human endocervical cells. *Cancer Research.* **52**(15), pp.4254-4260.

The Guardian. (2015). *Older women 'ignoring cervical cancer danger.'* [Online]. [Accessed 10 October 2015]. Available from: http://www.theguardian.com/society/2015/jun/15/older-women-ignoring-cervical-cancer-danger

Trent Cancer Registry and NHS Cancer Screening Programmes. (2012). *Profile of cervical cancer in England incidence: mortality and survival.* Sheffield: Trent Cancer Registry/NHS Cancer Screening Programmes.

Trevisan, J., Angelov, P.P., Carmichael, P.L., Scott, A.D. and Martin, F.L. (2012). Extracting biological information with computational analysis using Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst.* **137**(14), pp.3202-3215. doi:10.1039/c2an16300d.

Tsai, T-C. and Chen, S-L. (2003). The biochemical and biological functions of human papilloma virus type 16 E5 protein. *Archives of Virology*. **148**(8), pp.1445-1453.

UC Davies Chem Wiki. (no date). *Section 4.2: Infrared spectroscopy.* [Online]. [Accessed 12 April 2016]. Available from: http://chemwiki.ucdavis.edu/Core/Organic_Chemistry/Organic_Chemistry_With_a_Biological_Emphasis/Chapter_04%3A_Structure_Determination_I/Section_4.2%3A__Infrared_spectroscopy

Uthman, E. (2013). *LSIL (CIN 1), Cervical Biopsy (3776284166).* [Online]. [Accessed 10 October 2015]. Available from: https://commons.wikimedia.org/wiki/File:LSIL_(CIN_1),_Cervical_Biopsy_(3776284166).jpg?uselang=en-gb)

Uthman, E. (2007a). *Low-Grade_SIL_with_HPV_Effect.* [Online]. [Accessed 13 September 2015]. Available from: https://commons.wikimedia.org/wiki/File:Low-Grade_SIL_with_HPV_Effect.jpg?uselang=en-gb

Uthman, E. (2007b). *High-Grade SIL*. [Online]. [Accessed 11 September 2015]. Available from: https://commons.wikimedia.org/wiki/File:High-Grade_SIL.jpg?uselang=en-gb

Uthman, E. (2006a). *ThinPrep Pap smear HPV.* [Online]. [Accessed 15 October 2015]. Available from: https://commons.wikimedia.org/wiki/File:ThinPrep_Pap_smear_HPV.jpeg?uselang=en-gb

Uthman, E. (2006b). *Cervical intraepithelial neoplasia (3) CIN2.* [Online]. [Accessed 13 October 2015]. Available from:

https://commons.wikimedia.org/wiki/File:Cervical_intraepithelial_neoplasia_(3)_CIN2.jpg?uselang=en-gb

Uthman, E. (2006c). *Cervical intraepithelial neoplasia (4) CIN3.* [Online]. [Accessed 10 October 2015]. Available from: https://commons.wikimedia.org/wiki/File:Cervical_intraepithelial_neoplasia_(4)_CIN3.jpg?uselang=en-gb

Vaccarella, S., Lortet-Tieulent, J., Plummer, M., Franceschi, S. and Bray, F. (2013). Worldwide trends in cervical cancer incidence: Impact of screening against changes in disease risk factors. *European Journal of Cancer.* **49**(15), pp.3262–3273. doi:10.1016/j.ejca.2013.04.024.

Vicedo-Cabrera, A.M., Schindler, C., Radovanovic, D., Grize, L., Witassek, F., Dratva, J., Röösli, M. and Perez, L. (2016). Benefits of smoking bans on preterm and early-term births: a natural experimental design in Switzerland. *Tobacco Control.* pii: tobaccocontrol-2015-052739. doi: 10.1136/tobaccocontrol-2015-052739.

Vizcaino, A.P., Moreno, V., Bosch, F.X., Muñoz, N., Barros-Dios, X.M., Borras. J. and Parkin, D.M. (2000). International Trends in Incidence of cervical cancer: II squamous-cell carcinoma. *International Journal of Cancer.* **86**(3), pp.429-435.

Vizcaino, A.P., Moreno, V. and Bosch, F.X. (1998). International trends in incidence of cervical cancer: 1. adenocarcinoma and adenosquamous cell carcinomas. *International Journal of Cancer*. **75**(4), pp.536-545.

Wai, T.T. and Patil, D. (2008). Modern management of abnormal cervical smear. *British Journal of Medical Practitioners.* **1**(2), pp.18-22.

Walker, K-A.D., Morgan, C., Doak, S.H. and Dunstan, P.R. (2012). Quantum dots for multiplexed detection and characterisation of prostate cancer cells using a scanning near-field optical microscope. *PLoS ONE.* **7**(2), e31592. doi: 10.1371/journal.pone.0031592.

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. and Martin, F.L. (2007). ATR microspectroscopy with multivariate analysis segregates grades of exfoliative cervical cytology. *Biochemical and Biophysical Research Communications.* **352**(1), pp.213-219.

Williams, V.M., Filippova, M., Soto, U. and Duerksen-Hughes, P.J. (2011). HPV-DNA integration and carcinogenesis: putative roles for inflammation and oxidative stress. *Future Virology.* **6**(1), pp.45–57.

Women's Health and Education Centre. (2010). *Cervical cancer prevention: managing low-grade cervical neoplasia.* [Online]. [Accessed 5 June 2016]. Available from: http://www.womenshealthsection.com/content/gyno/gyno020.php3

World Health Organization. (2015). *Human papillomavirus (HPV) and cervical cancer; Fact Sheet Nº380, March, 2015.* [Online]. [Accessed 12 September 2015]. Available from: http://www.who.int/mediacentre/factsheets/fs380/en/

World Health Organization. (2013). *Latest world cancer statistics. Global Cancer burden rises to 14.1 million cases in 2012: Marked increase in breast cancers must be addressed; Press Release Nº223, 12 December, 2013.* [Online]. [Accessed 13 October 2015]. Available from: https://www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223_E.pdf

Yildiz-Arlan, S., Coon, J.S., Hope, T.J. and Kim, J.J. (2016). [Forthcoming]. Transcriptional profiling of human endocervical tissues reveals distinct gene expression in the follicular and luteal phases of the menstrual cycle. *Biology of Reproduction.* [Online]. [Accessed 29 May 2016]. Available from: http://www.biolreprod.org/content/early/2016/05/06/biolreprod.116.140327.long

Appendix A: Supplementary analyses

The full statistical methodology has been previously described (see Project One). In brief, the treated group used in the aforementioned project (cone depth removed) was segregated by smoking status, menstrual phase (luteal or follicular) and parity. These four groups were compared using means, SD and a student's *t* test. The same seven individual biomarkers at the following wavelengths were also extracted and compared using multiple t tests and corrected for using the Holm-Sidak method: ~1030 cm^{-1} (glycogen/collagen); ~1072 cm^{-1} (symmetric phosphate I of DNA); ~1170 cm^{-1} (glycomaterials and proteins); ~1223 cm^{-1} (asymmetric phosphate I of DNA); ~1470 cm^{-1} (lipids); ~1550 cm^{-1} (amide II) and ~1651 cm^{1} (amide I). The results for smoking status, parity, menstrual phase and COCP use are presented in Supplementary Figure 1, Supplementary Figure 2, Supplementary Figure 3, and Supplementary Figure 4, respectively.

Supplementary Tables 1 and 2 present the patient characteristics (smoking, parity, menstrual phase, and COCP use) for each group. Characteristics were tested for significance (alpha value = 0.05), using Fisher's exact test.

Characteristics	Comparison 1: Treated women with paired samples		*Fisher' exact test*	Comparison 2 : Normal post-treatment vs. normal untreated controls		*Fisher's exact test*	Comparison 3: Treated by cone depth/proportion vs. normal untreated controls		*Fisher's exact test*
	Pre-treatment	Post treatment	*P-value*	Normal post-treatment, (Cytology negative, HPV negative)	Normal controls (Cytology negative, HPV negative)	*P-value*	Post treatment	Controls (Cytology negative; HPV status ignored)	*P-value*
	(n= 29)	(n= 33)		(n = 39)	(n=20)		(n=58)	(n=27)	
Smoking status, n/N (%)			0.77			**0.04***			0.18
Non-smoker	23/29 (79)	25/33 (76)		27/39 (69)	19/20 (95)		41/58 (71)	23/27 (85)	
Current smoker	6/29 (21)	8/33 (24)		12/39 (31)	1/20 (5)		17/58 (29)	4/27 (15)	
Parity, n/N (%)			1.00			0.52			0.51
Nulliparous	21/29 (72)	24/33 (73)		32/39 (82)	15/20 (75)		45/58 (78)	22/27 (81)	
Parous	8/29 (28)	9/33 (27)		7/39 (18)	5/20 (25)		13/58 (22)	5/27 (19)	
Phase of menstrual cycle, n/N (%)			0.40			0.14			0.08
Luteal	15/29 (52)	18/33 (55)	0.60	21/39 (54)	6/20 (30)	0.23	29/58 (50)	7/27 (26)	0.12
Follicular	13/29 (45)	11/33 (33)		14/39 (36)	9/20 (45)		23/58 (40)	14/27 (52)	
Unknown	1/29 (3)	4/33 (12)		4/39 (10)	5/20 (25)		6/58 (10)	6/27 (22)	
Contraceptive use, n/N (%)									
COCP	14/29 (48)	17/33 (52)	1.00	20/39 (51)	3/20 (15)	**0.01****	28/58 (48)	4/27 (15)	**0.004****
Other	15/29 (52)	16/33 (48)		19/39 (49)	17/20 (85)		30/58 (52)	23/27 (85)	

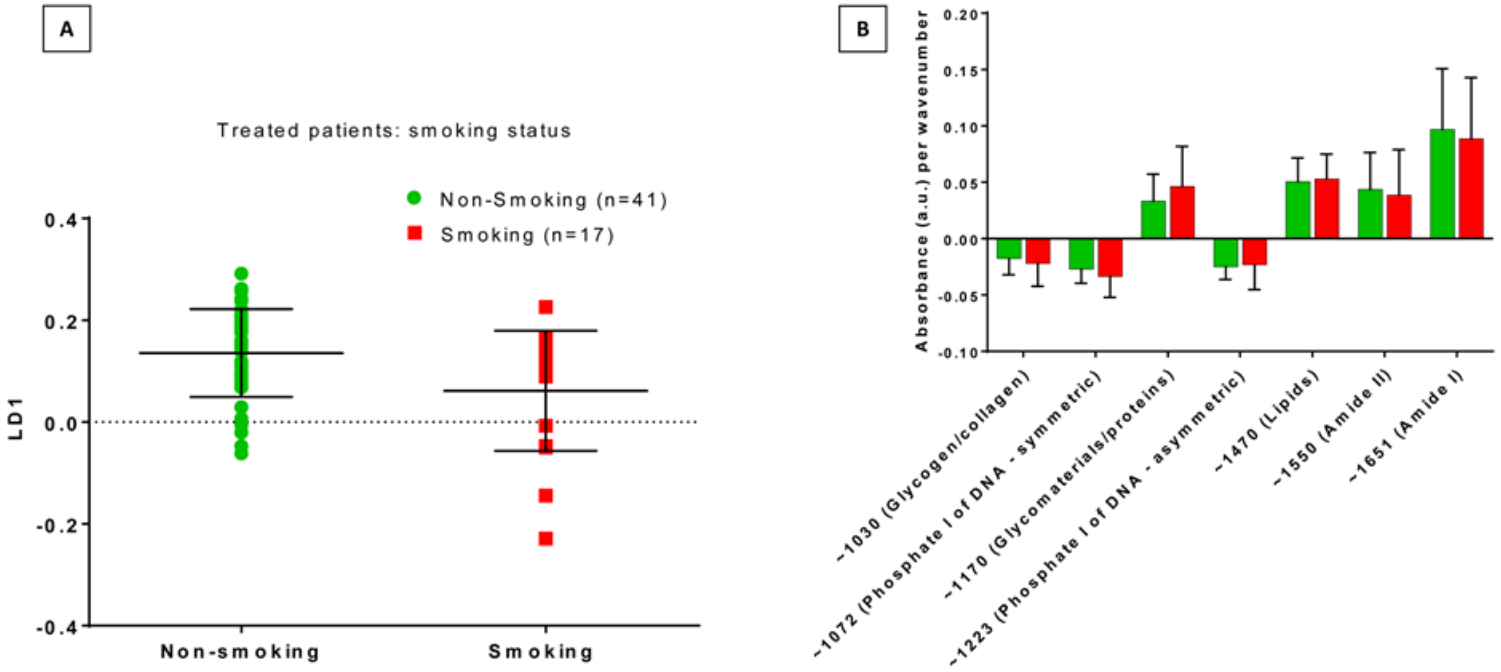
COCP: Combined oral contraceptive pill; HPV: Human papillomavirus.

Supplementary Table 1. Patient characteristics and Fisher's exact test for significance between characteristics.

Characteristics	**Normal controls (n=27)**	**Cone Depth[a] (n=58)**			**Percentage Excision[b] (n=53)**			
		Category 1: <10 mm (n=24)	**Category 2: 10-14 mm (n=24)**	**Category 3: >= 15 mm (n=10)**	**Category 2: 11-20% (n=9)**	**Category 3: 21-30% (n=22)**	**Category 4: 31-40% (n=15)**	**Category 5: >40% (n=7)**
Smoking status, n/N (%)								
Non-smoker	23/27 (85)	19/24 (79)	15/24 (63)	7/10 (70)	7/9 (78)	17/22 (77)	8/15 (53)	5/7 (71)
Current smoker	4/27 (15)	5/24 (21)	9/24 (37)	3/10 (30)	2/9 (22)	5/22 (23)	7/15 (47)	2/7 (29)
Parity, n/N (%)								
Nulliparous	22/27 (81)	19/24 (79)	20/24 (83)	6/10 (60)	8/9 (89)	18/22 (82)	12/15 (80)	5/7 (71)
Parous	5/27 (19)	5/24 (21)	4/24 (17)	4/10 (40)	1/9 (11)	4/22 (18)	3/15 (20)	2/7 (29)
Phase of menstrual cycle, n/N (%)								
Luteal	7/27 (26)	14/24 (58)	9/24 (38)	6/10 (60)	6/9 (67)	12/22 (55)	6/15 (40)	3/7 (43)
Follicular	14/27 (52)	8/24 (34)	12/12 (50)	3/10 (30)	3/9 (33)	10/22 (45)	6/15 (40)	3/7 (43)
Unknown	6/27 (22)	2/24 (8)	3/24 (12)	1/10 (10)	0/9 (0)	0/22 (0)	3/15 (20)	1/7 (14)
Contraceptive use, n/N (%)								
COCP	4/27 (15)	12/24 (50)	12/24 (50)	4/10 (40)	5/9 (56)	12/22 (54)	6/15 (40)	4/7 (57)
Other	23/27 (85)	12/24 (50)	12/24 (50)	6/10 (60)	3/9 (44)	10/22 (46)	7/15 (60)	3/7 (33)

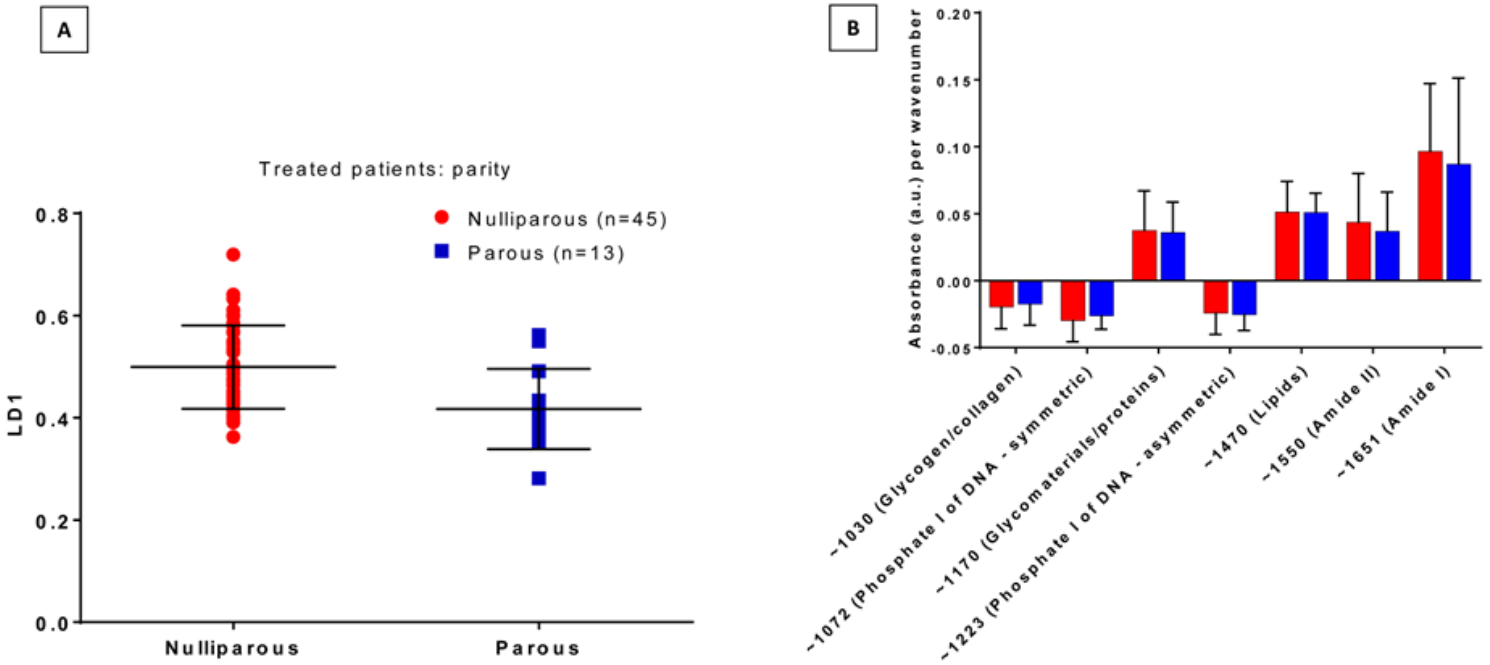
[a] Patients were included in the dimensional analyses if they had had treatment and had follow up data at 6 months after treatment. [b] There was insufficient patients in Category 1 to include in the analysis. No segregation was made based on histology, cytology or HPV testing. COCP: Combined oral contraceptive pill.

Supplementary Table 2. Patient characteristics for the dimensional groups (Comparison 3).



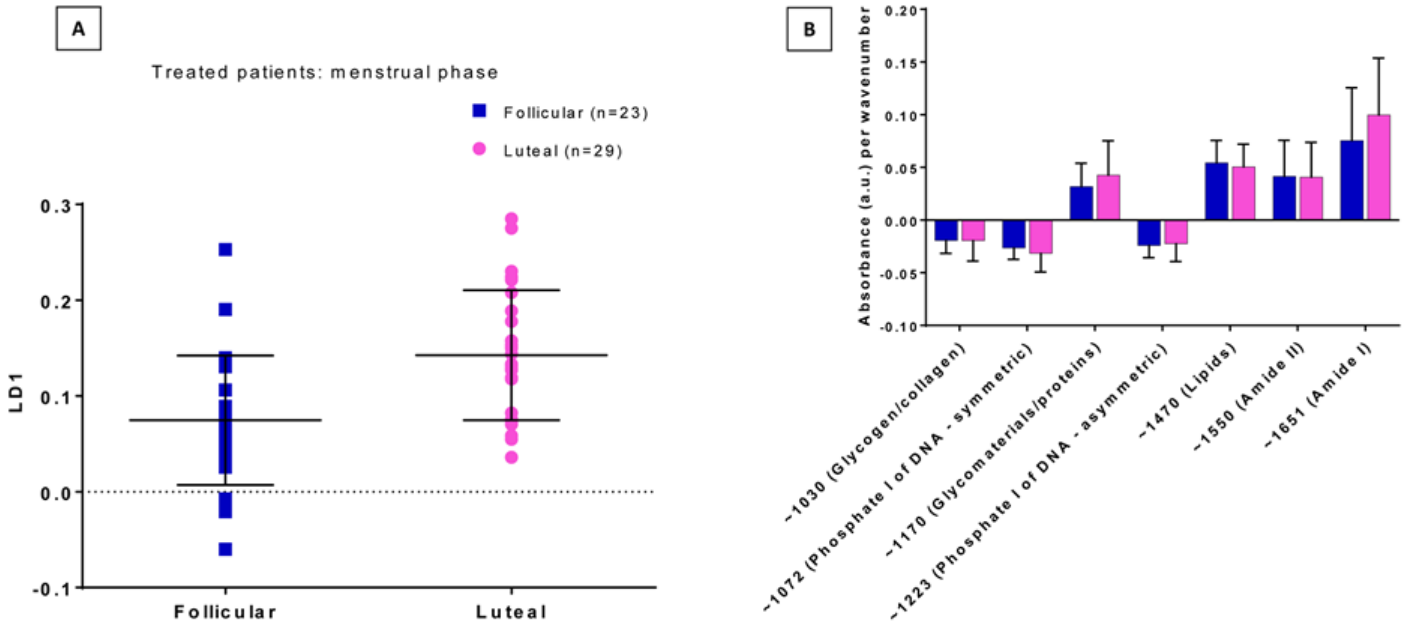
Supplementary Figure 1. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1/Absorbance (a.u.) per wavenumber: treated by smoking status.

The two groups were significantly different along LD1 (Mean/SD: Non-Smoking = 0.14/0.09; Smoking = 0.06/0.12; p=0.01, 95% CI = -0.013 to -0.02). There was no significant difference in absorbance associated with each of the seven biomarkers between non-smokers and smokers. ATR-FTIR: Attenuated total reflectance Fourier-transform; CI: Confidence interval; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis-Linear Discriminant Analysis cascade; SD: Standard deviation.



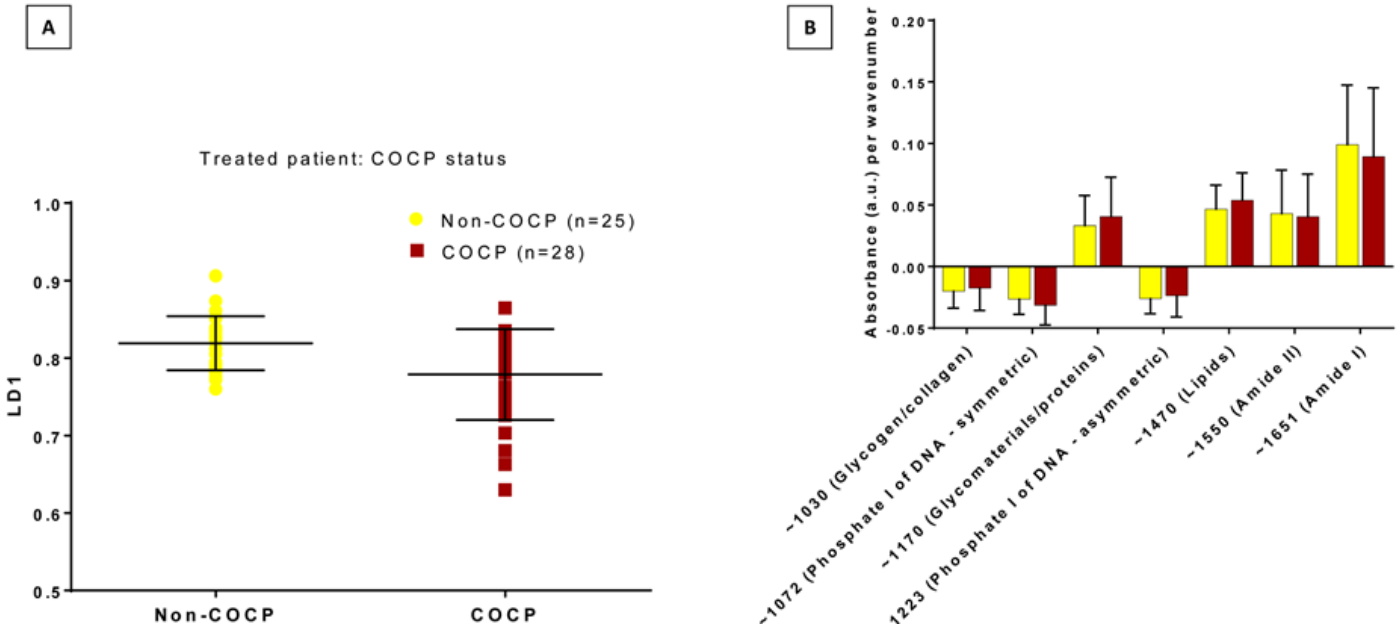
Supplementary Figure 2. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1/Absorbance (a.u.) per wavenumber: treated by parity.

The two groups were significantly different along LD1 (Mean/SD: Nulliparous = 0.50/0.08; parous = 0.42/0.08; p=0.002, 95% CI = -0.133 to -0.031). There was no significant difference in absorbance associated with each of the seven biomarkers between nulliparous and parous women. ATR-FTIR: Attenuated total reflectance Fourier-transform; CI: Confidence interval; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis-Linear Discriminant Analysis cascade; SD: Standard deviation.



Supplementary Figure 3. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1/Absorbance (a.u.) per wavenumber: treated by menstrual phase.

The two groups were significantly different along LD1 (Mean/SD: follicular phase = 0.07/0.07; luteal phase = 0.14/0.07; p=0.0007, 95% CI = -0.11 to -0.03). There was no significant difference in absorbance associated with each of the seven biomarkers between women in the luteal phase compared with women in the follicular phase. ATR-FTIR: Attenuated total reflectance Fourier-transform; CI: Confidence interval; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis-Linear Discriminant Analysis cascade; SD: Standard deviation. *** Menstrual phase was unknown for 6 patients; patients were excluded from analysis.**



Supplementary Figure 4. PCA-LDA scores plot of ATR-FTIR spectra with regards to LD1/Absorbance (a.u.) per wavenumber: treated by COCP status.

The two groups were significantly different along LD1 (Mean/SD: Non-COCP = 0.82/0.03; COCP = 0.78/0.06; p=0.004, 95% CI = -0.07 to -0.02). There was no significant difference in absorbance associated with each of the seven biomarkers between non-COCP compared with COCP. ATR-FTIR: Attenuated total reflectance Fourier-transform; CI: Confidence interval; COCP: Combined oral contraceptive pill; LD1: Linear Discriminant 1; PCA-LDA: Principal Component Analysis-Linear Discriminant Analysis cascade; SD: Standard deviation. *** Contraception status was unknown for 5 patients; patients were excluded from analysis.**

Appendix B: National Research Ethics Service Committee Approval (London – Fulham; Approval number 13/LO/0126)

NHS
Health Research Authority

National Research Ethics Service

NRES Committee London - Fulham
HRA NRES Centre Manchester
Barlow House
3rd Floor, 4 Minshull Street
Manchester
M1 3DZ

Telephone: 0161 625 7821
Facsimile: 0161 625 7299

03 April 2013

Dr Maria Kyrgiou
Subspecialty trainee in Gynaecologic Oncology - Honorary Clinical Lecturer
Imperial NHS Healthcare Trust
West London Gynaecologic Cancer Center - Dept of Obstetrics & Gynaecology - Dept of Surgery & Cancer
Queen Charlotte's & Chelsea - Hammersmith Hospital - Imperial College
Du Cane Road
W120HS

Dear Dr Kyrgiou

Study title:	**Proportion of excision volume and length after treatment for cervical intra-epithelial lesions: cervical healing, quality of regenerated tissue, immunology and pregnancy outcome**
REC reference:	**13/LO/0126**
Protocol number:	**01**
IRAS project ID:	**75802**

Thank you for your letter of 27 March 2013, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chairman.

We plan to publish your research summary wording for the above study on the NRES website, together with your contact details, unless you expressly withhold permission to do so. Publication will be no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to withhold permission to publish, please contact the Co-ordinator, Miss Shehnaz Ishaq, nrescommittee.london-fulham@nhs.net

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

The Committee has not yet been notified of the outcome of any site-specific assessment (SSA) for the non-NHS research site(s) taking part in this study. The favourable opinion does not therefore apply to any non-NHS site at present. We will write to you again as soon as one Research Ethics Committee has notified the outcome of a SSA. In the meantime no study procedures should be initiated at non-NHS sites.

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Evidence of insurance or indemnity	Imperial College London	30 July 2012
Evidence of insurance or indemnity	Imperial College London	30 July 2012
Investigator CV	Maria Kyrgiou	08 January 2013
Letter from Sponsor	Imperial College	03 January 2013
Participant Consent Form	01	02 October 2012
Participant Consent Form: Control Groups (Clean)	2.0	22 March 2013
Participant Consent Form: Control Groups (Tracked Changes	2.0	22 March 2013
Participant Consent Form: Treatment Group (Tracked Changes)	2.0	22 March 2013
Participant Consent Form: Treatment Group (Clean)	2.0	22 March 2013

Participant Information Sheet: Control Groups (Clean)	2.0	22 March 2013
Participant Information Sheet: Control Groups (Tracked Changes)	2.0	22 March 2013
Participant Information Sheet: Treatment Group (Tracked Changes)	2.0	22 March 2013
Participant Information Sheet: Treatment Group (Clean)	2.0	22 March 2013
Protocol	01	02 October 2012
REC application	75802/399961/1/809	14 December 2012
Response to Request for Further Information		27 March 2013

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

13/LO/0126 Please quote this number on all correspondence

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

Yours sincerely

Signed on behalf of:
Dr Charles Mackworth-Young
Chairman

Email: nrescommittee.london-fulham@nhs.net

Enclosures: "After ethical review – guidance for researchers"

Copy to: Miss Sadaf Ghaem-Maghami, Imperial College London

Ms Lucy Parker, Imperial College London and Imperial College Healthcare NHS Trust

Appendix C: Laboratory protocol (preparation of cervical LBC samples for ATR-FTIR spectroscopy)

Protocol for LBC Samples – lab preparation

Aim: This protocol aims to wash cervical cells collected at colposcopy clinic and immediately fixed into an alcohol-based fixative. The washing stage removes the alcohol (and therefore any spectral signature associated with it), rehydrates the cells and helps remove other cofounding factors, such as blood.

Safety:

Wear protective gloves/Howie/safety glasses if cutting.

Use the clinical bin for waste disposal; decant any collected suspension into the liquid waste.

Materials:

Bench top centrifuge with a 2000 rpm setting.

dH_2O.

Low-e slides (cut into 3) for economy.

Mini petri dishes.

Non-sterile pipette tips.

Pipettes.

Blue roll.

Marker pen.

Method:

1. Decide how many samples you wish to process and pre-label micro tubes and mini petri dishes with the sample number (label bottom of petri dish as well as top).

2. If cutting slides, ensure white tip is facing upwards and place a cut segment into each petri dish.

3. Agitate selected samples to re-distribute the cells.

4. Decant 500 µl from sample into clean, pre-labelled micro tube. Replace original samples into fridge.

5. Washing stages:

a. Wash #1: Centrifuge sample at 2000 rpm for 5 minutes to draw a pellet at the bottom of the tube.

b. Wash #2: Decant the ThinPrep from above the pellet using clean tip, taking care not to disturb it. Discard tip, replace with clean tip and add 500 µl of clean distilled water to tube. Repeat centrifuge step.

c. Wash #3: repeat step b.

6. Decant the water from above the pellet and add 100 µl of clean distilled water to tube.

7. Agitate sample to redistribute the cells.

8. Draw up cells using 100-200 µl setting, ensuring you have the entire sample.

9. Discharge sample onto slide in mini petri dish and allow to air dry for 24 hours. (Optional: leave lid ajar to promote faster drying). Discard micro tube into clinical waste.

10. For repeat samples where the pellet is small and therefore the number of cells less, dispense the final 100 µl as two x 50 µl aliquots, allowing a 24-hour drying period in between each application, and dispense the second aliquot directly on top of the first to promote a uniform spread of cells.

11. Once air dried, replace lids and collect into small batches and tape together. Label and place in a desiccator for a minimum of 24 hours to remove any remaining water.

12. Aim to prepare and collect spectra in within the same time window for each batch.

13. Once spectroscopy has been performed, replace samples back into desiccator.

Spectroscopy analysis

Pre-testing:

Click measure/advanced/Check signal – click save – dot should be in middle.

14. Humidity test each run should be zero% (click internet explorer/diagnostics/check detectors should be zero%.

15. However, 10% is acceptable. Click on diagnostics and detectors to check humidity – above 10% change desiccator (see below).

16. Close diagnostics down before running tests.

17. If prompted run tests but you need to remove unit.

18. Open video control panel – PW = ‘OPUS.’

19. Open video control panel to view diamond.

20. Clean diamond with distilled water to remove any debris and dry.

21. Click ‘Advanced measurement,’ then ‘background signal channel.’ Wait for the scan to finish.

22. Load slide – click ‘advanced’ tab and add in details of the sample name.

23. Change the pathway to the folder where you want to store your spectra.

24. Click ‘single channel scan’ to view the spectral image and adjust as necessary to get the water peak around 0.17-0.2. The amide groups should be clear. The trick is consistency. Use the same machine for all samples.

25. When you are ready to start, click ‘start measurement’ and it will take 32 scans containing 235 data points.

26. The slide needs to be touching the diamond. The water peaks should be peaks and not curves. If curves, return to desiccator for further drying.

27. When the scan has finished, move the slide a little and re-scan. You need to do this 10 times but as the count starts from zero, it will run from 0-9.

28. The area we are interested in is between 900-1800.

29. If the machine becomes stuck, unload you files first to send them to your folder. Switch off the machine and clean, then run background check again.

30. To unload files, click the number above image ‘0’, right mouse click, then click on last image, the right click to get unload menu, click unload. These images will then be sent to your folder.

31. You are looking to achieve consistency across the images. The density of the cells in contact with the diamond influence the resulting image. Better density improves the image output.

Changing the desiccator

32. Turn off machine at switch.

33. Use large Allen key to remove hood.

34. Remove desiccator tube and replace hood without locking.

35. If using grey pellets, these are reusable. Replace with dried our pellets kept warm in oven. Pour old desiccant into pot and replace to over.

36. Fill tube with fresh desiccator nearly to top, leaving about an inch.

37. Replace tube in holder and apply bolt with Allen key to hood.

38. Leave for about 15 minutes to settle. Re-run detectors. If 10% or under, ready to use.

Appendix D: Laboratory protocol (preparation of cervical LBC samples for SNOM-IR-FEL imaging)

Lancaster Mini Protocol – SNOM-IR-FEL experiment

Aim: This protocol aims to prepare cervical cells on a barium fluoride window as a single layer of cells with a minimum of debris.

Safety:

Wear protective gloves/Howie/safety glasses if cutting.

Use the clinical bin for waste disposal; decant any collected suspension into the liquid waste.

Materials:

LBC samples.

Barium fluoride windows.

Barium fluoride slide holders (to hold 3 windows per slide).

Tissue/slide cases.

Method

Two slides per sample

1. Wash the samples as described in Appendix C, but instead of suspending the final pellet in 100 µl of distilled water, if the pellet is large, re-suspend it in 1000 µl of dH_2O. For smaller pellets, re-suspend the pellet into in 500 µl of dH_2O.
2. Transfer the samples securely to the Pathology Department, Preston Hospital along with the appropriate number of barium fluoride windows, slide case and window holders.

3. The pathologist/cytologist will prepare each window and sample for centrifuging tin the cytospinner. The samples will be agitated to disperse the cells and then 5/6 drops will be added to the cytofunnel. The samples will be spun onto the barium fluoride windows using 3000 rpm for 5 minutes.

Method for mounting onto the SNOM

Mount duplicate slides at the same time and fix with blue tack.

Rotate through 1 area per slide (depending on time) collecting the following wavelengths:

- 8.16 um (DNA)
- 6.06 um (Amide I)
- 6.46 um (Amide II)
- 5.71 um (Lipids)
- Include 1 repeat scan in addition to above per area for quality control/repeatability
- If time allows, select a 2nd area on one of the slides.

Appendix E: Poster presented at The International Society for Clinical Spectroscopy (CLIRSPEC)

This poster was presented it at the "CLIRSPEC Summer School" conference in Windermere, 7-10th July 2015 as:

"CERVICAL CANCER: Can attenuated total reflection Fourier-transform infrared spectroscopy replace conventional cytology?"

Diane E. Halliwell, Georgios Theophilou, Pierre L. Martin-Hirsch, Maria Kyrgiou, and Francis L. Martin.

Contribution:

I created this poster based on my ongoing studies (at that time).

…… ……

Diane E. Halliwell Professor Francis L. Martin

CERVICAL CANCER: Can attenuated total reflection Fourier-transform infra-red spectroscopy replace conventional cytology screening?

LANCASTER UNIVERSITY
Centre for Biophotonics

Diane E. Halliwell[a], Georgios Theophilou[b], Pierre L. Martin-Hirsch[c], Maria Kyrgiou[d], Francis L. Martin[a]

[a] Centre for Biophotonics , Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster, LA1 4YW, UK.
[b] St James University Hospital, Leeds, West Yorkshire, LS9 7TF, UK.
[c] Lancaster Medical School, Furness Building, Lancaster University, Bailrigg, Lancaster, LA1 4YG.
[d] Faculty of Medicine, Department of Surgery & Cancer, Imperial College London, London, SW7 2AZ, UK.

Cervical Cancer Upcoming Research

Fast Facts

- Cervical cancer is the 4th most common cancer in women globally, with >500, 000 new cases diagnosed every year[2]
- It is caused by a sexually acquired infection with certain types of Human papillomavirus (HPV) [2]
- Global mortality rate remains high at 52%[2]
- Conventional cytology screening has been shown to be flawed, with inter- and intra-observer variability resulting in missed disease[3]
- Attenuated total refraction Fourier-transform IR (ATR-FTIR) spectroscopy can identify underlying disease more accurately than conventional cytology[3] and separate regressive *vs* static *vs* progressive disease[4]

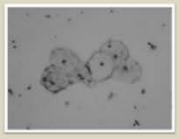
Multinucleated cell: invasive adenocarcinoma (~ 10% of all cervical cancers[1])

Research questions

1. Can ATR-FTIR spectroscopy identify changes in biomarkers (wavenumbers) following treatment for cervical cancer?
2. Does our research support current evidence that ATR-FTIR spectroscopy can separate grades of dysplasia more accurately than conventional cytology[4]?

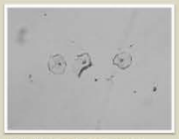
Quiz – How good is your cytology?

- Current cytology screening for cervical cancer is done by the human eye, looking for changes in cell morphology. In particular, changes in the nucleus (size, shape, number) and changes in cell shape
- The following shows a series of microscope images of cervical cells but their descriptions have been mixed up. *Can you match the correct microscope image to the correct description?*



#1: High-grade squamous intraepithelial lesion(s). Cell with early changes in nucleus size. NB: not *all* cells will be at the same stage in a sample

#2: Low-grade squamous intraepithelial lesion(s) showing mild dysplasia with slightly enlarged nuclei and changes to cell shape



#3: Normal squamous cells with small dense nuclei, regular shape and good adhesion to neighbouring cells

#4: Adenocarcinoma. A mix of lesions with enlarged nuclei, one showing vastly reduced cytoplasm and reduced cell size

Answers at the bottom of the poster

Sample population

- Cervical cell samples (n=231) were collected from women attending at a colposcopy clinic at the Hammersmith Hospital in London and were diagnosed on cytology
- Demographics included age at diagnosis, ethnicity, smoking history, contraceptive use, recent antibiotic use and history of pregnancy

Methods

- Samples were stored in 1000 µl of Thin-Prep®; washed and centrifuged 3 times. The resultant pellet was suspended in 100 µl of distilled water, mounted onto Low-E slides and desiccated for a minimum of 24 hours
- Ten spectra per slide are to be taken using ATR-FTIR
- Spectra are to be pre-processed by:
  - Cutting spectra to 900-1800 cm^{-1}
  - 1st order differentiation
  - Vector normalisation
- Multivariate analyses are to be conducted using cross-validated, PCA-LDA , followed by appropriate tests for significance between different groups and sub-groups

References
1) Pimenta JM, Galindo C, Jenkins D, Taylor SM. Estimate of the global burden of cervical adenocarcinoma and potential impact of prophylactic human papillomavirus vaccination. *BMC Cancer* 2013; 13(553): .http://www.biomedcentral.com/1471-2407/13/553 (accessed 25 June 2015).
2) World Health Organization. *Human papillomavirus (HPV) and cervical cancer; Fact Sheet No380. http://www.who.int/mediacentre/factsheets/fs380/en/ (accessed 18 May 2015).*
3) Gajjar K, Ahmadzai AA, Valasoulis G, Trevisan J, Founta C, Nasioutziki M, et al. Histology Verification Demonstrates That Biospectroscopy Analysis of Cervical Cytology Identifies Underlying Disease More Accurately than Conventional Screening: Removing the Confounder of Discordance. PLOS ONE 2014; 9(1; e82416):1-16.
4) Purandare N, Patel I, Lima K, Trevisan J, Ma'Ayeh M, McHugh A et al. Infrared spectroscopy with multivariate analysis segregates low-grade cervical cytology based on likelihood to regress, remain static or progress. Anal Methods. 2014;6(13):4576-4584.

Acknowledgements
Diane E Halliwell is a Research Masters Student under the supervision of Professor Frank Martin.
Attendance at this conference was supported by a bursary from Lancaster University's Graduate College.

Answers
#1: Normal cervical cells. #2: Adenocarcinoma (taken from same sample in Facts Facts).
#3: Low-grade squamous intraepithelial lesion(s). #4: High-grade intraepithelial lesion(s).